HENN
SARTORIUS
HELMCHEN
LAUTER
(EDITORS)

CONTEMPORARY PSYCHIATRY

Foundations of Psychiatry

Springer

Springer-Verlag Berlin Heidelberg GmbH

F. Henn · N. Sartorius
H. Helmchen · H. Lauter
(Editors)

Contemporary Psychiatry

Volume 1
Foundations of Psychiatry

With Contributions by
J. Aldenhoff · G. Andrews · J. Arboleda-Flórez · U. Baumann · T. Becker
G.E. Berrios · A. Bertelsen · J.M. Bertolote · B. Bogerts · W. Böker
K.H. Brisch · J.D. Brodie · A. Buchheim · M. Bullinger-Naber · S. Curran
A. Diefenbacher · H. Dilling · U. Ehlert · L. Eisenberg · H.M. Emrich
R.R. Engel · P. Falkai · A.R. Felthous · W. Gaebel · R. Gardner · J.A. Gingrich
G. Goldenberg · D. Hellhammer · H. Helmchen · D.R. Hemsley · R. Hen
A.S. Henderson · F. Henn · I. Heuser · R.J. Hitzemann · R. Jenkins
H. Kächele · R. Kessler · M. Knapp · A. Kraus · H.L. Kröber · P. Leaf
M. Linden · O. Lipp · P. McGuffin · W.T. McKinney · J. Mendlewicz
R. Michels · H.J. Möller · C. Mundt · D. Naber · M.C. O'Donovan · M. Perrez
D. Ploog · S. Poitras · C. Pull · N. Sartorius · H. Sass · W. Schiefenhövel
R. Schlösser · G. Schmücker · J. Scott · J. Siegrist · A. Sims · D. Souery
M. Spitzer · W. Strik · J.J. van Drimmelen-Krabbe · J. Vollmann · H.J. Walton

Springer

Prof. Dr. Dr.
Fritz Henn
Zentralinstitut für Seelische Gesundheit
P.O. Box 12 21 20
68072 Mannheim, Germany

Prof. Dr. Dr. Dr. h.c. mult.
Norman Sartorius
Hôpitaux Universitaires de Genève
Belle-Idée, Bâtiment Salève
2, chemin du Petit-Bel-Air
1225 Chêne-Bourg/Genève, Switzerland

Prof. em. Dr.
Hanfried Helmchen
Freie Universität Berlin
Psychiatrische Klinik
Eschenallee 3
14050 Berlin, Germany

Prof. em. Dr.
Hans Lauter
Technische Universität München
Klinikum rechts der Isar
Ismaninger Straße 22
81675 München, Germany

Library of Congress Cataloging-in-Publication Data
Contemporary psychiatry / F. Henn ... [et al.], editors.
p. ; cm.
Developed on the basis of the experience obtained with: Psychiatrie der Gegenwart, that appeared only in German.
Includes bibliographical references and indexes.
Contents: v. 1. Foundations of psychiatry – v. 2. Psychiatry in special situations – v. 3. Specific psychiatric disorders.

ISBN 978-3-642-64007-0 ISBN 978-3-642-59519-6 (eBook)
DOI 10.1007/978-3-642-59519-6

1. Psychiatry. I. Henn, Fritz A. II. Psychiatrie der Gegenwart.
[DNLM: 1. Mental Disorders. 2. Psychiatry – methods. WM 140 C761 2001]
RC454.C6541 2001
616.89–dc21 00-046344

Originally published by Springer-Verlag Berlin Heidelberg New York in 2001
Softcover reprint of the hardcvoer 1st edition 2001

Cover design: de'blik, Berlin
Typesetting: Scientific Publishing Services (P) Ltd, Chennai
Printed on acid-free paper SPIN: 10693473 26/3130SM – 5 4 3 2 1 0

Preface

Contemporary Psychiatry is an international textbook of psychiatry developed on the basis of the experience obtained with its first three editions, which only appeared in German. As such, this version incorporates the German tradition of phenomenology and diagnosis going back to Kraepelin, as well as other authors such as K. Schneider or Leonhard. However, for the first time, this tradition is combined with the Anglo-American tradition and the DSM-IV diagnostic system, making it a unique resource among psychiatric texts. Each edition has had a special focus. The first edition appeared some 15 years after the end of the Second World War and aimed to present the best of German psychiatrists' writing in order to help reestablish the links between psychiatry in Germany and elsewhere that were disrupted by the war. The second edition gave particular emphasis to social psychiatry and empirical research in psychiatry in the hope that they would become more familiar to German psychiatrists and be used by them in their research and in their administrative and clinical decisions. The third edition added a focus on the integration of psychodynamic approaches and psychotherapy in the practice of psychiatry. This present, fourth edition examines psychiatric problems systematically, combining scientific evidence and modern neurobiology with experience obtained in different sociocultural settings while paying particular attention to the insights gained by the different disciplines relevant to psychiatry.

While comprehensive in its coverage, *Contemporary Psychiatry* is not encyclopedic in the sense of listing all the issues and findings of the past two decades. It presents findings relevant to the key issues in psychiatry together with the paradigms that are necessary in order to gain a conceptual understanding of the current trends in psychiatric research, service, and training. In doing so, *Contemporary Psychiatry* naturally focuses on methods and approaches that are playing an increasingly important role in psychiatry, such as molecular biology, genetics, neuroimaging, and cognitive psychology. This method of presenting today's knowledge has been chosen because it is most likely to serve as a resource not only for psychiatrists, but also for medical students, those engaged in basic research relevant to psychiatry, and clinicians from other medical disciplines who are increasingly collaborating with psychiatrists in an effort to understand and manage psychiatric problems.

Although all the volumes of *Contemporary Psychiatry* are linked, each contains separate sections. The first volume presents the scientific basis of psychiatry and the contributions that disciplines overlapping with psychiatry can make to the resolution of psychiatric problems. Epidemiology, the neurosciences, psychology, genetics, and the social sciences are given particular prominence in this section.

The second part of Vol. 1 deals with general psychiatry. It begins with an analysis of classifications in psychiatry, paying particular attention to those included in the *International Classification of Diseases* (ICD) of the World Health Organization and in the *Diagnostic and Statistical Manual* (DSM) of the American Psychiatric Association. The chapters that follow deal with the skills necessary for psychiatric examinations, principles of psychiatric treatment, the structure of mental health services and their influence on the management of psychiatric problems, legal and ethical issues, and matters relating to undergraduate and postgraduate education in psychiatry. Volume 2 addresses the presentation of psychiatric problems and their management in special situations (such as that of a refugee status), in different cultures, and in combination with other mental and physical diseases. Part 2 of this Volume systematically presents knowledge about specific psychiatric disorders, including the dementia. Schizophrenias, affective disorders, personality disorders, neurotic disorders (such as anxiety disorders, obsessive-compulsive disorders), eating disorders, suicide, substance abuse, and dependence disorders are presented in Vol. 3. The epidemiological profile, genetic and pathophysiological characteristics, clinical form and course, treatment, and rehabilitation are presented for each of the disorders.

Contemporary Psychiatry combines the German tradition of psychiatry with the traditions of psychiatry in other countries, particularly that of the English-speaking countries. Its ecumenical strategy is exemplified by its approach to the classification of mental disorders, in that both clinical consensus and the guidelines contained in ICD-10 and DSM-IV have been used in organizing the parts dealing with specific diseases. The various chapters have been written by authors from 15 different countries, and versions of *Contemporary Psychiatry* in several other languages as well as in English and German are planned to be published, reflecting the aim of producing an international textbook that is useful in different settings throughout the world.

Fall 2000

F. Henn
N. Sartorius
H. Helmchen
H. Lauter

Contents

List of Contributors

ALDENHOFF, J., Prof. Dr., Christian-Albrechts-Universität, Klinik für Psychiatrie und Psychotherapie, Niemannsweg 14, 24104 Kiel, Germany

ANDREWS, G., Prof., Clinical Research Unit for Anxiety Disorders, 299 Forbes Street, Darlinghurst, NSW 2010, Australia

ARBOLEDA-FLÓREZ, J., Prof., Queens University, Faculty of Medicine, Department of Psychiatry, Kingston, Ontario K7L 3N6, Canada

BAUMANN, U., Prof. Dr., Institut für Psychologie der Universität Salzburg, Abteilung für Klinische Psychologie, Hellbrunnerstr. 34, 5020 Salzburg, Austria

BECKER, T., Prof. Dr., Universität Leipzig, Klinik und Poliklinik für Psychiatrie, Liebigstr. 22B, 04103 Leipzig, Germany

BERRIOS, G.E., Prof., University of Cambridge Clinical School, Addenbrooke's Hospital, Cambridge CB2 2QQ, United Kingdom

BERTELSEN, A., Dr., Aarhus Psychiatric Hospital, Department of Psychiatric Demography, 8240 Risskov, Denmark

BERTOLOTE, J.M., Department of Mental Health, World Health Organization, 1211 Geneva 27, Switzerland

BOGERTS, B., Prof. Dr., Medizinische Fakultät der Otto-von-Guericke-Universität Magdeburg, Klinik für Psychiatrie, Leipziger Str. 44, 38120 Magdeburg, Germany

BÖKER, W., Prof. em. Dr., Zähringerstr. 30a, 69115 Heidelberg, Germany

BRISCH, K.H., Dr., Klinikum der Universität Ulm, Abteilung Psychotherapie und Psychosomatische Medizin, Frauensteige 14a, 89075 Ulm, Germany

BRODIE, J.D., Prof. Dr., New York University Medical Center, Department of Psychiatry, 550 First Avenue, New York, NY 10016, USA

BUCHHEIM, A., Dipl.-Psych., Klinikum der Universität Ulm, Abteilung Psychotherapie und Psychosomatische Medizin, Frauensteige 14a, 89075 Ulm, Germany

BULLINGER-NABER, M., Prof. Dr., Universität Hamburg, Medizinische Klinik, Universitätskrankenhaus Eppendorf, Abteilung Medizinische Psychologie, Kollaustr. 67–69/B, 22529 Hamburg, Germany

CURRAN, S., Dr., St. James's University Hospital, Division of Psychiatry and Behavioural Sciences, Beckett Street, Leeds LS9 7TS, United Kingdom

DIEFENBACHER, A., Dr., Evangelisches Krankenhaus Königin Elisabeth Herzberge, Abteilung für Psychiatrie und Psychotherapie, Herzbergerstr. 79, 10362 Berlin, Germany

DILLING, H., Prof. Dr., Medizinische Universitätsklinik zu Lübeck, Klinik für Psychiatrie, Ratzeburger Allee 160, 23562 Lübeck, Germany

EHLERT, U., Prof. Dr., Universität Zürich, Klinische Psychologie, Zürichbergstr. 43, 8032, Zurich, Switzerland

EISENBERG, L., Prof. Dr., Harvard Medical School, Department of Social Medicine, 641 Huntington Avenue, Boston, MA 02115-6019, USA

EMRICH, H.M., Prof. Dr., Medizinische Hochschule Hannover, Psychiatrische Klinik der Universität, Carl-Neuberg-Str. 1, 30625 Hannover, Germany

ENGEL, R.R., Prof. Dr., Ludwig-Maximilians-Universität München, Psychiatrische Klinik und Poliklinik, Klinikum Innenstadt, Nußbaumstr. 7, 80336 Munich, Germany

FALKAI, P., Prof. Dr., Universität Bonn, Zentrum für Nervenheilkunde, Abteilung für Medizinische Psychologie, Sigmund-Freud-Str. 25, 53105 Bonn, Germany

FELTHOUS, A.R., MD, Chester Mental Health Center, Department of Psychiatry, P.O. Box 31, Chester, IL 62233–0031, USA

GAEBEL, W., Prof. Dr., Psychiatrische Klinik der Heinrich-Heine-Universität, Rheinische Landes- und Hochschulklinik, Bergische Landstr. 2, 40629 Düsseldorf, Germany

GARDNER, R., Prof. Dr., 214 Du Rose Terrace, Madison, WI 53705, USA

GINGRICH, J.A., MD, PhD, Columbia University, Department of Psychiatry, 1051 Riverside Drive Unit 40, New York, NY 10032, USA

GOLDENBERG, G., Prof. Dr., Städtisches Krankenhaus München-Bogenhausen, Abteilung für Neuropsychologie, Englschalkinger Str. 77, 81925 Munich, Germany

HELLHAMMER, D., Prof. Dr., Universität Trier, Fachbereich I – Psychologie, FPP, 54286 Trier, Germany

HELMCHEN, H., Prof. em. Dr., Freie Universität Berlin, Psychiatrische Klinik, Eschenallee 3, 14050 Berlin, Germany

HEMSLEY, D.R., Prof., University of London, King's College, Institute of Psychiatry at The Maudsley, De Crespigny Park, Denmark Hill, London SE5 8AF, United Kingdom

HEN, R., Columbia University, Center for Neurobiology and Behavior, P.I. Annex, 722 West 168th Street, New York, NY 10032, USA

HENDERSON, A.S., Prof. Dr., The Australian National University, Centre for Mental Health Research, Canberra, ACT 0200, Australia

HENN, F., Prof. Dr. Dr., Zentralinstitut für Seelische Gesundheit, Postfach 12 21 20, 68072 Mannheim, Germany

HEUSER, I., Prof. Dr. Dipl.-Psych., Psychiatrische Klinik am Zentralinstitut für Seelische Gesundheit, Postfach 12 21 20, 68072 Mannheim, Germany

HITZEMANN, R.J. Prof. Dr., SUNY at Stony Brook School of Medicine, Department of Psychiatry and Behavioral Science, Health Sciences Center T10, Stony Brook, NY 11794-8101, USA

JENKINS, R., Prof., WHO Collaborating Centre, Institute of Psychiatry, De Crespigny Park, Denmark Hill, London SE5 8AF, United Kingdom

KÄCHELE, H., Prof. Dr., Klinikum der Universität Ulm, Abteilung Psychotherapie und Psychosomatische Medizin, Frauensteige 14a, 89075 Ulm, Germany

KESSLER, R., Prof., Harvard Medical School, Department of Health Care Policy, 25 Shattuck Street, Boston, MA 02115, USA

KNAPP, M., Prof., Centre for the Economics of Mental Health, Institute of Psychiatry, De Crespigny Park, London SE5 8BB, United Kingdom

KRAUS, A., Prof. Dr., Psychiatrische Klinik der Ruprecht-Karls-Universität Heidelberg, Voßstr. 4, 69115 Heidelberg, Germany

KRÖBER, H.L., Prof. Dr., Freie Universität Berlin, UKBF, Institut für Forensische Psychiatrie, Limonenstr. 27, 12203 Berlin, Germany

LEAF, P., Dr., Johns Hopkins University School of Hygiene and Public Health, Department of Mental Hygiene, 624 North Broadway, Hampton House, Baltimore, MD 21205, USA

LINDEN, M., Prof. Dr., Reha-Klinik Seehof der BfA, Lichterfelder Allee 55, 14513 Teltow, Germany

LIPP, O., Dr., Hôpital Louis-H. Lafontaine, 7401 Hochelage E, Montréal, Québec H1N 3M5, Canada

MCGUFFIN, P., Prof., University of Wales College of Medicine, Division of Psychological Medicine, Heath Park, Cardiff CF14 4XN, United Kingdom

MCKINNEY, W.T., Prof. Dr., Northwestern University Medical School, The Asher Center, 303 East Chicago Avenue, Chicago, IL 60611-3008, USA

MENDLEWICZ, J., Prof. Dr. Dr., Clinical University of Brussels, Hôpital Erasme, Department of Psychiatry, 808 Route de Lennick, 1070 Brussels, Belgium

MICHELS, R., Prof., The New York Hospital-Cornell Medical Center, Department of Psychiatry, 418 East 71st Street, New York, NY 10021, USA

MÖLLER, H.J., Prof. Dr., Ludwig-Maximilians-Universität München, Psychiatrische Klinik und Poliklinik, Klinikum Innenstadt, Nußbaumstr. 7, 80336 Munich, Germany

MUNDT, C., Prof. Dr., Psychiatrische Klinik der Ruprecht-Karls-Universität Heidelberg, Voßstr. 4, 69115 Heidelberg, Germany

NABER, D., Prof. Dr., Universität Hamburg, Psychiatrische Klinik, Universitätskrankenhaus Eppendorf, Martinistr. 52, 20251 Hamburg, Germany

O'DONOVAN, M., Prof., University of Wales College of Medicine, Department of Psychological Medicine, Heath Park, Cardiff CF14 4XN, United Kingdom

PERREZ, M., Prof. Dr., Universität Fribourg, Psychologisches Institut, Rue de Faucigny 2, 1700 Fribourg, Switzerland

PLOOG, D., Prof. Dr., Max-Planck-Institut für Psychiatrie, Kraepelinstr. 2, 80804 Munich, Germany

Poitras, S., Technical Officer Policy and Legislation, Department of Mental Health, World Health Organization, 1211 Geneva 27, Switzerland

Pull, C., Dr., Centre Hospitalier de Luxembourg, Service de Psychiatrie, 4 rue Barblé, Luxembourg

Sartorius, N., Prof. Dr. Dr., Hôpitaux Universitaires de Genève, Belle-Idée, Bâtiment Salève, 2, chemin du Petit-Bel-Air, 1225 Chêne-Bourg/Genève, Switzerland

Sass, H., Prof. Dr., Medizinische Fakultät der RWTH Aachen, Klinik für Psychiatrie und Psychotherapie, Pauwelsstr. 30, 52074 Aachen, Germany

Schiefenhövel, W., Prof. Dr., Max-Planck-Institut, Forschungsstelle für Humanethologie, 82346 Andechs, Germany

Schlösser, R., Dr., Johannes Gutenberg-Universität Mainz, Psychiatrische Klinik und Poliklinik, Untere Zahlbacher Str. 8, 55131 Mainz, Germany

Schmücker, G., Klinikum der Universität Ulm, Abteilung Psychotherapie und Psychosomatische Medizin, Frauensteige 14a, 89075 Ulm, Germany

Scott, J., Prof., Gartnavel Royal Hospital, Department of Psychiatry, 1055 Great Western Road, Glasgow G12 0XH, United Kingdom

Siegrist, J., Prof. Dr., Medizinische Einrichtungen der Heinrich-Heine-Universität Düsseldorf, Institut für Medizinische Soziologie, Postfach 10 10 07, 40001 Düsseldorf, Germany

Sims, A., Prof., St. James's University Hospital, Division of Psychiatry and Behavioural Sciences, Beckett Street, Leeds LS9 7TS, United Kingdom

Souery, D., Clinical University of Brussels, Hôpital Erasme, Department of Psychiatry, 808 Route de Lennick, 1070 Brussels, Belgium

Spitzer, M., Prof. Dr. Dr., Psychiatrische Klinik der Universität Ulm, Leimgrubenweg 12–14, 89075 Ulm, Germany

Strik, W., Prof. Dr., Universitäre Psychiatrische Dienste Bern (UPD), Direktion Ost, Bolligenstr. 111, 3000 Bern, Switzerland

van Drimmelen-Krabbe, J.J., Dr., Av. E. van Becelaere 96, 1170 Brussels, Belgium

Vollmann, J., Prof. Dr. Dr., Freie Universität Berlin, Institut für Geschichte der Medizin, Klingsorstr. 119, 12203 Berlin, Germany

Walton, H.J., Prof., 38 Blacket Place, Edinburgh EH9 1RL, United Kingdom

Springer-Verlag Berlin Heidelberg GmbH

F. Henn · N. Sartorius
H. Helmchen · H. Lauter
(Editors)

Contemporary Psychiatry

Volume 2
Psychiatry in Special Situations

With Contributions by
P.B. Baltes · K. Beyreuther · H. Bickel · M. Blanchard · I.F. Brockington
E.J. Bromet · N. Cassem · M.F. Costantini-Ferrando · J. de Jong · R. Ferszt
M.M. Fichter · V. Folnegović-Šmalc · E. Fombonne · H. Förstl
H.J. Freyberger · G. Fricchione · T. Fuchs · I. Genefke · H.-J. Gertz
E.G.V. Giardina · A.H. Glassman · N. Graham · V.C. Hachinski · H. Helmchen
W. Hewer · C. Hock · A.J. Holland · J.C. Holland · A. Jablensky · W.G. Jilek
L. Jilek-Aall · S. Kanowski · M. Kastrup · N. Konrad · V. Krasnov · A. Kurz
M. Lanczik · H. Lauter · V. Lebedev · M. Maj · P. Martinez-Lage · H. Merskey
P. Monteleone · W.E. Müller · F. Müller-Spahn · F.M. Reischies · H. Remschmidt
R. Sandbrink · N. Sartorius · L.G. Schmidt · M.H. Schmidt · U.M. Staudinger
L. Tata Arcel · A. Tortorella · J. Wertheimer · E. West

Springer

Prof. Dr. Dr.
Fritz Henn
Zentralinstitut für Seelische Gesundheit
P.O. Box 12 21 20
68072 Mannheim, Germany

Prof. Dr. Dr. Dr. h.c. mult.
Norman Sartorius
Hôpitaux Universitaires de Genève
Belle-Idée, Bâtiment Salève
2, chemin du Petit-Bel-Air
1225 Chêne-Bourg/Genève, Switzerland

Prof. em. Dr.
Hanfried Helmchen
Freie Universität Berlin
Psychiatrische Klinik
Eschenallee 3
14050 Berlin, Germany

Prof. em. Dr.
Hans Lauter
Technische Universität München
Klinikum rechts der Isar
Ismaninger Straße 22
81675 München, Germany

Library of Congress Cataloging-in-Publication Data
Contemporary psychiatry / F. Henn ... [et al.], editors.
p. ; cm.
Developed on the basis of the experience obtained with: Psychiatrie der Gegenwart, that appeared only in German.
Includes bibliographical references and indexes.
Contents: v. 1. Foundations of psychiatry – v. 2. Psychiatry in special situations – v. 3. Specific psychiatric disorders.

ISBN 978-3-642-64007-0 ISBN 978-3-642-59519-6 (eBook)
DOI 10.1007/978-3-642-59519-6

1. Psychiatry. I. Henn, Fritz A. II. Psychiatrie der Gegenwart.
[DNLM: 1. Mental Disorders. 2. Psychiatry – methods. WM 140 C761 2001]
RC454.C6541 2001
616.89–dc21 00-046344

Originally published by Springer-Verlag Berlin Heidelberg New York in 2001
Softcover reprint of the hardcvoer 1st edition 2001

Cover design: de'blik, Berlin
Typesetting: Scientific Publishing Services (P) Ltd, Chennai
Printed on acid-free paper SPIN: 10693473 26/3130SM – 5 4 3 2 1 0

Preface

Contemporary Psychiatry is an international textbook of psychiatry developed on the basis of the experience obtained with its first three editions, which only appeared in German. As such, this version incorporates the German tradition of phenomenology and diagnosis going back to Kraepelin, as well as other authors such as K. Schneider or Leonhard. However, for the first time, this tradition is combined with the Anglo-American tradition and the DSM-IV diagnostic system, making it a unique resource among psychiatric texts. Each edition has had a special focus. The first edition appeared some 15 years after the end of the Second World War and aimed to present the best of German psychiatrists' writing in order to help reestablish the links between psychiatry in Germany and elsewhere that were disrupted by the war. The second edition gave particular emphasis to social psychiatry and empirical research in psychiatry in the hope that they would become more familiar to German psychiatrists and be used by them in their research and in their administrative and clinical decisions. The third edition added a focus on the integration of psychodynamic approaches and psychotherapy in the practice of psychiatry. This present, fourth edition examines psychiatric problems systematically, combining scientific evidence and modern neurobiology with experience obtained in different sociocultural settings while paying particular attention to the insights gained by the different disciplines relevant to psychiatry.

While comprehensive in its coverage, *Contemporary Psychiatry* is not encyclopedic in the sense of listing all the issues and findings of the past two decades. It presents findings relevant to the key issues in psychiatry together with the paradigms that are necessary in order to gain a conceptual understanding of the current trends in psychiatric research, service, and training. In doing so, *Contemporary Psychiatry* naturally focuses on methods and approaches that are playing an increasingly important role in psychiatry, such as molecular biology, genetics, neuroimaging, and cognitive psychology. This method of presenting today's knowledge has been chosen because it is most likely to serve as a resource not only for psychiatrists, but also for medical students, those engaged in basic research relevant to psychiatry, and clinicians from other medical disciplines who are increasingly collaborating with psychiatrists in an effort to understand and manage psychiatric problems.

Although all the volumes of *Contemporary Psychiatry* are linked, each contains separate sections. The first volume presents the scientific basis of psychiatry and the contributions that disciplines overlapping with psychiatry can make to the resolution of psychiatric problems. Epidemiology, the neurosciences, psychology, genetics, and the social sciences are given particular prominence in this section.

The second part of Vol. 1 deals with general psychiatry. It begins with an analysis of classifications in psychiatry, paying particular attention to those included in the *International Classification of Diseases* (ICD) of the World Health Organization and in the *Diagnostic and Statistical Manual* (DSM) of the American Psychiatric Association. The chapters that follow deal with the skills necessary for psychiatric examinations, principles of psychiatric treatment, the structure of mental health services and their influence on the management of psychiatric problems, legal and ethical issues, and matters relating to undergraduate and postgraduate education in psychiatry. Volume 2 addresses the presentation of psychiatric problems and their management in special situations (such as that of a refugee status), in different cultures, and in combination with other mental and physical diseases. Part 2 of this Volume systematically presents knowledge about specific psychiatric disorders, including the dementia. Schizophrenias, affective disorders, personality disorders, neurotic disorders (such as anxiety disorders, obsessive-compulsive disorders), eating disorders, suicide, substance abuse, and dependence disorders are presented in Vol. 3. The epidemiological profile, genetic and pathophysiological characteristics, clinical form and course, treatment, and rehabilitation are presented for each of the disorders.

Contemporary Psychiatry combines the German tradition of psychiatry with the traditions of psychiatry in other countries, particularly that of the English-speaking countries. Its ecumenical strategy is exemplified by its approach to the classification of mental disorders, in that both clinical consensus and the guidelines contained in ICD-10 and DSM-IV have been used in organizing the parts dealing with specific diseases. The various chapters have been written by authors from 15 different countries, and versions of *Contemporary Psychiatry* in several other languages as well as in English and German are planned to be published, reflecting the aim of producing an international textbook that is useful in different settings throughout the world.

Fall 2000

F. Henn
N. Sartorius
H. Helmchen
H. Lauter

Contents

List of Contributors

Baltes, P.B., Prof. Dr. Drs. h.c., Max-Planck-Institut
für Bildungsforschung, Lentzeallee 94, 14195 Berlin, Germany

Beyreuther, K., Prof. Dr., Ruprecht-Karls-Universität Heidelberg,
Zentrum für Molekulare Biologie, Im Neuenheimer Feld 282,
69115 Heidelberg, Germany

Bickel, H., Dr., Technische Universität München, Psychiatrische Klinik
und Poliklinik, Klinikum rechts der Isar, Ismaninger Str. 22,
81675 München, Germany

Blanchard, M., Dr., Queen Mary's, The Royal Free Hampstead NHS Trust,
Pond Street Division, 23 East Heath Road, London NW3 1DU,
United Kingdom

Brockington, I.F., Dr., The University of Birmingham,
Department of Psychiatry, Queen Elizabeth Psychiatric Hospital,
Mindelsohn Way, Birmingham B15 2Qz, United Kingdom

Bromet, E.J., Prof. Dr., New York University at Stony Brook,
Putnam Hall-South Campus, Stony Brook, NY 11794-8790, USA

Cassem, N., Prof. Dr., Massachusetts General Hospital, Harvard Medical School,
Department of Psychiatry, 75 Francis Street, Boston, MA 02115, USA

Costantini-Ferrando, M.F., Ph. D., Cornell University Medical School,
Memorial Sloan-Kettering Cancer Center, Department of Psychiatry and
Behavioral Sciences, 1275 York Avenue, New York, NY 10021, USA

de Jong, J., Dr. Dr., Transcultural Psychosocial Organization – TPO,
Keizersgracht 329, 1016 EE Amsterdam, The Netherlands

Ferszt, R., Prof. Dr., Freie Universität Berlin, Psychiatrische Klinik und
Poliklinik, Abteilung für Gerontopsychiatrie, Eschenallee 3,
14050 Berlin, Germany

Fichter, M.M., Prof. Dr., Medizinisch-Psychosomatische Klinik Roseneck,
Am Roseneck 6, 83209 Prien am Chiemsee, Germany

Folnegović-Šmalc, V., Prof. Dr., Neuropsihijatat, Psihijatrisjska Bolnica,
41090 Vrapce-Zagreb, Croatia

Fombonne, E., Dr., Kings's College London, Institute of Psychiatry at The Maudsley, Department of Child Psychiatry, De Crespigny Park, Denmark Hill, London SE5 8AF, United Kingdom

Förstl, H., Prof. Dr., Technische Universität München, Psychiatrische Klinik und Poliklinik, Klinikum rechts der Isar, Ismaninger Str. 22, 81675 München, Germany

Freyberger, H.J., Prof. Dr., Ernst-Moritz-Arndt-Universität, Klinik und Poliklinik für Psychiatrie und Psychotherapie, Rostocker Chaussee 70, 18407 Stralsund, Germany

Fricchione, G., Dr., Medical Psychiatry Service, Harvard Medical School, Division of Psychiatry, 75 Francis Street, Boston, Ma 02115, USA

Fuchs, T., PD Dr. Dr., Universitätsklinikum, Psychiatrische Klinik, Voßstr. 4, 69115 Heidelberg, Germany

Genefke, I., Dr. Dr., Rehabilitation and Research Centre for Torture Victims, Borgergade 13, 1014 Copenhagen K, Denmark

Gertz, H.-J., Prof. Dr., Psychiatrische Klinik der Universität, Liebigstr. 22, 04103 Leipzig, Germany

Giardina, E.G.V., Prof. Dr., Columbia University, College of Physicians & Surgeons, Center for Women's Health, 722 West 168th Street, New York, NY 10032, USA

Glassman, A.H., Prof. Dr., NYS Psychiatric Institute, Department of Clinical Psychopharmacology, 722 West 168th Street, New York, NY 10032, USA

Graham, N., Dr., Royal Free Hospital, School of Medicine, Academic Department of Psychiatry, Rowland Hill Street, London NW3 1DU, United Kingdom

Hachinski, V.C., Prof., University Hospital, 339Windermere Road, London, Ontario N5A 5A5, Canada

Helmchen, H., Prof. em. Dr., Freie Universität Berlin, Psychiatrische Klinik, Eschenallee 3, 14050 Berlin, Germany

Hewer, W., Dr., Zentralinstitut für Seelische Gesundheit, Postfach 12 21 20, 68072 Mannheim, Germany

Hock, C., Dr., Psychiatrische Universitätsklinik Basel, Wilhelm-Klein-Str. 27, 4095 Basel, Switzerland

Holland, A.J., Dr., University of Cambridge, Department of Psychiatry, Developmental Psychiatry Section, Douglas House, 18b Trumpington Road, Cambridge CB2 2AH, United Kingdom

Holland, J.C., M.D., Cornell University Medical School, Memorial Sloan-Kettering Cancer Center, Department of Psychiatry and Behavioral Sciences, 1275 York Avenue, New York, NY 10021, USA

Jablensky, A., Prof., The University of Western Australia, Department of Psychiatry, Medical Research Foundation Building, 50 Murray Street, Perth, WA 6001, Australia

Jilek, W.G., Prof. Dr., WPA, Transcultural Psychiatry Section, 571 English Bluff Road, Delta, British Columbia V4M 2M9, Canada

Jilek-Aall, L., Dr., WPA, Transcultural Psychiatry Section,
571 English Bluff Road, Delta, British Columbia V4M 2M9, Canada

Kanowski, S. Prof. Dr., Freie Universität Berlin,
Psychiatrische Klinik und Poliklinik, Abteilung Gerontopsychiatrie,
Eschenallee 3, 14050 Berlin, Germany

Kastrup, M., Dr., Hvidovre Hospital, Department of Psychiatry,
Broendbyostervej 160, 2650 Hvidovre, Denmark

Konrad, N., Prof. Dr., Krankenhaus der Berliner Vollzugsanstalten,
Abteilung für Psychiatrie und Psychotherapie,
Friedrich-Olbricht-Damm 17, 13627 Berlin, Germany

Krasnov, V., Prof., Russian Society of Psychiatrists, Poteshnaya, 3,
107076 Moscow, Russia

Kurz, A., Prof. Dr., Technische Universität München, Psychiatrische Klinik
und Poliklinik, Klinikum rechts der Isar, Ismaninger Straße 22,
81675 München, Germany

Lanczik, M., Dr., The University of Birmingham, Queen Elizabeth Psychiatric
Hospital, Department of Psychiatry, Mindelsohn Way, Birmingham B15 2QZ,
United Kingdom, and Psychiatrische Klinik mit Poliklinik der Universität
Erlangen-Nürnberg, Schwabachanlage 6, 91054 Erlangen, Germany

Lauter, H., Prof. em. Dr., Technische Universität München,
Klinikum rechts der Isar, Ismaninger Str. 22, 81675 München, Germany

Lebedev, V., Dr., Alturfjevskoe shosse, 95/B, apt. 399, 127572 Moscow, Russia

Maj, M., Prof., Secondo Università Napoli, Facolta di Medicina, Istituto di
Psichiatria, Lago Madonna delle Grazie, 80138 Napoli, Italy

Martinez-Lage, P., Dr., Hospital Virgen del Camino, Servicio Navarro de
Salud, Osasunbidea, Irunlarrea, 4, 31008 Pamplona, Spain

Merskey, H., Prof. Dr., LHSC-UC, 339 Windermere Road,
London, Ontario N6A 5A5, Canada

Monteleone, P., Prof., Secondo Università Napoli, Facolta di Medicina,
Istituto di Psichiatria, Lago Madonna delle Grazie, 80138 Napoli, Italy

Müller, W.E., Prof. Dr., Biozentrum Niederursel, Marie-Curie-Str. 9,
60439 Frankfurt, Germany

Müller-Spahn, F., Prof. Dr., Psychiatrische Universitätsklinik,
Wilhelm-Klein-Str. 27, 4025 Basel, Switzerland

Reischies, F.M., PD Dr., Freie Universität Berlin, Psychiatrische Klinik,
Eschenallee 3, 14050 Berlin, Germany

Remschmidt, H., Prof. Dr. Dr., Klinikum der Phillipps-Universität Marburg,
Zentrum für Nervenheilkunde, Klinik für Psychiatrie und Psychotherapie des
Kindes- und Jugendalters, Hans-Sachs-Str. 4–6, 35039 Marburg, Germany

Sandbrink, R., Dr., Schering AG, Klinische Entwicklung ZNS,
13342 Berlin, Germany

Sartorius, N., Prof. Dr. Dr. Dr. h.c. mult., Hôpitaux Universitaires
de Genève, Belle-Idée, Bâtiment Salève, 2, chemin du Petit-Bel-Air,
1225 Chêne-Bourg/Genève, Switzerland

SCHMIDT, L.G., Prof. Dr., Freie Universität Berlin, Psychiatrische Klinik, Eschenallee 3, 14050 Berlin, Germany

SCHMIDT, M.H., Prof. Dr., Zentralinstitut für Seelische Gesundheit, Kinder- und Jugendpsychiatrie, Postfach 12 21 20, 68072 Mannheim, Germany

STAUDINGER, U.M., Prof. Dr., Technische Universität Dresden, Institut für Pädagogische Psychologie und Entwickungspsychologie, Weberplatz 5, 01217 Dresden, Germany

TATA ARCEL, L., Prof., The International Rehabilitation Council for Torture Victims, Borgergade 13, 1014 Copenhagen K, Denmark

TORTORELLA, A., Prof., Secondo Università Napoli, Facolta di Medicina, Istituto di Psichiatria, Lago Madonna delle Grazie, 80138 Napoli, Italy

WERTHEIMER †, J., Prof., Hôpital Psychogériatrique, 1008 Prilly, Switzerland

WEST, E., Russian Society of Psychiatrists, Poteshnaya, 3, 107076 Moscow, Russia

Contemporary Psychiatry 3

Springer-Verlag Berlin Heidelberg GmbH

F. Henn · N. Sartorius
H. Helmchen · H. Lauter
(Editors)

Contemporary Psychiatry

Volume 3
Specific Psychiatric Disorders

With Contributions by
A. Batra · M. Bauer · H. Beckmann · M. Berger · R.J. Boland · D. Bremner
H.D. Brenner · G.W. Brown · G. Buchkremer · R. Budd · D. Chambers
D.S. Charney · P.J. Clayton · J.E. Cooper · P. Corcoran · F. Creed
J.M. Cyranowski · P.N. Dannon · S. Ecker · W.W. Fleischhacker · J.A. Fleming
J.S. Fowler · A. Francis · E. Frank · E. Franzek · T. Fuchs · W. Gaebel · J. Gibert
T.E. Goldberg · I.I. Gottesman · M. Griffiths · K.A. Halmi · A. Heinz · H. Heise
H. Helmchen · F. Henn · R.J. Hodgson · H. Hoffmann · F. Hohagen
H.H. Holcomb · I. Iancu · A. Jablensky · H.P. Kapfhammer · C.L.E. Katona
H.S. Keeley · M.J. Kelleher · M.B. Keller · H.D. Kleber · G. Kockott
A. Kopelowicz · P.B. Krabman · K.-T. Kronmüller · A.C. Lahti · M. Lawlor
J. Leff · F.R. Levin · R.P. Liberman · M. Linden · J. López-Ibor · W. Maier
K.F. Mann · R.A. Mayou · C. McAuliffe · D.R. Medoff · W.B. Mendelson
H.W. Moises · C. Mundt · M.S. Nobler · M. Rietschel · W. Rössler
H.A. Sackeim · H. Sass · Y. Sasson · H. Sauer · J.B. Saunders · L.G. Schmidt
S. Schwab · U. Schweiger · S.S. Shergill · L. Siegel · J.C. Simpson · S.P. Singh
T. Sitharthan · T.E. Smith · M. Soyka · C. Spanier · C.A. Tamminga · M.E. Thase
M.T. Tsuang · A. Uchtenhagen · E. Vermetten · U. Voderholzer · N.D. Volkow
D. von Zerssen · G.-J. Wang · T.W. Weickert · D.R. Weinberger · M. Weisbrod
S. Wessely · H.-U. Wittchen · J. Zohar · D. Zubrägel

Springer

Prof. Dr. Dr.
Fritz Henn
Zentralinstitut für Seelische Gesundheit
P.O. Box 12 21 20
68072 Mannheim, Germany

Prof. Dr. Dr. Dr. h.c. mult.
Norman Sartorius
Hôpitaux Universitaires de Genève
Belle-Idée, Bâtiment Salève
2, chemin du Petit-Bel-Air
1225 Chêne-Bourg/Genève, Switzerland

Prof. em. Dr.
Hanfried Helmchen
Freie Universität Berlin
Psychiatrische Klinik
Eschenallee 3
14050 Berlin, Germany

Prof. em. Dr.
Hans Lauter
Technische Universität München
Klinikum rechts der Isar
Ismaninger Straße 22
81675 München, Germany

Library of Congress Cataloging-in-Publication Data
Contemporary psychiatry / F. Henn ... [et al.], editors.
p. ; cm.
Developed on the basis of the experience obtained with: Psychiatrie der Gegenwart, that appeared only in German.
Includes bibliographical references and indexes.
Contents: v. 1. Foundations of psychiatry – v. 2. Psychiatry in special situations – v. 3. Specific psychiatric disorders.

ISBN 978-3-642-64007-0 ISBN 978-3-642-59519-6 (eBook)
DOI 10.1007/978-3-642-59519-6

1. Psychiatry. I. Henn, Fritz A. II. Psychiatrie der Gegenwart.
[DNLM: 1. Mental Disorders. 2. Psychiatry – methods. WM 140 C761 2001]
RC454.C6541 2001
616.89–dc21 00-046344

Originally published by Springer-Verlag Berlin Heidelberg New York in 2001
Softcover reprint of the hardcvoer 1st edition 2001

Cover design: de'blik, Berlin
Typesetting: Scientific Publishing Services (P) Ltd, Chennai
Printed on acid-free paper SPIN: 10693473 26/3130SM – 5 4 3 2 1 0

Preface

Contemporary Psychiatry is an international textbook of psychiatry developed on the basis of the experience obtained with its first three editions, which only appeared in German. As such, this version incorporates the German tradition of phenomenology and diagnosis going back to Kraepelin, as well as other authors such as K. Schneider or Leonhard. However, for the first time, this tradition is combined with the Anglo-American tradition and the DSM-IV diagnostic system, making it a unique resource among psychiatric texts. Each edition has had a special focus. The first edition appeared some 15 years after the end of the Second World War and aimed to present the best of German psychiatrists' writing in order to help reestablish the links between psychiatry in Germany and elsewhere that were disrupted by the war. The second edition gave particular emphasis to social psychiatry and empirical research in psychiatry in the hope that they would become more familiar to German psychiatrists and be used by them in their research and in their administrative and clinical decisions. The third edition added a focus on the integration of psychodynamic approaches and psychotherapy in the practice of psychiatry. This present, fourth edition examines psychiatric problems systematically, combining scientific evidence and modern neurobiology with experience obtained in different sociocultural settings while paying particular attention to the insights gained by the different disciplines relevant to psychiatry.

While comprehensive in its coverage, *Contemporary Psychiatry* is not encyclopedic in the sense of listing all the issues and findings of the past two decades. It presents findings relevant to the key issues in psychiatry together with the paradigms that are necessary in order to gain a conceptual understanding of the current trends in psychiatric research, service, and training. In doing so, *Contemporary Psychiatry* naturally focuses on methods and approaches that are playing an increasingly important role in psychiatry, such as molecular biology, genetics, neuroimaging, and cognitive psychology. This method of presenting today's knowledge has been chosen because it is most likely to serve as a resource not only for psychiatrists, but also for medical students, those engaged in basic research relevant to psychiatry, and clinicians from other medical disciplines who are increasingly collaborating with psychiatrists in an effort to understand and manage psychiatric problems.

Although all the volumes of *Contemporary Psychiatry* are linked, each contains separate sections. The first volume presents the scientific basis of psychiatry and the contributions that disciplines overlapping with psychiatry can make to the resolution of psychiatric problems. Epidemiology, the neurosciences, psychology, genetics, and the social sciences are given particular prominence in this section.

The second part of Vol. 1 deals with general psychiatry. It begins with an analysis of classifications in psychiatry, paying particular attention to those included in the *International Classification of Diseases* (ICD) of the World Health Organization and in the *Diagnostic and Statistical Manual* (DSM) of the American Psychiatric Association. The chapters that follow deal with the skills necessary for psychiatric examinations, principles of psychiatric treatment, the structure of mental health services and their influence on the management of psychiatric problems, legal and ethical issues, and matters relating to undergraduate and postgraduate education in psychiatry. Volume 2 addresses the presentation of psychiatric problems and their management in special situations (such as that of a refugee status), in different cultures, and in combination with other mental and physical diseases. Part 2 of this Volume systematically presents knowledge about specific psychiatric disorders, including the dementia. Schizophrenias, affective disorders, personality disorders, neurotic disorders (such as anxiety disorders, obsessive-compulsive disorders), eating disorders, suicide, substance abuse, and dependence disorders are presented in Vol. 3. The epidemiological profile, genetic and pathophysiological characteristics, clinical form and course, treatment, and rehabilitation are presented for each of the disorders.

Contemporary Psychiatry combines the German tradition of psychiatry with the traditions of psychiatry in other countries, particularly that of the English-speaking countries. Its ecumenical strategy is exemplified by its approach to the classification of mental disorders, in that both clinical consensus and the guidelines contained in ICD-10 and DSM-IV have been used in organizing the parts dealing with specific diseases. The various chapters have been written by authors from 15 different countries, and versions of *Contemporary Psychiatry* in several other languages as well as in English and German are planned to be published, reflecting the aim of producing an international textbook that is useful in different settings throughout the world.

Fall 2000

F. Henn
N. Sartorius
H. Helmchen
H. Lauter

Contents

List of Contributors

Batra, A., Dr., Universitätsklinik für Psychiatrie und Psychotherapie, Osianderstr. 24, 72076 Tübingen, Germany

Bauer, M., Prof. Dr. Dr., Neuropsychiatric Institute and Hospital, Department of Psychiatry and Behavioral Sciences, Mood Disorders Research Program, 300 UCLA Medical Plaza, Los Angeles, CA 90095-6968, USA

Beckmann, H., Prof. Dr., Universitäts-Nervenklinik, Psychiatrische Klinik und Poliklinik, Füchsleinstr. 15, 97080 Würzburg, Germany

Berger, M., Prof. Dr., Universitätklinik für Psychiatrie und Psychosomatik, Hauptstr. 5, 79104 Freiburg, Germany

Boland, R.J., Prof. Dr., Miriam Hospital, Department of Psychiatry, Providence, RI 02906, USA

Bremner, J.D., M.D., Yale University School of Medicine, Yale Trauma Research Program, 47 College Sreet, New Haven, CT 06519, USA

Brenner, H.D., Prof. Dr. Dr.,Universitäre Psychiatrische Dienste Bern (UPD), Laupenstr. 49, 3000 Bern 10, Switzerland

Brown, G.W., Prof. Dr., Royal Holloway, University of London, Socio-medical Research Centre, Department of Social Poilicy and Social Science, 11 Bedford Square, London WC1B 3RA, United Kingdom

Buchkremer, G., Prof. Dr., Klinik für Psychiatrie und Psychotherapie, Eberhard-Karls-Universität Tübingen, Osianderstr. 22, 72076 Tübingen, Germany

Budd, R., M.D., Cardiff Addiction Research Unit, Centre for Applied Public Health Medicine, University of Wales College of Medicine, Landsdowne Hospital, Cardiff CF1 8UL, United Kingdom

Chambers, D., M.D., National Suicide Research Foundation, 1 Perrot Avenue, College Road, Cork, Ireland

Charney, D.S., M.D., Department of Psychiatry, Yale University School of Medicine, Grace Education Building, 25 Park Street, New Haven, CT 06510, USA

Clayton, P.J., Prof. em. Dr., University of Minnesota Medical School, Department of Psychiatry, 420 Delaware Street SE, Minneapolis, MN 55455, USA

Cooper, J.E., Prof. Dr., Meadow Cottage, 25, Ireton Grove, Attenborough, Nottingham NG9 6BJ, United Kingdom

Corcoran, P., M.D., National Suicide Research Foundation, 1 Perrot Avenue, College Road, Cork, Ireland

Creed, F., Prof., Psychological Medicine, University Department of Psychiatry, Rawnsley Building, Manchester Royal Infirmary, Oxford Road, Manchester M13 9WL, United Kingdom

Cyranowski, J.M., Dr., University of Pittsburgh Medical Center, Western Psychiatric Institue and Clinic, Depression and Manic-Depression Prevention Program, 3811 O'Hara Street, Pittsburgh, PA 15213, USA

Dannon, P.N., M.D., Psychiatry Division, Sheba Medical Center, Tel Hashomer, Ramat Gan 52621, Israel

Ecker, S., Dr., Zentralinstitut für Seelische Gesundheit, J5, 68159 Mannheim, Germany

Fleischhacker, W.W., Univ.-Prof. Dr., Universitätsklinik für Psychiatrie, Abteilung für Biologische Psychiatrie, Anichstr. 35, 6020 Innsbruck, Austria

Fleming, J.A., Dr., Harvard Medical School, Department of Psychiatry, Massachusetts Mental Health Center, 74 Fenwood Road, Boston, MA 02115, USA

Fowler, J.S., M.D., Chemistry Department, Brookhaven National Laboratory, Building 555, Upton, NY 11973-5000, USA

Francis, A., Dr. Dr., SUNY Stony Brook, Health Science Center T-10, Stony Brook, New York, NY 11794, USA

Frank, E. Prof. Dr., University of Pittsburgh School of Medicine, 3811 O'Hara Street, Pittsburgh, PA 15213, USA

Franzek, E., PD Dr., Psychiatrische Klinik der Universität, Füchsleinstr. 15, 97080 Würzburg, Germany

Fuchs, T., PD Dr., Psychiatrische Klinik der Ruprecht-Karls-Universität, Voßstr. 4, 69115 Heidelberg, Germany

Gaebel, W., Prof. Dr., Psychiatrische Klinik der Heinrich-Heine-Universität, Rheinische Landes- und Hochschulklinik, Bergische Landstr. 2, 40629 Düsseldorf, Germany

Gibert, J., Prof., Department of Pharmacology, University of Cadiz, Facultad de Medicina, Plaza Fragela, s/n, 11003 Cadiz, Spain

Goldberg, T.E., Dr., National Institutes of Mental Health, 10 Center Drive, MSC 1379, Bethesda, MD 20892, USA

Gottesman, I.I., Prof. Dr., 245 Terrell Road, Charlottesville, VA 22901, USA

Griffiths, M., M.D., Cardiff Addiction Research Unit, Centre for Applied Public Health Medicine, University of Wales College of Medicine, Landsdowne Hospital, Cardiff CF1 8UL, United Kingdom

Halmi, K.A., Prof. Dr., Cornell Medical Center, Westchester Division, 21 Bloomingdale Road, White Plains, NY 10605, USA

Heinz, A., PD Dr., Zentralinstitut für Seelische Gesundheit, J5, 68159 Mannheim

Heise, H., Dr., Universitäre Psychiatrische Dienste Bern, Laupenstr. 49, 3000 Bern 10, Switzerland

Helmchen, H., Prof. em. Dr., Freie Universität Berlin, Psychiatrische Klinik, Eschenallee 4, 14050 Berlin, Germany

Henn, F., Prof. Dr. Dr., Zentralinstitut für Seelische Gesundheit, J5, 68159 Mannheim

Hodgson, R.J., Prof., Cardiff Addiction Research Unit, Centre for Applied Public Health Medicine, University of Wales College of Medicine, Landsdowne Hospital, Cardiff CF1 8UL, United Kingdom

Hoffmann, H., PD Dr., Universitäre Psychiatrische Dienste Bern (UPD), Laupenstr. 49, 3000 Bern 10, Switzerland

Hohagen, F., Prof. Dr., Klinik für Psychiatrie und Psychotherapie, Ratzeburger Allee 160, 23538 Lübeck, Germany

Holcomb, H.H., Dr., University of Maryland School of Medicine, Maryland Psychiatric Research Center, Baltimore, MD 21228, USA

Iancu †, I., M.D., Psychiatry Division, Sheba Medical Center, Tel Hashomer, Ramat Gan 52621, Israel

Jablensky, A., Prof. Dr., The University of Western Australia, Department of Psychiatry, Medical Research Foundation Building Level 3, 50 Murray Street, Perth, WA 6001, Australia

Kapfhammer, H.P., PD Dr. Dr., Psychiatrische Klinik der Ludwig-Maximilians-Universität, Nußbaumstr. 7, 80336 München, Germany

Katona, C.L.E., Prof. Dr., Institute of Psychiatry, Department of Psychological Medicine, De Crespigny Park, Denmark Hill, London SE5 8AF, United Kingdom

Keeley, H.S., M.D., National Suicide Research Foundation, 1 Perrot Avenue, College Road, Cork, Ireland

Kelleher †, M., M.D., National Suicide Research Foundation, 1 Perrot Avenue, College Road, Cork, Ireland

Keller, M.B., Dr., Butler Hospital, Sawyer Building, 345 Blackstone Boulevard, Providence, RI 02906, USA

Kleber, H.D., M.D., Columbia University, College of Physicians & Surgeons, Department of Psychiatry, 1051 Riverside Drive, New York, NY 10032, USA

Kockott, G., Prof. Dr., Psychiatrische Klinik und Poliklinik, Klinikum rechts der Isar, Ismaninger Str. 22, 81675 München, Germany

Kopelowicz, A., Dr., San Fernando Mental Health Clinic, 15535 San Fernando Mission Boulevard, Mission Hills, CA 91345, USA

Krabman, P.B., M.D., Western Sydney Drug and Alcohol Service, North Parramatta, NSW 2151, Australia

Kronmüller, K.-T., Dr., Psychiatrische Klinik der Ruprecht-Karls-Universität, Voßstr. 4, 69115 Heidelberg, Germany

Lahti, A.C., Dr., University of Maryland School of Medicine, Maryland Psychiatric Research Center, Baltimore, MD 21228, USA

Lawlor, M., M.D., National Suicide Research Foundation, 1 Perrot Avenue, College Road, Cork, Ireland

Leff, J., Prof., Medical Research Council, Social, Genetic and Developmental Psychiatry Research Centre, Social Psychiatry Section, Institute of Psychiatry, De Crespigny Park, London SE5 8AF, United Kingdom

Levin, F.R., M.D., Division of Substance Abuse, College of Physicians & Surgeons of Columbia University, 722 West 168th Street, New York, NY 10032, USA

Liberman, R.P., Prof. Dr., Clinical Research Center, West Los Angeles VA Medical Center, 11301 Wilshire Boulevard, Los Angeles, CA 90073, USA

Linden, M., Prof. Dr., Reha-Klinik Seehof der BfA, Lichterfelder Allee 55, 14513 Teltow, Germany

López-Ibor, J.J., Prof., Clinica López-Ibor, C. Nueva Zelanda, 44, 28035 Madrid, Spain

Maier, W., Prof. Dr., Rheinische Friedrich-Wilhelms-Universität, Klinik und Poliklinik für Psychiatrie und Psychotherapie, Sigmund-Freud-Str. 25, 53127 Bonn, Germany

Mann, K.F., Prof. Dr., Klinik für Abhängiges Verhalten und Suchtmedizin am Zentralinstitut für Seelische Gesundheit, J5, 68159 Mannheim, Germany

Mayou, R.A., Prof., University of Oxford, Department of Psychiatry, Warneford Hospital, Oxford OX3 7JX, United Kingdom

McAuliffe, M., M.D., National Suicide Research Foundation, 1 Perrot Avenue, College Road, Cork, Ireland

Medoff, D.R., Dr., University of Maryland School of Medicine, Maryland Psychiatric Research Center, Baltimore, MD 21228, USA

Mendelson, W.B., Prof. Dr., Sleep Research Laboratory, Department of Psychiatry, University of Chicago, 5841 S. Maryland Ave., Chicago, IL 60637, USA

Moises, H.W., Prof. Dr., Klinikum der Christian-Albrechts-Universität, Klinik für Psychiatrie, Niemannsweg 147, 24105 Kiel, Germany

Mundt, C., Prof. Dr., Psychiatrische Klinik der Ruprecht-Karls-Universität, Voßstr. 4, 69115 Heidelberg, Germany

Nobler, M.S., Prof. Dr., New York State Psychiatric Institute, Department of Biological Psychiatry, 1051 Riverside Drive, Unit 126, New York, NY 10032, USA

Rietschel, M., Dr., Rheinische Friedrich-Wilhelms-Universität, Klinik und Poliklinik für Psychiatrie und Psychotherapie, Sigmund-Freud-Str. 25, 53105 Bonn, Germany

Rössler, W., Prof. Dr., Psychiatrische Universitätsklinik, Militärstr. 8, 8021 Zürich, Switzerland

Sackeim, H.A., Prof. Dr., New York State Psychiatric Institute, Department of Biological Psychiatry, 1051 Riverside Drive, Unit 126, New York, NY 10032, USA

Sass, H., Prof. Dr., Klinik für Psychiatrie und Psychotherapie, Medizinische Fakultät der RWTH Aachen, Pauwelsstr. 30, 52074 Aachen, Germany

Sasson, Y., M.D., Psychiatry Division, Sheba Medical Center, Tel Hashomer, Ramat Gan 52621, Israel

Sauer, H., Prof. Dr., Klinikum der FSU, Klinik für Psychiatrie, 07740 Jena, Germany

Saunders, J.B., M.D., Department of Psychiatry, The University of Queensland, Royal Brisbane Hospital, Herston, QLD 4029, Australia

Schmidt, L.G., Prof. Dr., Psychiatrische Klinik und Poliklinik der Freien Universität Berlin, Eschenallee 3, 14050 Berlin, Germany

Schwab, S., Dr., Rheinische Friedrich-Wilhelms-Universität, Klinik und Poliklinik für Psychiatrie und Psychotherapie, Sigmund-Freud-Str. 25, 53105 Bonn, Germany

Schweiger, U., Dr., Klinik für Psychiatrie und Psychotherapie, Ratzeburger Allee 160, 23538 Lübeck, Germany

Shergill, S.S., Dr. Dr., Institute of Psychiatry, Department of Psychological Medicine, De Crespigny Park, Denmark Hill, London SE5 8AF, United Kingdom

Siegel, L., Dr., University of Pittsburgh School of Medicine, 3811 O'Hara Street, Pittsburgh, PA 15213, USA

Simpson, J.C., Dr., Harvard Department of Psychiatry, Community Support Services, Brockton/West Roxbury VA Medical Center, Massachusetts Mental Health Center, 74 Fenwood Road, Boston, MA 02115, USA

Singh, S.P., Dr., University Hospital, Queens Medical Centre, B Floor South Block, Nottingham NG7 2UH, United Kingdom

Sitharthan, T., M.D., Centre for Drug and Alcohol Studies, Department of Psychological Medicine, University of Sydney, Sydney, NSW 2006, Australia

Smith, T.E., Dr., White Plains, 21 Bloomingdale Road, New York, NY 10605, USA

Soyka, M., PD Dr., Psychiatrische Klinik der Ludwig-Maximilians-Universität, Nußbaumstr. 7, 80336 München, Germany

Spanier, C., Dr., University of Pittsburgh School of Medicine, 3811 O'Hara Street, Pittsburgh, PA 15213, USA

Tamminga, C.A., Dr. University of Maryland, Maryland Psychiatric Research Center, Department of Psychiatry, PO Box 21247, Baltimore, MD 21228, USA

Thase, M.E., Dr., University of Pittsburgh School of Medicine, 3811 O'Hara Street, Pittsburgh, PA 15213, USA

Tsuang, M.T., Dr. Dr., Massachusetts Mental Health Center, 74 Fenwood Road, Boston, MA 02115, USA

Uchtenhagen, A., Prof. Dr. Dr., Institut für Suchtforschung, Konradstr. 32, 8005 Zürich, Switzerland

Vermetten, E., M.D., Yale University School of Medicine, Yale Trauma Research Program, 47 College Street, New Haven, CT 06519, USA

Voderholzer, U., Dr., Albrecht-Ludwigs-Universität, Abteilung für Psychiatrie und Psychotherapie, Hauptstr. 5, 79104 Freiburg, Germany

Volkow, N.D., M.D., Chemistry Department, Brookhaven National Laboratory, Building 555, Upton, NY 11973-5000, USA

von Zerssen, D., Prof. Dr., Max-Planck-Institut für Psychiatrie, Kraepelinstr. 2, 80804 München, Germany

Wang, G.-J., M.D., Chemistry Department, Brookhaven National Laboratory, Building 555, Upton, NY 11973-5000, USA

Weickert, T.W., Ph.D., Clinical Brain Disorders Branch, National Institutes of Mental Health, National Institutes of Health, Building 10/Room 4C 215, Bethesda, MD 20892, USA

Weinberger, D.R., Prof. Dr., Clinical Brain Disorder Branch, DIRP, NIMH, Neurosciences Center at St. Elizabeth's Hospital, Washington, DC 20032, USA

Weisbrod, M., Dr., Psychiatrische Klinik der Ruprecht-Karls-Universität, Voßstr. 4, 69115 Heidelberg, Germany

Wessely, S., Prof., Guy's, King's and St Thomas' School of Medicine, Department of Psychological Medicine, 103 Denmark Hill, London SE5 8AZ, United Kingdom

Wittchen, H.-U., Prof. Dr., Max-Planck-Institut für Psychiatrie, Klinisches Institut, AG Klinische Psychologie, Kraepelinstr. 2–10, 80804 München, Germany

Zohar, J., Prof., Psychiatry Division, Sheba Medical Center, Tel Hashomer, Ramat Gan 52621, Israel

Zubrägel, D., Dipl.-Psych., Forschungsgruppe Ambulante Therapie, Psychiatrische Klinik und Poliklinik der Freien Universität Berlin, Eschenallee 3, 14050 Berlin, Germany

Part 1
Scientific Basis of Psychiatry

CHAPTER

1

C. Mundt, M. Spitzer

Psychopathology Today

Translator: S. Henn

1 Introductory Comment: Psychology in Transition

The following exposition continues the one presented in the 1989 edition of *Contemporary Psychiatry*. The predominant theme of that chapter was methodological dualism, which had thrown psychopathological research into disorder, eventually changing and diversifying it. The model of gradual changes proposed at the time has been taken up only occasionally (Frommer 1996). Instead, the prevalent tendency in the development of the field has been the diversification of research paradigms, clearly dominated by the fundamental principles of logical empiricism. Modern classification manuals have made a substantial contribution to this. Our analysis of the current situation and its inherent developmental trends will not repeat that debate on fundamental principles and the systematic examination of specialized psychopathology linked to it. Instead, it will turn directly to subsequent innovations and evaluate them from the perspective of the methodological problems described.

The goal and strategies of present research efforts in psychopathology are particularly evident in the results and limitations seen in two areas: classification research and experimental approaches that include imaging methods. These will therefore be dealt with in more detail.

2 Classification: Operational Diagnostic Systems

Although a separate chapter in this edition of *Contemporary Psychiatry* is devoted to classification (see Vol. 1, Part 2, Chap. 2), the present discussion of psychopathology also requires a critical look at this area of research.

The introduction to ICD-10 and DSM-IV and their present implementation in hospitals and research have revived the highly controversial discussion about operational diagnostic systems in the German-speaking psychiatric community (see Gross et al. 1997).

2.1 Anticipated Consequences – Reliability

As early as 1974, Spitzer and Fleiss proposed validity and reliability as fundamental research goals of psychopathological classification research. Reliability does not guarantee validity, but limits it. The association of this research goal with Kraeplin's medical model of mental illness (see Avenarius 1979) and with the reception of Jasper's descriptive psychopathology, restricted to symptom inventories, led to the desired abandonment of ideal types, perceived as integrated essences, in favor of mosaics composed of single symptoms, the combination of which could then be examined quantitatively and independently of all nosological, etiological, and pathogenetic biases. New findings and viewpoints regarding nosological categories had already been worked out in new versions of DSM and ICD, but the loss of old concepts, in part highly criticized, had to be accepted. As Mombour et al. (1990) stated, concentration on descriptive outcomes to the detriment of conceptions has been accepted in field studies; however, that was all that could be expected from those researchers who were motivated to participate in them. The authors of several reliability studies (Mombour et al. 1990; Freyberger et al. 1990; Sartorius et al. 1993; Dilling and Dittmann 1990; Mezzich 1992) found that a high degree of correspondence could be established between investigators and centers ($\kappa > 0.80$), except for in certain personality disorders, which only reached a κ value of 0.52 in the WHO study comparing different centers. Sartorius believes that a brief acquaintance with the new manual suffices to reach a high level of correspondence. Furthermore, ICD-10 is appropriate for diagnostic practices in European and non-European countries alike, especially those in the Third World, to which the ICD feels particularly committed. On the other hand, Sartorius' statement contradicts the fundamental criticism of the instruments which is increasingly present in the literature.

The reliability achieved in other field studies is far lower. Freyberger notes critically that values for personality disorders, Alzheimer's dementia, drug addiction, and depression are unsatisfactory. For a moderately severe depressive episode, a κ value of only 0.17 was reached among 134 hospitals and 10 centers in Germany. In addition, Freyberger points out that splitting up the category of dissociative disorders has caused particular difficulties in classifying corresponding phenomena.

The goal of improving the reliability of classification manuals has led a number of groups to work on developing a "family of instruments." These include symptom checklists, systematic self-assessment and objective assessment instruments with detailed instructions for use, and, finally, standardized interviews, which stipulate the diagnostic interview to be carried out, including the wording of questions. Thus, for example, the Schedules for Clinical Assessment in Neuropsychiatry (SCAN), which evolved from the Present State Examination (PSE) developed by Wing et al. (1990), was tested by Maurer et al. (1991). SCAN also contains instructions for analysis and evaluation which are derived from examined symptoms for ICD

and DSM diagnostic categories. A "modular system" is planned, which is intended to be polydiagnostic and to differentiate between data sources. As in many reports on reliability studies, the work by Maurer et al. (1991) is simply a critique inherent to the system: SCAN is still too long, the terse remarks for normal, or negative, findings are inadequate, and the statement should differentiate between question and commentary. Other things are also not considered, e.g., the effect of contact with the patient through a procedure focused on diagnosis rather than primarily upon therapy, such as concerns spontaneously presented by patients. Also principally excluded are problems of structuring, which are very important in the free interplay of hermeneutic acts between patient and therapist in psychotherapeutic diagnosis (see also Saß 1987: "Problems with data source remain unconsidered"). Personal experience with those that use standardized interviews where variable reliability was obtained shows that, even despite standardization and repeated questions, additional intuitive perceptions rank high and are included in the final diagnosis.

Another interview for assessing personality disorders, developed and preferred by WHO, is the International Personality Disorder Examination (IPDE). Dilling has added to this the Composite International Diagnostic Interview (CIDI). SCAN and CIDI have the disadvantage of much longer assessment times, a minimum of 1 h for normal results, otherwise 3 h or more. A lexicon explaining the psychopathological terms used in DSM and ICD and thus making a contribution to their use as standards is being prepared as a provisional conclusion in the "family of instruments."

Just as guaranteeing reliability is especially important in the spread and applicability of the instruments, so is feasibility. In this connection, Dilling and Dittmann (1990) point out that the new numbering system has room for 1000 classifications, about two thirds of which are now in use. Thus additional entries would not require new editions. Moreover, the approach to the DSM structure in view of expanded dementia concepts and the abandonment of qualitative elements of endogenous depression and some other older psychopathological concepts is viewed as providing "improved feasibility." Thus a high priority is given to reliability – especially comparability to the DSM-based diagnostic studies – while a discussion of reliability is more likely avoided by increasing categories such as "difficulties with the boss" on the ICD axis of psychosocial factors. Such classification requires some getting used to if we consider the highly theoretical level that had already been achieved in the conceptualization of psychosocial causal factors in German-speaking psychopathology (Kisker 1960; Tellenbach 1983). The question remains as to what knowledge is gained from empirical research into the frequency distribution of such categories over categories in axis I if the level of abstraction of a disorder cannot be conceptualized as a basic disorder. This was why stealing and kleptomania, for example, were not granted the status of a diagnosis (see Mundt 1986); they could, for example, be categorized as an expression of the fundamental disorder "frontal lobe disinhibition" or as an expression of a conflict situation in neurotic disorders, at best as a symptom. The interpretation necessary for differentiation was sacrificed to safeguard reliability.

2.2 Conceptual Changes – Problem of Validity

Dilling and Dittmann (1990) have worked out the conceptual changes in comparison with the traditional diagnostic concepts, which were largely shaped by authors of the Heidelberg School: the division into psychoses and neuroses was abandoned in order to free research from the etiological, pathogenetic, and nosological implications of these categories. The term "disorder" was used throughout, an inconsistency as DSM has been based on Kraepelin's medical model since the third edition (see Saß et al. 1994), so that it could be assumed from this that the concept of illness now more likely refers again to the biopsychosocial model of Adolf Meyer, which shaped DSM-I and DSM-II (Meyer 1960). In addition, this concept is now at the mercy of professional political battles between psychologists and psychiatrists. Baumann (1996) sees these terms in the manual as a "plan for distributing physicians' tasks," while Fydrich et al. (1996, p. 3) see the term "comorbidity" as a "reintroduction of the medical model through the backdoor." This results in a nominalistic overvaluation of terminology which obstructs research and serves professional political disputes.

The qualitative differentiation of depression categories into neurotic–endogenous or reactive–endogenous was abandoned in favor of a quantitative gradation. However, the assignment of degrees of severity has proven to be difficult (Dilling and Dittmann 1990). The severe form corresponds best in ICD to the earlier endogenous depression, and in DSM-IV to major depression with melancholic features; in this formulation, confusion with Tellenbach's melancholy type (1983) was taken into account in translation. Dysthymia corresponds best to the earlier neurotic depression, though enriched by the time criterion, but Saß (1987, 1994) considers this incorrect in relation to the corresponding DSM-IV category. Tendencies can be seen here of the various psychopathological schools incorporating their own traditions into the new system and not abandoning them.

The division of organic mental disorders into acute versus chronic and reversible versus irreversible disorders was also abandoned in favor of a purely descriptive procedure. To be consistent, ICD followed DSM in the broader dementia concept when the traditional criteria of irreversibility and lack of insight into illness no longer apply. Wieck's conception of the transitional syndrome (Wieck 1961), which was always disputed and was based upon clouded consciousness and reversibility, also no longer applies. There are now corresponding difficulties with the reliability of dementia diagnosis (Freyberger et al. 1990). As a result, special surveys are relied upon for research purposes (e.g., NADCA, AMCA; McKhann et al. 1984).

The corresponding DSM-IV chapter has also abandoned a solely descriptive approach and there is considerable confusion of etiological organizing principles, with the effect that neither of the two principles is really adhered to. This is related to the "loose" association of etiology and psychopathological appearance, to the generalization of exogenous reaction types according to Bonhoeffer (1910) and to their joint expression with nonorganic or mildly organic disorders, e.g., the earlier neuroses and psychopathies.

The schizophrenia chapter also now includes schizotypal and delusional disorders, which were formerly classified as psychopathic or personality disorders. New additions are undifferentiated schizophrenia and postschizophrenic depression. Dilling and Dittmann (1990) regret that in DSM-IV there is a difference in the time criterion of 1–6 months.

The ICD-10 chapter on "Neurotic, stress and somatoform disorders" also does not conform to the philosophy of the modern classification system insofar as it mixes etiological and descriptive elements to define the syndrome. Similarly, eating and sexual disorders are subsumed in the chapter on "Psychological and hormonal disorders."

Missing among the personality disorders is the old cyclothymic or thymopathic personality disorder, which now appears as cyclothymia under affective disorders. The accentuated personality was also omitted as a result of international advice, probably because of its stigmatizing character and difficult transcultural standardization. Newly included categories are pathological stealing, arson, and gambling, as well as artificial disorder. They were included in spite of unfavorable experiences in defining psychopathological syndromes for criteria of social behavior, e.g., the antisocial personality.

As symptomatic of the priority of reliability over validity, Dilling and Dittmann (1990) noted critically that, in spite of an exhaustive and accurate clinical description of the degree of mental retardation, the intelligence quotient is used as a definitive criterion, although this measure has been repeatedly questioned, even by psychologists.

2.3 Comorbidity Research

A major research goal, pursued by modern classification systems, was comorbidity research, which primarily separates symptom and personality levels, but also separates the symptom, or personality level on the one hand from psychosocial factors on the other. This led to the revival of the trauma theory, e.g., by discovering the increasing frequency of dissociative disorders following severe traumas. The more reliable diagnosis of personality disorders has resulted in a frequency of 10%–12% in nonclinical populations, about 27% in somatoform disorders and between 40% and 50% in anxiety and affective disorders (Fydrich et al. 1996), whereby very different clusters can be shown. Cluster C, the dependent, avoiding personality disorder, is most frequent in severe affective disorders, while cluster A, comprising the schizoid, schizothymic, and paranoid personality disorders, is least frequent. The fact that less than 30% of obsessive–compulsive syndromes (Fiedler 1995) are associated with compulsive personality disorder and, similarly, the association of dissociative disorders with histrionic personality disorders led to a differentiated view of their etiology. Studies on post-traumatic stress reactions (Fiedler 1995) have shown that dissociative phenomena are frequent, prognostically unfavorable consequences of a trauma without there necessarily having previously been a histrionic primary personality. In view of the abandonment of qualitative differentiation among depressions, a revival of this differentiation could develop from personality typology. A largely inconspicuous primary personality is mentioned in DSM-IV, namely, major depression with melancholy. This criterion had already been referred to by Weitbrecht (1966) and Schneider (1976) as a criterion for endogenicity in which, apart from signs of a weaker form, there is an autonomy of adaptation in relationships to self and the world that is very difficult to grasp.

The theoretical and strategic problem of modern classification systems is seen here quite clearly: a consistent restriction of the definition of depression to descriptive criteria also makes this diagnosis applicable to opiate addicts and cocaine users, for example. Thus the comorbidity of cocaine dependency and depression is the topic of a flood of publications in the United States. Comorbidity of depression and schizophrenia is also theoretically possible. However, inconsistently in the latter case, a separate category of post-schizophrenic depression has been introduced. The extensive discontinuation of principles of structuring, such as Jaspers' earlier stratification rule, which provided for a hierarchy according to the real severity of an illness, allows a certain disorder in category

formation to emerge to establish a new order, which is empirically better grounded but which is not, however, without assumptions. Fydrich et al. (1996) therefore advocate abandoning the personality disorder category and favor a dimensional-quantitative register of personality features with limits to the characteristic form of the groups, such that borderline cases between those personalities still socially acceptable and those recognized as pathological are included. They would then be defined solely according to categories of levels of symptoms, particularly interaction characteristics, and not according to simultaneous medical criteria, as was the case with the former concept of pseudo-psychopathy.

2.4 Critical Summary

The fact that modern systems of classification have made a fundamental contribution to the ability to compare epidemiological findings in therapy, research, administration, and quality control is indeed positive. New knowledge, gained through the philosophy of separate diagnostic axes, primarily affects the relativization of the connection of personality symptoms in former neuroses, the revival of the significance of trauma in research on post-traumatic stress disorder (PTSD), and a destigmatization of socially burdensome diagnoses. Although the operational classification systems are fundamental to an empirically meaningful and effective psychopathology and are needed in the future, critical objections prevail. They become more relevant against the background of expanded claims for the manuals.

Saß (1987, 1994) voiced the basic reservation that modern classification manuals represent a means by which one branch of psychiatry, psychopathology, is used in the service of another branch, biological psychiatry, with the consequent loss of a "pure psychopathology" (Janzarik 1988), i.e., organized thinking about structural and functional connections of the inner life.

A limitation in the differentiation of signs and symptoms occurs when reliability takes precedence over validity (see Spitzer and Fleiss 1974), and this has led in part to retrogression compared to the differentiation achieved earlier. Since behavioral and expressive symptoms can be classified more reliably than those relating to experience, an imbalance arises favoring behavioral symptoms and the corresponding methodology of objective procedures over the ideographically formed psychopathological tradition, which was more concerned with mental inner worlds. None of this would be a disaster, were it not for the inherent dynamics of the expanding claims for the validity of the manuals arising from practical considerations (see Blashfield and Fuller 1996).

The discrete character of many mental phenomena leads to a blurring of definitions of the terms (see Birley 1990), which tried to do justice to concepts such as that of ideal types (see Kraus 1991). This limited precision in defining terms should not lead to empty rituals of perfection and methodological overloading of procedures.

The neglect of conceptions derived from clinical evidence leads to an impoverishment of concepts and through editorial policy excludes one branch of psychopathology.

The "pseudo-lack of prerequisites" of logical empiricism has been repeatedly criticized, most recently by Birley (1990), who criticizes the manuals as being "essentialist scholastic wolves in the sheep's clothing of nominalists." He advocates a minimum of definition which would allow for a maximum of freedom of thought, whereby definitions would only be used for specific purposes. In view of the fact that before Kraeplin and after 100 years of the development of confusing nosological systems one was accustomed to regarding what was simple as being correct, the current increase in categories in the manuals makes one doubt advances in validation.

The manuals do not deal with the interests of psychotherapy, which requires a completely different point of view for its classification (see Mundt 1997). There is, however, a fundamental debate here among behavioral therapists in which one viewpoint with "disorder-specific", largely deindividualized techniques is given priority and the other is represented by an individual therapeutic approach oriented toward personality and biography rather than to nosology (for the controversy between Fiedler and Grawe, see Fiedler 1997). This discussion is taking place among behavioral therapists. From the outset psychoanalysts have followed and continue to follow their own diagnostic paths (see Janssen and Schneider 1994; Rudolf 1993).

With some innovations there is emotional resistance from the older generation to internalizing new ideas, not because they are diffident about accepting something new, but out of a lack of conviction and the feeling that something correct is being exchanged for something questionable.

The objections cited become more relevant due to the tendency of the manuals to claim generality and move toward radicalism. In spite of initial conciliations that the manuals were concerned with research criteria and contributed to administratively useful categories, it was first concealed, but then expressly stated, that they were intended as textbooks. The establishment of a "family of instruments" together with a textbook character, material which is the subject of examination, and editorial politics leads to a centralization and

uniformity of perception and thinking which did not previously exist.

It would therefore be desirable to relativize, though not to abolish, the system. A reappraisal of clinical evidence with a consensus of experts, an examination of quality criteria for psychopathological research as influenced by the humanities, and a revival of psychopathological research on concepts and structurally functional contexts should achieve a reversal of the instrumentalization of psychopathology through this recommendation for changeover steps: test the hypotheses by temporary blindness to meaning and return to concepts of ideal types.

3 Special Developments and Innovations in Classification Research

The studies by Klosterkötter (1988) in Huber's group of the transition from basic symptoms to fully developed schizophrenic-psychotic illness had already produced proof that the postulated functional connection exists. The process can be interpreted well using a vulnerability stress model or an intentionality model. Three stages in the development of basic symptoms were differentiated: the initial phase of nonspecific basic phenomena, the externalization phase leading to initial alienation and an "as if" character in acoustic perceptions, and, finally, the concretization phase in which there are increasing characteristic facial expressions with catathymic contexts. This differentiated result has great significance for prevention, since other basic symptoms are found more frequently in personality disorders and neurotic illnesses, especially in healthy adolescents and young adults, such as physiological instability and signs of alienation in developmental phases in which identity formation and role discovery are dealt with to a certain extent. In this respect, Klosterkötter et al. (1997) was able to show a broad predictive specificity for second-degree symptoms which already had a delusion-like and "as if" externalized character.

Good confirmation of Huber's long-term results were also shown by Marneros et al. (1991) and Häfner et al. (1992a). In a retrospective analysis of the long-term course of schizophrenics, Marneros et al. (1991) were able to compare course and outcome directly by including schizoaffective and affective psychoses. The tendency, already mentioned in the literature, was particularly clear in the slightly more favorable course for schizoaffective psychoses and still more favorable course for affective psychoses. Nevertheless, the familiar splitting into three courses also occurs in these groups, albeit at a generally better level. The primary personality is the chief predictor for all three groups. The study by Marneros et al. (1991) yields a wealth of psychopathological detail. However, to a certain extent it lacks basic symptomatological parameters of the syndrome, such as symptom linkages, number and duration of phases, as well as social aspects of course and outcome.

Häfner was able to confirm the findings clinically gathered by the Huber group with the survey instrument IRAOS (an instrument to assess the onset and early course of schizophrenia) developed by his group (Häfner et al. 1992b), in which he established that nearly all the patients who later developed schizophrenia reported such warning symptoms. He analyzed his findings to test different psychopathological concepts, e.g., the positive–negative dichotomy of Crow and Andreasen (see Häfner and Maurer 1991) and Conrad's stages model (see Hambrecht and Häfner 1993). His findings do not confirm the positive–negative model, since the precursor syndromes of his cohorts consist essentially of symptoms which can be judged as negative symptoms and which continually increase over a period of about 4 years until the first manifestation of positive symptoms, after which both positive and negative symptoms decrease. This synchronous rather than alternating pattern, or even the appearance of one or the other of the two syndromes, also described by other authors, speaks against the dichotomy. It is in this connection that the controversy about primary and secondary negative symptoms should be mentioned. Carpenter and colleagues (Carpenter et al. 1988; Kirkpatrick et al. 1989) had first been on the lookout for qualitative distinguishing features for this greater differentiation, which could not, however, be empirically validated (see Mundt et al. 1995; Barnett et al. 1996). The group introduced a time criterion that arbitrarily prescribed that negative symptoms which are long-lasting, appear at least 1 year before positive symptoms, and are not caused by medication, depression, psychotic anxiety, mistrust, or mental deficiency should be regarded as primary. This was the case for patients in the Häfner group, and even so the negative symptoms were in synchrony and did not alternate with positive symptoms.

In a study by Mundt et al. (1995) on the course of negative symptoms in a sample of schizophrenic and neurotic patients, comparable to those in Häfner's sample in which the group was between one to three hospitalizations and beginning to develop the illness, no greater differentiation into primary and secondary negative symptoms could be made in comparison with the neurotic patients. The controversy over interpretation of results (see also Amminger and Kirkpatrick 1997; Barnett et al. 1997) had placed the greatest focus of attention on special features of the samples. The characteristic affective rigidity found in many

schizophrenic patients in the postacute stage, in terms of the dynamic-structural model (dynamic constriction, as it is called by Kick 1991), indicates that disorders are more likely to be affective rather than structural-cognitive ones. In this respect, Kick's concept presents an interesting alternative to the cognitive-structural negative syndrome proposed by Crow (1980). According to Carpenter's definition of primary negative symptoms (Carpenter et al. 1988), the establishment in the group studied by Häfner et al. (1992a) of primary negative symptoms according to the time criterion would certainly be something to affirm. However, the short- to medium-term course in these patients does not show the expected positive-negative dichotomy. With Kick's dynamic constriction, it could be agreed with Carpenter that the formation of mental reaction in response to a high pressure of anxiety and processes of disintegration may possibly play a role in the development of negative symptoms, which would also be conceivable without the manifestation of positive symptoms. In this respect, Carpenter's definition of the time criterion does not fit with the absence of positive symptoms. In later publications, the Carpenter group (Carpenter et al. 1988; Kirkpatrick et al. 1989) incorporated the former qualitative-descriptive concept of the primary negative symptoms into an operationally defined deficit syndrome by introducing the time criterion. Thus Crow's syndrome of negative schizophrenia is reestablished, including the biological findings reported by the group.

Dynamic constriction also cannot be equated with the potentially composed posture of intentionality in apathy syndrome (Mundt 1985), which corresponds more to Carpenter's deficit syndrome with continuous negative symptoms and without chronic psychosis. Instead, the condition would be characterized by a tenseness and stiffening of intentional actions without a long-term effective level of relief being found or being necessary.

Late-onset schizophrenic and affective illnesses have recently received separate psychopathological consideration. Riecher-Rössler (1995) was able to show, on the basis of a large German and Danish group of representative patients, that late-onset schizophrenics do not differ in any of the major variables from early-onset patients. Psychopathological and sociodemographic variables, characteristics of the course of the illness, and biological variables suggest a continuum. Marneros, Jeste (1997), and other authors (see Marneros 1997) come to similar conclusions. An explanatory model for the high variance in age at onset of illness that is cited by most authors, particularly in the best methodological study by Riecher-Rössler (1995), is the vulnerability model, although it is still in an unsatisfactorily simplified form. The vulnerability concept is unsatisfactory for first-onset patients in that two aspects which are clinically indicated and were addressed in earlier studies of the psychopathology of first onset (see Kisker 1960) are not considered: the requirement that patients actively explore their world and develop in adolescence and early adulthood, and the distinctive interactions of the structural and dynamic constituents of personality (Janzarik 1988). Neither aspect is considered in the vulnerability model, which conceives of the individual as passively suffering. The models of structural dynamics and intentionality could enhance the vulnerability hypothesis in these points. There is good reason to believe that late-onset patients have more stable cognitive and affective structural patterns.

Once involution depression had been differentiated from affective illnesses in the interim between ICD-8 and ICD-9, the psychopathological discussion about the classification of late-onset affective disorders was focused, among other things, on the role of pseudodementia and the general interplay between factors causing depression and those causing biological dementia. A very extensive literature has now been established substantiating that late-onset depression with pseudodementia (see Morris and Rapoport 1990) carries an increased risk of actual development of cognitive deficits at a later stage; conversely, at least 30% of patients in early stages of dementia in whom no manifest disorders in self-perception and judgment have yet appeared develop depressive symptoms. A high proportion of late-onset affective disorder patients show corresponding brain morphology results.

In classification research with dementia patients, a greater differentiation in memory models has developed in a transitional area between psychopathology and neuropsychology which is also partially applicable in the theory of psychoses (see J. Schröder 1997). The differentiation developed by Squire (1986), Squire and Zola-Morgan (1991), and others between declarative and procedural memory along with the cerebral structures involved is now well established. In studies on the course of dementia using magnetic resonance imaging (MRI), Pantel et al. (1997) were able to show that the clinical psychopathological symptomatology, based essentially on disorders of declarative-explicative memory, is related in the early stage more to an atrophy of the hippocampus-amygdala complex than to an atrophy of the total area of the associative cortex. According to studies by Markowitsch (1995), procedural memory seems on the other hand to be connected to basal ganglia structures and to the cerebellum. Utilization of these findings and improved neuropsychological research methods for therapeutic training programs are still in the developmental stage.

A functional differentiation of dementia diagnosis using combined visual-spatial and semantic tests was also substantiated by Monsch et al. (1994). They were

able to show that patients with subcortical atrophy of the Binswanger type probably had disturbed access to semantic contents, though the contents themselves were not lost. With advanced cortical dementia of the Alzheimer's type, however, access is possible but the semantic contents of declarative memory are lost.

Schröder (see J. Schröder 1997) used the differentiation between declarative, procedural, and working memory as short or ultrashort storage along with further validation of psychopathological subtypes of schizophrenia as a bridge-building model to interpret patterns of cerebral activation seen in positron emission tomography (PET) and single photon emission computed tomography (SPECT). The three-factor solution, also replicated by him, of clustering of psychopathological symptoms of schizophrenics into a disorganized, a paranoid-hallucinatory, and an asthenic-apathetic subgroup corresponded to prefrontal hyperactivation, left hippocampal hypoactivation, and global hypofrontality. With disorders of mainly working, declarative, and procedural memory assigned to these subgroups, a neuropsychological model was suggested that is more concrete, better able to be tested, and less speculative in that it is closer to the results than the constructs of Liddle and Morris (1991) or Emrich (1990) of a disturbed central censor function for evaluating reality as the common "schizophrenia" of the three subtypes. The debate on subtypes continues to flare up on occasion. In addition to the most frequently replicated three-factor solution and the dichotomy model, four-, five-, and seven-factor solutions have recently been suggested (see Lindenmayer et al. 1995). In terms of clinical plausibility, statistics, patient selection, long-term stability, differences among survey instruments, and transcultural validity, the three-factor models, however, appear to be most convincing.

4 Increasing Importance of Developmental Psychopathology and Personality Research

Emotion research, coming largely from psychology, plays an increasingly important role in developmental psychopathology and in clinical research (Scherer 1993; Frijda 1986; Haviland and Lelwica 1987; Ekman 1993; Ekman and Friesen 1969; LeDoux 1989). Conceptual problems and the search for an adequate empirical methodological approach, which would at the same time do justice to phenomenology, currently dominate the field. Surveys of gestures and facial expressions and the establishment of physiological parameters, which are dependent on various emotional stimuli, are experimental and clinical parameters that are successfully used. The definition of emotion as a synchronization process of physiological oscillators (Scherer 1993) is not yet either clinically or experimentally convincing; whether basic emotions (Ekman and Friesen 1996; Machleidt et al. 1989) or a modular emotional conception are assumed still remains open. The extension of research to questions of cerebral localization and function is just beginning. As an example, Baxter et al. (1992) have shown that the emotional state connected to the manifestation of obsessive-compulsive phenomena leading to activation in the frontal caudate can be appreciably reduced and normalized by both medication and psychotherapy. This finding has led to speculations about the pathogenesis of the cerebral regulatory systems involved. Krause's group (Krause and Lütolf 1989; Krause et al. 1992), using gesture and facial expression paradigms, showed disturbed emotional interaction in communication between schizophrenic and healthy individuals and were able to clarify the significance of even the briefest and apparently meaningless facial stimuli as keys for controlling dialogue.

Apart from purely psychopathological research on personality disorders, aspects of personality and the psychopathology of their development already played an important role in the 1960s and 1970s in the interpretation of the psychopathology of endogenous psychoses. Tellenbach (1983) described the melancholic personality as a condition for clinical melancholy, and this was objectively validated by von Zerssen et al. (1994). In the course of their validation, the group was able to differentiate premorbid personality types for neurotic, schizophrenic, and affective patients (von Zerssen 1977; von Zerssen et al. 1996). Janzarik (1988) described the "preceding defect" for schizophrenia and Mundt (1985) the one for the premorbid affective personality. In the international literature, research focused on high-risk prospective studies, very largely directed at the spectrum of schizophrenic disorders, and mainly concentrated on discrete neurophysiological, neuropsychological, and psychomotor precursor symptoms of later schizophrenic illnesses.

Recently, there have been increasing numbers of studies in the field of empirically based developmental psychopathology on the etiology of depression. For example, analytical studies were carried out on the interaction of depressive parents with their children and their influence, as well as that of the entire constellation of their lives, on development of cognitive styles, regulation of self-esteem, social competence, and formation of social networks by these children. In particular, it was shown that interactions of depressive parents are generally reduced both quantitatively and in terms of their emotional breadth (see B. Schröder et al. 1996). In comparison to healthy parents, depressed ones showed more critical interactions with less involvement, less support, and more

anticipatory concern. The less beneficial styles of parental interaction were furthered by chronic burdens that the parents had – single parents, the double burden of career and household, chronic conflicts with partner or children (Brown 1996). Chronically depressive behavior patterns also have a negative effect on partnership and the partner's self-image. Brown (1996) has proven the prospective validity of these risk factors for developing a major depression in healthy adult women. Several groups have substantiated the risk of illness for children of depressive parents (Hops 1996; Hammen 1996). Depressive cognitive styles with tendencies to excessive self-criticism and self-negation, low self-esteem, and lack of social competence and acceptance in peer groups all work to bring about the development of the psychopathology. Children of insensitive mothers were shown by the Goodyear group to have retardation of cognitive, motor, and emotional development which can persist for a long time (Murray et al. 1996). The small children themselves appear to actively seek emotional rapport and attachment and, given the opportunity, to seek out responsive substitutes for their depressive mothers, if this is possible in the family.

This wealth of useful results relating to the pathopsychological and psychopathological development of manifest affective disorders provides a much clearer picture of possible risk constellations and protective factors than the previously typical features of personality that were acquired retrospectively and had a statistical/descriptive basis. Thus publications on the psychopathological development of affective disorders, particularly depression, emphasize pathopsychological models in which risk-bearing patterns of thinking, feeling, and behavior were transmitted in part by identification and in part by interaction. By comparison, results of preschizophrenic constellations are characterized more by fundamental organic disorders and the consequences of their inept handling in families, so that the impression arises sometimes that one is dealing with a special or abortive neuropsychological form of an organic brain disease.

The results of Walker and Lewine (1990) are worth mentioning in which home movies of small children who developed late-onset schizophrenia showed that discrete motor disorders were also the rule in early childhood. Resch (1996) summarized recent findings in his developmental psychopathology with the following ideas: tangential, cautious thinking, neuromotor deficits with disorders in speech development, attention deficit disorders, irritability of emotional control, and, finally, milder thought disorders describe the organic foundation. A rejecting, cold atmosphere, less active encouragement, low expectations, and active disqualification increase the developmental deficit and eventually lead to social incompetence and flattening of affect. Even for the development of a disorder which is primarily organic such as Alzheimer's dementia, Bauer et al. (1995) have shown a premorbid personality structure which is more strongly dependent in comparison to control probands. A corresponding interaction with comparable plausibility for organic/biological predispositions in patients at risk for affective disorders has not yet been developed. The enormous methodological difficulties of correlating the objective study of fundamental organic constellations and the hermeneutic study of a life history with its unfolding structure of meaning in a single comprehensible context has been recently pointed out by Bürgy and Mundt (1997) in their example of research on life themes.

Research on personality disorders as an independent nosological group, apart from continuing refinements and changes in survey inventories (see von Zerssen et al. 1996), has been enriched by the inclusion of biological aspects. Cloninger et al. (1995) have designed a repeatedly revised personality inventory, in which the categories "harm avoidance," "reward dependency," and "novelty-seeking" refer to an abnormal functional state of the dopaminergic, noradrenergic, and serotonergic systems. A genetic confirmation of this hypothesis, also indicated in view of the expression of the dopamine D_2 receptor (Svrakic et al. 1996), exists to date only for personality formations in alcoholism type 1 (late-onset, pronounced "harm avoidance") and type 2 (early-onset, pronounced "novelty-seeking"). Type 2 is thought to show a higher gene expression for the dopamine D_2 receptor than type 1. The as yet undeveloped validity of the inventory may have led to the sparse returns to date.

Herpertz and Saß (1997) have expanded personality research into experimental territory by examining the serotonin system using the fenfluramine test to attempt to stimulate affect in patients with impulsive personality disorder. They were able to show that dysfunctional affect as learned through experience is directly in conflict with impulsiveness, which is clearly biologically influenced and inherent in the personality, though cognitive performances and styles are not changed. The high degree of retention with increased control and explosive discharge of affect combined with distinctive affect change and differentiation play a role in subjective experience and have consequences for social ability.

Steinmeyer and Möller (1992) have made a valuable contribution to the statistical review of personality data generated from the inventories. Steinmeyer's facet theory analysis permits the modeling of clinically meaningful typologies gained by visualizing the overlap of clinical categories and those produced by statistical methods, e.g., factor analysis. The more or less pronounced convergence of these typologies with the empirically validated categories always remains evident.

Attempts from the field of pathopsychology (see below) to replace static-descriptive models of personality with functional models, which to some extent lack concrete ideas or empirical results, remain abstract and unclear (Becker 1995, 1996). Schmitt and Mundt (1991), Blöschl (1994), and Hops (1996) have investigated more closely the role of aggression in depressive personalities and were able to show relationships between suppression of aggression and low self-esteem as well as mechanisms of masked aggression among individuals with depression.

5 Influences of Phenomenology, Anthropology, and Philosophy on Psychopathology

The phenomenological-anthropological movement in psychopathology has a special tradition in Germany, the Latin countries, and Japan, and within Germany particularly at the hospital in Heidelberg and at the earlier centers in Marburg and Cologne. In a theoretical paper, Schmidt-Degenhard (1997a) named the two main interests of this research approach:

1. The ontic interpretation of psychopathological phenomena, i.e., explanation of the manner in which the being of the patients is adapted and its consequences for the relationship to self and the world. Husserl, Heidegger, and Sartre were all consulted as phenomenologists and existentialist philosophers concerning methodology (Husserl) and content.
2. The hermeneutic study of biographical, intentional, and mnestic structures of meaning guiding experience and action. This interest is also strongly influenced by psychoanalysis.

Schmidt-Degenhard's assertion that, because of its "existential harmony," the phenomenological-anthropological movement of psychopathology has special ethical support of higher standing than that of other research approaches should provoke criticism. He also directs this claim at "exploration as an open field" (Schmidt-Degenhard 1997b) for the effect of the patient upon the investigator and the establishment of intersubjectivity, an old demand made by von Baeyer (1955), which was already very controversial at the time. The implicit ethical devaluation of other traditions implied by such an assertion is problematic. The claim of taking the human condition as the object of investigation does not necessarily mean that it has been permanently improved. Historically speaking, important social psychiatrists have of course emerged from the phenomenological-anthropological tradition. However, a real improvement for psychiatric patients, especially since 1945, has come from the combination of many efforts, not least from the successes of psychopharmacology, evaluation of care for patients, psychotherapeutic research, clinical psychiatry and psychology, and social politics. The search for the gene for Huntington's chorea, for example, is as "in accordance with existence" as that for the transcendental organization of the ego (Blankenburg 1971).

Phenomenological-anthropological research has recently been given new stimuli in the German literature from Fuchs (1993, 1994a,b, 1995, 1996; Fuchs and Haupt 1994). In a survey of paranoid disorders in old age (Fuchs 1993), hearing losses in the speech frequency range were shown to be a possible precondition for the development of delusions in old age, as was the fate of having been refugees with particular vulnerability to intrusion into the psychophysical "inner space." Fuchs illustrated this experience with Kafka's *Bau* ("The Building") (Fuchs 1994b). The hearing deficits found in large numbers of the paranoid elderly point to an interruption of the gestalt circle, i.e., the possibility of constructing one's own reality by moving in a creative, circular process between sensory and motor interaction in the broadest sense. The modes which delusional patients use in modifying their existence, which Fuchs commented upon, and the consequent points of view for their stimulation and protection during the course of illness give important suggestions for psychotherapy with delusional patients (Mundt 1996a). Kraus (1992) recently studied the technical delusions of schizophrenics and the role-theory interpretation of affective disorders, which he interprets as the creation of a virtual reality in order to "normalize" pathological experiences that could not otherwise be categorized, an idea which Schmidt-Degenhard (1992) had already described with his work on oneiroids. According to Schmidt-Degenhard, delusion represents a modification of existence by disempowerment of the self. The lost world of socially binding common sense is replaced or added to by imagination. In a study of existential aspects of anxiety, Kraus (1996) takes up old dialectical concepts of existence and intention, of distance from self and alienation, a concept also discernible in a study on ego estrangement in melancholics, in whom there is a quasi-delusional self-accusation of lying against a culpable background. Mundt (1996b) was able to demonstrate the alienation in anxiety as a problem and difficulty in psychotherapeutic work with the compensational character of phobic patients. The predominance of the alienated nature of intentional acts in melancholia in contrast to a specific, more instrumental disorder of retardation was shown by Mundt et al. (1998) in a study involving a series of experiments with depressed patients on estimating and experiencing time.

Anglo-American phenomenological psychopathology, influenced by European phenomenologists and analytical philosophy of mind, studied the previously mentioned critique of operational classification (e.g., Radden 1994) and the discussion about popular psychology, i.e., lay psychological approaches to understanding mental disorders, among which psychopathological concepts (Radden 1994) are generally included by some American researchers (Harrison 1991). The polemic about these "eliminative materialists" (Schwartz 1991) appears sterile – the utility of the approaches will ultimately provide their legitimation, not the "metapsychopathological" dispute about them.

In the 1980s, great hopes were placed in qualitative, psychopathological research stimulated mainly by psychoanalysis, social psychology, and sociology. For a while, the leading research paradigm was the narrative interview, a research technique in which the subjective element of the interviewer involved in the hermeneutic act of determining meaning was to be restricted by better objectivity, e.g., by speech analysis of the narrative produced, i.e., the text that had not been prestructured by the investigator. The results are modest given the high cost in terms of methodology and time expended. As in many other cases, the use of more extensive methodology has confirmed what is already known (see Frommer 1996; Faller 1994), in this case, especially, the negative self-image of depressed persons, albeit with a significantly greater level of proof than the subjective evidence of individual clinicians indicates.

Bürgy and Mundt (1997) have pointed out the special methodological problems of biographical research. Methodological dualism (Mundt 1989) emerges here persistently again and again. It can probably only be overcome with an alternating model. Particularly research on life themes in the work of Thomae (1996) shows that a convergence and merger of hermeneutic and objective approaches is not possible. Instead, adherence to the subject and formulation of clearly distinguishable research paradigms is more successful. The flattening out that occurs with diary techniques (Laux and Weber 1987) or with objectified life-event techniques can be overcome by a hermeneutic examination of the results and can lead to renewed formulations, which can once again be subjected to objective investigation blinded to meaning.

6 Pathopsychology and Other Contributions of Clinical Psychology

Research on salutogenesis (positive or protective factors) has developed in recent years as a counterart to psychopathological pathogenetic research (Antonovsky 1987). It studies mechanisms of becoming and staying healthy, even under difficult conditions. In addition to classical stress research and coping concepts, recent studies on coping styles in illness and health maintenance among adolescents with and without difficult environmental influences, such as being raised in a foster home, have yielded interesting results (Bielefeld study). Earlier coping results were confirmed. Thus it was shown that active behavior, intelligence, prior experience of being effective, temperamental factors, and flexibility, including stable emotional relationships and models for crisis management, play a positive developmental role in adolescents from unfavorable circumstances and promise "resilience," i.e., elasticity (Lösel et al. 1992). Moreover, the differences in the healthy control group were slight in the Bielefeld study (see Becker 1995, 1996). It remains questionable whether a model that assumes threshold values for intelligence, for example, can explain the results better than a model of a balanced pattern of salutogenetic influential factors. Haltenhof and Vossler (1994) found distraction to be one of the most important factors for recovering from depression. Apart from aspects of severity of depression and autonomy, which must be taken into account in interpreting these findings, the methodological problem of salutogenetic research plays a greater role here than in other diagnostic groups, namely the confounding of preventive factors with variables of the current state of health as well as insufficient differentiation between short-term and long-term factors, and factors which positively influence the state of health and structure of salutogenesis.

The concepts of salutogenesis are related to research on quality of life (see Chap. 7, Vol. 2; Lauer 1997), and their development was also stimulated in psychiatry by cancer treatment. Just as in that field tumor diagnosis alone is no longer the criteria for therapeutic action, so increasingly in psychiatry psychopathological symptoms are supplemented by criteria that have quality of life as a therapeutic goal. An abundance of unsolved methodological problems still makes interpretation of results more difficult. This includes the gap between objective and subjective evaluation of quality of life. The latter is frequently judged by schizophrenics, for example, as excessively high, perhaps due to their experience of contrast and the "modesty" associated with autism. The result should by no means be allowed to lead to complacency in therapy, rehabilitation, and social policy. However, even self-determined life goals and standards, influenced by origins, can influence the subjective judgment of life quality by a realistic judgment of what is attainable or relative satisfaction based on overcoming past and more difficult circumstances. A generally accepted measure (see the so-called qualies) is still lacking.

7
Psychopathology and Cognitive Neuroscience

Cognitive neuroscience is a comparatively young discipline with roots in psychology, computer science, physiology, anthropology, philosophy, and linguistics. In this decade, the decade of the brain, extensive progress had been made in this field, which cannot go unnoticed by the science of psychopathology, particularly since cognitive neuroscience is about exactly those high-level mental functions whose pathology is at issue in psychiatry, e.g., thought, perception, affect, and volition. To give a further example, until just a few years ago, the use of the word "consciousness" in scientific discourse was taboo among neuroscientists. Today, many experimental studies are directly devoted to the neurobiological correlates of consciousness (Barinaga 1997, p. 1583). This demonstrates that the themes of psychopathology have finally become part of the scientific discourse.

To give yet another example, up until the late 1980s, the question of exactly where and when a mental image or an auditory hallucination can be localized in the brain appeared to be an error in categories rather than a serious goal of empirical research. This has changed fundamentally due to the development and use of new methods and new concepts.

The remaining parts of this chapter consists of a brief methodological section, after which some results, concepts, and interpretations regarding single psychopathological symptoms will be presented. The subsequent discussion will proceed from specific to general aspects and will close with remarks on psychopathology as a science.

7.1
Natural Sciences and Humanities: Methods

One hundred years ago, the field of brain research experienced a surge of popularity not unlike the present one. Unlike now, however, at that time new discoveries about neurons, neuron clusters, and brain modules could only be related highly speculatively to mental processes. Psychopathologists therefore distanced themselves and their discipline from the "mythologies of the brain." In doing so, they advocated the recognition of their field as belonging to the humanities, not to the natural sciences. This led to the separation of the brain and the mind in psychopathology, which was obviously a very unfortunate development, as nothing is more apparent in psychiatry than the immediate connection of mind and brain, as Andreasen (1997) recently and rightly emphasized. Like cognitive neuroscience, psychopathology is rooted both in the natural sciences and in the humanities, and it encompasses philosophy and psychology as well as computer science and neurobiology.

Given this diverse origin, it is no wonder that, from the very beginning of psychopathology, it faced major problems concerned its methodology. As Karl Jaspers (1957, p. 13) remarked in his autobiography, anyone speaking about the mind must know "what one knows, how one knows it and what one does not know." It is unfortunate that this dictum finds so little resonance in some present-day discussions of psychological issues. However, the recent *methodological* progress in the field of cognitive neuroscience enables psychiatrists to study psychopathology in new and unprecedented detail and depth.

There is no single "optimal method" for the study of mind and brain. It is rather the task of the researcher to find the right method for the question at hand. In most cases, several methods will have to be used in concert such that their respective strengths are combined and their respective weaknesses compensated for (see Posner and Raichle 1996). In order to find the components of complex mental processes, psychologists have used the procedure of measuring reaction times for more than 100 years. By constructing specific simple tasks to which the subject can respond rapidly, the nature of the mental processes that are at work when the task is performed can be inferred. In addition to such mental chronometry (Posner 1986), functional imaging procedures provide an answer to the question as to where exactly in the brain certain processes occur (Toga and Mazziota 1996). Electrophysiological procedures, in addition, can clarify exactly when, down to the millisecond, a certain mental process occurs (Heinze et al. 1994; Snyder et al. 1995). By studying the effects of psychoactive substances, the neurotransmitters and neuromodulators that participate in or modulate a process can be identified (see Hermle et al. 1996; Kischka et al. 1996; Spitzer et al. 1996).

7.2
Results of Research

In the following sections, findings of neuroscience research that are relevant to psychopathology are presented. This is not meant to be an exhaustive review. The topics have been chosen to exemplify the achievements of the cognitive neuroscience approach in psychopathology. In the subsequent sections, the map structure of the cortex and two basic concepts pertaining to the functionality of these maps, neuroplasticity and neuromodulation, are discussed.

7.2.1 Hallucinations

Jaspers defined hallucinations as deceptive and perceptual events (*Wahrnehmung*) that have to be strictly separated from mental imagery (*Vorstellung*). He attempted to support this definition by referring to phenomenology, i.e., the philosophical approach to the nature and the constituents of mental life founded by the German philosopher Edmund Husserl. This reference to Husserl's phenomenology has remained highly controversial (see C. Walker 1993; Wiggins and Schwartz 1996). As has been pointed out elsewhere (Spitzer 1988), it appears that Jaspers' definitions of psychopathological concepts do not refer to Husserl. They actually express nothing but the prejudices of a branch of eighteenth-century psychology and, furthermore, result from his quite limited clinical knowledge. (Jaspers worked only part time in psychiatry for only a few years, and he only dealt with special cases.) It is difficult to understand why Jaspers' definition of hallucinations has been maintained up until today. From a clinical point of view (see P. Schröder 1915), hallucinations can be anything from truly perception-like to faint and idea-like. Patients often identify their deceptive nature (and hence are not deceived), which is why the term "voices" was introduced to denote these special experiences. Detailed research into the subjective nature of hallucinations induced by various experimental procedures, from drugs to electrical brain stimulation and sensory deprivation, has further shown that Jasper's strict definition of hallucinations is of little use (Beringer 1927; for a summary, see Spitzer 1988).

The studies conducted by Shepard and Cooper in the early 1970s (for a summary, see Shepard and Cooper 1982) made it obvious that mental images can be investigated experimentally, despite their being an individual's most private, most "subjective" experiences. The results of a series of reaction time experiments on the mental rotation of imagined pictures and objects allowed inferences to be made regarding the nature of mental representations. These studies suggested that mental images are highly similar to perceptions. More recent studies in the field of cognitive neuroscience on perception and imagery conducted during the past 5 years strongly supported this view in a manner which would have been thought impossible several years ago: The processing of mental images apparently activates the very same central nervous structures that become active during visual perception. The content of a mental image does not arrive from the retina via the lateral geniculate body, but instead is provided by compressed information residing in long-term storage sites in the temporal lobe. This information is projected "backwards," i.e., downstream, to the "lower" visual cortical areas. This process is computationally intensive and hence takes time. Once the mental image is built up in the visual storage buffer (see Kosslyn and König 1992), it can be manipulated there (e.g., rotated or zoomed, i.e., its size enlarged or reduced), which also takes time, as can be measured by means of skillfully planned experiments. Functional MRI (fMRI) studies directly demonstrated the activation of visual cortical areas during mental imagery; in particular, the primary visual cortex is activated about half as much during imagery as it is in visual perception (O'Craven et al. 1997). Kosslyn and coworkers (1996) demonstrated in a PET study of normal subjects that the ability of the visual cortex to become activated correlated to the performance of subjects in visual imagery tasks.

These and a number of other findings indicate that perception and mental imagery are not two fundamentally different processes, but rather two aspects of the unitary function of visual information processing. Just as centers of "perception" are involved in mental imagery, stored information and hence mental images are involved in perception, as was demonstrated decades ago by gestalt psychologists. Perception is an active process of analysis of incoming information through synthesis using the available information already stored. The structure and gestalt formation processes that are involved in visual perception and imagery can be studied in rodents and in primates, including humans, down to the level of individual neurons.

Primates, like humans, are subject to visual illusions and ambiguous figures such as the Necker cube and the Rubin cup. Such ambiguity can also be provoked by projecting different images into either eye, which causes the effect of binocular rivalry, i.e., seeing one of the images for a few seconds and then the other one, but not both at the same time. In animal studies, this effect can provoked as follows: First the animals are shown unambiguous figures. Then neurons in the occipital and temporal lobes are identified which selectively become active to one of these figures. Finally, the monkey is shown a different figure in each eye. At any one time, the monkey will see one of the two figures (which can be experimentally depicted by its behavior), and the previously identified neurons will behave accordingly. The speed of change corresponds to that for human subjects, who under such conditions will also see one figure at a time. Further studies by Logothetis and coworkers (Leopold and Logothetis 1996; Logothetis et al. 1996) suggest that the "higher" the neurons rank in the visual information processing, the greater the extent to which their activity is influenced by the processes of mental image production. If 18% of the neurons in the primary visual cortex change their activity depending upon

which figure is currently being seen (i.e., which information arriving from the retina is being synthesized), this percentage is approximately 50% around the middle of visual information processing (mediotemporal cortical area), and almost 100% upon completion of the process (inferior temporal cortical area). In other words, the higher the level of visual information processing, the more the neurons will represent what is subjectively experienced or seen, and the less they will represent the physical environment. Moreover, these studies demonstrate that the transfer from the retinal image to the subjective experience does not take place at a specific point in the brain, but rather consists, neurobiologically speaking, of a gradual transition.

This view of visual perception as an active, synthetic process distributed over a large number of cortical areas renders the phenomena known from psychopathology – from the illusions driven by affect, up to the scenic, visual hallucinations – highly plausible. For example, if we take into account the fact that the serotonergic system has the highest density of projections in the visual cortex and that a number of hallucinogenic agents act upon the serotonergic system, it becomes apparent how top-down processes can come to outweigh bottom-up processes in the course of visual perception. Once this balance is disturbed, the mode of perception may change correspondingly to any psychopathological perceptual phenomenon, from illusory modification to full-blown hallucination.

The same applies to the auditory hallucinations of schizophrenic patients who hear voices. Using functional imaging, Silbersweig et al. (1995) recently provided direct evidence for the involvement of speech-processing cortical areas in patients who reported hearing voices. Of course, auditory hallucinations were suspected to involve language-related cortical areas on purely clinical grounds before the advent of functional imaging methods. T. Early (personal communication), for example, described a patient who became aphasic following an ischemic insult to the Broca area and who maintained that the voices she had heard had difficulty finding words. This observation directly implies that the same language areas that are involved in the production of speech are involved in the production of auditory hallucinations. Nonetheless, functional imaging delivers further confirmatory evidence.

7.2.2 Obsessions and Compulsions

Obsessive-compulsive thoughts and actions have been the focus of recent research, sparked by several neurobiological developments. This increased research interest has led to an increase in the number of papers on obsessive-compulsive disorder (OCD), which in turn led to an increased awareness of the disorder by clinicians. This corresponded to a considerable increase in the frequency of diagnosis of OCD in recent years (see Kaplan et al. 1994).

Most importantly, obsessions and compulsions were placed within the framework of frontal lobe cortico-striato-thalamocortical circuits, as described by Alexander and coworkers (1990). It is assumed that this type of frontal cortical circuitry leads to the integration of information from various cortical areas. The information-processing circuits from the cortex via subcortical regions and back to the cortex possibly help to stabilize the output of the information-processing system. This is important because the system always has to respond with a single response at any given point in time. If the system becomes stuck with a single output, however, the mechanism of stabilization overshoots, and this dysfunctional state may correspond to symptoms of OCD. It should be noted that OCD is not the only psychopathological syndrome which can be put in the framework of frontal cortical circuitry (see Kischka et al., in press).

In the case of OCD, the overshooting activity of one or more of the orbito-frontal-cortical and subcortical structures (striatum) was demonstrated by means of PET (Baxter et al. 1992). Most importantly, this pathologically increased activity was shown to become decreased and to return to normal in the course of successful pharmacotherapy and psychotherapy. Thus this study was the first to render the effects of psychotherapy directly visible through neuroimaging. In subsequent studies, obsessions and compulsions have been examined using fMRI, in which patients are scanned at rest and during the provocation of symptoms (Breiter et al. 1996). The results were similar to those described by Baxter in that those areas that had been found to be active in PET and normalized after successful therapy were also shown to light up during symptom provocation.

The studies by Baxter et al. (1992) and Breiter et al. (1996) illustrate the close connection between psychology and neurobiology. This connection pervades not only research, but also the therapy of obsessions and compulsions. Deficits of the patients in different areas of life often become apparent, and often for the first time, after OCD symptoms have been contained. These call for behavior therapy and/or insight-oriented psychotherapy. The neurobiologically proven close relationship between the psychological and biological levels in obsessions and compulsions corresponds to the closely meshed corresponding therapeutic strategies.

7.2.3 Formal Thought Disorders

Ever since the time of Bleuler (1911) and the inception of the concept of schizophrenia, formal thought

disorder has been conceived of as a disorder of associative processes. These processes have been investigated using experimental procedures concerning speech production and understanding. A number of studies indicate that in schizophrenic patients with formal thought disorder, the access to stored information is disturbed, whereas the store itself remains intact. This dysfunction has been characterized in detail by Spitzer and coworkers (Spitzer et al. 1997). A large number of behavioral data and electrophysiological experiments suggest that the focus of access to networks of stored information is wider in patients with formal thought disorder. This widening of the focus leads to less precise retrieval and has been both conceptually and experimentally related to the dopaminergic system.

The cognitive neuroscience view of formal thought disorders provides a parsimonious explanation for a number of nonrelated and unexplained clinical phenomena, including clinical and experimental association behavior, temporal aspects of spontaneous speech, the apparent contradiction of concretism and simultaneous overabstract thinking, the apparent "creativity" of some patients, and sound-related word associations (Spitzer 1997a). Within a cognitive neuroscientific framework, these phenomena can be related to each other and to underlying neurobiological processes. This allows for the further generation and testing of model-driven hypotheses and for the development of new diagnostic and therapeutic strategies (see Maier 1998).

7.3 Brain Maps: Structure and Dynamics

The term "brain maps" relates to a number of issues in brain research, some of them very old. Cortical areas can be differentiated according to various criteria, which led as early as at the turn of the century to the publication of different classification systems for the various parts of the cortex. The best-known mapping was done by Brodman (1909), who employed histological differences as distinguishing features. Flechsig (1929) differentiated cortical areas according to the time point of their myelinization. Zilles et al. (1995) used multiple biochemical markers.

Before the advent of noninvasive methods for mapping cortical areas in human beings, functional mapping of the cerebral cortex was done by means of animal stimulation and lesion experiments. In humans, the past 100 years of neuropsychological investigation of patients with lesions clearly demonstrated functional specialization in the human cortex. In addition, the invasive electrical stimulation studies conducted by Penfield and coworkers corroborated the functional map structure of some cortical areas more than six decades ago.

The new procedures that have been available for the past few years have opened up the path to an entirely new field of research into the map-like structure of the human brain. This functional explanation of the human cerebral cortex is a task which may take decades and is comparable in scope to the unraveling of the human genome (Albert 1997; Toga and Mazziotta 1996). The immensity of this task can be estimated by considering several plausible additional assumptions: We may begin by assuming that the cerebral cortex is weakly organized into modules (see Kosslyn and König 1992), i.e., that a particular mental performance can be achieved by the cooperation of different areas, possibly as many as one to two dozen. From a functional point of view, the cortical areas can be distinguished either on the basis of their selective response to certain stimuli or their activation by certain clearly defined actions.

In this regard, the visual system is the one that has been most studied. At least 32 different cortical areas are now known (Felleman and van Essen 1991), which are responsible for the function of vision. Most of these areas have been defined electrophysiologically in animal experiments with primates and have been proved in humans by means of functional imaging. The maps of animals and humans can thereby be compared and indeed have been shown to be quite similar (see Sereno et al. 1995).

The average size of the visual areas of primates is 170 mm^2. In humans these areas are probably larger. Thus the area V1 in a human is twice the size of the area V1 in a monkey. If we assume the size of the human cortex to be 250,000 mm^2, the number of different functional maps located in the human cortex and yet to be discovered can be estimated at 735 (see Spitzer et al. 1998). This figure is certainly the result of a rough estimate, but it illustrates the scope of the problem.

7.3.1 Structure of Brain Maps: Neuroplasticity

Functionally specific cortical areas are referred to as maps because the cortex, due to its internal structure, organizes incoming signals according to frequency and similarity, i.e., like maps. Importantly, this order is not static, but instead changes within certain limits according to the input the cortex receives. This input-driven change of cortical representations due to a rewiring of cortical connections is called neuroplasticity.

Generally speaking, neuroplasticity refers to neuronal reorganization that depends upon the signals from the outside world as well as from the organism.

Research into neuroplasticity conducted during the past two decades impressively documents that the brain is constantly reconstructing itself depending on the input to be processed. It does so by building up new neuronal connections in order to improve its processing of incoming signals.

It is well known that neurons are post-mitotic tissue, i.e., their cells have lost the capacity to divide. It follows that neuroplasticity does not involve the growth of new neurons, but instead the genesis of new links between existing neurons (neuronal sprouting). In addition, existing but nonfunctional links between neurons – so-called silent connections – may become functionally active. A large number of studies in primates and humans have impressively documented the continuous building of structures in cortical representations dependent upon experience (Merzenich and Sameshima 1993). For example, if a person has learned braille, the cortical area responsible for the tip of that person's right-hand index finger grows (Pascual-Leone and Torres 1993); if someone learns to play the violin, the representation of the left hand grows. If a synthetic inner ear (cochlea) is implanted, the auditory processing of speech is fundamentally reorganized on the basis of the completely novel input signals (temporal input patterns). The subjective hearing experience of the patients changes accordingly: Following the operation, they hear only an unpleasant rumbling upon listening to speech. A year later, the majority of patients are able to carry on telephone conversations, i.e., can understand speech without having to lip-read.

Due to the uniform structures of the neocortex (Nauta and Feirtag 1986), it can be assumed that neuroplasticity also occurs during more sophisticated mental operations, including language and thought. In his monograph on brain plasticity, Kolb stated this as follows:

> The idea that activity might change the heart or muscles is seldom questioned. The possibility that behavior could change the structure and function of the brain is seldom considered! Nevertheless, it is an important aspect of brain plasticity. Indeed *there is little doubt that even thought can change the brain* (Kolb 1995, author's italics).

7.3.2 Semantic Maps

When we speak or understand speech, we must learn to refer to learned information pertaining to language that has been stored in map-like associative networks. The representation of the meaning of a word in such a network can be conceived of as a "node" and by the relationships to other nodes. In the course of understanding a word, for example, a node becomes activated. This activation spreads more or less rapidly to other nodes (see Spitzer 1996, 1997b). Thus in word association tasks, for example ("Which word comes to mind when you hear the word ...?"), the answers given by subjects are often amazingly stereotypical. By a similar mechanism, in lexical decision tasks, the recognition time for a word is shortened if the word has been preceded a few hundred milliseconds earlier by a word whose meaning is represented by a supposedly adjacent node (semantic priming effect). Furthermore, a number of single case studies of neuropsychological patients have been carried out which indicate the presence of a selective naming deficit for specific categories. For example, a patient may selectively be impaired in naming animals, and another may find furniture or tools or fruits difficult to name. This can be interpreted in terms of a localized representation of the corresponding contents, as suggested by Farah and Wallace (1992). Studies of such high-level maps have now been carried out using functional imaging techniques (see Spitzer et al. 1995b; for a summary, see Spitzer and Kammer 1996).

Map-like semantic networks of the kind described above have also been generated in computer simulations of neuronal networks. Ritter and Kohonen (1989) fed sentences as input to a type of neuronal network, which is set up to model certain features of cortical connectivity. These networks are called self-organizing feature maps, as their very structure leads to the formation of representational maps of whatever input is given to them. Accordingly, when short sentences were given as input, the network built up an orderly representation of words according to their semantic similarity and grammatical category. Words that occur in the same context (this is the case, in particular, for opposite terms, such as "black" and "white") were represented close to one another (semantic mapping), and in addition, words were represented ordered according to whether they were verbs, adverbs, or nouns (see Spitzer 1996, 1997b). The network thus recognized the immanent regularity of the input and used it for the generation of map-like order. Most notably, this is achieved without the explicit application of rules.

In the case of higher cognitive functions, the input of the corresponding cortical areas arrives from other cortical areas. This shall briefly be outlined for the connection between the semantic lexicon and verbal working memory, as discussed by Just and Carpenter (1992): In order to comprehend and produce speech, it is necessary not only to have access to stored lexical information, but also to be able to keep this information "online," e.g., to construct the semantics of a grammatically complex sentence. People differ with

respect to verbal working memory, i.e., their capacity to keep words in mind and construct meaningful sentences from them. Moreover, this function is modulated by the dopamine system (see below). This results in considerable individual differences regarding the capacity to process complex spoken and written language.

The extent of neuroplasticity decreases over the lifespan of an organism. During infancy and development, the cortex is so plastic, as shown in studies in hamsters, that experimental detouring of visual input to the auditory cortex prompts development there of a visual cortex capable of functioning (Frost 1997). Later in life, after brain development has been completed, the structure of the stored maps nonetheless still changes throughout life according to experience. This occurs quite rapidly in younger organisms, slowing down with increasing age. This decrease in neuroplasticity with increasing age makes sense, as the following argument shows. Each organism has to learn as rapidly as possible. Once the most essential things have been learned, however, these memories must be protected from an all too rapid dismissal in the face of few newly learned opposite experiences. To quote an example from McClelland and coworkers (1995): If you have learned that fish can swim and that birds can fly, you should not become confused upon first contact with a penguin (i.e., a bird, and not a fish, that cannot fly but can swim). Thus it is necessary for a person who has already learned a great deal to learn more slowly and to integrate the contents to be newly learned with that which has already been learned. Computer simulations of neuronal networks indicated that this can be achieved through the circuitry of two networks with different features. This appears to be the case in our brains through the division of labor between the hippocampus and the cortex informing explicit memories of important events.

In a stable environment, older people are better adjusted than younger people to their environment. The decrease in neuroplasticity (and thus in learning capacity) with increasing age is a psychosocial problem only if the environment changes rapidly. This is the case within the current cultural situation. The elderly expert with his subtle experience is only seldom needed, stability is not valued or is not even perceived, and people can easily find themselves in a situation in which the capabilities and values they have learned are no longer needed and applicable. Thus, with regard to the psychopathology of aging, problems must necessarily arise when the speed of cultural change is viewed from a neurobiological perspective.

7.3.3 Neurobiologically Motivated Psychotherapy

Within psychiatry it is important that neuroplasticity by no means represents an empty psychopathological term. It is actually of great practical relevance. This shall be demonstrated by the following brief description of studies performed by Merzenich, Tallal, and coworkers (Merzenich et al. 1996; Tallal et al. 1996) on language-impaired children.

It has been estimated that 5%–8% of all children demonstrate a slowing of information processing in the area of early auditory signal processing. This minor deficit is aggravated during the development of language, as up to 85% of children may develop reading difficulties due to deficits in decoding certain rapid sounds (Barinaga 1996). According to the latest studies, the primary disorder can be diagnosed as early as the age of 2 years, i.e., even before speech has been completely developed and long before reading instruction begins. In most cases, though, it goes unnoticed. For example, the two syllables "ba" and "pa" differ only in their first consonant, which in turn only lasts a few milliseconds. If these two short consonants cannot be analyzed quickly, higher cortical analysis and synthetic processes are not receiving the "sharp" input they require. The child will have difficulty understanding spoken speech, although hearing is not disturbed. Instead, the analysis of a rapid sequence of temporal input patterns is disturbed.

It was shown that children were able to distinguish between "ba" and "pa" when the consonants were artificially temporally lengthened in an acoustics laboratory. The children were thus able, in principle, to decode relevant input patterns but needed more time to accomplish the task.

When the children practiced using temporally stretched input patterns for about 2 h per day for several weeks, their comprehension of normally spoken speech improved (Merzenich et al. 1996; Tallal et al. 1996). Training consisted of playing on a computer whereby the children had to learn to distinguish increasingly rapid temporal sound patterns. Skilled programming ensured that the children were offered only those input patterns (i.e., temporal phonetic images and other rapidly occurring temporal sequences) which they were able to process. In addition, the child-friendly design of the practice sessions guaranteed the willingness of the children to practice. Both are prerequisites for successful therapy, as shown in the studies by Mezernich.

The studies by Mezernich and Tallal are innovative in psychiatry insofar as here, for the first time, findings and ideas from experimental *neurobiology* concerning the determinants of cortical reorganization processes led to the development and successful application of

an exclusively *psychological* method of intervention. The example thus clarifies that neurobiological findings do not necessarily mean a reduction of therapeutic strategies to biological, i.e., mainly pharmacological, methods of intervention.

7.3.4 Dynamics: Neuromodulation

A further essential functional principle of brain maps is their modulation by the diffuse release of comparatively slow-acting substances (within the range of several hundred milliseconds). Neuromodulation is defined by Kaczmarek and Levitan (1987, p. 3) as "the ability of neurons to alter their electrical properties in response to intracellular biochemical changes resulting from synaptic or hormonal stimulation." Neuromodulation is different from neurotransmission, which underlies central nervous information processing. Neurotransmission denotes the exact, rapid transmission (in the range of a few milliseconds) of information by glutamatergic and other excitatory and inhibitory synapses. Neuromodulators are the very agents which are targeted by many psychotropic drugs, the most important being the monoamines (norepinephrine, dopamine), serotonin, and acetylcholine. They modulate, i.e., fine-tune, general parameters of information processing with regard to the demands of the environment and the organism. In addition to regulating the general level of activation, neuromodulators also affect the influence of bottom-up and top-down processes and another essential parameter of information processing, the so-called signal-to-noise ratio (see Cohen and Servan-Schreiber 1992, 1993). Neuromodulatory states correspond to subjective states of alertness, anxiety, apprehension, depressed or elevated mood etc. The neuromodulatory state that corresponds to relaxed awareness and lack of anxiety, for example, facilitates the flow of thought, including creative remote ideas, as can be measured by the word associating task. Results of experiments performed by Kischka et al. (1996) suggest that in such states dopaminergic modulation is somewhat downregulated. When the dopaminergic tone is experimentally increased, the availability of remote associations decreases. We already know that this is also the case in situations of immanent threat and danger, as under such circumstances lengthy "creative" considerations would be a sign of poor adjustment. The rapid action necessary in such situations requires the use of learned and proven sequences of behavior which can be rapidly unreeled. To sum up, the conditions of anxiety and stress decrease creativity and imagination, which appears to be advantageous in evolutionary terms. However, under the conditions of life in modern civilization, these physiological neuromodulatory effects can be disadvantageous, e.g., within the framework of taking examinations.

With regard to subjective experience, the issue of neuromodulation is hard to pin to any of the classical terms of psychopathology. As mentioned above, vigilance and affect are likely subjective features of mental processes, which can be taken as a function of neuromodulation. The term "dynamics," as defined by Janzarik, appears rather well suited to characterize the subjective side of neuromodulatory changes in the serotonergic, noradrenergic, and dopaminergic systems.

Recent studies show that neuromodulatory systems are not rigid and incapable of change, but respond to experience. Animal experiments with crayfish have shown that one single social interaction that ends either successfully or unsuccessfully is sufficient to reverse the postsynaptic response of a neuron to serotonin administered exogenously (Yeh et al. 1996). Thus, for the first time, social behavior was tracked to the level of individual neurons and subpopulations of receptors whose variable expression caused the different response to serotonin. If this line of thought is transferred to human beings, defined by Aristotle as social animals, it appears that neuromodulatory systems may change ways of responding to the environment, i.e., behavioral tendencies, according to experience. Thus temperament and character are at least in part shaped by the environment.

In addition, a genetically determined neuromodulatory setup of each individual has been assumed, which forms the basis of the temperament of that individual (Cloninger 1987). Individual differences in the availability of different neuromodulators, receptors, and receptor subtypes thus lay the foundation for behavioral dispositions and ways of responding to the environment, most notably the social environment. As shown in the above example, this foundation is not fixed through lifetime, but is rather itself subject to a certain amount of change driven by experience. The concept of neuromodulation thus yields not only the background for the therapy of acute schizophrenic and affective psychoses, but also paves the way for therapeutic access to temperament and character by means of pharmacological therapy and psychotherapy.

8 Psychopathology and Psychotherapy

Psychoanalysis was the dominant paradigm in psychotherapy in the German-speaking countries 10 years ago. This is well exemplified by the contributions of Bräutigam, Peters, Lang, Mundt (see Mundt 1989), and

many others to models of psychopathology for mental disorders. While psychoanalysis is still strong in French, Italian, and South American traditions of psychiatry, its importance has decreased both in Germany and in Anglo-American psychiatry, where it has been partly replaced by behavioral therapy.

Behavioral therapy has not only brought to psychiatry concepts of illness for well-known symptom patterns of psychopathology that are in part new or at least modified, it has also enriched the catalogue of symptoms with a phenomenology developed according to its reference framework, which has even made inroads into psychoanalysis. Examples of this are the theory of depression in behavior therapy and its descriptions of obsessions and compulsions and of phenomena related to anxiety. To give an example, in cognitive behavioral therapy, autonomized negative cognitions (as assessed by means of diaries), styles of causal attribution, subjective experiences of decreased intentional behavior, and lacking reward mechanisms in inactivity and anhedonic behavior have been intensively discussed. This has produced a new phenomenology that has made important contributions to the well-known inventory of symptoms for these disorders. There has been a similar development in both the area of obsessive–compulsive disorders and, with the refined analyses of avoidance behavior described in the literature of the 1960s, also for patients with anxiety disorders. Here, the descriptive approach to psychopathological states with no influence of any traditional schools of thought may, in the long run, help to develop hierarchies of symptoms for specific disorders such that they can be modified by psychotherapy. Now that Jaspers' stratification rule has been officially abandoned in recent psychiatric classification schemes, such hierarchies have been reconstructed in the schools of behavior therapy. This development may be exemplified by Linehan's borderline theory (Linehan 1993), where covert emotional irritability is regarded as the basis for overt secondary symptoms, such as reaction or coping. Similar distinctions have been proposed for schizophrenia and depression, and they are well suited for the development of cooperative models for psychotherapy and pharmacotherapy. The development of operationalized psychodynamic diagnostics within the field of psychoanalysis mentioned above in the section on classification has also led to the systematization of psychoanalytical structure and conflict diagnoses, which represent possible contributions to a descriptive psychopathology.

The importance of combined psychotherapy and pharmacotherapy and its evaluation with regard to the patient's psychopathological outcome is currently underscored more than ever by therapy efficacy studies. These studies indicate that persistent psychopathology – particularly psychotic and depressive states – itself acts pathogenetically, contributing to a poorer course the longer it exists. Within the neurobiological framework discussed above, in particular in the light of what we know about neuroplasticity, this epidemiological finding was to be expected. We know from controlled therapy studies that the combination of psychotherapeutic and pharmacotherapeutic methods is preferable to the use of either of the two methods alone. This finding can easily be reconciled by the neuromodulatory changes brought about via pharmacotherapy and the ensuing neuroplasticity-driven changes caused by normal experiences, made possible via pharmacotherapy. Thus an idea begins to emerge of how the various dynamic principles of therapeutic interventions interact, of their foundation, and of how they might possibly be optimized and tailored to the individual needs of the patients. In the following section, this view will be fleshed out by the example of therapeutic intervention in acute and chronic delusions.

8.1 Acute Delusions and Neuromodulation

Acute delusions are generally accompanied by anxiety, distrust, and increased apprehension and vigilance (delusional mood), which can result in the overinterpretation of what are actually unimportant events. Unfounded importance is given to things and events, and things are related to one another and to the deluded person without there being any reason to do so. Once delusional ideas have been formed and linked to the current motivational background of the patient, and once their contents determined by the patient's prior experience are built into the remaining normal experiences of the individual, they are then almost immune to doubt and become chronic. Patients work at their delusions and not only systematize their delusions but more or less weave them firmly into the rest of their life history. Chronic delusions are then not so much a certain mental state, but rather a part of the thoughts, values, world view, hopes, and goals of an individual. Acute delusions often respond well to neuroleptic treatment, whereas pharmacological treatment of chronic delusions is extremely difficult.

The dopamine hypothesis of schizophrenia has recently been modified. Positive symptoms such as (acute) delusions and hallucinations are assumed to be caused by dopaminergic hyperactivity, whereas negative symptoms, such as blunted affect, loss of motivation and drive, avolition, anergia, and anhedonia are thought to result from dopaminergic hypoactivity (see Davis et al. 1991). As rapid perceptions and thoughts can only be the product of fast neuro-

transmission, the question of how dopaminergic hyperactivity is related to distrust, delusional perception, and delusional ideation needs to be answered by the modulatory effect of dopamine on neurotransmission. It has been proposed that a decreased dopaminergic tone results in a decreased signal-to-noise ratio in glutamatergic and GABAergic neuronal networks. As mentioned above, the parameter of the signal-to-noise ratio modifies the width of the focus of the activation of information stored in neuronal networks. If the signal-to-noise ratio is low, the access to stored information is unreliable. This unreliability, however, bears the advantage that unusual associations may be activated. Thus the probability is increased that remote and sometimes even "creative" ideas may occur. As has been demonstrated by neural network simulations, neuromodulatory states of low signal-to-noise ratios lead to an increased probability of changes in the stored content (increased neuroplasticity; see Spitzer et al. 1995a). Conversely, if the signal-to-noise ratio is too high, although the retrieval of stored information will occur with great reliability, the probability of changes in the network will be small. Moreover, under conditions of high signal-to-noise ratios, small environmental signals may be amplified such that even meaningless events become subjectively experienced as meaningful. Once meaning has been attributed to minor events through such processes, it may further affect perceptual and thought processes. The high signal-to-noise state of modulation further prevents the activation of alternate hypotheses, interpretations of events, and "world views" and thus any possibility of their consideration. This state therefore causes limitations in the width of processing, a narrowing of the contents of consciousness. Finally, as discussed above, a high signal-to-noise neuromodulatory state does not favor change. Hence, delusions, once developed, will not easily be changed by experience.

In high signal-to-noise neuromodulatory states, the probability of unusual, indirect associations becoming active, i.e., the possibility of creative problem-solving, is decreased. This has been demonstrated directly in word association experiments: In normal subjects, stress and anxiety lead to an increase in standard associations (e.g., black/white, mother/father) and thus to a decrease in unusual associations (Mintz 1969).

8.2 Chronic Delusions and Neuroplasticity

Neuroplasticity occurs throughout the entire lifetime of higher organisms. As far as self-organizing maps, in which representations of the external world at different levels of abstraction have been constructed and maintained in an orderly fashion, are implemented in the cortex, chronic delusion can be conceived of as deformations of cortical, high-level, map-like representations.

The term "structural deformation," coined by Janzarik, is indeed used in an almost identical way by neurobiologists to label dynamic neuroplastic changes (see Merzenich et al. 1988, Pascual-Leon and Torres 1993). It must be assumed that - similarly to what has been proved for somatosensory maps - input-driven enlargement and diminution of certain representations occurs in high-level multimodal cortical areas.

Once such deformations have occurred, they cannot easily be reversed. As stored information, they are anchored as strongly as other memory content; however, due to the substantial emotional participation in their genesis, they are anchored even more strongly. Psychological research has long known that, in the face of contradictory evidence, even healthy individuals are slow to change their attitudes, if at all. Furthermore, it is also known from experience in clinical psychotherapy that important changes in our views of things are often accompanied by anxiety. This is particularly true for delusional individuals with their already strong tendency toward anxiety. As a consequence, chronically delusional patients experience anxiety in the face of confrontation, even after the acute neuromodulatory changes have faded away or have been successfully treated. This in turn leads to an increased rigidity of the maps, particularly during confrontation with contradictory evidence.

8.3 Psychotherapy and Pharmacotherapy

The discussion of acute and chronic delusional phenomena within the cognitive neuroscience framework in terms of the principles of neuromodulation and neuroplasticity has practical consequences for the therapy of these disorders. Thus it is generally known that acute delusions respond relatively well to neuroleptic treatment, whereas chronic delusions do not. This is due directly to the fact that neuroleptics influence only the changed neuromodulatory states without having any direct influence on the deformations of long-term representations stored in cortical maps. In order to change these representations, new experiences by the patient are needed. A dull psychosocial environment can contribute as much to chronicity as inadequate medication can. This therapeutic consequence results from the neurobiological view of chronic delusions discussed. It is important to note that the neurobiological framework suggests not only pharmacotherapy alone, but also "milieu therapy"

in order to ensure new experiences in an environment. These must furthermore be as free of anxiety as possible. In other words, according to the model proposed here, the chronically delusionally ill benefit from neuromodulatory changes (pharmacologically or psychologically induced) only if their social environment changes for the better such that new healthy and normal experiences become possible.

9 Conclusions and Outlook

Clinical psychopathology, as far as it works with intuitively generated concepts and descriptions, will become of historical interest only. It will be superseded by neurobiologically derived terms and categories. The polemic debates on scientific methodology prevalent in the 1970s have given way to a methodologial pluralism rather than dogmatic, pretentious statements. In addition to what may be called "classical" psychopathology, which might also be described as macropsychopathology, empirical research will pin down and standardize the signs and symptoms of neurobiologically distinct states. The focus of research has thus shifted from the context of the individual life history of the patient to functional neurobiological mechanisms of pathology. This development has brought psychopathology (Andreasen 1997; von Praag 1998) back to the science of medicine, with the consequence of improved communication between medical specialties at the expense of decreased contact with the humanities, as represented by historians of psychiatry (Berrios and Porther 1995; Ballerini and Stanghellini 1993). Nonetheless, the life history of the patient is still an issue in emotions research and other research areas at the border of psychiatry and psychology. Future developments in the area of "classical" psychopathology are hard to predict. It appears to have become somewhat stuck, but may enjoy a renaissance, e.g., in connection with psychotherapeutic evaluation research.

There is currently no doubt that the greatest enthusiasm is generated by the "cognitive neurosciences" as the successors to classical psychopathology. These sciences of the mind have surmounted the methodological dualism and have integrated biological function and subjective experience. Their goal is to understand brain functions and subjective experience, and their methods, in particular functional neuroimaging techniques, have great potential for innovation. Nothing can better express the enthusiasm of this research approach and the belief in the integrative power of its methods than a quote from Eric Kandel (1991, p. 1030):

The boundary between behavioral studies and biology is arbitrary and changing. It has been imposed not by the natural contours of the disciplines, but by the lack of knowledge. As our knowledge expands, the biological and behavioral disciplines will merge at certain points, and it is at these points that our understanding of mentation will rest on secure ground.

It remains for the future to show whether this enthusiasm can deliver what it has promised, and to what extent the field of psychiatry within medicine can account for all aspects of the mental life of human beings.

10 References

Albert MS (1997) The science of the mind (editorial). Science 275: 1547

Alexander GE, Crutcher MD, DeLong MR (1990) Basal ganglia-thalamocortical circuits: parallel substrates for motor, oculomotor, "prefrontal" and "limbic" functions. Prog Brain Res 85: 119–146

Amminger GP, Kirkpatrick B (1997) "Primäre und sekundäre Negativsymptome: eine sinnvolle Differenzierung?" Kommentar zum Beitrag von W. Barnett, Ch. Mundt und P. Richter. Nervenarzt 68: 438–439

**Andreasen NC (1997) Linking mind and brain in the study of mental illnesses: a project for a scientific psychopathology. Science 275: 1586–1593

*Antonovsky A (1987) Unravelling the mystery of health: how people manage stress and stay well. Jossey-Bass, San Francisco

Avenarius R (1979) Emil Kraepelin. Seine Persönlichkeit und seine Konzeption. In: Janzarik W (ed) Psychopathologie als Grundlagenwissenschaft. Enke, Stuttgart, pp 62–73

Ballerini A, Stanghellini G (1993) Some remarks on dysphoria from an anthropological point of view. Psychopathology 26: 189–194

Barinaga M (1996) Giving language skills a boost. Science 271: 27–28

Barinaga M (1997) Visual system provides clues to how the brain perceives. Science 275: 1583–1585

Barnett W, Mundt C, Richter P (1996) Primäre und sekundäre Negativsymptomatik: eine sinnvolle Differenzierung? Nervenarzt 67: 558–563

Barnett W, Mundt C, Richter P (1997) Erwiderung auf Amminger und Kirkpatrick. Nervenarzt 68: 440

*Bauer J, Stadtmüller G, Qualmann J, Bauer H (1995) Prämorbide psychologische Prozesse bei Alzheimer-Patienten und bei Patienten mit vaskulären Demenzerkrankungen. Z Gerontol Geriatr 28: 179–189

Baumann U (1996) Wissenschaftliche Psychotherapie auf der Basis der wissenschaftlichen Psychologie. Report Psychol 21: 686–699

**Baxter LR, Schwartz JM, Bergman KS et al (1992) Caudate glucose metabolic rate changes with both drug and behavior

therapy for obsessive-compulsive disorder. Arch Gen Psychiatry 49: 681–689

*Becker EP (1995) Seelische Gesundheit und Verhaltenskontrolle. Eine integrative Persönlichkeitstheorie und ihre klinische Anwendung. Hogrefe, Göttingen

Becker EP (1996) Persönlichkeit. In: Ehlers A, Hahlweg K (eds) Klinische Psychologie, Vol 1. Psychologische und biologische Grundlagen. Hogrefe, Göttingen, pp 465–534 (Enzyklopädie der Psychologie, Series 2)

Beringer K (1927) Der Meskalinrausch. Springer, Berlin

Berrios JE, Porther R (eds) (1995) A history of clinical psychiatry. Athlone, London

Birley JLT (1990) DSM III: from left to right or right to left? Br J Psychiatry 157: 116–118

Blankenburg W (1971) Der Verlust der natürlichen Selbstverständlichkeit. Ein Beitrag zur Psychopathologie symptomarmer Schizophrenien. Enke, Stuttgart

Blashfield RK, Fuller AK (1996) Predicting the DSM 5. J Nerv Ment Dis 184: 4–7

Bleuler EL (1911) Dementia praecox oder die Gruppe der Schizophrenien. Deuticke, Leipzig

Blöschl L (1994) Zur Rolle hostiler Tendenzen in der Depression: Verhaltensdiagnostische Aspekte. In: Bartussek D, Amelang M (eds) Fortschritte der differentiellen Psychologie und psychologischen Diagnostik. Festschrift zum 60. Geburtstag von Kurt Pawlik. Hogrefe, Göttingen, pp 259–267

Bonhoeffer K (1910) Die somatischen Psychosen im Gefolge von akuten Infektionen und inneren Erkrankungen. Deuticke, Leipzig

Breiter HC, Rauch S, Kwong KK, Baker JR, Weisskoff RM, Kennedy DN, Kendrick AD, Davis TL, Jiang A et al (1996) Functional magnetic resonance imaging of symptom provocation in obsessive compulsive disorder. Arch Gen Psychiatry 53: 595–606

Brodman K (1909) Vergleichende Localisationslehre der Großhirnrinde in ihren Prinzipien dargestellt auf Grund des Zellenbaues. Barth, Leipzig

*Brown GW (1996) Onset and course of depressive disorders: summary of a research programme. In: Mundt C, Goldstein MJ, Hahlweg K, Fiedler P (eds) Interpersonal factors in the origin and course of affective disorders. Gaskell, London, pp 51–167

Bürgy M, Mundt C (1997) Methodenprobleme der Biographieforschung über Lebensthemen. In: Saß H (ed) Lebensgeschichte und Psychopathologie. Fischer, Stuttgart

Carpenter WT, Heinrichs DW, Wagman AMI (1988) Deficit and nondeficit forms of schizophrenia: the concept. Am J Psychiatry 145: 578–583

Cloninger CR, Sigvardsson S, Przybeck TR, Svrakic DM (1995) Personality antecedents of alcoholism in a national area probability sample. Eur Arch Psychiatry Clin Neurosci 245: 239–244

Cloninger CR (1987) A systematic method for clinical description and classification of personality variants: a proposal. Arch Gen Psychiatry 44: 573–588

Cohen J, Servan-Schreiber D (1992) Context, cortex and dopamine: a connectionist approach to behavior and biology in schizophrenia. Psychol Rev 12: 45–77

*Cohen JD, Servan-Schreiber D (1993) A theory of dopamine function and its role in cognitive deficits in schizophrenia. Schizophr Bull 19: 85–104

Crow TJ (1980) Molecular pathology of schizophrenia: more than one disease process? Br Med J 280: 1–9

Davis KL, Kahn RS, Ko G, Davidson M (1991) Dopamine in schizophrenia: a review and reconceptualization. Am J Psychiatry 148: 1474–1486

Dilling H, Dittmann V (1990) Die Psychiatrische Diagnostik nach der 10. Revision der internationalen Klassifikation der Krankheiten (ICD-10). Nervenarzt 61: 259–270

**Ekman P (1993) Facial expression and emotion. Am Psychol 48: 384–392

Ekman P, Friesen WV (1969) The repertoire of nonverbal behavior. Categories, origins, usage and coding. Semiotica 1: 49–98

Emrich HM (1990) Psychiatrische Anthropologie. Pfeiffer, Munich

Faller H (1994) Das Forschungsprogramm "Qualitative Psychotherapieforschung". Versuch einer Standortbestimmung. In: Faller H, Frommer J (eds) Qualitative Psychotherapieforschung. Asanger, Heidelberg, pp 15–37

Farah MJ, Wallace MA (1992) Semantically-bounded anomia: implications for the neural implementation of naming. Neuropsychologia 30: 609–621

Felleman DJ, van Essen DC (1991) Distributed hierarchical processing in the primate cerebral cortex. Cereb Cortex 1: 1–47

*Fiedler P (1995) Persönlichkeitsstörungen, 2nd edn. Beltz Psychologie Verlagsunion, Weinheim

Fiedler P (1997) Therapieplanung in der modernen Verhaltenstherapie. Von der allgemeinen zur phänomen- und störungsspezifischen Behandlung. Verhaltensther Verhaltensmed 18: 7–39

Flechsig P (1929) Anatomie des menschlichen Gehirns und Rückenmarks auf myelogenetischer Grundlage. Thieme, Leipzig

Freyberger HJ, Dittmann V, Stieglitz RD, Dilling H (1990) ICD-10 in der Erprobung: Ergebnisse einer multizentrischen Feldstudie in den deutschsprachigen Ländern. Nervenarzt 61: 271–275

*Frijda NH (1986) The emotions. Cambridge University Press, Cambridge

Frommer J (1996) Qualitative Diagnostikforschung. Springer, Berlin Heidelberg New York

Frost DO (1997) Novel, surgically-induced neural connections: lessons for development, function and disease states. Lecture at the International Congress on Schizophrenia Research, Colorado Springs, 12 April 1997

Fuchs T (1993) Wahnsyndrome bei sensorischer Beeinträchtigung – Überblick und Modellvorstellungen. Fortschr Neurol Psychiatr 61: 257–266

Fuchs T (1994a) Uprooting and late-life psychosis. Eur Arch Psychiatry Clin Neurosci 244: 126–130

*Fuchs T (1994b) Die Welt als Innenraum. Kafkas "Bau" als Paradigma paranoider Räumlichkeit. Nervenarzt 65: 470–477

Fuchs T (1995) Auf der Suche nach der verlorenen Zeit – die Erinnerung in der Demenz. Fortschr Neurol Psychiatr 63: 38–43

*Fuchs T (1996) Leibliche Kommunikation und ihre Störungen. Z Klin Psychol Psychiatr Psychother 44: 415–428

Fuchs T, Haupt M (1994) Schutzmächte bei Altersparaphrenien. Nervenarzt 65: 345–349

Fydrich T (1996) Komorbidität psychischer Störungen. Empirische Untersuchung zu einem umstrittenen Konzept. Postdoctoral thesis, University of Heidelberg

Fydrich T, Schmitz B, Hennch C, Bodem M (1996) Zuverlässigkeit und Gültigkeit diagnostischer Verfahren zur Erfassung

von Persönlichkeitsstörungen. In: Schmitz B, Fydrich T, Limbacher K (eds) Persönlichkeitsstörungen: Diagnostik und Psychotherapie. Beltz Psychologie Verlagsunion, Weinheim, pp 91–113

Gross G, Huber G, Saß H (1997) Moderne psychiatrische Klassifikationssysteme: Implikationen für Diagnose und Therapie, Forschung und Praxis. Schattauer, Stuttgart

Häfner H, Maurer K (1991) Are there two types of schizophrenia? True onset and sequence of positive and negative syndromes prior to first admission. In: Marneros A, Andreasen NC, Tsuang MT (eds) Negative versus positive schizophrenia. Springer, Berlin Heidelberg New York, pp 134–159

Häfner H, Riecher-Rössler A, Maurer K, Fätkenheuer B, Löffler W (1992a) First onset and early symptomatology of schizophrenia. Eur Arch Psychiatry Clin Neurosci 242: 109–118

Häfner H, Riecher-Rössler A, Hambrecht M et al (1992b) IRAOS: an instrument for the assessment of onset and early course of schizophrenia. Schizophr Res 6: 209–223

Haltenhof H, Vossler A (1994) Coping der Depression: Wie begegnen depressive Patienten ihrer Erkrankung? Eine Literaturübersicht. Z Klin Psychol Psychiatr Psychother 42: 201–229

Hambrecht M, Häfner H (1993) "Trema, Apohänie, Apokalypse" – ist Conrads Phasenmodell empirisch begründbar? Fortschr Neurol Psychiatr 61: 418–423

Hammen C (1996) Stress, families, and the risk for depression. In: Mundt C, Goldstein MJ, Hahlweg K, Fiedler P (eds) Interpersonal factors in the origin and course of affective disorders. Gaskell, London, pp 101–112

Harrison PJ (1991) Are mental states a useful concept? Neural philosophical influences on phenomenology and psychopathology. J Nerv Ment Dis 179: 309–316

Haviland JM, Lelwica M (1987) The induced affect response: 10-week-old infants' responses to three emotional expressions. Dev Psychol 23: 97–104

Heinze HJ, Mangun GR, Burchert W et al (1994) Combined spatial and temporal imaging of brain activity during visual selective attention in humans. Nature 372: 543–546

Hermle L, Gouzoulis-Mayfrank E, Spitzer M (1996) Halluzinogeninduzierte psychische Störungen. Subjektives Erleben, Psychopathologie und Differentialdiagnose. Fortschr Neurol Psychiatr 64: 482–491

Herpertz S, Saß H (1997) Impulsivität und Impulskontrolle. Zur psychologischen und psychopathologischen Konzeptionalisierung. Nervenarzt 68: 171–183

Hops H (1996) Intergenerational transmission of depressive symptoms: gender and developmental considerations. In: Mundt C, Goldstein MJ, Hahlweg K, Fiedler P (eds) Interpersonal factors in the origin and course of affective disorders. Gaskell, London, pp 113–129

Janssen PL, Schneider W (eds) (1994) Diagnostik in Psychotherapie und Psychosomatik. Fischer, Stuttgart

**Janzarik W (1988) Die strukturdynamischen Grundlagen der Psychiatrie. Enke, Stuttgart

Jaspers K (1957) Philosophische Autobiographie. In: Schilpp PA (ed) Philosophen des 20. Jahrhunderts. Kohlhammer, Stuttgart, pp 1–79

Jaspers K (1965) Allgemeine Psychopathologie, 8th edn. Springer, Berlin Heidelberg New York

Jeste D (1997) Clinical aspects of late-onset schizophrenia. In: Marneros A (ed) Late onset mental disorders. Gaskell, London

Just MA, Carpenter PA (1992) A capacity theory of comprehension: individual differences in working memory. Psychol Rev 99: 122–149

Kaczmarek LK, Levitan IB (1987) Neuromodulation: the biochemical control of neuronal excitability. Oxford University Press, New York

Kandel E (1991) Genes, environmental experience, and the mechanisms of behavior. In: Kandel E, Schwartz J, Jessel T (eds) Principles of neural science. Elsevier, New York, pp 1009–1031

Kaplan HI, Sadock BJ, Grebb JA (1994) Synopsis of psychiatry, 7th edn. Williams and Wilkins, Baltimore

Kick H (1991) Psychopathologie und Verlauf der postakuten Schizophrenie. Springer, Berlin Heidelberg New York

Kirkpatrick B, Buchanan RW, McKenney PD, Alphs LD, Carpenter WT (1989) The schedule for the deficit syndrome: an instrument for research in schizophrenia. Psychiatry Res 30: 119–123

Kischka U, Kammer T, Weisbrod M, Maier S, Thimm M, Spitzer M (1996) Dopaminergic modulation of semantic network activation. Neuropsychologia 34: 1107–1113

Kischka U, Kammer T, Spitzer M (in press) Frontale subkortikale neuronale Schaltkreise. Fortschr Neurol Psychiatr

Kisker KP (1960) Der Erlebniswandel des Schizophrenen. Ein psychopathologischer Beitrag zur Psychonomie schizophrener Grundsituationen. Springer, Berlin Göttingen Heidelberg

*Klosterkötter J (1988) Basissymptome und Endphänomene der Schizophrenie. Eine empirische Untersuchung der psychopathologischen Übergangsreihen zwischen defizitären und produktiven Schizophreniesymptomen. Springer, Berlin Heidelberg New York

Klosterkötter J, Gross G, Huber G, Steinmeyer EM (1997) Sind selbstwahrnehmbare neuropsychologische Defizite bei Patienten mit Neurose- oder Persönlichkeitsdiagnosen für spätere schizophrene Erkrankungen prädiktiv? Nervenarzt 68: 196–204

Kolb B (1995) Brain plasticity and behavior. Erlbaum, Mahwah

**Kosslyn SM, König O (1992) Wet mind: the new cognitive neuroscience. Free Press/Macmillan, New York

Kosslyn SM, Thompson WL, Kim IJ, Rauch SL, Alpert NM (1996) Individual differences in cerebral blood flow in area 17 predict the time to evaluate visualized letters. J Cogn Neurosci 8: 78–82

*Kraus A (1991) Phänomenologische und symptomatologisch-kriteriologische Diagnostik. Fundam Psychiatr 5: 102–109

Kraus A (1992) Lügenmotiv und Depersonalisation in der Melancholie. In: Schmitt W, Hofmann W (eds) Phänomen – Struktur – Psychose. Roderer, Regensburg, pp 137–146

Kraus A (1994) Phenomenology of the technical delusion in schizophrenics. J Phenomenol Psychol 25(1): 51–69

Kraus A (1996) Spezifität melancholischer Verstimmung und Angst. In: Lang H, Faller H (eds) Das Phänomen Angst. Suhrkamp, Frankfurt am Main, pp 103–121

Krause R, Lütolf P (1989) Mimische Indikatoren von Übertragungsvorgängen – Erstuntersuchungen. Z Klin Psychol 18: 55–67

Krause R, Steimer-Krause E, Hufnagel H (1992) Expression and experience of affects in paranoid schizophrenics. Eur Rev Appl Psychol 42: 132–138

Lauer G (ed) (1997) Die Lebensqualität in der Psychiatrie. Deutsche und internationale Perspektiven. Enke, Stuttgart

Laux L, Weber H (1987) Erträgnisse biographischer Forschung im Bereich Streß und -Bewältigung. In: Jüttemann G,

Thomae H (eds) Biographie und Psychologie. Springer, Berlin Heidelberg New York, pp 285–298

LeDoux JE (1989) Cognitive-emotional interactions in the brain. Cogn Emotion 3: 267–289

Leopold DA, Logothetis NK (1996) Activity changes in early visual cortex reflect monkeys' percepts during binocular rivalry. Nature 379: 549–553

Liddle PF, Morris DL (1991) Schizophrenic syndromes and frontal lobe performance. Br J Psychiatry 158: 340–345

Lindenmayer JP, Bernstein-Hüman R, Grochowski S, Bark N (1995) Psychopathology of schizophrenia: initial validation of a five-factor model. Psychopathology 28: 22–31

Linehan M (1993) Cognitive-behavioral treatment of borderline personality disorder. Guilford, New York

Logothetis NK, Leopold DA, Sheinberg DL (1996) What is rivalling during binocular rivalry? Nature 380: 621–624

Lösel F, Kolip P, Bender D (1992) Stressresistenz im Multiproblemmilieu. Sind seelisch widerstandsfähige Jugendliche "Super-Kids"? Z Klin Psychol 21: 48–63

Machleidt W, Gutjahr L, Muegge A (1989) Grundgefühle. Phänomenologie, Psychodynamik, EEG Spektralanalytik. Springer, Berlin Heidelberg New York

Maier S (1998) Neuropsychologie und subjektive Krankheitsverarbeitung in der Rehabilitation. Ergebnisse einer Verlaufsstudie. Nervenarzt 69[Suppl]: 44

Markowitsch HJ (1995) Which brain regions are critically involved in the retrieval of old episode memory? Brain Res Rev 21: 117–128

Marneros A (ed) (1997) Late onset mental disorders. Gaskell, London

*Marneros A, Deister A, Rohde A (1991) Affektive, schizoaffektive und schizophrene Psychosen. Eine vergleichende Langzeitstudie. Springer, Berlin Heidelberg New York

Maurer K, Hillig A, Freyberger HJ, Velthaus S (1991) Erfahrungen mit den "Schedules for Clinical Assessment in Neuropsychiatry (SCAN)" im Rahmen einer multizentrischen Feldstudie. Schweiz Arch Neurol Psychiatr 142: 235–245

McClelland JL, McNaughton BL, O'Reilly RC (1995) Why there are complementary learning systems in the hippocampus and neocortex: insights from the success and failures of connectionist models of learning and memory. Psychol Rev 102: 419–457

McKhann G, Drachman D, Folstein M, Katzmann R, Price D, Stutlan EM (1984) Clinical diagnosis of Alzheimer disease. Neurology 34: 939–944

Merzenich MM, Sameshima K (1993) Cortical plasticity and memory. Curr Opin Neurol 3: 187–196

Merzenich MM, Recanzone G, Jenkins WM, Allard TT, Nudo RT (1988) Cortical representational plasticity. In: Rakic P, Singer W (eds) Neurobiology of neocortex. Wiley, Chichester, pp 41–67

Merzenich MM, Jenkins WM, Johnston P, Schreiner C, Miller SL, Tallal P (1996) Temporal processing deficits of language-learning impaired children ameliorated by training. Science 271: 77–81

Meyer A (1960) The thirty-fourth Maudsley lecture: emergent patterns of the pathology of mental disease. J Ment Sci 106: 785

Mezzich JE (1988) On developing a psychiatric multiaxial schema for ICD 10. Br J Psychiatry 152[Suppl 1]: 38–43

Mezzich JE (1992) Multiachsiale Diagnostik und internationale Klassifikation in der Psychiatrie. Fundam Psychiatr 6: 150–153

Mintz S (1969) Effect of actual stress on word associations. J Abnorm Psychol 74: 293–295

Mombour W, Spitzner S, Reger KH, von Cranach M, Dilling H, Helmchen H (1990) Summary of the qualitative criticisms made during the ICD 10. Field trial and remarks on the german translation of ICD 10. Pharmacol Psychiatry 23[Suppl]: 197–201

Monsch AU, Bondi MW, Butters N, Paulsen JS, Salmon DP, Brugger P, Swenson MR (1994) A comparison of category and letter fluency in Alzheimer's disease and Huntington's disease. Neuropsychology 8(1): 25–30

Morris P, Rapoport SI (1990) Neuroimaging and affective disorder in late life: review. Can J Psychiatry 35(4): 347–354

Mundt C (1985) Das Apathiesyndrom der Schizophrenen. Eine psychopathologische und computertomographische Untersuchung. Springer, Berlin Heidelberg New York

Mundt C (1986) Kleptomanie. In: Müller C (ed) Lexikon der Psychiatrie, 2nd edn. Springer, Berlin Heidelberg New York, pp 393–396

Mundt C (1989) Pychopathologie heute. In: Kisker KP, Lauter H, Meyer JE, Müller C, Strömgren E (eds) Psychiatrie der Gegenwart, 3rd edn, vol 9: Brennpunkte der Psychiatrie. Springer, Berlin Heidelberg New York, pp 147–184

Mundt C (1991) Constituting reality – its decline and repair in the course of schizophrenic psychoses: the intentionality model. In: Marneros A, Andreasen NC, Tsuang MT (eds) Negative versus positive schizophrenia. Springer, Berlin Heidelberg New York, pp 96–108

Mundt C (1996a) Zur Psychotherapie des Wahns. Nervenarzt 67: 515–523

Mundt C (1996b) Entfremdete und erlebte Angst im therapeutischen Prozeß. Überlegungen zur Problematik des Symptomcharakters von Angst anhand eines Falles von psychogenem Schwindel. In: Lang H, Faller H (eds) Das Phänomen Angst. Pathologie, Genese und Therapie. Suhrkamp, Frankfurt, pp 179–190

Mundt C (1998) Die Relevanz moderner Klassifikationssysteme für die psychotherapeutische Behandlung. In: Gross G, Huber G, Saß H (eds) Moderne psychiatrische Klassifikationssysteme: Implikationen für Diagnose und Therapie, Forschung und Praxis. Schattauer, Stuttgart New York, pp 220–229

Mundt C, Barnett W, Witt G (1995) The core of negative symptoms in schizophrenia. Affect or cognitive deficiency? Psychopathology 28: 46–54

Mundt C, Backenstrass M, Kronmüller KT, Fiedler P, Kraus A Stanghellini G (1997) Personality and endogenous/major depression: an empirical approach to typus melancholicus. 2. Validation of typus melancholicus core-properties by personality inventory scales. Psychopathology 30: 130–139

Mundt C, Richter P, van Hees H, Stumpf T (1998) Zeiterleben und Zeitschätzung bei endogen und neurotisch depressiven Patienten. Nervenarzt 69: 38–45

Murray L, Hipwell A, Hooper R (1996) The cognitive development of 5-year-old children of postnatally depressed mothers. J Child Psychol 37/8: 927–935

Nauta WJH, Feirtag M (1986) Fundamental neuroanatomy. Freeman, New York

O'Craven KM, Rosen BR, Kwong KK, Treisman A, Savoy RL (1997) Voluntary attention modulates fMRI activity in human MT-MST. Neuron 18: 591–598

Pantel J, Schröder L, Schad R et al (1997) Quantitative magnetic resonance imaging and neuropsychological functions in dementia of the Alzheimer type. Psychol Med 27: 221–229

Pascual-Leone A, Torres F (1993) Plasticity of the sensorimotor cortex representation of the reading finger in Braille readers. Brain 116: 39–52

Posner MI (1986) Chronometric explorations of mind. Oxford University Press, New York

Posner MI, Raichle M (1996) Bilder des Geistes. Spektrum, Heidelberg

Radden J (1994) Recent criticism of psychiatric nosology: a review. Philos Psychiatry Psychol 1: 193–200

**Resch F (1996) Entwicklungspsychopathologie des Kindes- und Jugendalters. Ein Lehrbuch. Beltz Psychologie Verlagsunion, Weinheim

Riecher-Rössler A (1995) Die Spätschizophrenie – eine valide Entität? Postdoctoral thesis, University of Heidelberg

Riecher-Rössler A (1997) 50 Jahre nach Manfred Bleuler. Was wissen wir heute über die Spätschizophrenien? Nervenarzt 68: 159–170

Ritter H, Kohonen T (1989) Self-organizing semantic maps. Biol Cybern 61: 241–254

Rudolf G (1993) Psychotherapeutische Medizin. Enke, Stuttgart

Sartorius N, Kaelber CT, Cooper JE et al (1993) Progress toward achieving a common language in psychiatry. Results from the field trial of the clinical guidelines accompanying the WHO-classification of mental and behavioral disorders in ICD 10. Arch Gen Psychiatry 50: 115–123

Saß H (1987) Die Krise der psychiatrischen Diagnostik. Fortschr Neurol Psychiatr 55: 355–360

Saß H (1994) Zur Problematik der operationalen Diagnostik in der Psychiatrie. In: Dilling H, Schulte-Markwort E, Freyberger HJ (eds) Von der ICD 9 zu der ICD 10. Neue Ansätze der Diagnostik psychischer Störungen in der Psychiatrie, Psychosomatik und Kinder- und Jugendpsychiatrie. Huber, Bern, pp 149–156

Saß H, Zaudig M, Houben I, Wittchen UH (1994) Einführung zur deutschen Ausgabe: Zur Situation der operationalisierten Diagnostik in der deutschsprachigen Psychiatrie. In: Saß H, Wittchen UH, Zaudig M (eds) Diagnostisches und statistisches Manual psychischer Störungen DSM-IV. Hogrefe, Göttingen, pp IX–XXIV

Scherer KR (1993) Studying the emotion-antecedent appraisal process: an expert system approach. Cogn Emotion 7(3,4): 325–355

*Schmidt-Degenhard M (1992) Die oneiroide Erlebnisform. Zur Problemgeschichte und Psychopathologie des Erlebens fiktiver Wirklichkeiten. Springer, Berlin Heidelberg New York

Schmidt-Degenhard M (1997a) Zur Standortbestimmung einer anthropologischen Psychiatrie. Fortschr Neurol Psychiatr 65: 435–480

Schmidt-Degenhard M (1997b) Die psychiatrische Exploration als offenes Feld zwischen Betroffensein und Verstehen. In: Jacobi RME (ed) Selbstorganisation. Duncker and Humblot, Berlin, pp 217–228 (Jahrbuch für Komplexität in den Natur-, Sozial- und Geisteswissenschaften, vol 7)

Schmitt W, Mundt C (1991) Zur Differentialtypologie von Patienten mit harten und weichen Suizidmethoden. Nervenarzt 62: 440–444

Schneider K (1976) Klinische Psychopathologie, 11th edn. Thieme, Stuttgart

Schröder B, Hahlweg K, Fiedler P Mundt C (1996) Marital interaction in couples with a depressed or schizophrenic patient. In: Mundt C, Goldstein MJ, Hahlweg K, Fiedler P (eds) Interpersonal factors in the origin and course of affective disorders. Gaskell, London, pp 257–276

*Schröder J (1997) Subsyndrome der chronischen Schizophrenie. Untersuchungen an bildgebenden Verfahren zur Heterogenität schizophrener Psychosen. Springer, Berlin Heidelberg New York

Schröder P (1915) Von den Halluzinationen. Monatsschr Psychiatr Neurol 37: 1–11

Schwartz MA (1991) Neural philosophy, psychopathology, and clinical psychiatric science. A commentary on Harrison's "Are mental states a useful concept?" J Nerv Ment Dis 179: 317–319

Sereno MI, Dale AM, Reppas JB et al (1995) Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. Science 268: 889–893

Shepard RN, Cooper LA (1982) Mental images and their transformations. MIT, Cambridge, MA

Silbersweig DA, Stern E, Frith C et al (1995) A functional neuroanatomy of hallucinations in schizophrenia. Nature 378: 176–179

Snyder AZ, Abdullaev YG, Posner MI, Raichle ME (1995) Scalp electrical potentials reflect regional cerebral blood flow responses during processing of written words. Proc Natl Acad Sci USA 92: 1689–1693

Spitzer M (1988) Halluzinationen. Springer, Berlin Heidelberg New York

**Spitzer M (1996) Geist im Netz. Modelle für Lernen, Denken und Handeln. Spektrum, Heidelberg

Spitzer M (1997a) A cognitive neuroscience view of schizophrenic thought disorder. Schizophr Bull 23: 29–50

Spitzer M (1997b) Neuronale Netzwerke und Psychopathologie. Nervenarzt 68: 21–37

Spitzer RL, Fleiss JL (1974) A re-analysis of the reliability of psychiatric diagnosis. Br J Psychiatry 125: 341–347

Spitzer M, Kammer T (1996) Combining neuroscience research methods in psychopathology. Curr Opin Psychiatry 9: 352–363

Spitzer M, Böhler P, Kischka U, Weisbrod M (1995a) A neural network model of phantom limbs. Biol Cybern 72: 197–206

Spitzer M, Kwong KK, Kennedy W, Rosen BR, Belliveau JW (1995b) Category-specific brain activation in fMRI during picture naming. Neuroreport 6: 2109–2112

Spitzer M, Thimm M, Hermle L et al (1996) Increased activation of indirect semantic associations under psilocybin. Biol Psychiatry 39: 1055–1057

Spitzer M, Winkler S, Maier S, Weisbrod M (1997) Ereigniskorrelierte Potentiale bei semantischen Sprachverarbeitungsprozessen schizophrener Patienten. Nervenarzt 68: 212–225

Spitzer M, Kammer T, Bellemann ME et al (1998) Funktionelle Magnetresonanztomographie in der psychopathologischen Forschung. Fortschr Neurol Psychiatr 66: 241–258

*Squire LR (1986) Mechanisms of memory. Science 232: 1612–1619

Squire LR, Zola-Morgan S (1991) The medial temporal lobe memory system. Science 253: 1380–1386

Steinmeyer EM, Möller HJ (1992) Facet theoretic analysis of the Hamilton-D-Scale. J Affect Disord 25: 53–62

Svrakic NM, Svrakic DM, Cloninger CR (1996) A general quantitative theory of personality development: fundamentals of a self-organizing psychobiological complex, development and psychopathology. Psychopathology 8: 247–272

Tallal P, Miller SL, Bedi G et al (1996) Language comprehension in language-learning impaired children improved with acoustically modified speech. Science 271: 81–84

Tellenbach H (1983) Melancholie, 4th edn. Springer, Berlin Heidelberg New York

**Thomae H (1996) Das Individuum und seine Welt. Eine Persönlichkeitstheorie, 3rd edn. Hogrefe, Göttingen

Toga AW, Mazziotta JC (1996) Brain mapping. The methods. Academic, San Diego

von Baeyer H (1955) Der Begriff der Begegnung in der Psychiatrie. Nervenarzt 26: 369–376

*von Praag HM (1988) Serotoninstörungen bei psychischen Erkrankungen. Funktionelle versus nosologische Interpretation. Adv Biol Psychiatr 17: 1–7

von Zerssen D (1977) Premorbid personality and affective psychoses. In: Burrows GD (ed) Handbook of studies on depression. Excerpta Medica, Amsterdam, pp 79–103

von Zerssen D, Pössl J, Gruben S, Tauscher R, Barthelmes H (1994) An operationalized procedure for the recognition of premorbid personality types in biographical case notes on psychiatric patients. Eur Arch Psychiatry Clin Neurosci 243: 256–272

*von Zerssen D, Barthelmes H, Black C et al (1996) Das biographische Persönlichkeitsinterview (BPI): Ein Forschungsinstrument zur Erfassung der prämorbiden Persönlichkeit. In: Möller HJ, Engel R, Hoff P (eds) Befunderhebung in der Psychiatrie: Lebensqualität, Negativsymptomatik und andere aktuelle Entwicklungen. Springer, Berlin Heidelberg New York, pp 303–307

Walker C (1993) Karl Jaspers as a Kantian psychopathologist. Hist Psychiatry 4: 209–238, 321–348

Walker E, Lewine RJ (1990) Prediction of adult onset schizophrenia from childhood home movies of the patients. Am J Psychiatry 147: 1052–1056

Weitbrecht HJ (1966) Die heutige Diskussion über das Wesen der endogenen Psychosen. Fortschr Neurol Psychiatr 34: 161–175

Wieck HH (1961) Zur klinischen Stellung des Durchgangssyndroms. Schweiz Arch Neurol Psychiatr 88: 4409–4419

Wiggins OP, Schwartz MA (1996) Chris Walter's interpretation of Karl Jaspers' phenomenology: a critique. Philos Psychiatry Psychol 2: 319–343

Wing JK, Babor T, Brugha T et al (1990) SCAN: Schedules for the Clinical Assessment in Neuropsychiatry. Arch Gen Psychiatry 47: 589–593

Yeh SR, Fricke RA, Edwards DH (1996) The effects of social experience on serotonergic modulation of the escape circuit of crayfish. Science 271: 366–369

Zilles K, Schlaug G, Matelli M et al (1995) Mapping of human and macaque sensorimotor areas by integrating architectonic, transmitter receptor, MRI and PET data. J Anat 187: 515–537

CHAPTER

2

A.S. Henderson

Central Issues in Psychiatric Epidemiology

Appreciation is expressed to my colleagues at the NHMRC Psychiatric Epidemiology Research Centre who critically read drafts of this chapter.

1 Introduction

Most of clinical psychiatry is about individuals. However, when these individuals are considered collectively, looking for general patterns, we enter the field of epidemiology. In his Herter Lectures at the Johns Hopkins University, Greenwood (1931) gave the most succinct definition of epidemiology as "the mass aspects of disease." The word comes directly from επιδημιοσ (*epidímios*), meaning a disease that is "prevalent among the people", compounded from επι (*epí*, upon or among) and δημος (*dímos*, the people in a district or country). It is the business of psychiatric epidemiology to determine the distribution of mental disorders in populations rather than in individuals, the factors determining that distribution and measures that may help in prevention. On the one hand, everyone is unique, each individual having his or her own genetic endowment and life experiences. Such an *idiographic* paradigm is balanced by the epidemiological one, which is a *nomothetic* paradigm because it searches for recurrent and predictable patterns in the whole of humankind. In this search, there are five fundamental requirements: a hypothesis, a population, specification of the psychiatric disorders to be investigated, specification of variables that may contribute to their onset or their outcome and scientifically adequate instruments for measuring these. The enquiry can be cross-sectional, examining the situation at one point in time, or it may be longitudinal, measuring change over a certain time period and trying to determine what brings about that change.

J.N. Morris (1957) described seven uses of epidemiology in his celebrated monograph bearing that title. Although this work has been repeatedly quoted, it has not yet become trite. It provides a framework for assessing the current state of *psychiatric* epidemiology in relation to current national and global conditions and to the experience of clinicians.

The first use of epidemiology is the historical study of the health of communities and of the rise and fall of diseases in the population, with the possibility of projections into the future, e.g. the incidence of schizophrenia may be declining and the disease may be becoming more benign in its clinical course; depressive disorder may have become more frequent in young persons (their suicide rate has indisputably increased in many industrialised countries); it is likely that eating disorders have increased in frequency; and it is certain that the use of heroin and the acquired immunodeficiency syndrome (AIDS) epidemic with its neuropsychiatric sequelae are new arrivals and will be a continuing burden.

A second purpose is community diagnosis, in which estimates of morbidity are obtained at the general population level, and not just in individuals reaching professional services. Recent examples are the National Comorbidity Study in the the United States (Kessler et al. 1994) and the Survey of Psychiatric Morbidity in the United Kingdom (Jenkins and Meltzer 1995; Jenkins et al. 1997a,b).

A third purpose is the study of the working of health services, starting by determining needs and resources, then analysing services currently in action and finally attempting to evaluate these services.

A fourth use is of direct value in clinical practice and consists in using the population data to hand to estimate an individual's chances of developing a particular disease, or of recovering or of relapsing, e.g. if the annual incidence rate for schizophrenia is known in a population, and if this information is age specific, it is possible to estimate the probability that a person aged, say, 20 years, will develop this disorder within the next 12 months. This is the base rate, before one starts to consider risk factors such as family history. Likewise, if the lifetime prevalence of schizophrenia is known, the chances of an individual developing the disease at some stage in life is also known. By aggregating data on the course of schizophrenia in treated sample populations, it is possible to estimate the chances of recovery or partial recovery for a person who has just had a first episode. The central issue is that data on large numbers of subjects are used to make a probability estimate for individuals.

The fifth use is a particularly attractive one: to help complete the clinical picture. By this, Morris meant knowing about all presentations and stages of a disease and then relating subclinical cases to fully developed ones. An excellent example here would be anxiety or depressive states that are to be seen in general practice or field surveys, compared to the more severe syndromes specified in the international criteria and encountered by mental health staff.

The sixth use of epidemiology is to identify syndromes by examining the distribution of clinical phenomena occurring in sections of the population. This fits well with recent experience of repetitive strain injury, chronic fatigue syndrome and post-traumatic stress disorder or its congeners.

The seventh use of epidemiology is the most appealing of all: "the search for causes of health and of disease, starting with the discovery of groups with high and low rates, studying these differences in relation to differences in ways of living; and, where possible, testing these notions in actual practice among populations."

To Morris's seven uses of epidemiology, a further use should be added, namely, prevention, which Ernest Gruenberg (1966) referred to as its "ultimate service."

Epidemiological research has had some successes. The more basic, so-called descriptive studies were valuable in demonstrating that mental disorders occur not only in institutional settings, but also to a substantial extent in all general populations. This has helped the governments of many countries to realise that better services are needed for people with mental disorders, particularly in the community, where the bulk of morbidity occurs. Some scientifically more advanced studies, known as analytic epidemiology, have made major contributions. For example, in schizophrenia, a slight excess of births in the late winter and early spring was observed. This was followed by an association being found between maternal influenza in the second trimester of pregnancy and subsequent development of schizophrenia in offspring. A separate observation was that the median age for the onset of schizophrenia is about 5 years later in women than in men. This led to a painstaking search for possible social and biological explanations, as reviewed by Jablensky (1995). Another remarkable achievement was the *International Pilot Study on Schizophrenia* (WHO 1973; Leff et al. 1992) and the subsequent follow-up entitled the *Determinants of Outcome in Severe Mental Disorder* (Jablensky et al. 1992). These international projects, conducted through the WHO across many regions, have shown that schizophrenia is ubiquitous; presentation is similar throughout the world, while its prognosis tends to be more favourable in less industrialised countries. In the epidemiology of Alzheimer's disease, identification of several risk factors was achieved through case-control studies. The association with Down's syndrome provided a lead for molecular geneticists to focus on chromosome 21 and the mechanism for formation of amyloid precursor protein in Alzheimer's disease.

It is very desirable for clinicians to be vigilant for unexpected opportunities to carry out epidemiological studies. These may arise when a population is recognised to have some relatively uncommon attribute that could throw light on aetiology or on outcome. Such attributes may be strikingly high or low rates for some group of disorders. Alternatively, the attributes may consist in a certain type of exposure, past or present, that putatively increases the incidence of a particular mental disorder. An example is the recent investigation by Susser and Lin (1992, 1994) on the incidence of schizophrenia in the offspring of women exposed to the Dutch Famine of 1944 or studies of communities exposed to an earthquake (Carr et al. 1997).

In the development of administrative policy and the need for services for the mentally ill, epidemiological studies have at times provided powerful evidence for advocacy purposes. The Epidemiologic Catchment Area (ECA) Studies (Robins and Regier 1991) and the later National Comorbidity Survey in the United States produced estimates of the frequency of mental disorders which have done much to promote mental health issues in many countries. The latter study found that nearly 50% of respondents had experienced at least one psychiatric disorder in their lifetime, and nearly 30% had had at least one in the previous 12 months. Of this group with a recent disorder, less than 20% had had some form of professional treatment. The most common conditions were anxiety and affective disorders, with 12-month prevalence rates of 17% and 11%, respectively. An important finding was the extent of comorbidity, here meaning the occurrence of two or more psychiatric disorders in an individual over his or her lifetime. It was found that the major burden of psychiatric disorder was concentrated in about one sixth of the population.

Epidemiology can be abused or misused. This happens under four broad circumstances. First, health administrators may sometimes mistakenly accord great value to obtaining local or regional data on the pattern of morbidity in their own community. When human and financial resources for mental health services are scarce in a region, it can rarely be justifiable for these to be deployed in obtaining data when the extent of unmet need is already obvious. The repetition of prevalence estimates in many countries is probably unnecessary. The second category is common to all research and is found when the scientific quality of research is so deficient that its findings are worthless. This can be called "feckless research" and is a waste of resources and time, including that of the subjects themselves. The third form of abuse occurs when an epidemiological study is limited to an enumeration of cases, a so-called "nose count", without testing any hypothesis. Such work can be described as "fancy that!" research, because the findings are unremarkable or even self-evident, doing nothing to advance knowledge. The fourth abuse consists in fieldwork being carried out in a clumsy manner, so that the individuals interviewed obtain no benefit from this unsolicited intrusion on their lives. Unlike patients, members of the general population who are interviewed in epidemiological surveys have not sought to be examined about their mental health. They therefore need to be competently interviewed with absolute guarantee of confidentiality (see below). It is nevertheless reassuring to know that the effect of mental health surveys on respondents has been systematically investigated. Only a tiny minority say they found the interview upsetting, and these were individuals who were already known to score highly on the trait of neuroticism. The overwhelming majority did not resent the experience, and many enjoyed it.

2
A Matrix for Organising Epidemiological Knowledge

In order to organise knowledge about the epidemiology of mental disorders, it is useful to use a matrix (A.S. Henderson 1988), in which the main categories of mental disorders are listed across the top to form the columns, while the rows are made up of those variables that may contribute to the onset or course of morbidity (Table 1). This matrix proves to be a tidy way of organising what information is already available; but it also acts heuristically by proposing associations that otherwise might not have been considered but which call for investigation. The variables can be placed in

Table 1. A matrix for organising epidemiological knowledge

		Clinical psychiatric disorders	Neurotic disorders	Affective disorders	Schizophrenia	Dementia	Personality disorders	Alcohol abuse	Other psychoactive substance abuse	Suicide	Parasuicide	Intellectual disability	Psychological well-being
Sociodemographic factors	Gender Age Marital status Social class Urbanisation Regional variation Secular change Other macrosocial factors												
Experimental factors	Seasonality of birth Birth order Parental age Childhood separation Parental style Cultural or subcultural beliefs and attitudes Adverse experiences/ events Extreme experiences Bereavement Expressed emotion Unemployment Social support Migration Noise Environmental toxins Dietary factors Comorbidity Personality and temperament Genetic factors Other biological factors Interaction of two or more factors												

categories: sociodemographic, experiential, intrapersonal (psychological) and biological.

The alert observer will notice that the matrix has a limitation, because it only has two dimensions, i.e. it does not display interactions between two or more variables, interactions which can be of the greatest importance. Examples are the interaction between age and gender in the onset of schizophrenia, or the combination of adverse parental style in childhood, adverse life events in adult life and a genetic vulnerability to affective disorders. In order to consider such interactions, users of the matrix need to select their own combinations. It is also desirable for the user to construct his or her own additional variables to suit the community in which the research work is being planned.

For many of the cells in Table 1, knowledge is already available in the literature, e.g. issues such as socio-economic class and schizophrenia, education and dementia or secular change and affective disorders. What is conspicuously lacking in psychiatric epidemiology is information on higher-level interactions such as gender, adverse experiences and genetic vulnerability.

Table 2. The three population levels

Level	Advantages	Disadvantages
Community surveys	Reflect the occurrence of morbidity free of some selection biases	Labour-intensive and costly Non-random refusals or losses
	Identify persons with a wide range of severity Provide data of administrative value	Low case yield for some disorders
General practices (family physicians)	Better yield of cases	Labour-intensive
	Opportunity to study recognition of cases and their outcome	Denominator may be uncertain Selection effects from help-seeking, doctors and refusals
Mental health services	Generate operational data of administrative value	Based only on individuals reaching treatment
	High yield of cases for most diagnostic groups	May lack data from private sector
	Can be linked to other databases	May be ethical problems in access
		Quality of data may be variable

3 Levels of Epidemiological Enquiry into Psychiatric Disorders

Psychiatric epidemiology conducts its enquiries at one of three levels: (1) morbidity as it occurs in the community or general population, (2) morbidity as it occurs in primary health care (including family physicians) and (3) morbidity as it occurs in mental health services. Each offers advantages for particular purposes, and these may be summarised as set out in Table 2.

From Table 2, it will be clear that research on aetiology may often be carried out with advantage at the community level rather than on treated series, because there are less likely to be biases in a community sample, particularly in relation to likelihood of exposure to putative risk factors. This can be important in avoiding Berkson's bias (Berkson 1946), whereby some other variable, unrelated aetiologically, may influence the chances of a person reaching services. The unaware researcher might then mistakenly conclude that this variable contributed causally. Such biases are absent in community samples. Such samples also cover a wide range of severity, whereas cases of depressive disorder, for instance, will be only of the severe type in hospital or clinic series. Nevertheless, community samples have their own problems. They are laborious and costly, requiring sufficient trained personnel to contact and interview hundreds and commonly thousands of individuals in their own homes. Invariably there are refusals and people who cannot be contacted, and it is likely that the very condition being investigated will be more common in such cases. This is called "non-ignorable non-response" and is itself a topic of some scientific and statistical interest, the issue being how to correct for it in computing estimates of prevalence or incidence or of outcome.

4 Commonly Used Designs

The main designs used in psychiatric epidemiology are cross-sectional, prospective longitudinal (cohort) and case-control studies. A cross-sectional study is often an excellent start to an enquiry, because it provides a picture of how much morbidity is present in a sample at one point in time and the variables most closely associated with this. However, because it is only a "snap-shot", the cross-sectional study can rarely reveal

much about aetiology. For example, suppose that data are collected from a community sample of a thousand adults and that a measure is obtained of the symptoms each has experienced in the previous month and the exposure they have had to adverse life events in the last year. Then suppose that the data show quite strongly that individuals who have had many adversities also tend to have more symptoms of anxiety or depression than those not so exposed. It would be unwise to conclude from this that adversity contributes to the onset of symptoms. This is because individuals with anxiety or depression may be more likely to have unpleasant things beset them, or they may be more likely to report that they have had many troubles, either because of a tendency to complain or as a result of "effort after meaning", whereby they can account for feeling psychologically unwell. It may also be that individuals who have certain personality traits or lifestyles may be more likely to have troubled lives with much adversity *and* be prone to anxiety or depression.

Such problems in methodology can be resolved to some extent by graduating to a prospective longitudinal design, also called a cohort study. In this, a population sample is assessed at the start, when most individuals are psychologically well. The information obtained from these subjects is therefore unlikely to be biased by their mood state or by selective recall. It may refer, for example, to their personality, lifestyle, recent or past exposure to adversity, past health and family history. They are then re-examined at least once after an appropriate interval. Some will have developed symptoms in the area of interest. The research question to be asked is whether the potential risk factors that were assessed at the start were more frequently present in those who developed the symptoms. A design of this type yields considerably more information about the processes likely to be at work, either those leading to mental disorders or those protecting against them. However, such a study design is obviously very demanding in human, administrative and financial resources. It also takes a long time to obtain all the data. For these reasons, epidemiologists often use the case-control method.

The case-control method has been under-used in psychiatric research, but can be a powerful strategy in trying to identify risk factors for a specified disorder. The essence of its design lies in obtaining data to complete a 2 × 2 table, defined by individuals with and without the condition in question and whether or not they have been exposed to a suspected pathogenic experience (Table 3). Notice that the cohort study is prospective, because it waits for morbidity to develop. By contrast, the case-control study is retrospective. The procedure involves finding a sample of individuals with the disorder in question, preferably new or incident cases, and a similar number of subjects who are matched for age, gender and other variables but who do not (or not yet) have the disorder. Both patients and controls, or their families, are then asked about the various possible risk factors. The question asked is whether there are more subjects in cell *a* than would occur by chance. We do not know the incidence of the disorder in all individuals in the population who were exposed to each risk factor, nor do we know the number not exposed. Likewise, we do not know how many people in the population have recently developed the disorder. As a consequence, we cannot compare the incidence in those exposed and not exposed *for the whole population*. All we have are the data from the patients examined, who are necessarily only a fraction of all incident cases in the population, and data from a fraction of all healthy subjects. However, we can proceed as follows. First, the relative risk is calculated from Table 3:

Table 3. A 2 × 2 table in a case-control study

		Case	
		Yes	No
Exposed	Yes	a	b
	No	c	d

Letters refer to number of individuals.

$$\text{Relative risk} = \frac{a}{a+b} \div \frac{c}{c+d}$$

By simple algebra, this becomes

$$\frac{a(c+d)}{c(a+b)}$$

Then something very helpful can be done. Where a disorder is fairly uncommon in the general population, *a* will be very small in comparison with *b* and *c* will be small in comparison with *d*. If we assume a negligible contribution by *a* in the term $(a+b)$, and by *c* in the term $(c+d)$, the relative risk will be nearly equal to:

$$\frac{a \times d}{b \times c}$$

This is the odds ratio, which is an expression of the strength of a risk factor. A more complete account of these issues can be found in Schlesselman (1982). By way of an overview, Anthony (1988) has given a lucid exposition of the use of case-control methods in psychiatry.

5
Sampling

5.1
Principles

There are few occasions when it is practicable to examine a total population. In studies at the general population level, this would mean carrying out an assessment of every individual in, say, an entire city or region. In a primary health care setting, it would mean assessing every person who consults staff in every facility during a specified period. In a case-control study in clinics or hospitals, it would mean examining every individual who presents with a specified disorder in all of the clinics across a whole country, as well as examining an equivalent number who do not have that disorder. The solution is to take a sample of the group under investigation. The essential principle is that everyone in the true denominator, i.e. the total population within a specified age-range, must have an equal probability of being included in that sample or, if there is to be weighting, a known probability. If this is not achieved, there is a distinct likelihood of bias, whereby the achieved sample may be different in ways that could be important in the analysis. For example, the sample of cases should not differ from *all* the incident cases in that population in attributes such as the severity or duration of the disorder or in level of education, age or likelihood of having been exposed to a candidate exposure or risk factor. Thus, in a study of the association between childhood sexual abuse and depression in adult life, the cases of depression should be representative – in severity, duration of depression and likelihood of exposure to abuse – of all those in the community with depression. The same principle of equivalence should apply to the controls without depression.

5.2
Non-Ignorable Non-Response

In field surveys, it has long been accepted that not everyone who is in the "target sample" will agree to be interviewed or will be available at the time the interviewer calls. It is common to find that only 70%–90% are actually assessed. Furthermore, those who refuse or are repeatedly not available are now known often to be more likely to have the mental disorder under investigation. Next, in many cohort studies of mental disorders in which a sample of individuals are followed over several years, it is recognised that those who die are more likely to have already had the condition under study at the beginning or to have developed it during the course of the study. This means that those who are successfully re-examined are a survival élite or are different in important ways from the original cohort. These distortions could lead to mistaken conclusions if the losses are not allowed for.

Statistical methods have been available for many years for estimating how much error may have occurred due to refusals and how to correct for this in the conclusions drawn. In cross-sectional surveys, these methods include the introduction of special incentives to increase the response rate and very brief psychiatric assessments that can be conducted on the doorstep. The information obtained then allows some correction through the application of weights to the data from the achieved sample (Kessler et al. 1995). Both in cross-sectional and in cohort studies, Bayes' theorem can be applied, whereby adjustments to the final estimates can be made on the basis of prior probabilities, some of which are known (Best et al. 1996).

6
Specifying Disorders

6.1
Diagnostic Categories

Whether the epidemiological study is conducted at the level of the community, primary health care or mental health services, it is obviously essential to specify which symptoms or diagnoses are to be studied. Traditionally, psychiatric research has focused on "cases", i.e. individuals who have been given a diagnosis. This is entirely appropriate in many circumstances. The task has been made much easier with the development of the diagnostic criteria now in wide international use. The first of these is the International Classification of Diseases (Tenth Revision) (ICD-10) with its Classification of Mental and Behavioural Disorders. This comes in two mutually complementary presentations: the *Clinical Descriptions and Diagnostic Guidelines* (WHO 1992) and the *Diagnostic Criteria for Research* (WHO 1993). These sets of diagnostic criteria have been prepared after very wide consultation with expert psychiatrists in some 40 countries. They therefore represent an international consensus. Furthermore, the *Diagnostic Criteria for Research* have been used as the basis for diagnostic instruments that have computer algorithms to apply the criteria, precisely and invariably, to the information obtained (see below).

Another system is presented in the *Diagnostic and Statistical Manual of Mental Disorders*, Fourth Edition (DSM-IV) of the American Psychiatric Association (1994). This is a national development that is used

throughout the United States and quite widely elsewhere. Through a process of extensive consultation that took place between the WHO and the architects of the DSM-IV, it is very similar to ICD-10, and "cross-walks" have been developed to allow movement between them for individual diagnoses. Together, ICD-10 and DSM-IV have brought a common language to clinical psychiatry and research. When authors submit a paper to a reputable scientific journal, it is now virtually mandatory that the disorders are described according to one or both systems. Indeed, it is highly desirable for investigators to use both simultaneously, because this carries distinct advantages for the analyses and for further knowledge to be acquired about the performance of ICD-10 and DSM-IV as diagnostic systems.

6.2 Continuous Measures of Morbidity

Case ascertainment might be assumed to be the sine qua non for any progress in the epidemiology of mental disorders, i.e. to recognise cases. However, using the traditional expression "case ascertainment" nicely illustrates the very problem that has to be rethought, because it implies a categorical structure in the morbidity that we wish to study. In a population, there are traditionally cases and non-cases. As expressed by Pickering (1968), "medicine in its present state can count up to two, but not beyond." He was referring to hypertension, but others (e.g. Rose 1993) have argued that most mental disorders are dimensional in their nature. The frequency distribution of their component symptoms, such as anxiety, depression or cognitive impairment, is usually a reversed J shape, with most people having none or only a few symptoms, and progressively fewer individuals having higher counts. A committee of clinicians in Geneva or Washington, whose experience is often limited to hospital and consulting room practice, have decided by consensus where the cut-off point should be placed for individuals to be "cases". While this is entirely appropriate for some purposes, it may not always be a true representation of the world. In statistical terms, it can sometimes lose information.

It is not disputed that disorders of mental health exist in categorical states or at least have some utilitarian value; a depressive episode, Alzheimer's disease, anorexia nervosa and alcohol dependency are clinically realistic entities, albeit not necessarily unitary ones. What is proposed here is that, in epidemiological studies at the general population level, hypotheses about aetiology are not well served unless very large numbers of respondents are interviewed, solely because the base rates for these conditions are not large. However, it is possible to identify individuals with *some* symptoms of depression, of cognitive impairment, of abnormal eating or of alcohol misuse, and a score of these symptoms can be made the dependent variable in an analysis of candidate risk factors. In other words, it can sometimes be statistically more powerful to look for associations between a putative risk factor and morbidity expressed as a continuous variable rather than as a dichotomy of cases and non-cases. Rose and colleagues confirmed that, when a continuous measure, such as the General Health Questionnaire (GHQ; Goldberg and Williams 1988) is applied to a population, a unimodal distribution curve is found, with no break between so-called cases and so-called normals. Rose went on to argue that this approach has three important consequences for studying morbidity. Firstly, a characteristic of the community as a whole emerges. This is the mean and standard deviation of its GHQ scores. Secondly, this collective characteristic may show large differences between men and women, geographic regions, social strata and income groups. These differences are based on shifts of the entire distribution. The third consequence is that differences between these groups in the prevalence of probable cases (those with a score above a threshold) are related to different average scores in these groups. As Rose put it, "the visible part of the iceberg (prevalence) is a function of its total mass (the population average)."

Rose went on to say that "psychiatrists, unlike sociologists, seem generally unaware of the existence and importance of mental health attributes of whole populations, their concern being only with sick individuals." It is an appealing notion that populations, while they are undeniably made up of individuals and as far as we know no other component, take on properties of their own, much as molecules have attributes not found in their constituent atoms. The concept of populations having different frequency distributions of morbidity, and *not just different prevalence rates for clinical cases*, carries with it the implication that one or more factors are shifting the overall distribution in some populations but not in others. The idea is that there is some pervasive force that is active in the biological or social environment and that promotes a disease or disorder. This notion is not far removed from Galen's atmospheric factor, the katastasis or miasma.

6.3 Disablement

There is a further advantage in considering morbidity as a continuum in a population. Morbidity usually refers to symptoms or disorders, but there is another universe of discourse closely linked to this, namely

disablement. This is the collective noun now used to refer to the impairment, disability and social role handicap in daily life that disorders bring with them. It is self-evident that the main categories of mental disorder, especially the psychoses, affective disorders and dementias, are almost always associated with substantial disablement. However, subclinical levels of mental disorders also carry with them a certain amount of disablement. From the point of view of a whole population, the amount of disablement from subclinical or milder conditions is cumulatively substantial. This is because such conditions have a high point prevalence. Therefore, from a public health perspective, the significance of milder mental disorders is not trivial.

7
Independent Variables: The Domains

The independent variables are traditionally those that are possibly associated with, or even contribute causally to the dependent variable, which is a psychiatric disorder or symptom score. Such variables can lie in several domains: sociodemographic, socio-environmental experiential (both past and current), psychological or biological. They can, of course, be a mixture of any of these, in which case the variables are said to show an interaction effect, either within or across domains. Psychiatric epidemiology has attended largely to sociodemographic, socio-environmental and experiential variables, as exemplified respectively by the celebrated studies of Faris and Dunham (1960) on the spatial distribution of patients with schizophrenia in Chicago, which showed a concentration in the central areas with cheap, single-room occupancy hotels; the Stirling County Study in Canada by Leighton et al. (1963a,b) with its concept of sociocultural disintegration; and the corpus of studies by Brown and colleagues on adverse life events, the presence of a confiding relationship, early childhood experiences and depression in women (Brown and Harris 1978). Psychological attributes such as personality or other indicators of vulnerability or resilience have been much less studied epidemiologically. Biological variables, including genes, are largely an untapped field, awaiting investigators with innovative hypotheses to integrate several of the above domains.

Recent years have seen unprecedented advances in molecular genetics with intense activity directed towards the genetic contribution in affective disorder, schizophrenia, alcohol dependency and Alzheimer's disease. There is good reason to believe that some of these advances, both in scientific understanding and in new methods for genetic analyses, may bring unparallelled opportunities to psychiatric epidemiology. It is therefore important that psychiatrists and epidemiologists now collaborate with experts in human genetics to construct testable hypotheses in which genetic measures can be considered alongside, and in interaction with, other markers of vulnerability. In this context, two complementary strategies could be followed. The first consists in a continuing search for genes associated with discrete disorders such as depression or schizophrenia, a search carried out at the general population level or in high-risk groups. The second strategy is quite different and involves searching not for genes that may cause the disorders, but for genes that confer vulnerability to them. Because such an approach is well suited to epidemiological designs, we shall look more closely at the thinking behind it.

Cloninger et al. (1996) have commented on the difficulties of replicating genetic associations with complex psychiatric disorders and have argued that "it may be more fruitful to map genes contributing to temperament, which has a relatively simple genetic architecture and can be quantified easily and reliably by questionnaire. Later, susceptibility to complex disorders like schizophrenia and alcoholism can be evaluated." One such attribute of temperament is the trait of neuroticism. This is a personality dimension characterised by a greater or lesser tendency to react emotionally to adverse stimuli, i.e. to get upset under stress. Because neuroticism is a relatively stable trait in contrast to the fluctuating nature of anxiety and depressive disorders, it provides a good basis for genetic investigation of the vulnerability to these disorders. Using longitudinal twin data, Kendler et al. (1993) found that approximately 55% of the genetic vulnerability to major depression is shared with neuroticism. The broad-sense heritability of neuroticism has been estimated to be about 0.42. There is some evidence that heritability is higher in women than in men and that it decreases from late adolescence to early adulthood.

While quantitative genetic studies have established a role for genetic factors in neuroticism, new developments in molecular genetics make it possible to go further and seek particular loci linked to such traits. Until fairly recently, molecular genetics techniques were mainly applied to categorical disorders thought to involve a single gene. However, psychological variables such as neuroticism form a continuum in the population and probably involve many genes, each of which is neither necessary nor sufficient for the trait. In recent years, there has been increasing interest in genes that contribute to variation in quantitative traits, of which neuroticism is one example. Quantitative trait loci (QTL) may vary in the size of their effect on a trait from infinitesimally small to modest. The aim of

current QTL research is to find the loci which have the largest effect.

There are a number of techniques available for the study of QTL, including linkage analysis in families, allele-sharing methods between relatives, association studies in population samples and experimental crosses in laboratory animals. The allelic association approach has greater statistical power than either conventional linkage analysis or allele-sharing methods to detect genes with small effect and, despite some limitations discussed below, it may be the most effective approach for QTL detection. It is argued that allelic association studies are at present the strategy of choice for detecting quantitative trait loci. Because a large number of candidate markers are being investigated, there is a risk of type I errors. In order to reduce these, the search for allelic associations can be carried out with an original community sample comprising the top and bottom 5% of the total distribution and a replication sample covering the top and bottom 1%.

The first studies on QTL for personality traits have now been reported. Using an allelic association methodology, Ebstein et al. (1996) found an association between the trait of novelty seeking and the D4 dopamine receptor gene. Benjamin et al. (1996), using both population and sib-pair association methods, found an association between extraversion and the same gene. Other candidate genes are now being examined. This work is a landmark in the study of personality, demonstrating for the first time that there is a specific biological basis for variation in a personality trait. It also marks a new era in psychiatric epidemiology in which it should be possible simultaneously to include biological, intrapersonal and social variables in aetiological research.

8 Instruments

Instruments fall into two types: symptom scales and psychiatric examinations. In any one study, either or both may be used, depending on the investigator's aims and resources.

8.1 Symptom Scales

The more simple type of instrument is a symptom scale, which can be completed by respondents themselves or administered by the interviewer. Widely used examples are the GHQ (Goldberg and Williams 1988), the Hopkins Symptom Checklist (HSCL; Derogatis et al. 1974) and the Mini-Mental State Examination (MMSE; Folstein et al. 1975). The GHQ and HSCL cover symptoms of anxiety and depression and are therefore useful for obtaining a continuous measure of these, expressed as a score; alternatively, they can be used as a screening instrument to identify those individuals in a sample who have a high, or low, probability of having a mental disorder. In an analogous manner, the MMSE can provide a score for a person's cognitive function or, by applying a previously established cut-off point to that score, it can be used to identify those who are likely to have a dementia (Brayne and Calloway 1990; A.S. Henderson et al. 1994).

When a questionnaire or self-rated instrument is used in research, its psychometric properties should be known beforehand, and these are often specific to the population being studied. Variables such as cultural patterns and education can have a significant effect on the instrument's performance. The two properties to be established are its sensitivity and specificity, both expressed as a percentage. These refer to the instrument's performance when compared with a criterion or "gold standard", such as a comprehensive psychiatric examination of established validity or a consensus diagnosis amongst experts. In Table 4, a sample is shown of individuals examined both by the screening instrument and a full examination. We consider the number of subjects who are "cases" according to both instruments (a), according to one but not the other (b or c) and according to neither (d).

Sensitivity is the proportion of individuals who screen positive who are indeed "cases" according to the criterion, i.e.:

$$\frac{a}{a+c}$$

and specificity is the number who screen negative who are indeed not "cases":

$$\frac{d}{b+d}$$

The sensitivity and specificity of a test will vary according to where the cut-off point is placed on the scores. As sensitivity increases, specificity tends to decrease, so that an appropriate balance between the two has to be determined by the investigator. For some

Table 4. Criterion

		Yes	No	
Cases by screening test	Yes	a	b	a+b
	No	c a+c	d b+d	c+d

Cases by full psychiatric assessment.

purposes, such as in screening for depression, it is more important to identify as many as possible of the true cases, but it does not matter if there are quite a few "false positives", because these can be corrected in a second stage by more extensive examination. Under these conditions, one would want a highly sensitive screening test that placed most of the true cases in cell *a* and few in *c*. It matters rather less if quite a few of the true non-cases are mistakenly placed in *b*.

8.2 Standardised Psychiatric Interviews

Even with the best-designed scales or questionnaires with the best psychometric properties, these cannot be a substitute for a psychiatric interview, which can yield something approaching a diagnosis according to the international criteria. Symptom scales do not purport to make a diagnosis. Progress of fundamental significance for all of psychiatry has been made since the late 1960s with the development of standardised examinations. These examinations are standardised in two mutually complementary ways. Firstly, the questions asked or the ratings of behaviour are not left to the idiosyncrasies of the interviewer, which would lead to different information being obtained across interviewers and across studies. Instead, this so-called information variance is reduced by having interviewers ask about symptoms in the same way. Secondly, criterion variance is reduced, the symptoms or signs elicited being assembled, like building bricks, in exactly the same way, both within and across studies. This is achieved by applying to the data an algorithm, which can easily be computerised. The algorithm is a precise expression of the diagnostic criteria in ICD-10 or DSM-IV.

There are two types of standardised psychiatric examination. This is for the good reason that, although full clinical interviews in the field by experienced clinicians may be the ideal, such resources are often not available. The first type of instrument is mainly for use by research clinicians after some training in its use. These instruments allow some flexibility in questioning and enable the clinician to use his or her judgement when making a rating about the presence or absence of each symptom or behaviour. The Schedule for Clinical Assessment in Neuropsychiatry (SCAN)[1] is such an instrument (see below). There is also another group of instruments for use by laypersons. These are said to be "fully scripted", where the questions asked are invariable and must be strictly adhered to, with only very few of the items calling for any judgement. The Diagnostic Interview Schedule (DIS) and the Composite International Diagnostic Interview (CIDI) are of this type (see below).

The SCAN is the successor to the ground-breaking Present State Examination (PSE) developed over two decades ago by Wing, Cooper and Sartorius and now revised by Wing et al. (1990) for the WHO (1992). The SCAN is a clinician's instrument, because it requires familiarity with the phenomenology of mental disorders and assumes that the interviewer is comfortable in examining individuals with a mental disorder. In complete contrast to interviews for use by laypersons, the clinician asks the main question, but is allowed to probe with further questions, if necessary, before deciding whether a symptom is present or not. The correct use of the SCAN requires formal training in one of the designated centres around the world. The SCAN has a number of modules, each dealing with a group of disorders such as anxiety states, affective disorders, substance abuse or psychoses.

The Clinical Interview Schedule, Revised (CIS-R) was developed from an earlier version by Goldberg et al. (1970), largely so that it can be used by lay interviewers (Lewis et al. 1992). It was successfully used in the National Survey of Psychiatric Morbidity in the United Kingdom (Jenkins et al. 1997a,b). The CIS-R generates diagnoses by ICD-10 and DSM-IV, but is also able to give scores for symptoms. It is known to have been well received by respondents and to have been user-friendly for interviewers. It is considerably less elaborate than the CIDI, for example.

The DIS was developed by Robins et al. (1981) as a psychiatric interview that could be administered for research purposes by non-clinicians after training. It was the instrument used in the ECA studies. Its current version allows diagnoses to be made according to DSM-IV. There is also an equivalent assessment for children, the Diagnostic Interview Schedule for Children (DISC).

The CIDI was developed by Robins et al. (1988) for the WHO and the U.S. Alcohol, Drug Abuse and Mental Health Administration Task Force on Psychiatric Assessments. It combines questions from the DIS with questions designed to elicit PSE and SCAN items. Like the DIS, the CIDI can be used by laypersons after only a few days' training. Its performance in clinical field trials has been established, and Wittchen (1994) has given a comprehensive account of its reliability and validity in studies to date. The CIDI is fully scripted and has been automated for administration on laptop computers by Peters and Andrews (1995). It is a powerful tool for epidemiological research, having a number of available modules which cover the main categories of mental disorder and generating both ICD-10 and DSM-IV diagnoses as well as symptom scores.

[1]Copies may be obtained from Publications, World Health Organization, Avenue Appia, 1221 Geneva 27, Switzerland.

Table 5. Some prevalence estimates (%)

Reference	Country	Time period	Depressive disorders		Anxiety disorders		Alcohol abuse or dependence		Schizophrenia or non-affective psychosis		Instrument used
			M	F	M	F	M	F	M	F	
Bebbington et al. (1981)	England	1 month	4.8	9.0	1.0	4.5	–	–	–	–	PSE
S. Henderson et al. 1981	Australia	1 month	2.6	6.7	4.1	3.0	–	–	–	–	GHQ/PSE
Regier et al. 1993	USA	1 month	3.5	6.6	4.7	9.7	5.0	0.9	0.7	0.7	DIS
von Korff et al. 1985	–	Lifetime	–	–	–	–	–	–	0.4	0.8	DIS, PSE
Kessler et al. 1994	USA	12 months	8.5	14.1	11.8	22.6	14.1	5.3	0.5	0.6	CIDI
Jenkins et al. (1997b)	UK	1 week	–	–	–	–	–	–	–	–	CIS-R
Australian Bureau of Statistics (1998) Mental Health & Well-Being	Australia	12 months	4.2	7.4	7.1	12.1	9.4	3.7	–	–	CIDI-A

PSE, Present State Examination; GHQ, General Health Questionnarie; DIS, Diagnostic Interview Schedule; CIDI, Composite Internation Diagnostic Interview; CIS-R, Clinical Interview Schedule, revised.

8.3 Instruments for Assessing Dementia and Depression in the Elderly

Epidemiological research on mental disorders in late life calls for specially designed instruments to detect cognitive impairment, cognitive decline, depressive states and performance in the activities of daily living. Three such instruments will be mentioned here, all of which include a section for assessing elderly people and a separate section for an informant, usually a relative or close friend. The first to be developed was the Geriatric Mental State Examination (GMS; Copeland et al. 1976), which has now been widely used in several countries. The second is the Cambridge Examination for Mental Disorders of the Elderly (CAMDEX; Roth et al. 1986), which was intended as a clinician's instrument, but which can be used by laypersons after some training. A combination of parts of the GMS and the CAMDEX has recently been used in a large study of cognitive impairment and dementia in the United Kingdom. The third instrument is the Canberra Interview for the Elderly (CIE; Social Psychiatry Research Unit 1992).[2] The CIE was developed "from the bottom up", in that the authors started with the diagnostic criteria in both ICD-10 and DSM-III-R (now DSM-IV) and then constructed items to tap each element in these criteria. The CIE is available in English, German and French versions. From the CIE, Jorm and Mackinnon (1995) developed a set of compact scales, the Psychogeriatric Assessment Scales (PAS)[2] to enable non-clinicians working with the elderly to detect depression, cognitive decline and stroke in a standarised manner (Jorm et al. 1995, 1997). The PAS are now in wide use internationally.

8.4 Typical Prevalence Estimates

There are now a large number of surveys reporting the prevalence of the main mental disorders in the general population. Some of the above instruments have been used in these surveys for case-finding. Examples are shown in Table 5.

8.5 Disablement

In assessing the epidemiology of any mental disorders, it is often very useful to estimate the amount of disablement that these entail. This should be done

[2]Copies may be obtained from the Secretary, Centre for Mental Health Research, Australian National University, Canberra, A.C.T. 0200, Australia.

separately, but in a complementary way to the measurement of symptoms. There are a number of instruments available to measure disablement associated with mental disorders. The most comprehensive is the Disability Assessment Schedule (DAS; WHO 1988), which assesses an individual's functioning across a hierarchy of self-care and functioning in daily life. Another is the Groningen Social Disabilities Schedule (Wiersma et al. 1988), which gives a comprehensive assessment of functioning. A shorter, self-rated instrument is the SF-36 and its even more brief version, the SF-12.

8.6 Aetiological and Other Associated Variables

8.6.1 Sociodemographic Variables

Few epidemiological studies are likely to omit a certain amount of information on the age distribution, gender, marital status, ethnic background and socio-economic or educational level of the sample population. It is best to obtain such information using methods that are already established and known to perform well in the investigator's region, rather than to invent one's own items. This also allows comparisons with other studies to be made with more confidence.

8.6.2 Social Environment

Social environment can be considered in two parts: (1) the individual's immediate social environment – what the sociologist Cooley (1909) called the primary group – made of those around a person with whom he or she has both interaction and commitment and (2) the wider social environment, the community, with its beliefs and values. Plausibly, both may have some influence on the incidence of mental disorders and on their course and prognosis. For an account of instruments to measure the individual's immediate social environment, and the support it may afford, reference may be made to A.S. Henderson (1988).

8.6.3 Experiential Variables

Some epidemiological studies may wish to test hypotheses about depressive disorder in adults, for instance, and the experiences they report having had in childhood and adolescence or their exposure to recent adversity.

Parker (1983, 1992) investigated exposure to a specific style of behaviour in parents and the subsequent risk of non-melancholic depression in the offspring some decades later. He developed the Parental Bonding Instrument (PBI) to obtain self-reported information on how affectionate and how controlling a person's mother and father had each been. The "toxic" exposure was found to be affectionless control, i.e. parents who had been highly controlling, but not at all affectionate or caring. Parker (1992) reviewed the evidence for inadequate parental care as a risk factor in adult depression, integrating this with the evidence on parental loss and, importantly, on compensating or mitigating factors. The latter may take place through adequate parental care from a substitute source or through strong affectional relationships in adulthood.

More recently, there has been considerable interest in the effect of exposure to sexual or other abuse in childhood and the risk of mental disorders in adult life. Here, the findings point to the many adverse experiences that commonly accompany childhood sexual abuse, including physical violence, unstable and untrustworthy relationships with parents and emotional deprivation. This is a good illustration of the problem of confounding variables, in which a candidate causal variable is likely to be associated with others which have an established link with the condition being studied.

There is a wide choice of instruments for assessing a person's exposure to adverse life events and experiences. These range from quite brief checklists to the comprehensive assessment provided by the interview developed by G.W. Brown and colleagues (Brown and Harris 1978).

8.6.4 Personality Variables

The hypotheses under investigation may refer to some markers of vulnerability to a particular group of mental disorders. Some such markers may lie within the domain of personality. Among these, one of the most promising is the construct of neuroticism, as measured by the Eysenck Personality Questionnaire (EPQ). A high score on this trait confers vulnerability to anxiety or depression. Neuroticism and some other personality traits can be reliably measured in only a few minutes by a self-completion questionnaire and may provide useful predictor variables in longitudinal studies. They are therefore often worth inclusion among the independent variables.

8.6.5 Genetic Factors

Sufficient DNA for some analyses can be obtained non-invasively by asking individuals to rub a cotton wool bud on the inside of their cheek for about 1 min. This is then placed in a sterile plastic container and labelled. It is usually not necessary to keep it cold, provided that it can be delivered to the genetics laboratory within

12 h. If the study calls for larger amounts of DNA, a sample of venous blood is preferable.

9
Analysis of Epidemiological Data

9.1
Strategies

After the field data have been edited and cleaned by removal of all obvious errors in coding, it is always best to start the analysis using the most simple methods, such as inspection of frequency distributions, means and standard deviations, cross-tabulations and correlations. Later, it may be appropriate to carry out multivariate analyses, typically with the measure of morbidity as the dependent variable. These methods include log-linear analysis and multiple linear regression for continuous variables such as symptom scores. All these methods allow the possible interaction effects between independent variables to be explored.

9.2
Application of Recent Advances in Statistics

There are sometimes opportunities to apply to psychiatric epidemiology some of the techniques developed in other disciplines, such as education and econometrics. For example, latent trait analysis was originally developed in statistical research on the measurement of intelligence before being imported to psychiatry by Duncan-Jones et al. (1986). Latent trait and latent class analysis are highly attractive methods for exploring the properties of an instrument and its individual items or of assessing the behaviour of diagnostic criteria when these are applied to population-based data. Where large data sets on symptoms in general population samples have been obtained, these can be used to explore the fundamental structure of psychiatric classification by the methods of numerical taxonomy. When the sample size is large enough to allow it, testing of hypotheses about aetiology or the course of disorders over time can be undertaken using structural equation modelling. In this, the existence of causal links between a set of variables can be tested by determining the strength of the path coefficients between them.

10
Ethical Issues in Psychiatric Epidemiology

As with other areas in the health sciences, psychiatric epidemiology has had to confront several ethical issues. These should be anticipated in every instance and presented to the Institutional Ethics Committee in the research worker's hospital or university. Some ethical issues concern the research team's access to data that already exist, such as hospital records, case registers for a whole geographic region or country, health insurance companies or police records. Where individuals could be identified, these data are made available to the researchers only if the individuals concerned have given informed consent. Another issue is the researcher's obligation to guarantee, and then be able to honour, complete confidentiality regarding the personal information obtained in a research interview or from any special investigations, such as the sensitive areas of human immunodeficiency virus (HIV) testing or genotype. While the researcher may genuinely state that confidentiality will be absolute, provision should always be made for eventual destruction of the data and for the information to be protected from other agencies, such as the courts. Contemporary public attitudes to privacy and confidentiality have made it difficult at times for epidemiologists to be allowed to develop databases containing information on people's health and to have access to those registers that already exist. This is an example of the tension that exists between the need for privacy and the public benefit that may be gained from using such information for medical and psychiatric research.

A further problem is already arising in genetic research conducted at the community level. In the course of research studies, if an individual has a genetic test showing an increased risk of developing a particular disorder such as cancer, Huntington's disease or, in the near future, depression or schizophrenia, a problem arises for that individual if he or she knows about the finding and then seeks life insurance. When applying for insurance, it is usual for the individual to be required to divulge any information that might prejudice his or her health. If this is divulged, the insurance company may refuse to underwrite the policy; if the information is not divulged, the company can refuse payment in the event of a claim. No satisfactory resolution to these matters has been reached.

In the course of a community survey, the research interviewer occasionally finds a person who is in profound distress or is a danger to self or others. While there is no problem if that individual agrees to be immediately contacted by a health professional, there can be an ethical problem if he or she refuses. This can sometimes be resolved by having an expert clinician available by telephone to the field staff at all times during the fieldwork, so that the interviewers are relieved of responsibility.

11 Conclusion

Research on the epidemiology of psychiatric disorders is at an interesting stage in its evolution. There is an abundance of testable hypotheses on aetiology and on the course of mental disorders. Instruments are now available for obtaining data of good quality both on symptoms and abnormal behaviour and on social or experiential variables. Of particular significance is the emergent capability to advance our understanding of the contribution made to aetiology by specific genes in interaction with environmental exposures.

12 References

American Psychiatric Association (1994) Diagnostic and statistical manual of mental disorders, 4th edn. (DSM-IV). American Psychiatric Association, Washington

Anthony JC (1988) The epidemiologic case-control strategy, with applications in psychiatric research. In: Henderson AS, Burrows GD (eds) Handbook of social psychiatry. Amsterdam, Elsevier, pp 157–171

Bebbington P, Hurry J, Tennant C, Sturt E, Wing JK (1981) Epidemiology of mental disorders in Camberwell. Psychol Med 11: 561–579

Benjamin J, Li L, Patterson C, Greenberg BD, Murphy DL, Hamer DH (1996) Population and familial association between the D4 dopamine receptor gene and measures of novelty seeking. Nat Genet 12: 81–84

Berkson J (1946) Limitations of the application of fourfold table analysis to hospital data. Biometrics Bull 2: 47–53

Best NG, Spiegelhalter DJ, Thomas A, Brayne CEG (1996) Bayesian analysis of realistically complex models. J R Statist Soc Aust 159: 323–342

Brayne C, Calloway P (1990) The association of education and socioeconomic status with the Mini-Mental State Examination and the clinical diagnosis of dementia in the elderly. Age Ageing 19: 91–96

Brown GW, Harris TO (1978) Social origins of depression: a study of psychiatric disorder in women. Tavistock, London

Carr VJ, Lewin TJ, Webster RA, Kenardy JA (1997) A synthesis of the findings from the Quake Impact Study: a two-year investigation of the psychosocial sequelae of the 1989 Newcastle earthquake. Soc Psychiatry Psychiatr Epidemiol 32: 123–136

Cloninger CR, Adolfsson R, Svrakic NM (1996) Mapping genes for human personality. Nat Genet 12: 3–4

Copeland JRM, Kelleher MJ, Kellett JM, Gourlay AJ, Gurland BJ, Fleiss JL, Sharpe L (1976) A semi-structured clinical interview for the assessment of diagnosis and mental state in the elderly: the Geriatric Mental State Schedule. I. Development and reliability. Psychol Med 6: 439–449

Cooley CH (1909) Social organization: a study of the larger mind. Scribner's Sons, New York

Derogatis LR, Lipman RS, Rickels K, Uhlenhuth EH, Covi L (1974) The Hopkins Symptom Checklist (HSCL): a self-report symptom inventory. Behav Sci 19: 1–15

Duncan-Jones P, Grayson DA, Moran PAP (1986) The utility of latent trait models in psychiatric epidemiology. Psychol Med 16: 391–405

Ebstein RP, Novick O, Umansky R, Priel B, Osher Y, Blaine D, Bennett ER, Nemanov L, Katz M, Belmaker RH (1996) Dopamine D4 receptor (D4DR) exon III polymorphism associated with the human personality trait of novelty seeking. Nat Genet 12: 78–80

Faris REL, Dunham HW (1960) Mental disorders in urban areas. An ecological study of schizophrenia and other psychoses. Hafner, New York

Folstein MF, Folstein SE, McHugh PR (1975) 'Mini-Mental State': a practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res 12: 189–198

Goldberg DP, Williams P (1988) A user's guide to the GHQ. NFER Nelson, London

Goldberg DP, Cooper B, Eastwood MR, Kedward HB, Shepherd M (1970) A standardized psychiatric interview for use in community surveys. Br J Prev Soc Med 24: 18–23

Greenwood M (1931) Epidemiology, historical and experimental. Hopkins, Baltimore

Gruenberg EM (1966) Epidemiology of mental illness. Int J Psychiatry 2: 78–134

Henderson AS (1988) An introduction to social psychiatry. Oxford University Press, Oxford

Henderson AS, Jorm AF, Mackinnon A, Christensen H, Scott LR, Korten AE, Doyle C (1994) A survey of dementia in the Canberra population: experience with ICD-10 and DSM-III-R criteria. Psychol Med 24: 473–482

Henderson S, Byrne DG, Duncan-Jone P (1981) Neurosis and the social environment. Academic, Sydney

Jablensky A, Sartorius N, Ernberg G, Anker M, Korten A, Cooper J, Day R, Bertelsen A (1992) Schizophrenia: manifestations, incidence and course in different cultures. A World Health Organization ten-country study. Psychol Med Suppl 20: 1–97

Jablensky A (1995) Schizophrenia: Recent epidemiologic issues. Epidemiol Rev 17: 10–20

Jenkins R, Meltzer H (1995) The National Survey of Psychiatric Morbidity in Great Britain. Soc Psychiatry Psychiatr Epidemiol 31: 1–4

Jenkins R, Bebbington P, Brugha T, Farrell M, Gill B, Lewis T, Meltzer H, Pettigrew M (1997a) The National Psychiatric Morbidity Survey of Great Britain – strategy and methods. Psychol Med 27: 765–774

Jenkins R, Lewis G, Bebbington P, Brugha T, Farrell M, Gill B, Meltzer H (1997b) The National Psychiatric Morbidity Survey of Great Britain – initial findings from the household survey. Psychol Med 27: 775-789

Jorm A, Mackinnon A (1995) Psychogeriatric Assessment Scale. User's guide and materials, 2nd edn. Anutech, Canberra

Jorm A, Mackinnon A, Henderson AS, Scott R, Christensen H, Korten AE, Cullen JS, Mulligan R (1995) The Psychogeriatric Assessment Scales: a multi-dimensional alternative to categorical diagnoses of dementia and depression in the elderly. Psychol Med 25: 447–460

Jorm A, Mackinnon AJ, Christensen H, Henderson AS, Jacomb PA, Korten AE (1997) The Psychogeriatric Assessment Scales (PAS): further data on psychometric properties and validity

from a longitudinal study of the elderly. Int J Geriatr Psychiatry 12: 93–100

Kendler KS, Kessler RC, Neale MC, Heath AC, Phil D, Eaves LJ (1993) The prediction of major depression in women: toward an integrated etiologic model. Am J Psychiatry 150: 1139–1148

Kessler RC, McGonagle KA, Zhao S, Nelson CB, Hughes M, Eshleman S, Wittchen H-U, Kendler KS (1994) Lifetime and 12-month prevalence of DSM-III-R psychiatric disorder in the United States: results from the National Comorbidity Survey. Arch Gen Psychiatry 51: 8–19

Kessler RC, Little RJA, Groves RM (1995) Advances in strategies for minimizing and adjusting for survey nonresponse. Epidemiol Rev 17: 192–204

Leff J, Sartorius N, Jablensky A, Korten A, Ernberg G (1992) The international pilot study of schizophrenia: five-year follow-up findings. Psychol Med 22: 131–145

Leighton DC, Harding JS, Macklin DB, Hughes CC, Leighton AH (1963a) Psychiatric findings of the Stirling County study. Am J Psychiatry 119: 1021–1026

Leighton DC, Harding JS, Macklin DB, Macmillan AM, Leighton A (1963b) The character of danger. Basic, New York

Lewis G, Pelosi AJ, Araya R, Dunn G (1992) Measuring psychiatric disorder in the community: a standardized assessment for use by lay interviewers. Psychol Med 22: 465–486

Morris JN (1964) Uses of epidemiology. Williams and Wilkins, Baltimore

Parker G (1983) Overprotection: a risk factor in psychosocial development. Grune and Stratton, New York

Parker G (1992) Early environment. In: Paykel ES (ed) Handbook of affective disorders, 2nd edn. Guilford, New York, pp 171–183

Pickering GW (1968) High blood pressure. Churchill, London

Peters L, Andrews G (1995) Procedural validity of the computerized version of the Composite International Diagnostic Interview (CIDI-auto) in the anxiety disorders. Psychol Med 25: 1269–1280

Regier DA, Farmer ME, Rae DS, Myers JK, Kramer M, Robins LN, George LK, Karno M, Locke BZ (1993) One-month prevalence of mental disorder in the United States and sociodemographic characteristics: the Epidemiologic Catchment Area Study. Acta Psychiatr Scand 88: 35–47

Robins LN, Regier DA (1991) Psychiatric disorders in America. Free Press, New York

Robins LN, Helzer JE, Croughan J, Ratcliff KS (1981) National Institute of Mental Health Diagnostic Interview Schedule. Its history, characteristics, and validity. Arch Gen Psychiatry 38: 381–388

Robins LN, Wing J, Wittchen H-U, Helzer JE, Babor TF, Burke J, Farmer A, Jablensky A, Pickens R, Regier DA, Sartorius N, Towle LH (1988) The Composite International Diagnostic Interview. Arch Gen Psychiatry 45: 1069–1077

Rose G (1993) Mental disorder and the strategies of prevention. Psychol Med 23: 553–555

Roth M, Tym E, Mountjoy CQ, Huppert FA, Hendrie H, Verma S, Goddard R (1986) CAMDEX. A standardised instrument for the diagnosis of mental disorder in the elderly with special reference to the early detection of dementia. Br J Psychiatry 149: 698–709

Schlesselman JL (1982) Case-control studies: design, conduct, analysis. Oxford University Press, New York

Social Psychiatry Research Unit (1992) The Canberra Interview for the Elderly: a new field instrument for the diagnosis of dementia and depression by ICD-10 and DSM-III-R. Acta Psychiatr Scand 85: 105–113

Susser EZ, Lin SP (1992) Schizophrenia after prenatal exposure to the Dutch hunger winter of 1944–1945. Arch Gen Psychiatry 49: 983–988

Susser EZ, Lin SP (1994) Schizophrenia after prenatal exposure to the Dutch Hunger Winter of 1944–1945. Reply. Arch Gen Psychiatry 51: 333–334

von Korff M, Nestadt G, Romanoski A, Anthony J, Eaton W, Merchant A, Chahal R, Kramer M, Folstein M, Gruenberg E (1985) Prevalence of treated and untreated DSM-III schizophrenia. J Nerv Ment Dis 173: 577–581

Wiersma D, DeJong A, Ormel J (1988) The Groningen Social Disabilities Schedule. Its development in the context of the ICIDH and its use in various populations of patients with mental disorders. Int J Rehabil Res 11: 213–224

Wing JK, Babor T, Brugha T, Burke J, Cooper JE, Giel R, Jablensky A, Regier D, Sartorius N (1990) SCAN. Schedules for Clinical Assessment in Neuropsychiatry. Arch Gen Psychiatry 47: 589–593

Wittchen H-Ü (1994) Reliability and validity studies of the WHO-composite international diagnostic interview (CIDI): a critical review. J Psychiatr Rev 28: 57–84

WHO (1973) Report of the International Pilot Study of Schizophrenia. World Health Organization, Geneva

WHO (1988) Psychiatric Disability Assessment Schedule (WHO/DAS). World Health Organization, Geneva, pp 1–88

WHO (1992) The ICD-10 Classification of Mental and Behavioural Disorders. Clinical descriptions and diagnostic guidelines. World Health Organization, Geneva

WHO (1993) The ICD-10 Classification of Mental and Behavioural Disorders. Diagnostic criteria for research. World Health Organization, Geneva

CHAPTER

3

O. Lipp, D. Souery, J. Mendlewicz

Population Genetics

1 Introduction

Population genetics consists in studying genetic transmission in families and populations. The role of genetic factors in transmitting psychiatric disorders has been documented in a wide range of studies. Different strategies such as family, twin, and adoption studies have been used to investigate genetic risk factors in psychiatric disorders. Their objectives are to investigate the genetic and nongenetic (environmental) causes of a disease and to estimate morbidity risks for relatives of an affected proband.

Epidemiological genetic studies have convincingly shown the involvement of genetic factors in bipolar affective disorder (BPAD) and schizophrenia. It is thus expected that molecular genetic studies will soon identify the gene or genes involved in the etiology of these disorders. Epidemiological strategies have also demonstrated the implication of genetic vulnerability factors in the etiology of other disorders such as unipolar affective disorder (UPAD), anxiety disorders, alcoholism, obsessive–compulsive disorder (OCD), dementia, and some personality disorders.

The limitations and methodological problems of genetic strategies will first be discussed in this chapter. The most recent or most relevant literature on the epidemiological genetic aspects of psychiatric disorders will be summarized. The interplay between epidemiological and molecular genetics research will also be highlighted.

2 Genetic Epidemiological Strategies

2.1 Family Studies

The first step when studying genetic epidemiological factors is to conduct family studies. Their aim is to detect the presence of familial aggregation, which is defined by the occurrence of a disorder at a higher frequency in relatives of affected individuals than in the general population. Familial aggregation is related to the presence of vulnerability factors, either genetic and/or environmental. Indeed, the first-degree relatives (mother, father, siblings, children) are more likely to develop a disorder than the second-degree relatives (e.g. uncle, nephew, cousins) and than the general population if that disorder has a genetic etiology. Family studies begin by identifying a proband, i.e. an affected subject, and by studying his or her relatives.

The main limitation of family studies is that shared environment and not only shared genes might be responsible for the observed familial aggregation (Ottman et al. 1991). Indeed, family studies cannot distinguish between genetic, family environment, and nonfamilial environmental factors. The relative contribution of genetic and environmental factors in disease etiology is therefore assessed by studying strategic populations such as twins and adoptees.

A second goal of family studies is to estimate the morbidity risk (MR), i.e. the probability that an individual will be affected at the age of risk. MR provided by family studies are highly relevant to genetic counseling. However, comparison of MR throughout several studies is often misleading, since different diagnostic criteria have been used. In order to reach standardized and reliable diagnoses, standardized instruments were developed such as the Research Diagnostic Criteria (RDC) (Spitzer et al. 1978), the Schedule for Affective Disorders and Schizophrenia–Lifetime Version (SADS-L), DSM-IV (American Psychiatric Association 1994), and ICD-10.

Family studies can also help to determine which clinical entities are transmitted together or independently within family members. If two disorders have distinct genetic etiologies, they will be transmitted independently in families. They provide additional evidence to validate new or controversial diagnostic entities (Feighner et al. 1972). In addition, the overlap of different diseases is examined through family studies.

New methodological standards for such studies have been well reviewed by Weissman et al. (1986a). Adequate family studies should make adjustment for potential disparities such as family size and years at risk of the relatives. Results from the majority of family studies were reported using the age-corrected "lifetime risk" or "morbid risk." Using this correction, the expected lifetime prevalence is given with all first-degree relatives having passed the age of risk of the disease. Several age-correction methods have been used (Gershon et al. 1982; Crowe and Smouse 1977), and these can also lead to discrepant results. Sex distribution in the relatives is another possible bias. For instance, the rate of affective disorder is higher in women than in men. An unequal distribution of the sex ratio in the groups of relatives compared can thus affect the results. Sample variations, differences between populations in the prevalence of genetic vulnerability, or cultural factors in the perception and description of hypomania and depression can also modify rates of MR (Gershon et al. 1982).

Data collection should be performed blind to the diagnosis of the proband. Two methods of data collection have been used to assess the presence of psychiatric disorder in the relatives: family history

data or family study data. In the latter method, all available relatives are evaluated by direct interview. Although structured interview and questionnaires such as the Family History - Research Diagnostic Criteria (FH-RDC) were designed to obtain more reliable familial information by interviewing only the proband (Andreasen et al. 1977), family studies using only family history data are considered less accurate. Indeed, probands tend to underestimate the prevalence of affective illness and other psychiatric disorders in their relatives and to overestimate the age at onset of illness in their ill relatives (Mendlewicz et al. 1975).

2.2 Twin Studies

Twin studies were designed to assess the relative contribution of genetic and environmental factors to disease development in the presence of familial aggregation. It has been traditionally accepted that monozygotic (MZ) twins are genetically identical. A discordance in the status of their disease would thus result from environmental factors, either biological or psychological. In twin studies, the difference in concordance rates for a trait between sets of MZ and dizygotic (DZ) twin pairs is analyzed. MZ twins behave genetically as identical individuals. DZ twins share only half of their genes on average and thus behave as sibs, except that they shared the same uterine environment before birth. Both types of twins are expected to share a similar environment. A genetic factor is suspected as the etiologic agent of a disease when the concordance rate in MZ twins is higher than in DZ twins.

Other designs can be used such as the study of twins reared apart since early childhood. This powerful strategic population provides another opportunity to delineate the effects between genetic and environmental factors.

A main limitation of twin studies is derived from the "equal environment assumption." The twin method has to assume that MZ and DZ twins have had the same environmental influences in order to prove a genetic contribution. Two recent studies (Kendler et al. 1994; Morris-Yates et al. 1990) supported the validity of the equal environment assumption. However, parents usually treat MZ twins more equally than they treat DZ twins. Moreover, environmental factors can differ in MZ twins as soon as their intrauterine life begins. After implantation, two thirds of MZ twins share a single chorion and placental circulation, and because of imbalance in their shared blood supply, they may differ more than DZ twins in their intrauterine experience and in birth weight (MacGillivray et al. 1975). Kendler and colleagues (Kendler 1983; Kendler et al. 1993) comprehensively reviewed studies assessing the validity of the twin method with regard to the equal environment assumption.

Other potential sources of bias in twin research exist. Probands described in several early twin studies were recruited from inpatient units (hospital-based studies), which could lead to an overidentification of concordant and more severe cases. In addition, unreliable psychiatric diagnoses for the concordance evaluation, zygosity misclassification (MZ pairs incorrectly labeled as DZ), and inadequate age at onset correction may lead to over- or underidentification of concordant cases (Kringlen 1995).

2.3 Adoption Studies

Adoption studies provide another opportunity to separate the interacting roles of heredity and environment on disease etiology. Several methodologies are used in such studies. The prevalence of the disease in adoptive and biological parents of affected adoptees can be compared. In the adoptees' design, the prevalence of disease in adopted-away children of affected parents is compared with the adopted-away children of normal control parents. Using the cross-fostering method, the prevalence of the disorder is compared between offspring of ill and offspring of normal biological parents, both reared by ill foster-parents. The prevalence of disease in biological and adoptive relatives of ill adoptees can also be compared to relatives of normal adoptees (adoptees' family study).

A limitation of adoption studies, when trying to separate the genetic and environmental endowments, is the time spent with the biological parents before adoption. Well-designed adoption studies should make adjustments for potential disparities between groups regarding family size, years at risk (up to current age, age at death, or age at diagnosis), and sex distribution in the relatives (Vieland et al. 1995). A potential confounding bias in adoption studies is that the disorder afflicting the affected biological parent may influence the likelihood and social location of the adoptive placements of their offspring. Conversely, adoptive parents represent a highly selected population, having been screened for high social stability and good mental health by the adoption agencies (von Knorring et al. 1983).

2.4 Segregation Analysis

Segregation analysis is used to determine the mode of genetic transmission or mode of inheritance of a disorder in affected families. It consists of analyzing

trait and phenotype distributions in families (phenotypes are observable features of a subject). The heritability of a particular trait is studied from one generation to the next and among members of the same generation. Such studies based on transmission probabilities are complex, requiring sophisticated statistical support. They can explore qualitative (discrete) or quantitative traits or phenotypes using different models of transmission (Weiss 1993).

Changes in prevalence or in age at onset during the last century, considered as a birth cohort effect (Gershon et al. 1987), can affect segregation analysis, which requires estimation of the prevalence of a disease in the general population. Ascertainment biases and assortative mating also have to be considered since they can alter results of segregation analyses. Indeed, assortative mating, which is a tendency for mated pairs (a couple) to be more similar for some phenotypic trait than would be expected if they were chosen at random, was observed in several psychiatric disorders and psychological traits (Merikangas 1982).

3 Bipolar Affective Disorder

Most family studies have shown that BPAD tends to cluster in families (Mendlewicz 1988). The lifetime risk for developing the disease in relatives of bipolar probands is significantly higher than in the general population. Table 1 shows that the familial rates of affective disorders averaged 5.7%–8% of BPAD and 10.2%–21.9% of UPAD in relatives of probands with BPAD. Exact rates are, however, difficult to obtain, as different diagnostic criteria or narrowed definitions of affective disorders were used throughout studies.

The MR for affective disorders were consistently higher in the relatives of bipolar than unipolar patients (Leonhard 1959; Perris 1968; Maier et al. 1993). This indicates that bipolar patients have a greater genetic loading for affective disorder than the unipolar ones. This observation generated worldwide interest from several research teams and prompted them to map genes involved in affective disorders, mainly for the bipolar form of the disease.

In a large-scale family study, the prevalence of bipolar II disorder was substantially increased (8.2%) among the relatives of individuals with bipolar II disorder, while only 1.1% of them had bipolar I disorder (Andreasen et al. 1987). This finding suggests that the disorder may breed true, providing some partial support for the independence of bipolar II disorder. No significant difference in familial aggregation was, however, observed between these two forms of BPAD by other investigators (Gershon et al. 1982). On the other hand, several studies suggest that bipolar II is closer to bipolar I than to UPAD (Dunner 1983; Endicott et al. 1985).

Early age at onset of probands with BPAD is also associated with a higher MR in relatives (Rice et al. 1987). Furthermore, BPAD patients with a positive family history of bipolar illness exhibit a more severe course of illness and have more manic episodes than patients without family history (Mendlewicz and Rainer 1974). On the other hand, no significant increase in familial aggregation of rapid-cycling BPAD is found compared with nonrapid cyclers (Lish et al. 1993; Coryell et al. 1992). However, the small sample size of these two studies does not permit definitive conclusions to be drawn.

Interestingly, a relationship between the effectiveness of lithium therapy and family history was observed. Responders to lithium stabilization more

Table 1. Morbid risks (MR) of bipolar (BPAD) and unipolar affective disorder (UPAD) in relatives of BPAD and UPAD probands

Probands	Study	At risk (*n*)	MR of BPAD (%)	MR of UPAD (%)
BPAD	Review by Gershon et al. (1982)[a]		8.0	10.2
	Gershon et al. (1982)[b]	441 (422)[c]	8.6	14.0
	Rice et al. (1987)[b]	557	5.7	
	Maier et al. (1993)[b]	389	7.0	21.9
UPAD	Review by Gershon et al. (1982)[a]		0.6	6.5
	Gershon et al. (1982)[b]	138 (133)[c]	3.0	16.6
	Rice et al. (1987)[b]	823	1.1	
	Maier et al. (1993)[b]	697	1.8	21.6
General population			0.2–2.0	4.0

[a]Included 12 studies on BPAD (1966–1980) and five studies on UPAD (1966–1980).
[b]Age-adjusted rates.
[c]Number at risk for BPAD (UPAD) according to their mean age at onset.

frequently have relatives affected with BP disorder than nonresponders (Mendlewicz et al. 1973; Grof et al. 1994). Recently, this observation was taken into account to design molecular studies with lithium responder patients in order to obtain a less heterogeneous population (Turecki et al. 1996).

A higher rate of affective illness, including BPAD, UPAD, and schizoaffective (SA) disorder, was observed in parents genetically related to BPAD probands compared with parents who adopted and raised them (Mendlewicz and Rainer 1977). Moreover, the degree of psychopathology in the biological parents of BPAD adoptees was similar to that in the parents of the nonadopted BPAD patients, while the rate of psychiatric disorder in the adoptive parents of the experimental group was similar to the rate in the adoptive parents of the normal offspring group. Cadoret (1978) studied adopted-away offspring of biological parents with affective illnesses. These probands had significantly more depressive disorders (mainly UPAD) in adulthood than did adoptees whose biological parents were well or had other psychiatric conditions.

There is a large amount of literature reporting a significantly higher concordance rate for BPAD in MZ twins than in DZ twins. The concordance rates in MZ twins vary between 50% and 92.5% (mean, 69.3%), compared with 0%–38.5% in DZ twins (mean, 20%). Although concordance rates were higher in earlier studies, the contribution of genetic factors has also been consistently noticed in the more recent ones. The differences seen across all these studies may result from diagnostic misclassifications and ascertainment biases in samples of the earlier ones.

In order to delineate the effects between genetic and environmental factors, the outcome of 12 identical pairs of twins reared apart since early childhood has been studied (Price 1968). Among these pairs, eight (67%) were concordant for the presence of the disease, a rate close to that for MZ twins reared together. This finding suggests that predisposition to BPAD will usually express itself regardless of the early environment.

The exact mode of inheritance is still not well understood in affective disorders. Some models of transmission can, however, be discarded. Thus the considerable number of families with two- and three-generation transmission of the illness is incompatible with autosomal recessive inheritance (Mendlewicz 1988). The two leading hypotheses arising from data analyses in BPAD patients are thus the single major locus (SML) model with incomplete penetrance (Strömgren 1938; Spence et al. 1995) and the polygenic model with a possible interaction between two or more loci (Perris 1972; Slater et al. 1972). Segregation data on affective disorders have been well reviewed by Rice et al. (1987), who reported the results of a very rigorous segregation analysis of BPAD. A mixed model, which allows for a SML with a multifactorial background, gave evidence for the presence of a major locus when controlling for the effects of birth cohort age at onset. This result was recently echoed by Spence et al. (1995), but neither this study nor the one by Rice et al. (1987) was large enough to test a polygenic/multifactorial hypothesis.

The SML model can also be tested by linkage analysis. For instance, linkage analysis provided evidence of X linkage in some families with BPAD (Mendlewicz et al. 1972; Baron et al. 1987). Despite exclusion of X linkage in other families, the X-linked dominant model appears to be a possible mode of transmission in some families with BPAD. Indeed, a recently published molecular study has again reported a strong linkage for BPAD with a DNA marker on chromosome X (Pekkarinen et al. 1995). This marker is closely located to the F9 locus where Mendlewicz et al. (1987) first provided evidence of a possible involvement of chromosome X in manic depression. Genetic heterogeneity and polygenic inheritance are likely to be present in BPAD, and this may be translated by the involvement of a high number of potential genes. Indeed, linkage, sib pair, and association methods have produced variable results in BPAD with loci on chromosomes 5, 11, X, and more recently 4, 18, and 21 (Souery et al. 1997).

Table 2 summarizes empirical risks for relatives of BPAD probands. Children and sibs of BPAD probands constitute high-risk groups. In second-degree relatives, the rates are smaller, suggesting that disease risks became smaller as the degree of consanguinity decreased, as expected if a genetic component explains the etiology of the disease.

4 Unipolar Affective Disorder

Most family studies have shown familial aggregation in UPAD, which favors a genetic component. The familial rates of UPAD averaged 6.5%–22% in relatives of probands with UPAD. In most studies, the relatives of UPAD patients had a higher lifetime prevalence of UPAD compared with the prevalence in the relatives of normal controls (around 6%). Early on, Leonhard (1959) and the Berlin School proposed the differentiation between bipolar and unipolar subtypes in mood disorder. Indeed, results of most family studies suggest that BPAD and UPAD have different genetic etiologies (Gershon et al. 1982; Perris 1968; Mendlewicz 1988;

Table 2. Estimated morbid risks in healthy subjects having relatives with bipolar affective disorders (BPAD), unipolar affective disorder (UPAD), and schizophrenia

	BPAD[a] (%)		UPAD (%)		Schizophrenia (%)	
	Range	Median	Range	Median	Range	Median
Monozygote twins	50–92.5	69.3	50	–	33–78	48
Dizygote twins	15–38.5	20	15–20	20	10–28	15
One affected parent	15–30	30	15–20	20	2–13	12
Two affected parents	50–75	75	40–50	50	40–50	40
Sibling	15–25	25	10–20	20	8–18	10
Parent second degree	3–7	7	3–5	5	3–5	4
General population	0.2–2	–	4	–	0.2–2	1

Adapted from Feinberg (1994) and from Gottesman and Shields (1982).
[a]The table provides the estimated risk of a healthy subject to develop BPAD if that subject has a relative (one or both parents, sibling) with BPAD.

Winokur et al. 1995). Table 1 shows the MR associated with each disorders. First, relatives of unipolar probands had no significantly elevated risk for BPAD (0.6%–3%) compared with controls (0.2%–2%), suggesting that the majority of unipolar probands are genetically distinct from BPAD (Andreasen et al. 1987; Winokur et al. 1995; McGuffin and Katz 1989). Second, as previously mentioned, family studies have consistently showed that bipolar patients have a greater genetic loading for affective disorder than the unipolar ones.

On the other hand, reports of discordant MZ pairs in which one individual developed BPAD while the other developed UPAD suggested that, in some instances, the same genotype can lead to both forms of the disease (Perris 1974; Bertelsen et al. 1977). Some UPAD patients can be misclassified and their illness can develop into BPAD, which can explain this discrepancy. This implies a different age at onset of disease for such MZ pairs, suggesting other vulnerability factors in addition to genetic ones. Moreover, a strong tendency for twin pairs to also be concordant for the affective illness type (UPAD or BPAD) was observed in one of the largest twin studies on affective disorder (Bertelsen et al. 1977). In addition, a higher concordance rate for MZ probands with BPAD (79%) was observed compared with MZ probands with UPAD (54%), also showing the greater genetic loading of the BPAD form of the disease.

Familial genetic differences in MR between early- and late-onset forms of unipolar illness were also observed. UPAD patients with early-onset disease had a greater familial morbidity for depression, alcoholism, and sociopathy than those with late-onset disease (Mendlewicz and Baron 1981). An excess of unipolar depression was also observed in female relatives of early-onset unipolars when compared to late-onset probands, regardless of the proband's sex.

The hypothesis of the "depression spectrum disease," a subtype of major depression characterized by families in which male relatives are alcoholic and females are depressed, originated from family studies (Winokur et al. 1971). Although independent family studies yielded negative results (Merikangas et al. 1985), an adoption study recently reexamined the hypothesis (Cadoret et al. 1996). Major depression in female was predicted by an alcoholic diathesis only when combined with a disturbed adoptive parent variable. This suggests that gene–environment interaction can be an etiologic factor in the depression spectrum disease. These interactions will be further discussed below.

The clinical and etiological heterogeneity of UPAD (Kupfer et al. 1975; Weissman et al. 1986b) represents a major difficulty when studying genetic aspects in UPAD, either at the epidemiological or the molecular level. Some patients can be correctly classified as having UPAD according to the DSM-IV classification system after a second episode of major depression following major stressful events. Whether all these patients represent "true genetic cases" is uncertain. On the other hand, no significant difference in familial aggregation was observed between "endogenous" and "nonendogenous depression" (Weissman et al. 1986; Andreasen et al. 1986; McGuffin et al. 1988). Interestingly, an increased risk of depression and of life events was observed in relatives of depressed probands (McGuffin et al. 1988). However, no significant association between life events and depression was observed, suggesting shared familial factors that predispose to both conditions. Although these results argue against the acceptance of earlier classifications of depression into "endogenous" and "reactive" subtypes, they cannot discard the concept of "endogenous depression." Indeed, family data is only one method of validating a classification system.

5 Schizophrenia

The presence of familial aggregation in schizophrenia was confirmed by Kendler et al. (1985). Schizophrenia was significantly more prevalent in 723 first-degree relatives of schizophrenic probands as compared to the first-degree relatives of 1056 surgical controls. Table 2 summarizes the estimated risks for relatives of schizophrenic probands. From Table 2, it is tempting to conclude that a higher difference in MR is observed between sibs and parents of schizophrenic probands than in those of probands with BPAD and UPAD. However, this difference may be related to a statistical artifact. Schizophrenia is associated with a marked reduction in reproductive fitness, and parents are therefore underrepresented, reducing the "risk" of illness in parents as compared with siblings and offspring of affected individuals (Essen-Moller 1955; Kendler et al. 1993). Family studies can examine whether the clinical manifestations of schizophrenia are correlated in affected sibling pairs. Global course, outcome, and all major symptoms except hallucinations were modestly, but significantly correlated in 256 sibling pairs concordant for DSM-III-R schizophrenia (Kendler et al. 1997).

Genetic studies suggest that there is a spectrum of disorders which could be similar to schizophrenia and could share the same genes. These disorders have been referred to as the "schizophrenia spectrum disorders." A recent large study has refined the boundaries of this spectrum. A total of 384 psychotic patients, 150 controls, and 1753 relatives were analyzed in the Roscommon family study (Kendler et al. 1993). The risk of developing schizophrenia was found to be higher in the relatives of probands with schizophrenia, but also in relatives affected with SA disorder, schizotypal and paranoid personality disorders, and other nonaffective psychotic disorders.

Family, adoption (Kendler et al. 1981, 1982), and twin studies (Torgersen et al. 1993) have documented the increased prevalence of schizotypal personality disorder among the biologic relatives of schizophrenic patients (Tsuang and Faraone 1994). Considering these studies, this disorder may be viewed as a milder form of schizophrenia. However, in the study by Kendler (1993), as in some previous family studies, schizotypal probands were hospitalized in psychiatric institutions. These patients may thus represent a sample of individuals biased in the direction of a familial relationship with schizophrenia.

The validity of Kraepelin's division into two major groups (dementia praecox/schizophrenia and manic-depressive illness) has been frequently challenged. Genetic studies offer the opportunity to investigate the overlap between these disorders and to evaluate the validity of Kraepelin's system. Most recent studies support this view (Kendler et al. 1993). Schizophrenia and the BPAD disorders were also transmitted independently in a recent study on 559 probands and 2845 relatives (Maier et al. 1993). However, schizophrenia probands had an increased familial risk for UPAD, indicating a familial relationship between the predisposition to schizophrenia and to major depression.

A review of the literature on twins revealed that a higher concordance rate was observed in all studies in MZ twins than in DZ twins with schizophrenia (Kendler 1983). Concordance rates varied from 18% to 28% for DZ twins and from 33% to 78% for MZ twins. Although the differences in concordance rates were larger in the earlier studies, a significant difference between MZ and DZ twins has still been observed in the more recent ones, suggesting a role for genetic factors in their etiology. However, using new methodological standards, the concordance rate for MZ twins remains under 50% (Onstad et al. 1991), suggesting interaction with other vulnerability factors. Gottesman and Bertelsen (1989) studied offspring of twins recruited from the Danish survey (Fischer 1973). Interestingly, the MR for schizophrenia of offspring of affected MZ twins (16.8%) was close to that of unaffected MZ co-twin's offspring (17.4%). Regarding offspring of discordant DZ twins, the MR were 17.4% for the affected DZ twin and 2.1% for the unaffected co-twin. Taken together, these data favor the hypothesis of an incomplete expression of a predisposing genotype in schizophrenia. A number of recent studies have also been designed to analyze discordant MZ pairs in order to explore developmental or environmental problems, such as maternal infections and obstetric complications (Davis and Bracha 1996). Such efforts might clarify nongenetic influences (Mednick et al. 1994) in the etiology of severe psychiatric disorders.

Large-scale adoption studies are available in schizophrenia. The extension of the Danish adoption study of schizophrenia throughout Denmark (Kety et al. 1994) confirmed the significant concentration of chronic schizophrenia (9.9%) and "latent schizophrenia" (5.7%) in the biological relatives of chronic schizophrenic adoptees as compared to adoptive relatives of schizophrenic adoptees and of normal controls. This suggests the presence of heritable genetic factors in the susceptibility to schizophrenia.

"Vertical cultural transmission" is a nongenetic factor that could explain the presence of familial aggregation in a behavioral syndrome such as schizophrenia (Kendler 1983). Vertical cultural transmission implies that a given characteristic is "learned" by the offspring from their parents. Adoption studies can provide insights on this issue. Hence children of schizophrenic parents reared by nonschizophrenic

adoptive parents should have a lower risk for schizophrenia if this form of transmission is operating in the etiology of the disease. Moreover, children of a nonschizophrenic biologic parent reared by an adoptive schizophrenic parent should have a higher risk for schizophrenia.

Although most adoption studies suggested that simple vertical cultural transmission is not likely to play a major role in schizophrenia (Kendler 1983), some of them suggested gene–environment interaction in the development of the disease. Tienari et al. (1994) observed a significantly higher rate of both psychosis and other severe diagnoses such as severe personality disorders when studying 155 offspring of schizophrenic patients compared to 186 matched control adoptees. Interestingly, notable differences between these two groups emerged only in the families rated as disturbed, suggesting gene–environment interaction.

Studies on the transmission mode of schizophrenia were well reviewed by Baron (1986). The two main transmission hypotheses are the SML and the multifactorial polygenic model. A mixed model with an SML transmission with a polygenic background is also compatible with the data (Risch and Baron 1984). Further segregation analyses using large sample size in order to test this polygenic model of inheritance are awaited. Several chromosomal regions may play a role in the etiology of schizophrenia, including genes on chromosomes 5, 6, 11, and 22 (Gurling 1996).

6 Schizoaffective Disorder

It is now well established that genetic factors are involved in the etiology of SA disorder (Zerbin-Rüdin 1988). The concept and diagnostic criteria of SA have varied tremendously during the two past decades, being either closer to the criteria for schizophrenia or to those for affective disorders. Family, twin, and adoption studies have thus generated variable results, mostly due to changes in diagnostic criteria. Although results from family, twin, and adoption studies are divergent, most support a separate classification of broadly defined SA psychoses as possibly being phenotypical variations or expressions of genetic interforms between schizophrenia and affective psychoses (Mendlewicz et al. 1980; Endicott et al. 1986; Bertelsen and Gottesman 1995). Molecular studies will hopefully determine whether SA disorders are separate entities or subtypes of affective disorders and schizophrenia. Meanwhile, it seems most appropriate to retain SA disorder as a distinct syndrome (Bertelsen and Gottesman 1995).

The overall lifetime risks for affective illness in first-degree relatives were investigated in a cohort of 172 probands and 1254 relatives (Gershon et al. 1982). Lifetime prevalences of major affective disorder (including BPAD, UPAD, and SA) were higher in relatives of probands with SA (37%), BPAD type I (24%), BPAD type II (25%), and UPAD (20%) compared to relatives of normal controls (7%). The highest overall MR of affective disorders was thus found in relatives of probands with SA, supporting the concept of SA disorder as a variant and genetically virulent form of affective illness (Tsuang et al. 1977; Gershon et al. 1982).

A model using a continuum of vulnerability factors was developed by Reich et al. (1975) in which the vulnerability of affective disorders is viewed as linear. In this model, SA disorder is expressed when vulnerability is the highest, BPAD when vulnerability is lower, and UPAD at a still lower threshold value. However, we and others (Shopsin et al. 1976; Rice et al. 1987; Winokur et al. 1995) have not reproduced these results, and the notion of a continuum of vulnerability factors remains controversial.

Similarly, probands with SA disorder had more schizophrenic first-degree relatives than other probands, which also suggests a relationship with schizophrenia (Mendlewicz et al. 1980; Gershon et al. 1982). In summary, a significant familial aggregation is observed between SA disorder and either BPAD or schizophrenia, while little familial aggregation is present between BPAD and schizophrenia. In addition, some studies suggest that SA disorder (bipolar subtype) may be more related to pure affective disorder, while SA disorder (depressed type) may be closer to schizophrenia (Andreasen et al. 1987).

7 Anticipation and Major Psychiatric Disorders

Segregation analyses in families with affective disorder and schizophrenia have not proved beyond doubt the SML or polygenic model of inheritance. Other complex modes of inheritance have been investigated in these disorders. A potential mode of inheritance is related to the concept of anticipation. Anticipation implies a progressively earlier age of onset and increased severity of disease in successive generations. It may explain the deviations from the mendelian inheritance pattern observed in some heritable diseases.

The phenomenon of anticipation has been observed in several neurological diseases such as myotonic dystrophy, fragile X syndrome, and Huntington disease (Fischbeck and Paulson 1996). Anticipation in these disorders has been correlated to specific mutations: expansion of a trinucleotide repeat sequence. These sequences are unstable and may expand in size within family members when transmitted from one

generation to the other, thereby increasing the severity of the disorder.

An exciting development in psychiatry research has been initiated by the emergence of studies assessing anticipation in psychiatric disorders. Several studies described anticipation in schizophrenia (Thibaut et al. 1995), although the study by Asherson et al. (1994) provided less conclusive results. Anticipation was also described in bipolar spectrum disorder (McInnis et al. 1993). However, this study included a wide definition of the phenotype of the 34 pairs studied, grouping together bipolar, unipolar, SA disorder (manic type), and major depressive disorder (single episode). Since the median age at onset in these four disorders differs widely, it is questionable whether these phenotypes can be compared through different generations.

Anticipation was described in 16 BPAD–BPAD pairs (Nylander et al. 1994) and in 16 UPAD–UPAD pairs (Engström et al. 1995). A total of 62 BPAD–BPAD pairs and 48 UPAD–UPAD intergenerational pairs were analyzed in a larger study on affective disorders involving 340 affectively ill patients recruited from 53 families. Changes in age at onset and episode frequency between generations consistent with anticipation were observed (Lipp et al., submitted).

However, it is not yet possible to conclude that true anticipation is present in mood disorders and schizophrenia because of various selections biases that could arise in such studies (Penrose 1948). Nevertheless, the clinical observation of anticipation in mood disorders and schizophrenia has prompted the search for dynamic mutations such as CAG repeats. We and others have observed an association between CAG repeats detected by the repeat expansion detection (RED) method and BPAD illness (Lindblad et al. 1995; O'Donovan et al. 1996). Moreover, we have recently observed a significant trinucleotide CAG repeat expansion between parental and offspring generation when the phenotype changed in severity from major depression (single episode) or UPAD to BPAD (Mendlewicz et al. 1997). These reports provide preliminary evidence for the implication of such mechanisms in the etiology of BPAD.

8
Genetic Factors and Other Aspects of Mental Disorders

8.1
Anxiety Disorders

It has been suggested that genetic factors are involved in the pathophysiology of anxiety disorders. A greater risk of developing anxiety disorders in first-degree relatives of patients with panic disorder has been observed in family studies (Crowe et al. 1983; Mendlewicz et al. 1993). Conversely, no increased familial loading was found for generalized anxiety disorder in probands with panic disorder (Noyes et al. 1987). These results validate the concept of generalized anxiety disorder as being an illness distinct from panic disorder. It also supports the hypothesis of a genetic contribution to predisposition to panic disorder. However, a twin study examining anxiety disorders came to a different conclusion (Skre et al. 1993). In a sample of 20 MZ and 29 DZ co-twins of probands with anxiety disorder, the prevalence of anxiety disorders was compared. For both panic disorder and generalized anxiety disorder, the MZ–DZ concordance ratio was more than 2:1. From these results, it may be concluded that a genetic contribution is likely to be present in both disorders. Clearly, more work needs to be done using larger sample sizes to evaluate the genetic contribution in generalized anxiety disorder and to define its relationship with panic disorder.

8.2
Obsessive–Compulsive Disorder

In several twin studies, the concordance rates for OCD ranged from 53% to 87% for MZ twins and from 22% to 47% for DZ twins, depending on the sample and the diagnostic criteria (Pauls et al. 1995). The diagnostic criteria and methodology of assessment of OCD were not properly established until the beginning of the 1990s. Interpretation of family studies performed before this era may therefore be misleading. None of the ten family studies of OCD reviewed by Black et al. (1992) used standardized assessments, a control group, and a blind data collection together. The paper by Black and colleagues reported the results of a study using new methodological standards of family studies, but no significant increase in MR of OCD in relatives of OCD probands was observed. However, the risk for a more broadly defined OCD (including relatives with obsessions and compulsions not meeting criteria for OCD) was increased among relatives of OCD probands.

Conversely, the recent family study by Pauls et al. (1995) on 100 probands and 466 relatives, using DSM-II-R and the Yale-Brown Obsessive Compulsive Scale (Y-BOCS) severity criteria, showed familial aggregation in OCD. The MR for OCD was significantly higher in OCD probands (10.3%) than in normal controls (1.9%). Data collection used all available information (both direct interview and multi-informant family history information), something which was not done by Black and colleagues (1992) and may explain their

negative result. On the other hand, investigators in the paper by Pauls et al. (1995) were not blind to the proband's diagnosis. Further family studies will be needed in order to legitimate familial aggregation in OCD.

8.3 Alcoholism and Substance Abuse

The concordance in diagnosis of alcoholism was 26% in MZ twins and 12% in DZ twins in a large sample of 715 male twins (Hrubec and Omenn 1981). Using two different types of statistical models, a segregation analysis was reported in 35 multigenerational families with alcoholism (Yuan et al. 1996). The hypothesis of an SML with strictly mendelian transmission was rejected. This analysis suggested a major effect with or without additional multifactorial effects.

Drug abusers without antisocial personalities were more likely to have a biologic parent with a background of alcohol problems, as shown by the adoption studies carried out by Cadoret et al. (1995). This suggested the possibility of two genetic pathways to drug abuse/dependency, one starting with alcoholism in biologic parents and the other from an antisocial biologic parent, proceeding through aggressive variables of adoptees such as conduct disorder and eventually ending in drug abuse/dependency.

8.4 Personality Disorders

Assortative mating, clinical heterogeneity, uncertain validity of phenotypes described by diagnostic criteria, and poor interrater reliability in assessing most personality disorders hamper genetic epidemiological studies. Despite these limitations, new data favor the involvement of genetic factors in some personality disorders. A weighted mean pairwise concordance rate for adult criminality of 51% in MZ twins and 22% in DZ twins was observed in seven twin studies pooled together (McGuffin and Gottesman 1985). As previously mentioned, results of adoption studies suggest a role for genetic factors in antisocial behavior (Cadoret et al. 1995). Furthermore, there is increasing interest in genetic research for dimensional aspects of personality traits (Cloninger et al. 1987). An association between characteristics of temperament (novelty seeking) and the seven-repeat allele in the locus for dopamine receptor D4 gene (DRD4) has been observed in a group of 124 unrelated Israeli normal subjects (Ebstein et al. 1996), a finding replicated by Benjamin et al. (1996).

8.5 Suicide

The role of genetic factors in transmitting suicidal behavior has been documented in twin (Juel-Nielsen and Videbech 1970), adoptees (Schulsinger et al. 1979), and family studies (Egeland and Sussex 1985). In the Amish population study, 26 suicides have been reported (Egeland and Sussex 1985). The majority of these cases (92%) were diagnosed with a major affective disorder and were situated in multigenerational families with heavy loading for BPAD, UPAD, and other affective spectrum illnesses. It is worth noting that, in other Amish families, in spite of a strong genetic loading for affective disorder, suicidal behavior was not expressed. Furthermore, liability to suicidal behavior might be transmitted familally as a trait independent of axis I and II disorders (Brent et al. 1996). These studies have provided evidence of genetic factors in both suicide and affective disorders. The distribution of depressed patients with a lifetime history of attempted suicide was also compatible with a polygenic inheritance of suicidal behavior (Papadimitriou et al. 1991). An association between tryptophan hydroxylase polymorphism and suicidality was observed (Nielsen et al. 1994). This observation was, however, not replicated (Abbar et al. 1995).

8.6 Dementia

Although results of family studies have suggested the implication of genetic vulnerability factors in the etiology of Alzheimer's disease, these studies are difficult to carry out in such disorders. The final diagnosis of Alzheimer's disease requires histopathological confirmation, the mean age at onset is very late, and the validity of the family history method is very low (Heun et al. 1996), which all constitute limitations in genetic studies. Nevertheless, it is estimated that 25%–40% of cases of Alzheimer's disease are familial, i.e. with at least one other case of dementia in the family (Van Broeckhoven 1996). Moreover, large families have been identified in which the disease was transmitted through several successive generations. Results of twin studies are rather inconsistent (Nee et al. 1987). Hence, further twin and adoption studies are required in order to separate genetic and environmental endowments in dementia. Using molecular genetic studies, four genes have been identified that predispose to Alzheimer's disease on chromosome 1, 14, 19, and 21 (Van Broeckhoven 1996).

9 Gene–Environment Interaction

The etiology of mood disorders is likely to be multifactorial, and both genes and environment may have a pivotal role as etiologic or risk factors. Indeed, several studies have pointed out the role of environmental and psychosocial factors in the pathogenesis, onset, and outcome of these diseases (Bauwens et al. 1991; Pardoen et al. 1993). Environmental factors may also include biological factors such as infections, which may affect gene expression. However, the precise contribution of genetic and environmental factors to disease etiology is not yet firmly understood in psychiatric disorders, and nor is the way in which these factors may interact.

How genes and environmental risk factors interact to produce a disease is a very complex task to investigate. True gene–environment interaction requires variations in both the environment and the genetic predisposition. So far, most of the published literature has used an "additive" model, in which genetic and environmental factors combine to produce a disease. There is, however, growing evidence that other models should be considered (Kendler and Eaves 1986). In a second model, genes do not directly alter the probability of illness. Rather, they control the degree to which the individual is sensitive to the risk-increasing or risk-reducing aspects of the environment. In this model, one allele conveys sensitivity and another one insensitivity to the environment.

In a third model put forward by Kendler and Eaves, genes control exposure to the environment. For instance, a gene can produce personality traits such as impulsiveness, personal stability, and frustration tolerance. The likelihood that this individual will experience more life events (predisposing environment) such as job and relationship changes is higher. This gene can thus predispose an individual to depression, but only by increasing its chance of experiencing life events. Moreover, to add to this complexity, genetic effects can result from the interaction of several loci, and the environmental effects are likely to be more complicated than being either "protective" or "predisposing."

The quantitative aspects of psychosocial or environmental factors can confer an advantage for genetic studies. Diseases such as hypertension and coronary heart disease are characterized by multiple intermediate quantitative traits or risk factors that are likely to play important roles in the susceptibility of individuals to develop such illness (Moldin 1994). Such quantitative traits are correlated with liability to affection and provide much more information than grouping individuals into affected or unaffected classes.

None of the molecular genetics studies performed so far has properly taken into account psychosocial factors. This prompted the elaboration of a large collaborative project investigating gene–environment interaction (BIOMED: EC Concerted Action on Affective Disorders: Interaction Between Genetic and Psychosocial Vulnerability Factors, led by J. Mendlewicz). The quantitative effects of certain relevant psychosocial vulnerability factors, such as social adjustment, the level of self-esteem, and other dimensional aspects of personality, on gene expression are evaluated through association, sib-pair, and linkage analyses. To verify the gene–environment interaction hypothesis, a model testing the relationship among selected variables, genes, and psychosocial factors has been defined. A specific statistical approach, defined as structural equation modeling (SEM) (Neale and Cardon 1990), will be used in order to test the validity of the model.

10 Future Research

Improved clinical strategies are needed to carry out genetic studies in psychiatry in order to minimize problems arising from clinical heterogeneity. New epidemiological genetic studies should focus on clinical definitions of major psychiatric disorders. Unreliability is often viewed as a major problem in psychiatric diagnosis, but this might be overestimated (Kringlen 1993). A criteria-based diagnostic system such as DSM-IV or ICD-10 can lead to reliability or agreement between clinicians. The central problem is, however, validity. How true or meaningful are these diagnostic classifications? These classifications have not solved the problems of clinical heterogeneity, and they inevitably lead to the inclusion of phenocopies.

The relationship between SA disorder, schizophrenia, affective disorders, and personality traits such as schizotypal personality disorder should be further investigated. The border between the "schizophrenia spectrum" and the "bipolar spectrum" has to be further refined. This is mandatory, since molecular linkage analyses need to specify the affected status of each family member. Hence a wrong specification of the affected or unaffected status limits the power of any linkage analysis. Inclusion of phenocopies as "genetic cases" in linkage analysis weakens the evidence for linkage. The exact mode of transmission is still unknown in the majority of psychiatric disorders. Misspecification of the genetic parameters of the phenotype may also lead to errors in linkage studies (Ott 1991). Segregation analyses using a large sample

size to test the polygenic model of inheritance still need to be performed.

Finally, genetic epidemiological studies have been traditionally used to validate nosology, diagnostic criteria, or classification. These methodologies will be useful to validate new proposed diagnostic categories (DSM-IV) such as premenstrual dysphoric disorder, simple deteriorative disorder (simple schizophrenia), and binge-eating disorder. The overlap between these new proposed diagnostic categories and depression, schizophrenia, affective disorders, and eating disorders can also be well analyzed by family, twin, and adoption studies.

Investigation of gene–environment interactions is one of the most promising areas dealing with complex and multifactorial diseases such as psychiatric disorders. Elaboration of new dimensional instruments to assess behavioral quantitative traits of psychiatric disorders, personality, and environmental factors should be a primary goal in order to refine the next generation of linkage, sib pair, and association studies. Gene–environment interactions need to be further investigated in large-scale collaborative epidemiological and molecular studies.

11 Conclusion

Strategies such as family, twin, adoption, and segregation studies have been the main source of genetic epidemiologic data in psychiatric disorders. These strategies have consistently demonstrated the implication of genetic vulnerability factors in the etiology of major psychiatric disorders. The results have prompted a fascinating worldwide effort to map the genes involved in these diseases. Furthermore, findings from genetic epidemiological studies, such as segregation data compatible with a sex-linked inheritance in some families or anticipation in BPAD and UPAD, have generated interesting molecular testable hypotheses. Hence genetic epidemiologic studies should remain in the forefront of genetic research, since they provide the rationale and clues for molecular research.

12 References

Abbar M, Courtet P, Amedeo S (1995) Suicidal behaviors and the tryptophan hydroxylase gene. Arch Gen Psychiatry 52(10): 846–849

American Psychiatric Association (1994) Diagnostic and statistical manual of mental disorders, 4th edn (DSM-IV). American Psychiatric Association, Washington, DC

Andreasen NC, Endicott J, Spitzer RL Winokur G (1977) The family history method using diagnostic criteria. Arch Gen Psychiatry 34: 1229–1235

Andreasen NC, Sheftner W, Reich T et al (1986) The validation of the concept of endogenous depression: a family approach. Arch Gen Psychiatry 43: 246–251

Andreasen NC, Rice JP, Endicott J et al (1987) Familial rates of affective disorder: a report from the National Institute of Mental Health Collaborative Study. Arch Gen Psychiatry 4: 461–469

Asherson P, Walsh C, Williams J et al (1994) Imprinting and anticipation. Are they relevant to genetic studies of schizophrenia? Br J Psychiatry 164: 619–624

Baron M (1986) Genetics of schizophrenia. 1. Familial patterns and mode of inheritance. Biol Psychiatry 21: 1051–1066

Baron M, Rish N, Hamburger R et al (1987) Genetic linkage between X-chromosome markers and bipolar affective illness. Nature 326: 289–292

Bauwens F, Tracy A, Pardoen D et al (1991) Social adjustment of remitted bipolar and unipolar out-patients. Br J Psychiatry 159: 239–244

Benjamin J, Li L, Patterson C (1996) Population and familial association between the D4 dopamine receptor gene and measures of novelty seeking. Nat Genet 12(1): 81–84

Bertelsen A, Gottesman II (1995) Schizoaffective psychoses: genetical clues to classification. Am J Med Genet (Neuropsychiatr Genet) 60: 7–11

Bertelsen A, Harvald B, Hauge M (1977) A Danish twin study of manic-depressive disorders. Br J Psychiatry 130: 330–351

Black DW, Noyes R, Goldstein RB, et al (1992) A family study of obsessive-compulsive disorder. Arch Gen Psychiatry 49: 362–368

Brent DA, Bridge J, Johnson BA, Connolly J (1996) Suicidal behavior runs in families – a controlled family study of adolescent suicide victims. Arch Gen Psychiatry 53: 1145–1152

Cadoret RJ (1978) Evidence for genetic inheritance of primary affective disorder in adoptees. Am J Psychiatry 134: 463–466

Cadoret RJ, Yates WR, Troughton E (1995) Adoption study demonstrating two genetic pathways to drug abuse. Arch Gen Psychiatry 52(1): 42–52

Cadoret RJ, Winokur G, Langbehn D et al (1996) Depression spectrum disease. I. The role of gene-environment interaction. Am J Psychiatry 153: 892–899

Cloninger CR (1987) A systematic method for clinical description and classification of personality variants. Arch Gen Psychiatry 44: 573–588

Coryell W, Endicott J, Keller M (1992) Rapid cycling affective disorder: demographics, diagnosis, family and course. Arch Gen Psychiatry 49: 126–131

Crowe RR, Smouse PE (1977) The genetic implication of age dependent penetrance in manic-depressive illness. J Psychiatr Res 13: 273–285

Crowe RR, Noyes R, Pauls DL, Slymen D (1983) A family study of panic disorder. Arch Gen Psychiatry 40: 1065–1069

Davis JO, Bracha HS (1996) Prenatal growth markers in schizophrenia: a monozygotic co-twin control study. Am J Psychiatry 153(9): 1166–1172

Dunner DL (1983) Subtypes of bipolar affective disorder with particular regard to bipolar II. Psychiatr Dev 1: 75–86

Ebstein RP, Novick O, Umansky R et al (1996) Dopamine D4 receptor (D4DR) exon III polymorphism associated with the human personality trait of Novelty-Seeking. Nat Genet 12: 78–80

Egeland JA, Sussex JN (1985) Suicide and family loading for affective disorders. JAMA 254: 915–918

Endicott J, Nee J, Andreasen NC et al (1985) Bipolar II: combine or keep separate? J Affective Disord 8: 17–28

Endicott J, Nee J, Coryell W et al (1986) Schizoaffective, psychotic, and nonpsychotic depression: differential familial association. Compr Psychiatry 27(1): 1–13

Engström C, Johansson EL, Langström M et al (1995) Anticipation in unipolar affective disorder. J Affective Disord 35: 31–40

Essen-Moller E (1955) The calculation of morbid risk in parents of index cases, as applied to a family sample of schizophrenics. Acta Genet 5: 334–342

Feighner JR, Robins E, Guze SB et al (1972) Diagnostic criteria for use in psychiatric research. Arch Gen Psychiatry 26: 57–63

Feinberg S (1994) Genetic counseling issues in affective disorders: the orthodox Jewish community. In: Papolos DF, Lachman HM (eds) Genetic studies in affective disorders. Wiley, New York

Fischbeck KH, Paulson HL (1996) Trinucleotide repeats in neurogenetic disorders. Annu Rev Neurosci 19: 79–107

Fischer M (1973) Genetic and environmental factors in schizophrenia. Acta Psychiatr Scand Suppl 238: 1–151

Gershon E, Hamovit JH, Guroff JJ et al (1982) A family study of schizoaffective bipolar I, bipolar II, unipolar, and normal control probands. Arch Gen Psychiatry 39: 1157–1167

Gershon E, Hamovit J, Guroff J, Nurnberger J (1987) Birth cohort changes in manic and depressive disorders in relatives of bipolar and schizoaffective patients. Arch Gen Psychiatry 44: 314–319

Gottesman II, Bertelsen A (1989) Confirming unexpressed genotypes for schizophrenia: risks in the offspring of Fischer's Danish identical and fraternal discordant twins. Arch Gen Psychiatry 46: 867–872

Gottesman II, Shields J (1982) Schizophrenia: the epigenetic puzzle. Cambridge University Press, Cambridge

Grof P, Alda M, Grof E et al (1994) Lithium response and genetics of affective disorders. J Affective Disord 32: 85–95

Gurling H (1996) The genetics of schizophrenias. In: Mendlewicz J, Papadimitriou GN (eds) Clinical psychiatry, genetics of mental disorders. Bailliere, London, pp 15–46

Heun R, Hardt J, Burkart M, Maier W (1996) Validity of the family history method in relatives of gerontopsychiatric patients. Psychiatr Res 62(3): 227–238

Hrubec Z, Omenn GS (1981) Evidence of genetic predisposition to alcoholic cirrhosis and psychosis: twin concordances for alcoholism and its biological ends points by zygosity among male veterans. Alcohol Clin Exp Res 5: 207–215

Juel-Nielsen N, Videbech TV (1970) A twin study of suicide. Acta Gen Med Gemellol (Roma) 19: 307–310

Kendler KS (1983) Overview: a current perspective on twin studies of schizophrenia. Am J Psychiatry 140: 1413–1425

Kendler KS (1993) Twin studies of psychiatric illness: current status and future directions. Arch Gen Psychiatry 50: 905–915

Kendler KS, Eaves LJ (1986) Models for the joint effect of genotype and environment on liability to psychiatric illness. Am J Psychiatry 143(3): 279–289

Kendler KS, Gruenberg AM, Strauss JS (1981) An independent analysis of the Copenhagen sample of the Danish adoption study of schizophrenia. III. The relationship between paranoid psychosis and the schizophrenia spectrum disorder. Arch Gen Psychiatry 38: 985–987

Kendler KS, Gruenberg AM Strauss JS (1982) An independent analysis of the Copenhagen Sample of the Danish Adoption Study of Schizophrenia: the relationship between major depressive disorder and schizophrenia. Arch Gen Psychiatry 39: 639–642

Kendler KS, Gruenberg AM Tsuang MT (1985) Psychiatric illness in first-degree relatives of schizophrenic and surgical control patients: a family study using DSM-IV criteria. Arch Gen Psychiatry 42: 770–779

Kendler KS, McGuire M, Gruenberg AM et al (1993) The Roscommon family study I, II and III. Arch Gen Psychiatry 50: 527–540, 645–652, 781–788

Kendler KS, Neale MC, Kessler RC et al (1994) Parental treatment and the equal environment assumption in twin studies of psychiatric illness. Psychol Med 24(3): 579–590

Kendler KS, Karkowski-Shuman L, Oneill FA et al (1997) Resemblance of psychotic symptoms and syndromes in affected sibling pairs from the Irish study of high-density schizophrenia families: evidence for possible heterogeneity. Am J Psychiatry 154(2): 191–198

Kety SS, Wender PH, Jacobsen B et al (1994) Mental illness in the biological and adoptive relatives of schizophrenic adoptees: replication of the Copenhagen Study in rest of Denmark. Arch Gen Psychiatry 51: 442–455

Kringlen E (1993) Genes and environment in mental illness. Perspectives and ideas for future researches. Acta Psychiatr Scand Suppl 370: 79–84

Kringlen E (1995) Twin studies in mental disorders. In: Mendlewicz J, Papadimitriou GN (eds) Genetics of mental disorders. I. Theoretical aspects. Bailliere Tindall, London, pp 47–62

Kupfer DJ, Pickar D, Himmelhoch JM, Detre TP (1975) Are there two types of unipolar depression? Arch Gen Psychiatry 32: 866–871

Leonhard K (1959) Aufteilung der endogenen Psychosen. Akademie, Berlin

Lindblad K, Nylander PO, De Bruyn A et al (1995) Expansion of trinucleotide CAG repeats detected in bipolar affective disorder by the RED method. Neurobiol Dis 2: 55–62

Lish JD, Gyulai L, Resnick SM et al (1993) A family history study of rapid-cycling bipolar disorder. Psychiatr Res 48: 37–45

Lipp O, Souery D, Mahieu B. New evidence for anticipation in bipolar and unipolar affective disorders (submitted)

MacGillivray I, Nylander PPS, Corney G (1975) Multiple reproduction. Saunders, Philadelphia

Maier W, Lichtermann D, Minges J et al (1993) Continuity and discontinuity of affective disorders and schizophrenia: results of a controlled family study. Arch Gen Psychiatry 50: 871–883

McGuffin P, Gottesman II (1985) Genetic influences on normal and abnormal development. In: Child psychiatry: modern approaches, 2nd edn. Blackwell, London

McGuffin P, Katz R (1989) The genetics of depression: current approaches. Br J Psychiatry 155[Suppl 6]: 18–26

McGufn P, Katz R, Bebbington P et al (1988) The Camberwell Collaborative Depression Study. 3. Depression and adversity in the relatives of depressed probands. Br J Psychiatry 152: 775–782

McInnis MG, McMahon FJ, Chase GA et al (1993) Anticipation in bipolar affective disorder. Am J Hum Genet 53: 385–390

Mednick SA, Huttunen MO, Machon RA (1994) Prenatal influenza infections and adult schizophrenia. Schizophren Bull 20(2): 263–267

Mendlewicz J (1988) Population and family studies in depression and mania. Br J Psychiatry 153[Suppl 3]: 16–25

Mendlewicz J, Baron M (1981) Morbidity risks in subtypes of unipolar depressive illness: differences between early and late onset forms. Br J Psychiatry 139: 463–466

Mendlewicz J, Rainer JD (1974) Morbidity risk and genetic transmission in manic-depressive illness. Am J Hum Genet 26(6): 692–701

Mendlewicz J, Rainer JD (1977) Adoption study supporting genetic transmission in manic-depressive illness. Nature 268(5618): 327–329

Mendlewicz J, Fleiss J, Fieve RR (1972) Evidence for X-linkage in the transmission of manic-depressive illness and schizophrenia. JAMA 222: 1627

Mendlewicz J, Fieve RR, Stallone F (1973) Relationship between the effectiveness of lithium therapy and family history. Am J Psychiatry 130: 1011–1013

Mendlewicz J, Fleiss JL, Cataldo M, Rainer JD (1975) Accuracy of the family history method in affective illness, comparison with direct interviews in family studies. Arch Gen Psychiatry 32: 309–314

Mendlewicz J, Linkowski P, Wilmotte J (1980) Relationship between schizoaffective illness and affective disorders or schizophrenia. J Affective Disord 2(4): 589–602

Mendlewicz J, Simon P, Sevy S et al (1987) Polymorphic DNA marker on chromosome and manic-depression. Lancet 1(8544): 1230–1232

Mendlewicz J, Papadimitriou G, Willmotte J (1993) Family study of panic disorder: comparison with generalized anxiety disorder, major depression and normal subjects. Psychiatr Genet 3: 73–78

Mendlewicz J, Souery D, Lindblad K et al (1997) Expanded trinucleotide CAG repeats in families with bipolar affective disorder. Biol Psychiatry 42(12): 1115–1122

Merikangas K (1982) Assortative mating for psychiatric disorders and psychological traits. Arch Gen Psychiatry 39: 1173–1180

Merikangas K, Leckman J, Prusoff B et al (1985) Familial transmission of depression and alcoholism. Arch Gen Psychiatry 42: 367–372

Moldin SO (1994) Indicators of liability to schizophrenia: perspectives from genetic epidemiology. Schizophren Bull 20(1): 169–184

Morris-Yates A, Andrews G, Howie P, Henderson S (1990) Twins: a test of the equal environment assumption. Acta Psychiatr Scand 81: 322–326

Neale MC, Cardon LR (1990) Methodology for genetic studies of twins and families. Kluwer, Dordrecht

Nee L, Eldridge R, Sunderland T et al (1987) Dementia of the Alzheimer type: clinical and family study of 22 twin pairs. Neurology 37: 359–363

Nielsen DA, Goldman D, Virkkunen M et al (1994) Suicidality and 5-hydroxyindoleacetic-acid concentration associated with a tryptophan hydroxylase polymorphism. Arch Gen Psychiatry 51(1): 34–38

Noyes JR, Clarkson C, Crowe RR et al (1987) A family study of generalized anxiety disorder. Am J Psychiatry 144(8): 1019–1024

Nylander PO, Engstrom C, Chotai J et al (1994) Anticipation in Swedish families with bipolar affective disorder. J Med Genet 31: 686–689

O'Donovan GC, Craddock N, Murphy KC et al (1995) Expanded CAG repeats in schizophrenic and bipolar disorder. Nature Genet 10: 380–381

Onstad S, Skre I, Torgersen S et al (1991) Twin concordance for DSM-III-R schizophrenia. Acta Psychiatr Scand 83: 395–401

Ott J (1991) Analysis of human genetic linkage, 2nd edn. John Hopkins University Press, Baltimore

Ottman R, Susser E, Meisner (1991) Control for environmental risk factors in assessing genetic effects on disease familial aggregation. Am J Epidemiol 134: 298–309

Papadimitrou GN, Linkowski P, Delarbre C et al (1991) Suicide on the paternal and maternal sides of depressed patients with a lifetime history of attempted suicide. Acta Psychiatr Scand 83: 417–419

Pardoen D, Bauwens F, Tracy A et al (1993) Self-esteem in recovered bipolar and unipolar out-patients. Br J Psychiatry 163: 755–762

Pauls LP, Alsobrook JP, Goodman W et al (1995) A family study of obsessive-compulsive disorder. Am J Psychiatry 152(1): 76–84

Pekkarinen P, Terwilliger J, Bredbacka P-E et al (1995) Evidence of a predisposing locus to bipolar disorder on Xq24-q27.1 in an extended Finnish pedigree. Genome Res 5: 105–115

Penrose LS (1948) The problem of anticipation in pedigrees of dystrophia myotonica. Ann Eugen 14: 125–132

Perris C (1968) Genetic transmission of depressive psychoses. Acta Psychiatr Scand Suppl 203: 45–52

Perris C (1972) Abnormality on maternal and paternal sides: observations in bipolar manic-depressive and unipolar depressive psychosis. Br J Psychiatry 118: 207–210

Perris C (1974) The genetics of affective disorders. In: Mendels J (ed) Biological psychiatry. Wiley, New York, pp 385–415

Price J (1968) The genetics of depressive behaviors. In: Coppen A, Walk A (eds) Recent developments in affective disorders. Br J Psychiatry Special Publ no. 2

Reich T, Cloninger CR, Guze SB (1975) The multifactorial model of disease transmission. I. Description of the model and its use in psychiatry. Br J Psychiatry 127: 1–10

Rice J, Reich T, Andreasen NC et al (1987) The familial transmission of bipolar illness. Arch Gen Psychiatry 44: 441–447

Risch N, Baron M (1984) Segregation analysis of schizophrenia and related disorders. Am J Hum Genet 36: 1039–1059

Schulsinger F, Kety SS, Rosenthal D (1979) A family study of suicide. In: Schou M, Strömgren E (eds) Origin prevention and treatment of affective disorders. Academic, New York, pp 277–287

Shopsin B, Mendlewicz J, Suslak L et al (1976) Genetics of affective disorders. II. Morbidity risk and genetic transmission. Neuropsychobiology 2: 28–36

Skre I, Onstad S, Torgersen S, Lygren S, Kringlen E (1993) A twin study of DSM-III-R anxiety disorders. Acta Psychiatr Scand 88: 85–92

Slater E, Maxwell J, Price JS (1972) Distribution of ancestral secondary cases in bipolar affective disorders. Br J Psychiatry 118: 215–218

Souery D, Lipp O, Mahieu B et al (1997) Molecular genetics of mental disorders with particular reference to affective disorders. Eur Psychiatry 12[Suppl 2]: 63–69

Spence A, Flodman P, Sadovnick AD et al (1995) Bipolar disorder: evidence for a major locus. Am J Med Genet (Neuropsychiatr Genet) 60: 370–376

Spitzer RL, Endicott J, Robins E (1978) Research diagnostic criteria: rationale and reliability. Arch Gen Psychiatry 35: 773–782

Strömgren E (1938) Beitrage zur psychiatrischen Erblehre. Munksgaart, Copenhagen

Thibaut F, Martinez M, Petit M et al (1995) Further evidence for anticipation in schizophrenia. Psychiatr Res 59: 25–33

Tienari P, Wynne L, Moring J et al (1994) The Finnish adoptive family study of schizophrenia: implications for family research Br J Psychiatry 164[Suppl 23]: 20–26

Torgersen S, Onstad S, Skre I et al (1993) 'True' schizotypal personality disorder: a study of co-twins and relatives of schizophrenics probands. Am J Psychiatry 150: 1661–1667

Tsuang MT, Faraone SV (1994) The genetic epidemiology of schizophrenia. Comp Ther 20(2): 130–135

Tsuang MT, Dempsey GM, Dvoredsky A et al (1977) A family history of schizoaffective disorder. Biol Psychiatry 12: 331–338

Turecki G, Alda M, Grof P et al (1996) No association between chromosome 18 markers and lithium-responsive affective disorders. Psychiatry Res 63: 17–23

Van Broeckhoven C (1996) Genetics of Alzheimer's disease. In: Papadimitriou GN, Mendlewicz J (eds) Clinical psychiatry, genetics of mental disorders, part II. Bailliere, London

Vieland V, Susser E, Weissman M (1995) Genetic epidemiology in psychiatric research. In: Papadimitriou GN, Mendlewicz J (eds) Clinical psychiatry, genetics of mental disorders. I. Theoretical aspects. Bailliere, London

von Knorring AL Cloninger R, Bohman M, Sigvardsson S (1983) An adoption study of depressive disorders and substance abuse. Arch Gen Psychiatry 40: 430–434

Weiss KM (1993) Genetic variation and human disease: principles and evolutionary approaches. Cambridge University Press, Cambridge

Weissman MM, Merikangas KR, John K et al (1986a) Family-genetic studies of psychiatric disorders: developing technologies. Arch Gen Psychiatry 43: 1104–1116

Weissman MM, Merikangas KR, Wickramaratne P et al (1986b) Understanding the clinical heterogeneity of major depression using family data. Arch Gen Psychiatry 43: 430–434

Winokur G, Cadoret R, Dorzab J, Baker M (1971) Depressive disease: a genetic study. Arch Gen Psychiatry 24: 135–144

Winokur G, Coryell W, Keller M, et al (1995) Family study of manic-depressive (bipolar I) disease. Arch Gen Psychiatry 52: 367–373

Yuan H, Marazita ML, Hill SY (1996) Segregation analysis of alcoholism in high density families: a replication. Am J Med Genet (Neuropsychiatr Genet) 67: 71–76

Zerbin-Rüdin E (1988) Schizoaffective and other atypical psychoses: the genetic aspect. In: Marneros A, Tsuang MT (eds) Schizoaffective psychoses. Springer, Berlin Heidelberg New York, pp 225–231

CHAPTER

4

M.C. O'Donovan, P. McGuffin

New Approaches and Findings in the Molecular Genetics of Major Mental Disorder

1 Introduction

The relatively new disciplines of molecular genetics and molecular biology have revolutionized the study of biology and medicine, and at least a passing familiarity with these subjects is now a requirement for reading and comprehending the contents of almost any major biomedical journal. Every week, disease genes are mapped or identified, and it is expected that in the near future this plethora of positional and structural genetic information will be translated into detailed functional data that will unravel the aetiology of otherwise poorly understood diseases and ultimately lead to improvements in their treatments.

After a period of false optimism in which dramatic (but ultimately unfounded) claims were made about the mapping of major genes for schizophrenia and bipolar disorder, psychiatric genetics is also now on the threshold of major advances as a result of the application of new paradigms, molecular techniques and statistical methods. In this chapter, we present a guide to the theory and rationale behind the most common approaches in molecular genetics and illustrate their utility, giving examples from studies of schizophrenia and bipolar disorder.

2 From Genotype to Phenotype

Until relatively recently, molecular genetic research was mainly concerned with observable characteristics (phenotypes) with simple patterns of transmission in which there is a straightforward relationship between genetic endowment (genotype) and phenotype. For example, if a single copy of a mutation in one gene is sufficient to cause disease, then inheritance is dominant, but if both copies of a gene are required to carry a mutation for a phenotype to be observed, then transmission is recessive. Huntington's disease and phenylketonuria are, respectively, classical examples of dominant and recessive traits of psychiatric relevance. Sex linkage represents a further form of simple transmission, complicated by the fact that males only have a single X chromosome, which is always transmitted by the mother, and females carry two copies of the X chromosome, one of which is largely inactivated.

While there are numerous examples of simply transmitted phenotypes, most common familial diseases (e.g. heart disease, diabetes, arthritis, asthma, inflammatory bowel disease and most psychiatric disorders) display more complex patterns of transmission. The expression of these phenotypes may require the inheritance of mutations within a small number of genes (oligogenic inheritance), numerous polymorphisms within numerous genes (polygenic inheritance) or a combination of both DNA variation and environmental factors (multifactorial inheritance). Common diseases may well display a mixture of the above. Thus there may be uncommon forms of a disease that are caused by single genes, some by the action of a few genes and others by a combination of a few genes of moderate effect operating in a multifactorial background. Under these circumstances, the concept of genetic susceptibility to disease is more useful than that of simple genetic causation.

3 From Phenotype to Genotype

The ultimate goal of molecular genetics is a complete understanding of the molecular basis of phenotypes, but the first task is to establish the precise nature of the DNA sequence variation responsible. Since this may be simply a single base pair change out of the 6×10^9 base pairs in the human diploid genome, this is a daunting task. The task may be simplified by clues offered by an association between gross chromosomal abnormalities and the phenotype, but commonly no such clues are available. Under this circumstance, the challenge is to identify pathogenic or susceptibility sequences purely on the basis of their map positions in the genome. The power and appeal of this approach, called positional cloning, lies in the fact that disease genes can be identified and their functions dissected even where the pathophysiology of the disease is completely unknown.

Although there are variations in detail, the first step in positional cloning is to identify the approximate chromosomal location of the disease genes by linkage analysis. Once this has been achieved, the region is narrowed to a more manageable size using more detailed linkage analysis or linkage disequilibrium methods. Finally, the genes that map within this relatively small area are identified and their sequences scanned for variations that may be responsible for the phenotype. These approaches will be described in more detail below.

4 DNA Markers

DNA markers have largely replaced the classical genetic markers (e.g. ABO blood typing) that were first used in human linkage and association studies. DNA markers are sequences with a defined map

position in the genome which show common sequence variation (polymorphism) in the general population. At present, the most commonly used markers are di-, tri- or tetranucleotide repeats, also known as simple sequence repeat markers which consist of a repetitive sequence, e.g. the dinucleotide CA, flanked by a known specific sequence. Such simple sequences often display allelic variation in the number of repeats units, e.g. a particular marker may have four possible alleles containing 16, 17, 18 or 19 repeats. Although in the past it was a cumbersome business, measurement of repeat number can now be relatively easily performed using the polymerase chain reaction (PCR), which allows chosen sequences of DNA to be amplified to detectable amounts. The size of the PCR products, and hence the number of repeat units, can then be simply determined by its relative mobility during size fractionation by electrophoresis in a gel matrix.

5 Recombination and Linkage

During meiosis, recombination events result in independent segregation of *most* of the genome, and this applies even to sequences on the same chromosome. However, within limits, the probability that a recombination event will occur between two markers is roughly proportional to the physical distance (in base pairs) between the markers, and it follows that markers located close to each other will not be randomly separated by recombination. Such markers will therefore tend to be inherited together, deviating from the law of independent segregation. This tendency for co-inheritance is called linkage. The approximate relationship between physical distance and recombination frequency gives rise to the concept of genetic distance, which is usually expressed in centimorgans (cM). A genetic distance of 1 cM between two markers means that they are separated in one meiosis out of 100, which corresponds to a physical distance of approximately one million base pairs.

From the discussion so far, it should be apparent that, in general terms, the shorter the distance between two markers, the greater the tendency for their co-inheritance and the tighter the linkage between them. However, there are limits to this principle. Recombination occurs between homologous chromosomes when they exist in a quadruple structure (after DNA replication, each chromosome is duplicated into pairs of chromatids "pinched" in at the centre) and only involves one of the pairs of chromatids for each chromosome. Therefore, the maximum recombination frequency that can be observed is 50%, because there will also be a non-recombinant copy to be transmitted, even if markers are separated by such a large distance that at least one recombination event will always take place between them. Another source of discrepancy between genetic and physical distance occurs because some chromosomal regions have greater or less than average tendencies for recombination. Markers flanking these regions will therefore be separated by genetic distances that are respectively greater and less than would be expected by their physical separation. Similarly, recombination events are more common in the female than the male germline, although the physical size of the chromosomes do not differ.

The principles of genetic linkage apply readily to simple traits exhibiting classical patterns of inheritance, because the simple transmission patterns make it easy to infer the presence or absence of the (as yet unknown) disease genotype from the phenotype. Disease and non-disease status can therefore simply be treated as different alleles of a genetic marker, and the task is then a simple one of finding mapped markers (i.e., markers of known position in the genome) that are co-transmitted with disease status more often than by chance in large pedigrees with multiple affected members.

In practice, it is possible to systematically map a dominant disease gene fairly rapidly and surely by linkage using an evenly spaced set of as few as 200 genetic markers about 20 cM apart covering the whole genome, since the sex-averaged genome length is approximately 3700 cM.

However, there are problems of simply applying linkage to complex traits with unknown modes of inheritance such as schizophrenia or manic depression. The method is most powerful when there is a simple causal relationship between the presence of a disease mutation and a phenotype. However, if only *susceptibility* to disease is transmitted, an individual may well inherit the susceptibility mutation and its linked genetic marker, yet never express the disease. Alternatively, in the absence of a test to determine susceptibility prior to the onset of disease, an individual at the time of study may be rated as unaffected only to become affected several years later. In the face of genetic heterogeneity, even if genes of fairly large effect do exist, different families may transmit disease genes on different chromosomes. All these factors conspire to create errors in the inference of genotype and therefore make it much more difficult to detect linkage.

The obstacles described above can be partly addressed by computer programmes for linkage analysis if the mode of inheritance is approximately known and if there are genes of at least moderate effect involved. However, if the mode of inheritance of the disease being studied is unknown, the factors above cannot be accurately controlled for and this leads to significant loss of power. This is particularly

problematic in studies of large pedigrees with affected individuals in multiple generations, because errors of inferred genotype made in one generation have a cumulative effect on the analysis of subsequent generations. The effect can be so severe that a change in diagnosis of a relatively small number of individuals can drastically reduce the evidence favouring linkage.

Because of these problems with classic linkage approaches, researchers investigating complex traits are turning their attention to smaller families. In this approach, linkage is detected by the transmission of a marker allele to pairs of affected siblings more often than expected by chance. This overcomes the problem of individuals that are currently unaffected yet destined to become affected; in addition, there is no knock-on effect in further generations, and smaller genetic effects can be detected. These features make this type of so-called affected sib-pair analysis more robust to diagnostic error than large pedigree approaches. More importantly, affected sib-pair methods are robust to misspecification of the mode of transmission. Furthermore, it can be argued that families with two affected individuals are more typical of complex diseases such as schizophrenia and bipolar disorder than large multi-generational families with multiple affected individuals. Therefore, the results obtained by studying affected sib-pairs may be more generalizable to the aetiology of the diseases. The main disadvantage of this approach is that it is much less powerful than classical linkage analysis when it comes to detecting genes of major effect that segregate in a mendelian fashion.

6 Linkage Disequilibrium

The apparent success of newer, more robust linkage methods in mapping disease genes that contribute to complex traits such as diabetes, asthma and multiple sclerosis gives cause for optimism that the approximate map positions of genes conferring susceptibility to major mental disorders will soon be known. However, because of the difficulties discussed above, it will be difficult to firmly establish fine mapping by linkage. Indeed, it is likely that regions defined by linkage studies for most complex traits will be so large that they may contain hundreds or thousands of genes (Schuler et al. 1996), and it is therefore essential to obtain a more precise localization before targeting individual genes for analysis.

Currently, the best solution to this problem takes advantage of the phenomenon of linkage disequilibrium. When two loci map within a small chromosomal region, they are only infrequently separated by recombination during meiosis. Because of this limited genetic shuffling, a small chromosomal segment surrounding the disease mutation is likely to contain a non-random distribution of marker alleles. Consider a marker system containing two alleles A_1 and A_2, each of which is present with a population frequency of 0.5 (i.e. 50% of chromosomes carry A_1 and 50% A_2). Subsequently, a disease mutation occurs on a chromosome at a genetic distance from the marker allele A_1 such that it is only separated by 1 cM. Because the mutation is so close to the marker, even after many generations most chromosomes carrying the pathogenic mutation will also carry marker allele A_1, which will therefore in turn be more common in affected individuals than the allele frequency of 0.5 in controls. This phenomenon, known as linkage disequilibrium, can therefore be detected in genetic studies as a difference in allele frequencies between patients and controls. Because the two loci must be positioned closely to each other for disequilibrium to last long in the population, this approach allows a much more precise localization than linkage. Why then undertake linkage studies at all? The answer is that the very property of linkage disequilibrium that allows precise mapping also means that linkage disequilibrium cannot be detected at large distances. Therefore, if the candidate region is not previously defined by linkage, a very dense marker set probably containing at least 2000 markers would be required to screen the genome. While such a map exists, the labour involved in genotyping is staggering, although this may soon be surmountable using DNA pooling. Pooling strategies allow DNA from perhaps up to 1000 individuals to be mixed and subjected to PCR in a single tube. The relative abundance of each PCR product then reflects the allele frequencies. Although there are still technical challenges, when these are overcome, the improvement in throughput will certainly make genome-wide linkage disequilibrium studies feasible.

A common problem that confounds linkage disequilibrium studies is that the control group and the sample of affected individuals must be tightly ethnically matched, because allele frequencies may vary dramatically between populations. This is called population stratification and may lead to a false appearance of linkage disequilibrium simply because there is a systematic difference in the genetic background (and therefore the allele frequencies) of the patients and controls. To overcome stratification, an important advance in study design has been to use the parental genotypes of affected individuals as a control. Disequilibrium is now identified by the preferential transmission of a particular allele to affected individuals more often than would be expected by chance. This is tested statistically by the haplotype relative risk (the relative risk conferred by each haplotype) or by

the transmission disequilibrium test (Spielman and Ewens 1996). Since the parents are a perfect control for the genetic background, this approach removes population stratification as a confounder, but does of course increase the difficulty with which a large sample can be obtained because of the requirement to have living parents willing to participate.

7 Statistical Considerations

The application of statistics to linkage studies of simple traits is relatively straightforward. Most commonly, support for linkage is assessed by a measurement called the LOD score, which is a numerical value corresponding to the logarithm of the odds that the trait and marker are inherited in a pattern consistent with linkage versus independent transmission of trait and marker. In conventional studies, where the phenotype can be designated unambiguously and the mode of inheritance is fully known, a LOD score of 3 (or an odds ratio in favour of linkage of 1000 to 1) is considered as indicating that the evidence for linkage is firm. A LOD score of minus 2 (odds of 100 to 1 against linkage) is taken as evidence for exclusion of linkage. Although these odds ratios sound extremely conservative, they correspond to conventional significance levels following a genome-wide search. Thus a LOD score of 3 corresponds to a reliability of 95% that linkage is present.

For complex traits, the situation is much more uncertain. Inflated rates of false-positive results may occur because of the use of multiple diagnostic classifications (e.g. exploring the effect of a broad versus narrow definition of the disorder) and the exploration of multiple genetic models of transmission. The net effect of these is that a LOD score of 3 cannot be viewed as conservative evidence for linkage. How then can we decide whether linkage is significant or not? In the case of a single report, the LOD score should be greater than 3, although there is no consensus on exactly how much greater this should be (Lander and Kruglyak 1995). One solution is to undertake simulation computations to estimate how frequently LOD scores of any given value are likely to be observed by chance. However, irrespective of the LOD score results and the simulation data, until findings are independently replicated, all reports of linkage concerning complex diseases should be viewed as preliminary.

Affected sib-pair linkage studies are usually evaluated by the more familiar p value, where p is the probability that allele sharing in affected siblings has occurred by chance. Again, because of the low prior probability that any linkage observed is true, some authors advocate a very conservative p value of 0.00002 (Lander and Kruglyak 1995). However, even then, replication must be regarded as essential. In practice, many groups pursue much weaker evidence for linkage in the pragmatic hope that more significant values will be generated or that suggestive reports of linkage will be generated in a sufficient number of data sets to provide strong evidence for linkage. The requirement for replication, while necessary to avoid false-positive results, carries its own problems, because even relatively large studies carry low power to detect genes of small or modest effect. Furthermore, replication of a linkage finding where multiple loci are involved in general requires a replication sample that is much larger than that used in the original study. Therefore, even a string of failures to replicate a finding may be the result of type II error.

The statistical interpretation of linkage disequilibrium studies follows the same broad rules as for linkage. Highly stringent requirements are made for results to be viewed as significant, because the prior probability of any marker being in linkage disequilibrium is small. Again, all findings are provisional until replicated.

8 Positional Cloning – The End Game

After a manageable region of interest (approximately 10^6 bases or less) has been confirmed, the next step is to identify the specific sequence variations in that region that confer susceptibility to disease. A number of complementary strategies may be used. However, most first require the creation of a complete or almost complete representation of the candidate region as cloned DNA. In brief, DNA from the whole genome, a single chromosome or even a fragment of a chromosome is broken into fragments. These are then randomly inserted into "cloning vectors" that allow replication of the human fragment in a host organism, usually a bacterium or yeast. For example, genomic DNA may be cut or physically broken into pieces of approximately 10^6 bases in length, which are then randomly inserted into yeast artificial chromosomes (YACs). The YACs are then inserted into a host strain of yeast which replicates the human DNA along with its own. After many cycles of reproduction, each single yeast organism that received a YAC will have grown into a yeast colony containing many copies of that single YAC. The YAC DNA can then be extracted, yielding large amounts of that specific human DNA fragment for further laboratory manipulation. A number of cloning systems are available, each with their own advantages and disadvantages; for example, YACs

and bacterial artificial chromosomes (BACs) carry relatively large fragments of DNA. Intermediate sizes of DNA are carried by P_1 artificial chromosomes (PACs), cosmids and some phage, and relatively small fragments of DNA are carried by phage and plasmids. Phage vectors are derived from natural viruses that infect bacteria; plasmids are circular DNA inclusions that occur naturally in bacteria (conferring antibiotic resistance, for example); and cosmids are hybrids derived from plasmids and phage. All these vectors have been modified to allow replacement of some of the original vector DNA with human fragments, allowing insertion and replication of the vector-human recombinant in bacteria such as *Escherichia coli.*

The next step is to isolate a series of individual clones containing DNA fragments originating from the candidate region until the whole region of interest is represented in the test tube. In fact, the procedures briefly outlined here are highly complex, but several genome mapping centres have now established a series of overlapping clones representing an almost complete genome. In principle, therefore, a rough but imperfect representation of almost any candidate region can be ordered "by catalogue." There is, however, still a considerable amount of work required in putting together a map with fine enough resolution for the next phase of work.

Most of the genome has no known function and is sometimes referred to (perhaps erroneously) as "junk" DNA. The next stage therefore usually involves identifying the relatively small proportion of the DNA within the region which corresponds to genes and is transcribed to messenger RNA. Again, work in a number of centres is underway to establish a complete map containing the position of every human gene, and at present 16,000 of the total anticipated 50,000–100,000 genes have been mapped (Schuler et al. 1996). When this work is complete, it will be possible to identify all the genes within a candidate region by consulting a database. Until then, the human gene map can be supplemented by maps from other species using the phenomenon of synteny. Regions of synteny are large stretches of DNA in which the order and identity of genes is conserved between species despite major differences in the organization of their genomes. Therefore, if two genes are known to map to a candidate region in humans, and their map positions are conserved in a different species, e.g. mouse, a database of the mouse genome can be consulted to see whether there are any other genes within the syntenic mouse region which may also map to the candidate region for the human disease.

At present however, individual research groups must usually identify coding sequences from their own candidate regions. This is most commonly done by elegant techniques such as "exon trapping" or "direct cDNA selection." cDNA is DNA which has been synthesized from mRNA and therefore consists of relatively "junk"-free DNA. The detail of these methods is beyond the scope of this chapter, but briefly, in exon trapping, small genomic fragments from the candidate region are inserted into mammalian cells in a special cloning vector which allows the host cells to excise non-coding (intronic) sequences, while expressing the coding sequences (exons). In direct cDNA selection, fragments of genomic DNA are mixed in solution with cDNA synthesized from a tissue of interest. By base pairing, the cDNAs corresponding to the coding parts of a specific genomic DNA fragment adhere to it and can then be "fished" out of solution (Lovett 1994).

However identified, the next stage involves looking at affected individuals and controls for differences in the sequences of genes from the candidate region. The sequence of genes can be compared definitively by sequencing in patients and controls, but while this is technically feasible it is still laborious. Therefore, short-cut methods for detecting sequence variation are usually used. These methods rely upon inferring differences in sequence from indirect measures, e.g. differences in mobility, sensitivity to enzymatic degradation and differences in thermal stability. These methods have the advantage of being much faster than sequencing, but share the common disadvantage that they are not usually 100% sensitive.

Having identified sequence variations within genes in the candidate region, the final stage in positional cloning is to determine which of these are involved in the disease. Where there is a simple relationship between genotype and phenotype, this process is relatively straightforward, since the sequence variation is not expected to be present in controls but will be present in affected individuals (for recessive conditions, two copies may be required in patients and one may be present in some controls). Where sequence variations merely confer susceptibility, they may only be present at a higher frequency in patients than controls. However, as we have seen earlier, polymorphisms flanking the relevant mutation may be in linkage disequilibrium, and therefore many that are not relevant to pathogenesis may be present at a higher rate in patients than controls.

The task of distinguishing between sequence variation resulting from linkage disequilibrium and that directly involved in a disease is not straightforward. However, as a guide, a sequence variation that has a direct involvement in disease should:

1. Have a demonstrable impact upon the abundance, function, sequence or distribution of the encoded protein and therefore be of biological importance

2. Be demonstrable in affected individuals from different populations, (because linkage disequilibrium is minimal across genetically diverse populations)

Definitive proof that a polymorphism can contribute to the pathogenesis of a disorder requires further molecular analysis, e.g. the demonstration of disturbed cellular physiology in a manner compatible with the disease process following the insertion of the abnormal gene into cellular systems in vitro. Similarly, the observation of relevant pathophysiological or behavioural changes following the "knocking out" (disabling) of homologous genes in laboratory animals is highly suggestive that at least the important sequence variation is within that particular gene. Finally, compelling evidence is provided by the "rescuing" of animals displaying relevant pathophysiological features by the insertion of "normal" (wild-type) forms of the gene.

9
Positional Cloning Studies in Schizophrenia and Bipolar Disorder

Despite the difficulties discussed in the preceding paragraphs, it now seems that progress is finally being made in the search for susceptibility genes for major mental disorders. In the discussion that follows, we will restrict ourselves mainly to a discussion of positive findings that have so far achieved some measure of support from more than one research group. The reason for this is that, while it is important to report all data (Lander and Kruglyak 1995), failures to replicate findings are not particularly convincing evidence against linkage.

9.1
Bipolar Disorder

Great interest has recently been directed at chromosome 18 following a report of linkage to markers for this chromosome and bipolar disorder (Berrettini et al. 1994a). Although no significant LOD scores were obtained in the complete sample, modest LOD scores were obtained in some families if different models of transmission were applied to different families. For reasons discussed above, a model free sib-pair analysis was then undertaken, and suggestive evidence for allele sharing was found in the complete data set. The following year, another group reported data for markers on chromosome 18 (Stine et al. 1995), but the results are difficult to interpret. Although some excess allele sharing was found within the region suggested by Berrettini and colleagues, the strongest evidence for linkage was obtained families in which transmission was restricted to the paternal side of the family. In these families, a LOD score of 3.51 and allele sharing ($p = 0.00002$) were achieved with marker D18S41, which lies a considerable distance from the area of maximum linkage identified by Berrettini and colleagues. Interestingly, the data set in which linkage to chromosome 18 was first reported (Berrettini et al. 1994a) has now been examined, and the pedigrees divided according to the gender of the transmitting parent. No pedigrees were found in which the disease was transmitted by males alone. Instead, this group was able to separate the pedigrees into those in which transmission occurred through females and those in which transmission occurred through both parents (Gershon et al. 1996). Significant allele sharing was observed primarily in the families in which both father and mother transmitted the disease. Whereas the area of maximal linkage in this study was on the short arm of chromosome 18 (Gershon et al. 1996), the previous study (Stine et al. 1995) found the maximum area of linkage to be on long arm of chromosome 18. Nevertheless, there was some overlap in the regions in which there was some evidence for linkage between studies, and it is conceivable that both groups have identified evidence for the same locus.

Another group has recently reported the results of a genome scan in Costa Rican pedigrees. Although no marker gave convincing evidence for linkage, three of the six markers giving the greatest LOD scores were located on chromosome 18 (Freimer et al. 1996). Because the Costa Rican bipolar pedigrees come from an isolated population and the disease is believed to originate from a few founders going back only about seven generations or so, large areas of linkage disequilibrium surrounding a disease locus can be expected due to the relatively small number of meiotic events separating all family members. In their study, Freimer et al. (1996) found that most of the families shared similar allelic variants across an 8-cM region of chromosome 18, and using the more familiar case control association approach, evidence for linkage disequilibrium was obtained between some markers in this region and bipolar disorder. Unfortunately, while this region may possibly overlap with the area of linkage in one of the previous studies (Stine et al. 1995), it does not overlap with the particular candidate region that is shared by Stine and Berrettini. Therefore, although several reports suggest that chromosome 18 contains a susceptibility gene for bipolar disorder, the results are as yet inconclusive.

The second region of the genome that has continued to provide controversy is the X chromosome. Reports suggestive of linkage of bipolar disorder to the X chromosome date back to the 1960s, when two large pedigrees were reported in which bipolar disorder

appeared to be co-inherited with colour blindness (Reich et al. 1969). Despite the absence of any formal evidence for X linkage in the pattern of transmission of bipolar disorder, positive reports of linkage continued to emerge, e.g. to a marker for blood clotting factor IX (F9) and also glucose-6-phosphate dehydrogenase (G6PD) deficiency (reviewed in McGuffin et al. 1994). Unfortunately, the distance between some of these markers is so large that the same manic depression gene cannot be linked to all these loci. A re-evaluation of G6PD in the linked families using DNA markers (rather than the less informative and imprecise measures of G6PD enzyme activity) has considerably reduced the previous evidence for linkage (Baron et al. 1993). Furthermore, evidence for linkage to F9 has also diminished after a re-analysis of the other main data set that had provided strong evidence for linkage (Mendelbaum et al. 1995). However, just as the tide appeared to be turning decisively against X linkage, a maximum LOD score of 3.54 has recently been obtained with markers close to F9 (Pekkarinen et al. 1996). It is therefore premature to regard the X-linkage hypothesis of bipolar disorder as disproven.

While chromosome 18 and the X chromosome are the regions of potential linkage that have sparked the greatest interest, a number of other regions are currently considered "hot spots". These include the long arm of chromosome 21, the long arm of chromosome 12 and the short arm of chromosome 4 (Risch and Botstein 1996). The latter result (Blackwood et al. 1996) appears to meet criteria for significant linkage, but as yet it does not meet the more important criterion of replication.

9.2 Schizophrenia

The first result in schizophrenia of continued interest originated from a conventional genome search in large families which yielded weak evidence for linkage to the long arm of chromosome 22. Despite an initial multicentre "replication" study that was negative, weakly positive data from this region was produced by a number of independent groups, and a meta-analysis of data from 11 centres using the more robust sib-pair approach suggested ($p = 0.001$) that there is a locus for schizophrenia in this region (Schizophrenia Collaborative Linkage Group 1996).

Linkage disequilibrium methods have been applied in an attempt to refine the putative linkage to chromosome 22. In addition to the encouraging reports of linkage, this region has been targeted because a dysmorphic deletion syndrome (velo-cardio-facial syndrome) which may carry a high risk of psychotic illness maps adjacent to this region. Two groups were able to demonstrate significant allelic association between schizophrenia and alleles of the marker D22S278 (Vellada et al. 1995; Moises et al. 1995). However, the findings of Moises and colleagues were not statistically significant after correction for multiple testing. Further studies of linkage disequilibrium in this region are required for definitive results.

A second putative gene for schizophrenia has been mapped to chromosome 6p24-p22 in two overlapping data sets consisting of more than 250 small Irish pedigrees (Wang et al. 1995; Straub et al. 1995). The data suggest that either a single major locus for schizophrenia maps to this region in 15%–30% of families or, alternatively, that there is oligogenic inheritance. Three independent groups have subsequently reported suggestive, but not significant linkages to 6p (O'Donovan and Owen 1996b). Although the exact location of the strongest evidence for linkage varies between studies, there is an overlap in the linked area, and it seems likely there is a susceptibility gene for schizophrenia in this region, albeit of small effect. It is interesting that this area also contains the HLA complex, which was explored prior to the advent of DNA markers for possible involvement in schizophrenia (McGuffin et al. 1994).

As other genome searches approach completion, more regions can now be classified as "hot spots" or potential areas of linkage. These include the short arm of chromosome 8 and the long arm of chromosome 13 (O'Donovan and Owen 1996b). None of these meet stringent levels of significance, and attempts to extend them are under way using collaborative data sets from multiple centres.

10 Candidate Gene Approaches

In the preceding discussion, we have considered strategies for identifying genes where the pathophysiology of the disease is completely unknown. However, where there are some clues to the aetiology of a disorder, a different approach is possible. Instead of identifying susceptibility genes on the basis of their map position, genes can be selected on the grounds that their function suggests a possible role. The associations between HLA genotype and several diseases in which "auto-immune" mechanisms are thought to be involved, such as ankylosing spondylitis, diabetes and rheumatoid arthritis, are probably the best known examples of these.

Initially, polymorphisms in or around candidate genes are identified in patients as described earlier, and then evidence for either linkage or linkage disequilibrium is sought using the identified sequence variations as genetic markers. The potential benefit of

this approach is that most studies benefit from a considerable increase in power if the markers are located close to the pathogenic sequence. Furthermore, if one chooses the candidate gene with skill or good fortune, disease genes can be identified without the whole complex procedure of positional cloning.

In practice, the grounds for choosing candidate genes for mental disorders are unsatisfactory because of uncertainty concerning aetiology. The more obvious candidates are genes involved in dopaminergic and serotonergic transmission in schizophrenia and genes involved in adrenergic and serotonerigic transmission in affective disorders. A large number of such studies have now been undertaken, but the findings are largely unimpressive and, at the time of writing, there are no firm findings of allelic association in bipolar disorder. There is, however, still some interest in inconclusive reports of an association between bipolar disorder and tyrosine hydroxylase (Leboyer et al. 1990), and more recently between both unipolar and bipolar disorder and the serotonin transporter (Ogilvie et al. 1996). Interestingly, a functional polymorphism in the promoter region that affects expression of the latter appears to be associated with neuroticism, although this awaits replication (Lesch et al. 1996).

In schizophrenia, there are some results that have rather more substance. The first concerns a polymorphism in the dopamine D3 receptor gene (DRD3). Several groups have now reported increased rates of homozygosity for this polymorphism in schizophrenic patients and, despite a number of negative reports, a recent meta-analysis of 23 data sets supports the findings (Williams et al. 1998). Moreover, the meta-analysis suggests that the effect is present primarily in males. A second finding of interest concerns an association between a polymorphism in the 5HT2a receptor gene and schizophrenia. This was first observed in a small Japanese study, but has now been replicated in a large European collaborative study (Williams et al. 1996). Should these findings ultimately be unambiguously confirmed, it should be noted that sequence variations in these genes make only a small contribution to the overall susceptibility to schizophrenia. Furthermore, the criteria suggestive of direct involvement in disease have not been met, because neither variation has any apparent functional consequences. The results of functional studies are therefore keenly awaited.

11
Novel Types of Inheritance

Until now, we have considered DNA sequences that (with the exception of uncommon new mutations) are transmitted unaltered between generations. However, recent evidence suggests that other mechanisms may operate in some psychiatric disorders, namely dynamic mutations and genomic imprinting.

11.1
Dynamic Mutations

Expanded trinucleotide repeats were first described in the gene responsible for fragile X mental retardation syndrome and have subsequently been discovered as the mutation mechanism underlying several neurological and neuropsychiatric disorders, such as myotonic dystrophy, Huntington's disease and a number of spinocerebellar atrophies (O'Donovan and Owen 1996a). Trinucleotide repeats are simple sequences consisting of three bases (e.g. CAG) repeated consecutively several times. Such sequences are sometimes polymorphic in the general population but, as far as is known, most of this variation is without any phenotypic consequences. However, in some circumstances, where the repeat sequence is associated with a gene, the possession of a repeat number beyond a certain threshold may cause disease. For example, Huntington's disease is caused when the CAG repeat number is greater than 36 or 37 repeats in what is now known as the *huntingtin* gene on chromosome 4p.

Pathogenic expanded trinucleotide repeats are sometimes called dynamic mutations because the repeat number may increase or decrease during transmission from parent to offspring. Another important feature of dynamic mutations is that, in most cases, as the repeat number increases, the corresponding phenotype is either more severe or is expressed at an earlier age. Consequently, changes in repeat size between generations may result in phenotypic variation between parent and offspring. Usually, there is a tendency for repeat number to increase in subsequent generations, leading to a progressively more severe phenotype or earlier age of onset in a pedigree. This progressive change is called anticipation, and its presence suggests the operation of an expanded trinucleotide repeat mechanism, although there may also be other as yet unknown biological explanations for this phenomenon.

Several recent studies have suggested that anticipation occurs in the transmission of both schizophrenia and bipolar disorder, but it is still not certain whether this is a real phenomenon or whether it results from sampling biases (O'Donovan and Owen 1996a). Nevertheless, independent evidence has recently been obtained suggesting that expanded CAG or CTG trinucleotide repeats may contribute to the pathogenesis of both disorders. Using a technique called repeat expansion detection (RED), which measures

the largest trinucleotide repeat in genomic DNA, several groups have found that patients afflicted with either disorder have, on average, larger CAG/CTG repeats than controls (O'Donovan et al. 1996). This opens the possibility that some of the apparently complex patterns of inheritance of these disorders (e.g. marked variation in phenotypic severity and a high prevalence of "sporadic" cases) might be attributable to expansion and contraction of a CAG repeat. Recently, a large European multi-centre study has provided further strong support for the involvement of CAG/CTG repeats in both disorders, but definitive proof will require the identification of the specific repeats associated with the disorders (O'Donovan et al. 1996).

11.2 Genomic Imprinting

Certain DNA sequences have different properties depending upon whether they are transmitted by the father or the mother. This phenomenon, called genomic imprinting, is illustrated perhaps most dramatically by Prader-Willi and Angelman's syndromes. Although these syndromes are very different phenotypically (McGuffin et al. 1994), they are both usually the result of micro-deletion of the long arm of chromosome 15. However, Prader-Willi syndrome occurs from a deletion of the paternal chromosome, whereas Angelman's syndrome is the result of a deletion in the maternal chromosome. Less commonly, these syndromes may be caused by uniparental disomy, which means that both copies of a chromosome are inherited from one or other parent. Prader-Willi syndrome then arises when both copies are from the mother (equivalent to a complete deletion of the paternal chromosome), and Angelman's syndrome when both are from the father (equivalent to a complete deletion of the maternal chromosome). The precise molecular mechanism underlying genomic imprinting is not completely known, but it seems likely that one mechanism is that some genes are differentially inactivated or activated during gametogenesis by the addition or removal of methyl groups from regulatory regions of DNA.

Genomic imprinting is suggested at the phenotypic level by a sex difference in the likelihood of transmission of a disease in the absence of a difference in the prevalence of the disease by gender (distinguishing it from sex chromosome linkage). Such a pattern has been observed in the families of one of the studies suggesting linkage of bipolar disorder to chromosome 18 (Stine et al. 1995), although, as discussed above, these findings are not conclusive.

12 Animal Models

As we have seen, even diseases with apparently complex inheritance are now susceptible to dissection using modern approaches to genetic analysis. However, it is also clear that major difficulties remain, particularly as the size of the genetic effect decreases. One way to overcome these difficulties is to use animal models of the phenotype. Animal models confer a number of major advantages. Firstly, most experimental animals, e.g. mice, produce large litters in a short period of time, thus providing large numbers for analysis. Secondly, environmental variables can be strictly controlled and therefore minimized (or maximized) in animal experiments. Thirdly, animals can be subjected to specific breeding programmes designed to maximize the detection of genes of small effect. Several breeding paradigms have been successfully used to map complex traits such as stroke, cancers, epilepsy, obesity and diabetes (Frankel 1995). Here, we will just mention one that has recently been used to map genes of neuropsychiatric interest; the F_2 intercross.

The first step is to identify inbred strains of animals (usually mice) that have high (strain A) and low (strain B) scores for the phenotype of interest. It is a characteristic of inbred strains that members of each strain are effectively genetically identical and are homozygous for all alleles in the genome. Thus strain A has genotype AA at every map position, and strain B possesses genotype BB. The two inbred strains are then crossed to produce an F_1 generation. Each F_1 member carries a single copy of every allele from both progenitor strains and is also effectively genetically identical (genotype AB at every locus). Members of the F_1 generation are then bred or intercrossed, which allows recombination to take place and results in F_2 offspring. At any given locus, an F_2 animal may possess one of the following genotypes: homozygous AA, heterozygous AB or homozygous BB. If progenitor A has high scores for the phenotype, then animals in the F_2 generation that have high scores for the phenotype should tend to be homozygous for the A alleles, and low scorers should carry homozygous alleles from strain B (ignoring any dominant gene effects). Moreover, because there is effectively only one generation during which recombination has taken place (the F_1 generation, being genetically identical, does not count), large chromosomal regions surrounding the mutations of interest will be in linkage disequilibrium, and therefore fairly distant markers will also tend to display the genotype AA or BB around the mutation of interest in high and low scorers, respectively. The task is then one of testing for allelic association, but

because of the large areas of linkage disequilibrium, only a few hundred markers are required.

Regions of putative linkage can then be more precisely defined by examining animals that are more distantly related, e.g. by examining outbred animals strains. This is similar to a case control association study in humans. Finally, the identification of the genes responsible for the phenotype in the animals is, in principle, similar to end-point positional cloning in humans. These genes then become candidate genes in human linkage disequilibrium studies.

Although this approach is very appealing, there are few obvious models for major mental illness. No animal model can expect to fully emulate the complexity of behavioural traits such as schizophrenia and manic depression. However, even modelling some of the features of a disease affords the possibility of uncovering some of the genes that contribute to it. At present, models of alcoholism, drug abuse, anxiety and depression are available and have been used. For example, an approach which modelled "emotionality" has recently mapped a number of loci (Flint et al. 1995), and similar approaches have been successful in mapping loci that may be related to drug abuse and alcoholism (Berrettini et al. 1994b). However, it remains to be seen whether the genes associated with the phenotypic traits in rodents contribute to the modelled behaviour in humans.

13 Concluding Remarks

In contrast with the dramatic "advances and retreats" of the past (O'Donovan and Owen 1992), research into the molecular genetics of mental disorders is now entering a period of slow, consolidating advances (O'Donovan and Owen 1996b). Tentative, but suggestive findings of linkage are emerging for both schizophrenia and bipolar disorder, and as more genome scans reach fruition, more of these can be expected. The main challenge remains how to move from these results to precisely delineated regions of linkage. This will almost certainly require the use of large samples of families containing affected relative pairs and of unrelated individuals for linkage disequilibrium mapping. A second challenge will be to develop methodologies to facilitate the genotyping of large samples, and several promising strategies are now on the horizon, including DNA pooling and DNA chip technologies. Data from genetic studies of animal models may be useful here, both in the primary identification of potential regions of linkage by synteny, but also of specific genes within candidate regions by homology. Finally, methods are required that will allow the selection and mutation screening of the large number of candidate genes that are expected to map to a particular candidate region. Again, we can be optimistic that data from the Human Genome Project and technical advances in mutation detection will mean that these difficulties should not prevail.

14 References

Baron M, Freimer NF, Risch N et al (1993) Diminished support for linkage between manic depressive illness and X-chromosome markers in three Israeli pedigrees. Nat Genet 3: 49–55

Berrettini WH, Ferraro TN, Goldin LR et al (1994a) Chromosome 18 DNA markers and manic-depressive illness: evidence for a susceptibility gene. Proc Natl Acad Sci USA 91: 5918–5921

Berrettini WH, Ferraro TN, Alexander RC (1994b) Quantitative trait loci mapping of three loci controlling morphine preference using inbred mouse strains. Nat Genet 7: 54–58

Blackwood DHR, He L, Morris SW et al (1996) A locus for bipolar affective disorder on chromosome 4p. Nat Genet 12: 427–430

Flint J, Corley R, DeFries JC, et al. (1995) A simple genetic basis for a complex psychological trait in laboratory mice. Science 269: 1432–1435

Frankel WN (1995) Taking stock of complex trait genetics in mice. Trends Genet 11: 471–477

Freimer NB, Reus VI, Escamilla MA et al (1996) Genetic mapping using haplotype, association and linkage methods suggests a locus for severe bipolar disorder (BPI) at 18q22-q23. Nat Genet 12: 436–441

Gershon ES, Badner JA, Detera-Wadleigh SD et al (1996) Maternal inheritance and chromosome 18 allele sharing in unilineal bipolar illness pedigrees. Am J Med Genet (Neuropsychiatr Genet) 67: 202–207

Lander E, Kruglyak L (1995) Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. Nat Genet 11: 241–247

Leboyer M, Malafosse A, Boularand S et al (1990) Tyrosine hydroxylase polymorphisms associated with manic-depressive illness. Lancet 335: 1219

Lesch KP, Bengel D, Heils A et al (1996) Association of anxiety-related traits with a polymorphism in the serotonin transporter gene regulatory region. Science 274: 1527–1531

Lovett M (1994) Fishing for complements: finding genes by direct selection. Trends Genet 10: 352–357

McGuffin P, Owen MJ, O'Donovan MC, Thapar A, Gottesman II (1994) Seminars in psychiatric genetics. Gaskell, London

Mendelbaum K, Sevy S, Souery D (1995) Manic-depressive illness reanalysis in the Xq27-Xq28 region of chromosome X. Neuropsychobiology 31: 58–63

Moises HW, Yang L, Li T et al (1995) Potential linkage disequilibrium between schizophrenia and locus D22S278 on the long arm of chromosome 22. Am J Med Genet (Neuropsychiatr Genet) 60: 465–467

O'Donovan MC, Owen M (1992) Advances and retreats in the molecular genetics of major mental illness. Ann Med 24: 171–177

O'Donovan MC, Owen MJ (1996a) Dynamic mutations and psychiatric genetics. Psychol Med 26: 1–6

O'Donovan MC, Owen MJ (1996b) The molecular genetics of schizophrenia. Ann Med 28: 541–546

O'Donovan MC, Guy C, Craddock N et al (1996) Confirmation of association between expanded CAG/CTG repeats and both schizophrenia and bipolar disorder. Psychol Med 26: 1145–1153

Ogilvie AD, Battersby S, Bubb VJ (1996) Polymorphism in serotonin transporter gene associated with susceptibility to major depression. Lancet 347: 731–733

Pekkarinen P, Terwilliger J, Bredbacka P-E et al (1996) Evidence of a predisposing locus to bipolar disorder on Xq24-q27.1 in an extended Finnish pedigree. Genome Res 5: 105–115

Reich T, Clayton PJ, Winokur G (1969) Family history studies. V. The genetics of mania. Am J Psychiatry 125: 1358–1369

Risch N, Botstein D (1996) A manic depressive history. Nat Genet 12: 351–353

Schizophrenia Collaborative Research Group (1996) A combined analysis of D22S278 marker alleles in affected sib-pairs: support for a susceptibility locus for schizophrenia at chromosome 22q12. Am J Med Genet (Neuropsychiatr Genet) 67: 40–45

Schuler GD, Bosuski MS, Stewart EA et al (1996) A gene map of the human genome. Science 274: 540–546

Spielman RS, Ewens WJ (1996) The TDT and other family-based tests for linkage disequilibrium and association. Am J Hum Genet 59: 983–989

Stine OC, Xu J, Koskela R et al (1995) Evidence for linkage of bipolar disorder to chromosome 18 with a parent-of-origin effect. Am J Hum Genet 57: 1384–1394

Straub RE, MacLean CJ, O'Neill FA et al (1995) A potential vulnerability locus for schizophrenia on chromosome 6p24-24: evidence for genetic heterogeneity. Nat Genet 11: 287–293

Vellada H, Curtis D, Sham PC et al (1995) Chromosome 22 markers demonstrate transmission disequilibrium with schizophrenia. Psychiatr Genet 5: 127–130

Wang S, Sun CE, Walczak CA et al (1995) Evidence for a susceptibility locus for schizophrenia on chromosome 6pter-p22. Nat Genet 10: 41–46

Williams J, Spurlock G, McGuffin P et al (1996) Association between schizophrenia and T102C polymorphism of the 5-hydroxytryptamine type 2a-receptor gene. Lancet 347: 1294–1296

Williams J, Spurlock G, Holmans P et al (1998) A meta-analysis and transmission disequilibrium study of association between the dopamine D3 receptor gene and schizophrenia. Psychiatr Genet 3: 141–149

CHAPTER

5

J.A. Gingrich, R. Hen

Using Transgenic Mice to Probe the Role of Specific Genes in Behavior

1 Introduction

This chapter will describe transgenic techniques that permit experimental manipulation of gene expression in the mouse. These techniques create specific and permanent changes to the germline of experimental mice, and these strategies have been fruitfully applied to the study of oncology, immunology, development, and more recently neurobiology and behavior. Specifically, in the area of neurobiology, modification of gene expression in the mouse has permitted direct assessment of the role of specific genes in the development of neural physiologic processes and in regulating behavior. Since this technique and its application to mouse behavior may seem far removed from the subject of human psychiatric disorders, we will first begin with a review of some basic concepts in molecular biology, followed by a brief introduction that will place the transgenic mouse strategy in the context of other genetic approaches to studying behavior.

2 Molecular Basis of Gene Expression

2.1 Chromosomes and Genes

Figure 1 will be a helpful aid to the reader in this section. A chromosome is a linear piece of DNA that is comprised of elements called genes. The human genome is comprised of 23 chromosome pairs that are packaged in the nucleus (mice have 20 chromosome pairs). Genes contain two parts: a coding region that specifies the sequence of a messenger RNA (mRNA), and a promoter region that specifies the temporal and spatial expression properties of the mRNA. Consistent with these two aspects of a gene, the chromosomal DNA serves two functions: it acts as a stable record of an organism's genetic information, and it is a dynamically regulated template that allows for selective expression of genes. The information-carrying function of DNA is intrinsic to its double-stranded structure and its mode of replication. The dynamic aspects of gene regulation are provided by the promoter. The promoter region confers the ability of a gene to be expressed during certain times in development and to be expressed in some cells but not others. This regulatory function is crucial, because each cell in an organism (except for the gametes) contains exactly the same genetic information – a complete copy of all genes. What determines the fate of a cell is the genes that it expresses. Thus a neuron is a neuron and not a kidney cell because of the repertoire of genes that it expresses. Each cell expresses approximately 15%–30% of the available genes (15,000–30,000 of an estimated 100,000 genes in the human genome). In Fig. 1, the promoter is shown to be comprised of an enhancer (that confers temporal and cell-specific expression) and a TATA box (where the RNA polymerase will begin transcription). A gene becomes transcriptionally active when the enhancer elements of a gene loop toward the RNA polymerase binding site and activate the RNA polymerase. Thus an RNA copy of one strand of the gene is made (transcription).

2.2 Ribonucleic Acid

An RNA intermediate between genes and protein is required because chromosomes stay in the nucleus, but proteins are made in the cytoplasm of cells. Thus an intermediate step is needed to get a copy of the gene out to the cytoplasm. Enzymes in the nucleus (RNA polymerases) create copies of genes in the nucleus in a process called transcription. The nuclear RNA transcript is a "raw" copy of the gene, containing both exons and introns.

2.3 Exons and Introns

As shown in Fig. 1, the coding region of the gene is comprised of exons and introns. The coding information for a protein is carried on exons. Exons are separated by stretches of sequence called introns. Only eukaryotic organisms have introns that interrupt the exon sequences, and their function is not known (prokaryotes do not). As shown, a gene always begins with an exon and ends with an exon. The first and last exons of a gene will contain some sequence that will never be translated into protein (therefore called the 5′- and 3′-untranslated region of the mRNA).

2.4 RNA to Protein

Before an RNA can be turned into a protein, it must be modified before it leaves the nucleus. First, the exons must be joined together to make a contiguous stretch of RNA. This means that the introns must be spliced out. The spliced RNA transcript is then modified with a "cap" on one end and a polyadenosine (polyA) tail on the other. These modifications confer stability to the RNA, which in its modified form is called messenger RNA or mRNA. The mRNA is then transported to

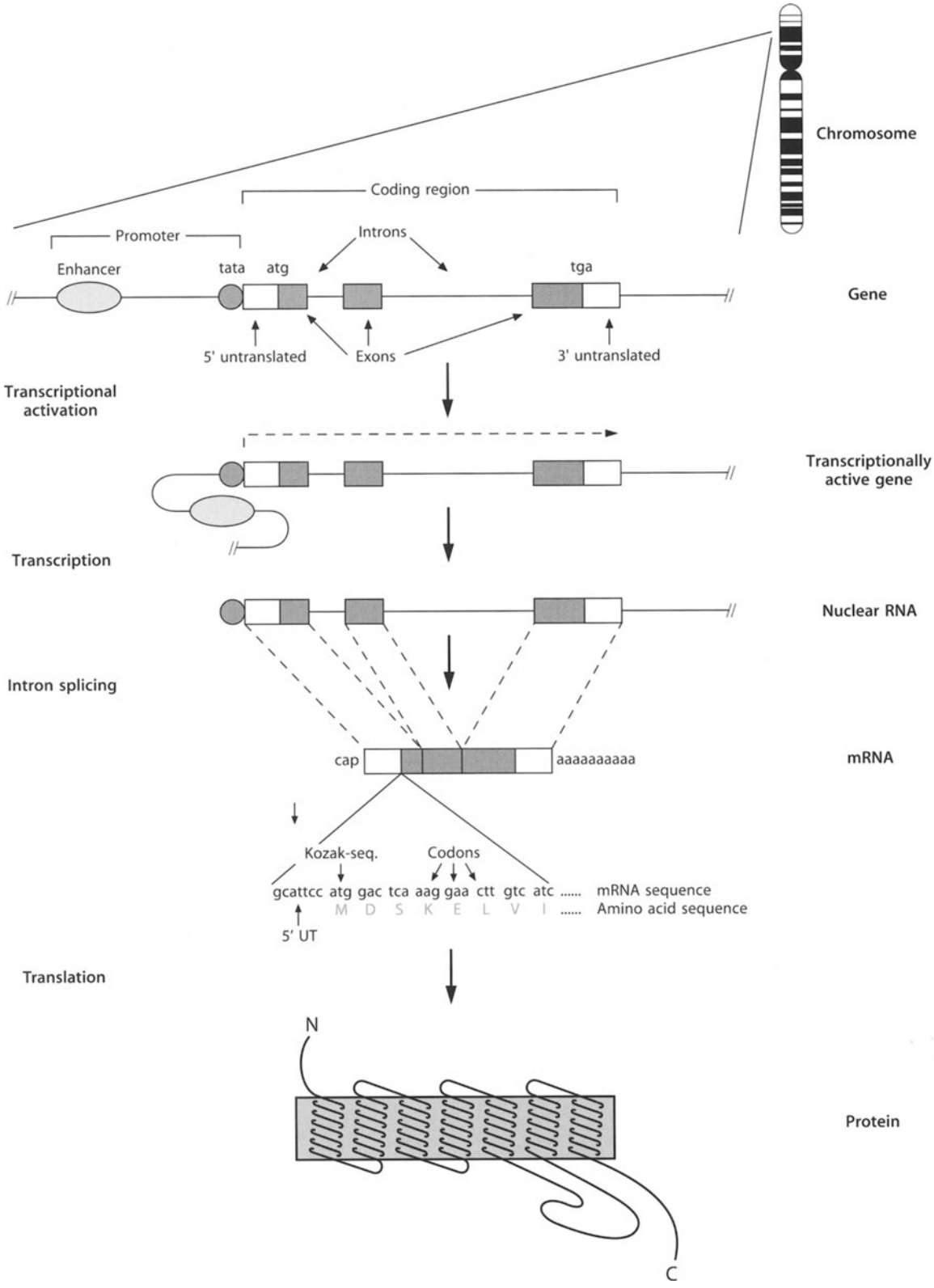


Fig. 1. The path from DNA to protein. See text for details

the cytoplasm of a cell, where ribosomes decode the sequence of the mRNA to synthesize a polypeptide chain or protein. The process of generating protein from mRNA is called translation. The ribosome begins translation of an mRNA from a sequence element called the Kozak site (named after Marilyn Kozak, who defined it). Within the Kozak site is the first codon, ATG, which begins every eukaryotic protein and codes for the amino acid methionine. The information that allows the ribosome to "decode" the mRNA is the codon. Codons are three nucleotide bases in length. Nucleotides are of four types: thymidine, adenosine, cytosine, and guanine (or T, A, C, and G as they are commonly abbreviated). Any three nucleotide bases can form a codon, and $4^3 = 64$ different possible codons can therefore be generated. Different codons specify one of 20 possible amino acids or a stop signal for the ribosome. In the example shown in Fig. 1, a G protein-coupled receptor is being made. This family of receptors comprises integral membrane proteins that traverse the lipid bilayer seven times.

3 DNA Engineering

Over the last 30 years, the technology has been developed that allows researchers to isolate, manipulate, and reengineer DNA in almost any form imaginable. A basic understanding of the techniques of DNA engineering will be helpful in understanding the more technical parts of this chapter.

3.1 Cloning

Cloning is the process of isolating a fragment of DNA that represents a gene or a DNA copy of an mRNA (so-called cDNA). In order to clone a gene, it is necessary to know something about it: some partial nucleotide sequence or a function that can be easily assayed in cells. If the sequence of a gene is known, the gene can be cloned by making a probe using the sequence. The probe can be used to find the gene in a collection of clones known as a library. The ability of a probe to identify a gene depends on the property of single-stranded DNA sequences to bind or hybridize to complementary strands. Once a clone is identified in a library by a probe, the clone can be amplified to generate large amounts of this DNA. This amplification is possible because isolated DNA fragments can be propagated conveniently in bacteria in the form of plasmids (circular forms of DNA that grow in bacteria). Plasmid DNA can be easily grown in large quantities in bacteria (amplification) and purified to yield pure plasmid DNA that is free from the bacteria's genomic DNA or RNA. In this pure form, the plasmid DNA can be manipulated with different enzymes.

3.2 Polymerase Chain Reaction

Another way to generate large amounts of specific DNA sequences without the need for cloning uses a process called the polymerase chain reaction (PCR). Successful PCR also depends on some knowledge of the sequence of a gene. In this case, the sequence information is used to generate small synthetic pieces of DNA called oligonucleotide primers. The primers are chosen to flank the part of the gene that is to be amplified. When a double-stranded DNA template is heated to 95 °C, the two strands separate (denature). Upon cooling, the primers will bind to their complementary sequence on each strand, and a DNA polymerase added to the reaction then synthesizes another complementary strand from the primer (which is why they are called primers). This process of heating and cooling is called a cycle. In this one cycle, the amount of DNA between the primers has doubled. If another such cycle is performed, the doubled DNA is doubled again. Repeating this process with 30 or 50 cycles can amplify a given DNA sequence from 2^{30}- to 2^{50}-fold (or in excess of a billionfold). This technique has been optimized over the last 10 years to incorporate a thermostable DNA polymerase that can withstand the repeated cycles of high temperature needed to denature double-stranded template. PCR now permits amplification of specific sequences from the DNA contained within a single cell. The revolution that this technique has created in molecular biology cannot be overstated.

3.3 Restriction Enzymes

DNA can be cut very specifically using enzymes known as restriction endonucleases (or restriction enzymes). These restriction enzymes recognize specific sequences in DNA from four to eight nucleotides in length and cleave the DNA strand within or nearby this recognition sequence. Examples of such enzymes are *Eco*RI, *Bam*HI, *Hin*dIII, and *Xho*I. The ability of these enzymes to cut a gene allows a type of map to be created known as a restriction map. DNA fragments generated in this way can be ligated to other similarly cut DNA fragments in a process that resembles cutting and pasting. In this way, genes can be cloned,

modified, and used to make targeting constructs for making knockout mice.

4 Behavioral Genetics

The field of behavioral genetics is the study of the heritable component of behavior and personality. The goal of this research is no longer to establish that behavior is influenced by heritable factors, but to precisely determine the genes involved, the nature of their interactions, and their interaction with the environment. In humans, genetic factors are believed to influence as much as 30%–50% of the variance of intelligence, cognition, and personality traits such as extroversion (sociability, impulsiveness, and liveliness), neuroticism (moodiness, anxiousness, irritability), agreeableness (likability, friendliness), conscientiousness (conformity, will to achieve), and culture (openness to new experience). Most behavioral and psychological traits exist on a continuum and thus appear to have multiple genes that contribute to the "phenotype" of an organism.

Like normal personality characteristics, psychiatric disorders also appear to have a genetic component. The genetic contribution to illnesses such as schizophrenia, autism, and bipolar disorder are more substantial than that measured for personality traits. Disorders such as depression and anxiety also have a genetic contribution, but less than that found for the aforementioned psychiatric conditions. It has also become accepted recently that complex behaviors and disorders such as schizophrenia are not the result of a single aberrant gene, but rather the result of the interaction of multiple genes. Thus geneticists are no longer looking for the schizophrenia gene, but rather one of the many genes that may contribute to this disorder. The influence of genes on normal behavior and on abnormal clinical psychiatric conditions is therefore complex and likely involves complex interactions such as gene–gene, gene–environment, and environment–gene interactions. The difficult problem facing behavioral genetics is how to identify the genes involved and understand their contribution to behavior, personality traits, and psychiatric disorders. To address this problem, two conceptually distinct approaches are taken: so-called forward genetics, which begins with a phenotype and searches for the gene or genes that play a role, and so-called reverse genetics, which creates a specific genotypic change (mutation) in an organism and searches for the phenotypic changes that result from the mutation. The transgenic strategy falls into the latter category and will be discussed in detail. The concept of forward and reverse strategies is outlined in Fig. 2.

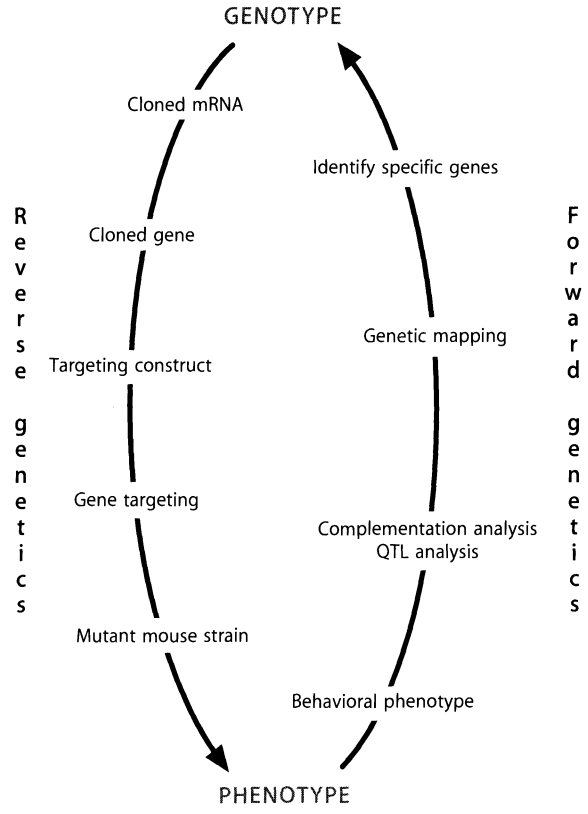


Fig. 2. Forward and reverse genetic strategies. See text for details

4.1 Forward Genetic Strategies

Other chapters in this volume describe different strategies used to identify the genes that contribute to a given behavior. Most of these methods fall into the category of so-called forward genetics. The massive efforts to identify genes that contribute to disorders such as schizophrenia and bipolar disorder are an example of this approach. In these studies, investigators begin with a rigorously defined phenotype (usually DSM-VI or Research Diagnostic Criteria for a disorder) and attempt through various strategies to identify genetic loci that are associated with the disorder.

These types of searches are simplified when there is only a single gene involved with nearly complete penetrance of its phenotype. Many of the neuropsychiatric disorders such as Huntington's disease, fragile X syndrome, dystonia syndromes, muscular dystrophy, and even some cases of Alzheimer's disease have been found to be single-gene disorders. The search for

genes in more complex, multigene disorders has moved less rapidly, but the first such genes for disorders such as schizophrenia and bipolar disorder are beginning to appear and be replicated independently.

In the nonhuman animal realm, a more controlled approach to this type of problem is possible due to the existence of genetically homogeneous strains that exhibit different behaviors. Genetic analysis of such strain differences (quantitative trait loci or QTL analysis) can identify the number of genetic loci that contribute to a given behavior (as well as its contribution relative to the other loci). Similarly, new strains of animals can be created by breeding successive generations for their difference on a given behavioral trait. Once such lines have been sufficiently separated phenotypically and genetically, and stabilized by inbreeding, a similar approach can be undertaken to identify loci that contribute to the behavior of interest. Such QTL approaches have been only one of the trends over the past 30 years that has marked the convergence of efforts of quantitative geneticists and molecular geneticists to identify particular genes that contribute to a given behavior.

4.2
Reverse Genetic Strategies

A more direct approach to studying the effect of genes on behavior uses techniques that mutate selective genes. This approach has been applied extensively in simple organisms such as bacteria, *Caenorhabitis elegans*, and *Drosophila* using chemical mutagens and selection procedures to find organisms that differ on a given behavior. Such approaches have led to the identification of dozens of genes that contribute to chemotactic behavior in organisms such as bacteria, paramecium, and *C. elegans*. This approach in *Drosophila* has identified genes that are involved in more complex behaviors such as learning, locomotion, and sexual behavior.

Animal models continue to be essential tools for the study of neuropsychiatric disorders. For example, specific animal models have been most widely used to screen for potential psychotropic medications. In addition, animal preparations have been crucial to investigate mechanisms of brain function (such as development, plasticity, neurotransmission), behavioral states (such as anxiety, depression, and aggression), and behavioral processes (such as learning, appetite control, locomotion, and biological rhythmicity). These models take many forms, and thorough discussions of this topic can be found elsewhere in this volume.

5
Techniques of Genome Manipulation in Mammals

This section describes the methods that have been developed to manipulate the genome of experimental animals – primarily mice. Some confusion arises among those unfamiliar to this field regarding the differences between the different techniques, e.g., the difference between a transgenic mouse and a so-called knockout mouse. In fact, both types are technically considered transgenic given that the definition of a transgenic mouse is an animal that has foreign DNA stably integrated into its genome and passable to its offspring (germline transmission). There are three established techniques for introducing foreign DNA into the mouse germline: (1) retroviral vectors, (2) direct injection of DNA into pronuclei of fertilized eggs, and (3) use of embryonic stem cells that have been transfected with foreign DNA. Each of these methods will be considered.

5.1
Infection of Developing Embryos with Retroviral DNA

This technique of infecting developing embryos with retroviral DNA is useful for studying development, including neurodevelopment, generating mutant animals by insertional mutagenesis, and to a lesser degree as a method to introduce functional genes into the genome. This last use has been largely supplanted by the other techniques that we will discuss in this chapter. The retroviral method has been reviewed elsewhere (Rossant 1990; Babinet et al. 1989) and will not be discussed further in this chapter. The two remaining methods of creating transgenic mice that will be discussed in more detail include random introduction of foreign or engineered transgenes into the mouse genome and techniques that permit site-specific integration of foreign or engineered DNA into the mouse genome.

5.2
Random Integration of Transgenes

This approach depends on the direct injection of foreign DNA into the pronuclei of fertilized eggs. This technique was first demonstrated in mice in 1980 and has subsequently been applied to the creation of numerous mutant mouse lines. The concept of this method is to use DNA engineering to create a fusion construct of a functional promoter region with a gene to be expressed. This construct is then injected directly into the pronuclei of a fertilized single cell embryo.

The engineered DNA construct then inserts randomly into the germline of the mouse. Typically, multiple copies of the transgene are integrated in tandem head-to-tail arrays.

The choice of the promoter and the gene used to make the construct (i.e., transgene) provides almost limitless possibilities for expression strategies. For example, promoters have been characterized that drive gene expression in limited anatomic or tissue-specific patterns or that provide strong expression in many different tissues. The gene that is transcribed can be of any variety, including genes normally expressed in the animal, mutated genes, or so-called antisense genes that express the complementary strand of an existing gene. How these various strategies can and have been applied to modeling neuropsychiatric disorders will be discussed in the next section.

5.3 Specific Gene Targeting Strategies

The specific gene targeting strategy, commonly known as the "knockout" approach, differs from the transgene approach in that the mutation occurs in a specified gene locus, modifying or ablating an existing gene. This technique depends on the property of cells to undergo a process called "homologous recombination." This process is presumably similar to the process of recombination that occurs between two sister chromosomes during meiosis: namely, two homologous stretches of DNA can recombine, essentially replacing one DNA sequence with another. For the purpose of the knockout strategy, the experimenter introduces an engineered version of the gene to be disrupted into an embryonic cell line. The targeting construct possesses two "recombinogenic arms" that will "target" the engineered construct to the native gene locus. Different types of recombination events can occur depending on the nature of the targeting construct. As shown in Fig. 3a, when an "omega" targeting construct undergoes homologous recombination with its target gene, a substitution of the targeting construct for the endogenous gene occurs. If the targeting construct is engineered to delete a crucial exon, introduce a frame-shift mutation, or interrupt the coding sequence with another gene (such as a gene that confers neomycin resistance to cells), the targeted gene is effectively inactivated. This strategy also permits introduction of additional DNA sequences without loss of the endogenous gene (so-called knock-in strategy; Fig. 3b). This is an important strategy for creating inducible or tissue-specific knockouts, which will be discussed in the next section. In Fig. 3b, the example shown simply places a reporter gene, *lacZ*, into the locus of the gene of interest. In this example, neurons that would normally express the gene of interest will now express the *lacZ* gene, whose gene product can stain the neurons a blue color for tracing purposes. Such a construct has been elegantly used to study the organization of the olfactory bulb in relation to olfactory receptors (Mombaerts et al. 1996).

The second type of targeting construct that has been used (but less so recently) is the "O-type" construct shown in Fig. 3c. When such a construct undergoes homologous recombination, a duplication event occurs (Bronson and Smithies 1994). Even if the two ends of the targeting construct have a large gap between them, endogenous processes in the cell permit this gap to be filled. Thus even a very large gene and its regulatory (promoter) regions can be duplicated using a targeting construct that contains only a small portion of the total gene. This strategy creates the opposite of a knockout, namely an increase in the function of a given gene (Smithies and Kim 1994). This strategy allows the experimenter to increase gene dosage rather than to decrease it. Although less commonly performed, such mutations may prove to be an important compliment to the knockout approach.

Although we have explained the concept of homologous recombination and its differences from the transgene approach, we have not discussed the methods by which a targeting construct can be introduced into a viable animal. Whereas the transgene is injected directly into a fertilized embryo, the ratio of homologous recombination to random integration is too low (often greater than 1:500) for this approach to be viable. The number of embryos that would need to be injected in order to obtain the desired recombination event would not be practical. To address this problem, a new strategy needed to be developed. Thus nearly all homologous recombination experiments are performed in a cell line derived from pluripotent embryonic stem (ES) cells. These cells can be grown in standard cell culture conditions and be propagated in an undifferentiated state. The targeting construct is introduced into these cells by a process of electroporation in which a high-voltage shock temporarily renders these cells permeable to a solution containing the linearized DNA targeting construct. The DNA enters only a small portion of the cells, and only a small portion of the cells that take up the construct will integrate the DNA into their genome. Of those cells that integrate the targeting construct into the genome, only a small fraction will undergo the process of homologous recombination (1/10 to 1/1000) that integrates the targeting construct in the intended locus. To appreciate the low probability of this process, in a typical experiment the researcher electroporates 10 million cells, of which perhaps 5000 cells will integrate the DNA randomly in their genome and 50 cells will undergo homologous recombination. Thus

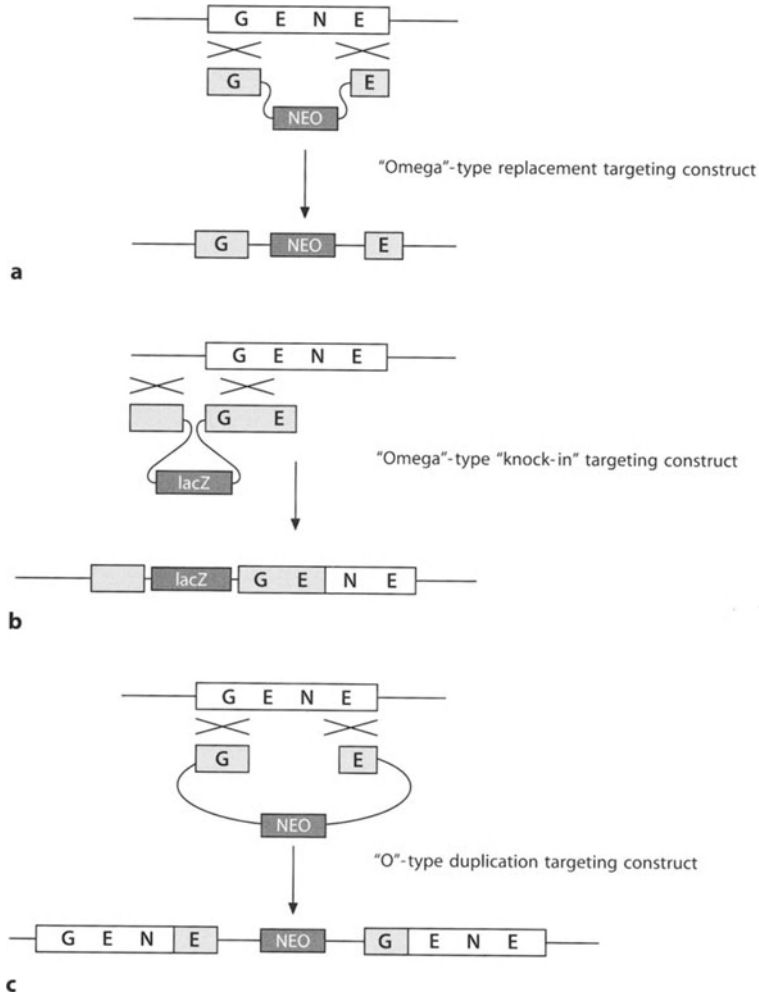


Fig. 3a–c. Different strategies for gene targeting. **a, b** An omega-targeting construct is used to generate either a "knockout" or a "knock-in." **c** An O-type construct is used to create a gene-duplication event. See text for details. (Adapted from Smithies and Kim 1994)

the task for researchers making a knockout is to find those 50 cells in the 10 million cells from which their experiment started. For this purpose, a selection procedure is used to distinguish cells that have integrated DNA from those which have not. The most commonly used selectable marker is the gene from bacteria that confers resistance to the toxin neomycin. Thus growth of electroporated cells in neomycin-containing medium eliminates the vast majority of cells that did not integrate the target DNA (conveying neomycin resistance), and we are left with the more reasonable task of finding the homologous recombination events within this background of nonspecific integration events.

Two strategies are used at this step: (1) a brute force approach in which the experimenter screens large numbers of clonal cell lines (200–1000 lines) using techniques such as Southern blot analysis or PCR to distinguish specific from nonspecific integration, or (2) a second selectable marker, thymidine kinase (TK), can be incorporated into the targeting construct on one extremity of the construct. By incorporating TK on the outside portion of a homologous arm of the construct, it will be lost in all homologous recombination events but be retained in nearly all random integration events. Thus clonal cells grown in the presence of neomycin and gancyclovir will eliminate most unwanted clones (gancyclovir in the presence of TK is toxic to cells) and simplify the process of screening.

When a desired clone possessing a homologous recombination event is obtained, it is amplified in number and then hundreds of the mutated cells are injected into the hollow core of a developing blastocyst. The "host" blastocyst is derived from mice of the C57/B6 strain. Importantly (as will be seen), C57/B6 mice possess a black coat color. The ES cell line is derived from a male embryo of the C129 mouse strain. In contrast to the C57/B6 strain, 129 mice possess a yellow coat color (agouti). The injected blastocysts, now filled with mutant 129 cells, are transferred to a foster mother. As the embryo develops, the mutant cells and the wild-type cells of the blastocyst will both

be integrated into the young mouse. The resulting pups are chimeric or mosaic in the sense that a portion of their body is derived from the mutant cells and a portion from the wild-type C57 strain. This chimerism is reflected in their coat color, which is typically a mix of yellow (agouti) and black. Based on the predominant coat color, the percentage of mutant cells that comprise the resulting animal can be estimated. Researchers are interested in chimeras that will be able to transmit the mutation created in the ES cells onto its progeny. This ability is called "germline transmission." When sufficiently mature, male chimeras are mated with wild-type C57/B6 females (remember that the ES cells were male, and thus the chimeras most likely to transmit the mutation will be male). Because the agouti coat color is dominant over the black coat of the wild-type mice, offspring that bear the mutation will necessarily also have the yellow coat. Despite their uniform yellow coat, these offspring will nonetheless have a mixed genetic (129/C57) background. This heterogeneity can confound behavioral experiments, and thus it is often preferred to have the mutation on a pure genetic background. To achieve this, the chimeric males must be mated with 129 wild-type females. Heterozygote offspring will necessarily be of a pure 129 genetic background (since if they transmitted the mutation, the germline must be derived from the 129 strain). Since coat color cannot be used to follow the mutation, a sample of DNA has to be obtained from the offspring (usually a short snip of the tip of the tail) and PCR or Southern blot analysis used to determine the genotype. The heterozygote offspring can of course be of either gender, and once males and female heterozygotes are obtained, they can be mated to produce offspring homozygote for the desired mutation.

5.4
Advantages and Problems of the Knockout Strategy

The power of the transgenic technique lies in its specificity and the absolute nature of the lesion created. Even in genes of unknown function, or receptors for which no known antagonist exists, an experimenter may investigate the function of the gene and its resulting protein using the transgenic approach. In this sense it represents a new and complementary approach to pharmacological techniques.

5.5
Problem of Inferring Function from Dysfunction

From the point of view of understanding the role of a gene in the functioning of an adult animal, the knockout approach has several drawbacks. First, function has to be inferred from looking for the dysfunction that occurs in an animal in the absence of the gene. Using an automobile analogy, if we tried to infer the function of a steering wheel by trying to see how a car functioned without it, several conclusions might be reached. The car will appear normal from the outside. It will start and idle normally, but problems would quickly become apparent if someone tried to drive the car. If the person operating the car drove slowly to compensate for the lack of directional control, an outside observer might incorrectly conclude that the steering wheel must influence the speed of the vehicle and its absence results in a much slower car. If we were trying to understand the function of the fuel pump by studying car function in its absence, we would quickly conclude that the car does not start or move without it. However, unless we had more detailed assays to perform on the car, the function of the fuel pump would remain largely mysterious. Thus one way of appreciating the knockout strategy is that an experimenter creates a "broken" animal and tries to understand the role of the mutated gene by observing and testing the animal bearing the mutation under a variety of conditions.

5.6
Lack of Tissue Specificity in Knockouts

One reason that the dysfunction can be difficult to interpret is the fact that the gene is inactivated everywhere. Often, the same gene is expressed in different tissues and can serve different functions depending on where it is expressed. For example, tryptophan hydroxylase (TpH) is found in serotonergic neurons and the pineal gland. In serotonergic raphe neurons, TpH is the rate-limiting enzyme for serotonin (5-HT) synthesis. In the pineal gland, TpH helps turn tryptophan into melanin. An animal without a functional TpH gene may thus exhibit mixed effects (melatonin and serotonin absence) of the gene deletion in these two different processes.

5.7
Developmental Problems and Compensation by Other Genes

Another problem that often confounds the interpretation of a dysfunction seen in a mutant animal is the question of the role that such a gene plays in development. Remember that the gene in question has been disrupted in the germline and consequently will never be present. This means that the gene will be absent during crucial periods of ontogeny; if the gene

is involved in development, this could have disastrous consequences for the adult animal (even creating the ultimate phenotype, nonviability). Again, returning to the automobile analogy, if the part being removed is crucial to the subsequent installation of other parts, or if it contains crucial instructions for the ongoing assembly of the automobile, the resulting car could be so badly assembled that examining the finished product would be almost useless for the purposes of understanding the role of the part in a normal car. Likewise, if a deleted gene is an important gene in development, the study of an adult may reveal abnormalities that indicate very little about the role of the ablated gene in adult behavior. Rather, the phenotype is predominated by the developmental defects of the mouse. In such cases, the development of embryos needs to be studied to understand at what point the normal developmental pattern went awry. A similar, but somewhat different problem is the seeming ability of an organism to "compensate" for a missing gene by enhancing expression of normally dormant genes or overexpression of other active genes. A familiar example of this is seen in patients with sickle cell disease or thalassemia, who express high amounts of fetal hemoglobin to compensate for the absence or poor functioning of adult hemoglobin. If such a compensatory process occurs following a gene knockout, such an animal may have no discernible phenotype, i.e., it appears completely normal. In such a case, it is difficult to determine whether the ablated gene simply has no function or whether compensation has obscured the phenotype. Another related possibility is that, in the process of compensating for an absent gene, the animal may create another unexpected phenotype (similar to the steering wheel analogy given earlier). Again, this can cause confusion regarding the role of the gene in adult animals.

5.8
New Technologies to Address the Limitations of the Classical Knockout Approach

To overcome the problems of compensation and development, many groups are working to develop an "inducible knockout" strategy. The idea is to give the experimenter temporal control over the inactivation of the gene. This would allow the animal to develop with the gene present. When the animal reaches adulthood, the gene could then be inactivated. Many strategies are being developed to allow such temporal control of gene activity and are beyond the scope of this chapter. However, the interested reader is referred to the discussion of this topic by Lucas and Hen (1995).

The other new technology being developed will allow for regional or tissue-specific loss of a gene. This strategy involves creating somatic mutations in mice resulting in loss of a gene in certain tissues or brain regions but not others. Again, the technology involved in this "tissue-specific" knockout strategy is beyond the scope of this chapter, but some early successes have been achieved in this area. For example, a group was successful in eliminating the tyrosine hydroxylase gene only in dopaminergic neurons while sparing catecholaminergic neurons (Zhou and Palmiter 1995). In another example, researchers were able to explore the role of the *N*-methyl-D-aspartate (NMDA) receptor in a specific region of the hippocampus by selective disruption of a receptor subunit only in CA1 cells (McHugh et al. 1996; Tsien et al. 1996a,b; Wilson and Tonegawa 1997). Of course, the "holy grail" of transgenic researchers is the development of technology that will allow both temporal and regional anatomic control over the process of gene inactivation. At the time of writing, this remains a topic of active investigation, but without definitive success.

5.9
Knockouts as Models of Human Genetic Disorders

It is worth emphasizing that most of the knockout mutations that are created in mice have no known analogue in humans. However, there are many known single-gene mutations in humans that could be (and have been) effectively modeled in transgenic mice. In the cases in which a knockout mouse is modeling a known human gene mutation, any developmental abnormalities in the mouse will be an interesting feature to study, not necessarily a confounding variable. Indeed, to fully mimic the human gene disorder, a classical knockout needs to be studied. Tissue-specific and inducible knockouts will be valuable in helping to dissect the phenotype that is observed in a classical knockout animal. Indeed, the classical knockout and the other types of knockouts (tissue-specific and inducible) should be seen as complementary approaches that are useful for answering specific biologic questions.

6
Using Knockouts to Study Aggressive Behavior

6.1
Knockouts and Aggression

We have chosen to discuss aggression and knockouts because we believe it illustrates both the promises and the problems associated with the knockout strategy. As we will see, aggression is an important behavioral state

that can be studied in laboratory animals and humans. However, the biology and genetics of aggressive behavior are poorly understood. The transgenic approach holds the promise of uncovering genes that influence aggressive behaviors. Indeed, the study of aggression in different transgenic and knockout strains has led to the discovery of different genes and transmitter systems whose contribution to aggression had been previously unappreciated. These mutant mice have opened new avenues of research into aggression and its control and will prove valuable tools for ongoing studies.

The problems posed by the transgenic approach in this area are the same ones discussed above: namely confounding aspects of development, compensatory processes, and lack of tissue specificity. We will begin this section with a brief overview of the biology of aggression and then discuss individually different transgenic mice that have alterations in their aggressive behavior. In each case, we will endeavor to highlight the strengths and weaknesses of the resulting mouse to illustrate the points we have made above.

6.2 Aggressive Behavior in Mice and Humans

Several functions of aggressive behaviors have been suggested for various species, including mice. These include establishment of territoriality, establishment of dominance (to secure access to food and female mating partners), predatory violence (attacking other species to provide nourishment), and defensive aggression (responding and repelling attacks by predators or other members of the same species). In addition, aggression in mice can be elicited by various stressful stimuli (social isolation, foot shock, food deprivation), but the adaptive value of this type of aggression is less clear. Some of these categories are similar to some forms of human aggression, while others are less so. Different behavioral tests have been designed to study aggression in mice (and other rodents). These include the resident–intruder test (test of territorial aggression), isolation-induced or pain-induced aggression (stress-induced aggression), mouse defensive test battery (defensive aggression), maternal aggression (mother attacks intruders entering nest area during lactation), and predatory aggression (attacking and killing insects).

6.2.1 Genetics of Aggression in Mice

There is a strong genetic contribution to aggressive behavior in mice that can be demonstrated either by examining inbred stains (Guillot and Chapouthier 1996) or by selectively breeding mice for different forms of aggressive behaviors (Lagerspetz and Lagerspetz 1971; Sandnabba 1996). Different functional types of aggressive behavior appear to be under distinct genetic control when inbred stains are examined (Popova et al. 1993). In contrast, breeding studies have found that different functional aggressive behaviors (e.g., isolation-induced and predatory aggression) segregated together during selective breeding (Sandnabba 1995). Not surprisingly, environmental factors greatly influence the expression of aggressive behavior in mice. In particular, early exposure to aggression enhanced later expression of this trait (Sandnabba 1996). The role of the preweaning maternal environment is controversial. In selective breeding studies, the influence of the mother appeared to have little influence on the development of aggression (Hoffmann et al. 1993; Sandnabba 1996), while cross-fostering certain inbred strains with one another was able to influence aggressive behavior in some (Southwick 1968), but not others (Fredericson 1952; Geinsburg and Allee 1942).

Using the kind of forward genetic strategies that were outlined earlier in this chapter (Fig. 2), different research groups have attempted to identify loci that contribute to aggressive phenotypes. One focus of attention has been the Y chromosome, since large sex differences in aggression have been noted (Roubertoux et al. 1994). Other efforts to delineate the loci that contribute to the strain-to-strain differences in aggression are proceeding, but currently no other loci have been identified.

6.2.2 Genetics of Aggression in Humans

In humans, the contribution of genetic factors has been much more difficult to establish. Adoption and twin studies have revealed both environmental and genetic modulators for aggressive behavior in humans (Coccaro et al. 1997; Miles and Carey 1997). In addition, there is evidence that, among the DSM-IV axis II disorders, antisocial personality disorder has a strong genetic contribution (Dahl 1993). Criminal behavior also has a considerable genetic component, but paradoxically this association only holds true for nonviolent or property crime (Brennan et al. 1996). Of course there is a large environmental contribution to aggressive behaviors, but studies have suggested that environment contributes more to attributes such as hostility and less to attributes such as angry aggression and irritability (Gustavsson et al. 1996). Environmental factors such as perinatal problems and not genetic transmission appear to influence the development of violent criminal behavior (Mednick and Kandel 1988).

6.3 Brunner's Syndrome

Thus far, the only characterized mutation that might be associated with violent behavior is a point mutation in the monoamine oxidase A (MAOA) gene that changes a glutamine codon to a stop codon, leading to an inactive, truncated form of the enzyme (Brunner et al. 1993). The MAOA gene codes for a protein that metabolizes serotonin and other bioamines. MAOA appears to be a major pathway by which synaptic serotonin is inactivated. Individuals lacking a functional MAOA gene were found to suffer from borderline mental retardation and altered behavior including impulsive aggression, arson, attempted rape, and exhibitionism. This mutation and the accompanying behaviors have been called "Brunner's syndrome."

The MAOB gene is functionally and structurally related to the MAOA gene. It metabolizes dopamine and phenethylamine derivatives in the brain. The MAOA and MAOB genes are found in a tandem array on the X chromosome, and thus inheritance of a single mutated or deleted copy of either gene leads to complete deficiency of the gene in males but not females. Since the initial report of subjects deficient in MAOA, microdeletions of the X chromosome in this locus have been characterized that affect either MAOA, MAOB, or both. Double mutations lead to severe mental retardation, whereas selective MAOB deficiencies lead to no abnormal behavior or mental retardation (Lenders et al. 1996). The study of such individuals is providing much information about the role of these enzymes in brain development and behavior (Lenders et al. 1998), but an animal model of these disorders would allow a level of inquiry not possible with human subjects. Specifically, such models would allow a detailed study of the behavioral effects of altered monoamine metabolism.

In a striking example of serendipity, a group of researchers who introduced a transgene for interleukin (IL)-β into the mouse genome noted that one of their mouse strains appeared significantly more aggressive than other strains. Further analysis of these aggressive mice found that one of the copies of the IL-β gene had inserted into a crucial portion of the coding region of the MAOA gene. As this gene is found on the X chromosome, a single mutation was sufficient to create a complete absence of MAOA activity in male mice. These mice had striking abnormalities in their bioamine metabolism and behavior. In pup brains, serotonin concentrations were increased by up to ninefold, and norepinephrine by twofold. In adults, serotonin and 5-hydroxyindoleacetic acid (5-HIAA) levels returned to normal, but adults displayed a specific set of behavioral alterations, including enhanced intermale aggressive behavior, increased attempts to mount nonreceptive females, and decreased immobility in a forced swim test (a test used to screen for antidepressant activity). The enhanced aggressive behavior displayed by mice lacking MAOA parallels the finding of Brunner's subjects lacking a functional MAOA. The ability to reproduce the aggressive phenotype in transgenic mice suggests that their unusual behavior may therefore be a consequence of the MAOA mutation rather than the result of some unusual environmental or genetic background in that family. Consistent with the characterization of individuals with microdeletions affecting the MAOB gene, knockout mice lacking the MAOB gene did not display overt changes in aggressive behavior (Grimsby et al. 1997). No report of a double MAOA/B mutant has been described.

6.4 Serotonin and Aggression

To many researchers, the finding of a hyperaggressive phenotype in the MAOA mutant mice and in Brunner's syndrome came as a surprise. The MAOA mutation in each case increases brain serotonin, and aggression is normally associated with low brain serotonin levels. Indeed, one of the more robust findings in psychiatry has been the association of suicide (Mann et al. 1990), type II alcoholism, and violent behavior (Linnoila and Virkkunen 1992) with low brain serotonin levels. Low cerebrospinal fluid (CSF) serotonin metabolites are found in aggressive nonhuman primates (Mehlman et al. 1994), and serotonin depletion in the brains of rodents yields hyperaggressive animals in tests of predation (Valzelli et al. 1981), offensive aggression (Vergnes et al. 1986), and stress-induced aggression (Sheard and Davis 1976). Moreover, certain serotonin agonists have antiaggressive properties in animal models (Olivier et al. 1995). MAOA inhibitors, which were widely used as antidepressants, have not been reported to induce aggressive behavior.

The MAOA knockout mouse is a good example of a classical knockout providing a surprising phenotype that would not have been predicted from pharmacology alone. For example, it is not widely observed that inhibitors of MAOA can cause increased aggressive behavior either in human subjects or in animal models. In fact, if MAOA inhibitors have any effect on aggression in adults, it would be to decrease aggression. Thus the MAOA knockout mouse mimicked Brunner's syndrome likely because of the developmental and compensatory effects of not having this gene during brain development. For example, to compensate for high serotonin levels, postsynaptic serotonin receptors might be downregulated, which could in turn result in a phenotype similar to low serotonin activity. Consistent

with this possibility has been the unexpected finding that in children who come from a cohort of adjudicated siblings, the children with higher indices of serotonin activity have higher rates of behavioral aggression (Pine et al. 1997). Thus it may be speculated that, similar to the MAOA knockout mouse, low serotonin activity in adults with aggression may be the result of elevated activity in early life. While such an interpretation is very speculative, we suggest it in order to illustrate how findings in a mouse model may eventually inform hypotheses to be tested in human disorders.

Another intriguing parallel between Brunner's syndrome and the MAOA mutant mice is the possibility of aberrant neurodevelopment in each case. Serotonin is also known to be an important modulator of growth and morphogenesis (Lauder and Krebs 1978; Moiseiwitsch and Lauder 1995; Yavarone et al. 1993). If MAOA mutations lead to dramatic changes in serotonin in the brain, this might produce structural changes in the brain. Indeed, mice lacking MAOA display neuroanatomic abnormalities such as an absence of the cortical barrels field that are the somatosensory representation of the vibrissae (Cases et al. 1996). This is a situation where inducible or tissue-specific knockouts may help disentangle a complex but relevant phenotype.

6.5 Aggressive Behavior and the 5-HT1B Receptor Knockout

As discussed in the last section, there appears to be an important role of serotonin in controlling aggression and impulsivity in several species, including humans. One of the serotonin receptors that is thought to be involved in the effect of serotonin on aggression is the 5-HT1B receptor. The 1B receptor functions as an autoreceptor and a postsynaptic heteroreceptor that controls the release of other neurotransmitters such as γ-aminobutyric acid (GABA) or glutamate in response to serotonin (Saudou and Hen 1994). Pharmacological compounds that stimulate the 1B receptor have antiaggressive properties (Olivier et al. 1995). To better understand the role of the 5-HT1B receptor in aggression, mice lacking 5-HT1B receptors were generated by gene targeting (Ramboz et al. 1996; Saudou et al. 1994) in our laboratory.

The homozygous mutants developed normally and did not differ from their wild-type litter mates anatomically. In a number of behavioral tests aimed at measuring feeding, locomotion, and anxiety, either no differences or only small differences were found between the wild-type and the mutant mice. However, when analyzed in the isolation-induced aggression test, the male homozygous mutants displayed increased aggressive behavior (Saudou et al. 1994). In this test, wild-type or mutant males that have been isolated for 4 weeks are confronted with a male intruder that has not been isolated. The wild-type mice attack the intruder only after an initial investigative period that includes sniffing and aggressive displays such as tail rattling. In contrast, the mutant mice attack immediately or after a short latency. In addition, the number and the intensity of the attacks were significantly higher in the case of the mutants. The short latency of attack displayed by these mutant mice is reminiscent of the impulsive behavior often associated with deficits in central serotonin in primates.

To investigate whether the aggressive phenotype extends beyond isolation-induced aggression (a stress-induced aggression), we subjected 1B knockout mice to other tests of aggression, including territorial aggression and maternal aggression tests, and discovered that the 1B knockout mice did indeed display increased aggression in these functional types of aggression as well.

Unlike the example of the MAOA mutant mouse, where the phenotype was not predicted by pharmacology, 5-HT1B agonists have been shown to decrease aggressive behavior in a number of rodent models, including the isolation-induced aggression test. It is therefore consistent that mice lacking this receptor display the opposite phenotype. However, it is still possible that developmental or compensatory changes are responsible for the aggressive behavior of these mutant mice. Indeed, we have observed compensatory changes in the expression of various genes, including dopaminergic genes, serotonergic genes, and genes in the *fos* family. We are in the process of developing mice that will have either temporal control over 1B gene expression, tissue-specific control, or both. We plan to use these animals to understand more fully the aggressive phenotype that is observed in the classical 1B knockout.

Finally, we would again like to emphasize that, unlike the MAOA mutation, which is found in at least some human families, a mutation in the human 5-HT1B receptor gene has not been reported. Thus the 1B knockout mouse is an example where we have created our own "accident of nature" to investigate its potential role in aggressive behavior. Despite the absence of a human counterpart, this mutant mouse has been and will continue to be a useful tool in investigating the biology of aggression and in screening for new medications that can help remedy some of the problems posed by human aggressive behavior.

6.6 Aggressive Behavior in Other Knockout Strains

A number of other knockout mice have been reported that display changes in aggressive behavior compared

to their wild-type litter mates. Different examples of these mice will be discussed in the context of the promise of knockouts to open new avenues of research as well as the potential confounding factors that can affect interpretation of their phenotype.

6.6.1 Neuronal Nitric Oxide Synthase

Nitric oxide (NO) has recently been implicated as a nearly ubiquitous signaling molecule that controls many renal, gastrointestinal, cardiac, and brain functions (Zhang and Snyder 1995). NO is synthesized by two different isoenzymes: neuronal and endothelial NO synthases. Neuronal NOS (nNOS) is expressed widely both in the developing and in the adult nervous system. Mice lacking nNOS display an increase in offensive aggressive behavior and altered sexual behavior (Nelson et al. 1995). In the resident–intruder paradigm, mutant males attack male intruders more often and more intensely than wild-type males, while the latency to the first attack is the same in both groups. The male nNOS knockout mice also have inappropriate sexual behavior in that they tried to mount unreceptive females. The absence of prior reports linking the NO system to aggression led many to speculate that the aggressive phenotype was the result of neurodevelopmental or compensatory mechanisms in these mice. However, a follow-up study demonstrated that nNOS inhibitors can produce hyperaggressive behavior in two different models of aggression to a degree found in the nNOS knockout mice (Demas et al. 1997). The mechanism by which NO can modulate aggression is not known. A parsimonious hypothesis would tie together an effect of NO on known modulators of aggression such as serotonin. In fact, nNOS is commonly colocalized with bioamine neurons, and a large percentage of serotonergic neurons in the raphe that project to the cortex coexpress nNOS and serotonin (Blottner et al. 1995). Although intriguing, the effect of NO on serotonin target neurons is unknown. In summary, the nNOS knockout mouse is an excellent example of how transgenic techniques were able to open new avenues of exploration in an area such as aggression. However, questions still remain regarding the mechanism of NO actions on aggression, the brain structures involved in this response, and the ontogeny of this response.

6.6.2 Calcium Calmodulin Kinase II

Mutant mice that are heterozygous for a mutation in the gene encoding the enzyme CaMKII display increases in defensive aggression but not in offensive aggression in the resident–intruder paradigm (Chen et al. 1994). This hyperaggressive phenotype came as a surprise, because, like nNOS, this enzyme had not previously been linked to aggression. However, if we look for connections to the serotonin system, we can see that these mice display decreased serotonin release as measured by extracellular and whole-cell patch-clamp recordings from brainstem slices. This may be explained by the role of CaMKII-mediated phosphorylation in the activation of TpH, the rate-limiting enzyme for serotonin synthesis. A reduction in TpH activity might therefore explain the aggressive phenotype of the CaMKII heterozygous mice. As an example of how a phenotype can be confounded by other factors, it was found that homozygous mutants for the CaMKII gene are not aggressive. However, these animals display a large number of other behavioral abnormalities. The homozygote mutant animal may be an example of developmental abnormalities affecting even a phenotype seen in heterozygote mice.

6.6.3 Preproenkephalin Gene Knockout

Enkephalins and endorphins are two endogenous peptides that appear to function as natural analgesics in animals. A well-known phenomenon in different species including humans is the development of a profound analgesia during periods of stress and danger (Miczek et al. 1982). In humans, such profound analgesia has allowed wounded soldiers to continue to fight without awareness of their wounds. This stress-induced analgesia has both opiate and nonopiate mechanisms (Grisel et al. 1993). In mice, the ability to develop an opiate-analgesic response during aggressive encounters is thought to influence the type of strategy the animal pursues to defend itself and may therefore influence social hierarchy (Miczek et al. 1994). The enkephalins are thought to be the primary mediator of stress-induced analgesia (Raab et al. 1985). It might have been predicted that, in the absence of enkephalins, stress-induced analgesia would be impaired and the mice would exhibit less aggression. Although this hypothesis appears sound, the problem was that enkephalin-deficient mice had a normal analgesic response to stressful situations (Konig et al. 1996). In addition, an increase in offensive aggressive behavior was observed compared to wild-type mice. These knockout mice were also found to be more anxious than wild-type mice, but the contribution of this anxiety state to aggression is unclear. Thus, the enkephalin-deficient mice have a surprising phenotype in many respects. They will prove to be an interesting model to study the link between the enkephalin system and aggressive behaviors.

6.6.4 Neurokinin 1 Receptor Knockout

Substance P is known to mediate transmission of some painful stimuli and to modulate inflammatory responses. It is also found in the limbic regions and basal ganglia of the brain. The main receptor for substance P is the neurokinin (NK) 1 receptor. When the gene for this receptor was disrupted in mice, the resulting animals were found to have alterations in nociception, stress-induced analgesia, and aggression (De Filipe et al. 1998). Since the NK1 receptor appears to signal pain, as expected, these NK1 knockout mice had a reduced response to painful stimuli. In contrast, NK1 knockout mice had impaired stress-induced analgesia. Based on our reasoning outlined for the preproenkephalin knockout, we predicted that, based on their inability to develop analgesic responses during stress, these animals may be less apt to engage in aggressive encounters. Consistent with that prediction, NK1 knockout mice displayed decreased aggression compared to wild-type animals. NK1 knockout mice did not display any differences in a test of anxious behavior.

6.6.5 Adenosine A_2 Receptor Knockout

The purine adenosine is thought to act as a neurotransmitter in the brain and the periphery. Adenosine receptors are found in the brain in high concentrations and appear to be involved in a number of different processes. The A_{2a} adenosine receptor subtype is found in the basal ganglia and appears to be the major target for the stimulant properties of caffeine. Although caffeine, an antagonist at A_{2a} receptors, can potentiate aggressive behavior, and agonists at this receptor can have serenic-like effects, the purinergic system has not been classically implicated in the modulation of aggressive behavior. Nonetheless, A_{2a} knockout mice were found to be more aggressive in resident–intruder tests of aggression, a finding consistent with the proaggressive effects of A_{2a} antagonists such as caffeine. Like the enkephalin knockout mice, these animals were also found to be more anxious than their wild-type litter mates, but the connection of anxiety to aggression is not clear. Indeed, other knockout mice with elevated anxiety (see Table 1) have not been observed to be more aggressive than nonmutants. Perhaps more relevant to understanding the aggressive phenotype was the finding that these A_{2a} knockout mice were less sensitive to painful stimuli than control mice. If the ability to develop pain insensitivity during stress is important for expression of aggressive behavior, the low pain sensitivity seen in A_{2a} knockout mice may partially explain the aggressive phenotype. As in our previous examples, the implication of adenosine as a major modulator of aggression was unexpected and has generated another tool with which to study the biology of aggression.

6.6.6 Estrogen Receptor Knockout

Aggressive behavior and social dominance in animals is often associated with elevated testosterone levels. Social defeat typically leads to long-lasting decreases in serum testosterone (Bonson et al. 1994; Koolhaas et al. 1997; Lisciotto et al. 1990). As estrogen often acts functionally to antagonize the effects of testosterone, it might be expected that mice lacking an estrogen response would be hypermasculine. However, the aggressive behavior of male mice was dramatically decreased with an almost complete absence of male-typical offensive attacks (Ogawa et al. 1997). At first, this result would seem surprising but , it has been demonstrated that in many animal species there is a paradoxical affect of estrogen receptors on masculinization. In rats, the hypothalamus contains the enzyme aromatase, which converts testosterone to estrogen. It is aromatized testosterone in the hypothalamus that induces masculine behavior in the rat via estrogen receptors. Thus the absence of the estrogen receptor in male mice may lead to paradoxical feminization due to the failure to stimulate these hypothalamic estrogen receptors. In contrast to male estrogen receptor knockout mice, female mice exhibited increased aggression against other females and mothers tended to cannibalize their pups more often. Interestingly, female estrogen receptor knockout mice were treated as "male" intruders by resident male mice and were thus attacked rather than mounted (Ogawa et al. 1996). Thus ablation of the estrogen receptor has complex behavioral effects that are paradoxical in males (demasculinization, decreased aggression) but are more expected in the female (masculinization, increased aggression).

6.6.7 Oxytocin and Oxytocin Receptor Knockout

The hormone oxytocin performs a role in a variety of functions, including milk ejection, birth, and behavior, especially affiliative and social behaviors and aggression (Young et al. 1997). In particular, it has been observed that monogamous prairie voles and nonmonogamous montane voles have different patterns of oxytocin receptor expression that are thought to play a role in social behaviors. To further explore the role of oxytocin in behavior, knockouts of the oxytocin

Table 1. Various behavioral phenotypes of knockout and transgenic mice with potential relevance to psychiatry

Mutation (genotype)	Phenotype	Reference
Aggressive mice strains		
5-HT1B Receptor (−/−)	Increased offensive aggression Increased maternal aggression	Sadou et al. 1994 Unpublished
MAOA (−/−)	Increased intermale aggression Inappropriate sexual behavior	Cases et al. 1995
nNOS (−/−)	Increased offensive aggression Inappropriate sexual behavior	Nelson et al. 1995
αCAM kinase II (+/−)	Increased defensive aggression No change in offensive aggression	Chen et al. 1994
Preproenkephalin (−/−)	Increased offensive aggression	Konig et al. 1996
Oxytocin receptor (−/−)	Decreased intermale aggression	DeVries et al. 1997
NK1 receptor (−/−)	Decreased intermale aggression	Babinet et al. 1989
Adenosine receptor 2a (−/−)	Increased isolation-induced aggression	Ledent et al. 1997
Estrogen receptor (−/−)	Decreased intermale aggression Increased interfemale aggression	Ogawa et al. 1997 Ogawa et al. 1996
NCAM (−/−)	Increase offensive aggression	Stork et al. 1997
TGF-α	Increased offensive aggression Low brain 5-HIAA	Hilakivi-Clarke et al. 1993
Memory and learning-impaired strains		
PKC-γ (−/−)	Impaired hippocampal LTP Impaired spatial learning Impaired contextual learning	Abeliovich et al. 1993a Abeliovich et al. 1993b
αCaMKII (−/−)	Impaired hippocampal LTP	Silva et al. 1992
CREB (−/−)	Impaired long-term memory Intact short-term memory Impaired hippocampal LTP Intact PPF and PTP	Bourtchuladze et al. 1994
PKA Cβ1 (−/−)	Impaired mossy fiber LTP	Huang et al. 1995
PKA R1β (−/−)		Huang et al. 1995
nNOS (−/−)	Normal LTP, learning	O'Dell et al. 1994
eNOS × nNOS (−/−)	Impaired LTP CA1	Son et al. 1996
Type I adenylyl cyclase (−/−)	Impaired CA1 LTP Impaired spatial memory	Wu et al. 1995
mGluR1 (−/−)	Impaired hippocampal LTP Impaired associative learning Impaired cerebellar LDP Impaired motor learning	Aiba et al. 1994a Aiba et al. 1994b
αCaMKII inducible	Impaired LTP Impaired spatial learning Impaired fear conditioning	Mayford et al. 1996
αCaMKII-activated mutation		Mayford et al. 1995; Rotenberg et al. 1996
PKA R(AB) (−/−)	Impaired L-LTP Impaired spatial learning	Abel et al. 1997
Anxiety and stress-altered mice strains		
CRF1 receptor (−/−)	Decreased anxiety Decreased stress response	– –
5-HT1A receptor (−/−)	Increased anxiety	Unpublished
Serotonin transporter (−/−)	Decreased anxiety	Unpublished

Table 1 (*Continued*)

Mutation (genotype)	Phenotype	Reference
Transgenic strains		
CRF overexpression	Increased anxiety	Stenzel-Poore et al. 1994
Type II glucocorticoid antisense transgenic	Decreased stress response	Pepin et al. 1992
Mice strains altered in catecholamine systems		
Dopamine D2R (−/−)	Parkinson-like motor impairment	Baik et al. 1995
Dopamine D3R (−/−)	Hyperactivity in novel environment	Accili et al. 1996
Dopamine D1R (−/−)	Hyperactivity in novel environment Cocaine-induced hypolocomotion Reduced cocaine-induced stereotypies	Xu et al. 1994b Xu et al. 1994a Moratalla et al. 1996
Tyrosine hydroxylase (−/−)	Embryonic lethality E11.5–E15.5 Rescued by L-dopa	Zhou et al. 1995
Tyrosine hydroxylase (DA specific)	Severe hypoactivity, adipsia, aphagia Rescued by L-dopa	Zhou and Palmiter 1995
Dopamine transporter (−/−)	Increased spontaneous locomotion Impaired locomotor response to cocaine and amphetamine Slow weight gain Early death Impaired maternal behavior	Giros et al. 1996
DBH (−/−)	Embryonic lethality Rescued by dihydroxyphenyl serine	Thomas et al. 1995
α2C adrenergic receptor (−/−)	Impaired clonidine hypothermia	Link et al. 1995; Sallinen et al. 1997
α2B adrenergic receptor (−/−)	Impaired hemodynamic effects of α2 agonists	Link et al. 1996

MAOA, monoamine oxidase A; NCAM, neural cell adhesion molecule; TGF, transforming growth factor; CaMK, calcium calmodulin kinase; NK, neurokinin; 5-HIAA, 5-hydroxyindoleacetic acid; PKC, protein kinase C; LTP, long-term potentiation; PKA, protein kinase A; CREB, cAMP response element binding protein; nNOS, neuronal nitric oxide synthase; eNOS, endothelial NOS; CRF, corticotropin-releasing factor; DA, dopamine; DBH, DA-β-hydroxylase.

peptide and its receptor were both created (DeVries et al. 1997; Nishimori et al. 1996). Interestingly, the oxytocin receptor knockout mice were much less aggressive in two different tests of aggression than either wild-type or heterozygous mutants. The mutation of the oxytocin gene (OKO) exhibited no behavioral phenotype except for an inability to nurse pups (defective milk ejection) that was correctable with exogenous oxytocin injections. The reduced aggressiveness seen in the oxytocin receptor knockout mice is consistent with the findings that this peptide can play a role in social behaviors.

7 Conclusion

In conclusion, we have attempted in this chapter to demonstrate the power of transgenic techniques to more directly understand the connections between gene function and behavior. The advantage of this technique is the power to create mutations that are not found in nature and to study their effects on animals. These mutations are absolute and precise. We have also attempted to enlighten the reader concerning some of the confounding problems that complicate interpretation of some knockout experiments. In the future, more precise control of the temporal and anatomic specificity of the mutation may be possible and will help address many of the criticisms of current knockout techniques.

We have chosen to illustrate many of these principles using the phenotype of aggression to demonstrate that many different genes can contribute to the modulation of aggressive behavior in mice. Mutations in mice have been used to mimic a human mutation in the MAOA gene to create other mutations in the serotonin system that also affect aggression (5-HT1B receptor gene) but for which no human counterpart exists. Lastly, we have described several other knockout mice that have alterations in their aggressive

behavior. These animals are interesting in several respects. First, they have often implicated new genes and neurotransmitter systems in the control of aggression that had not been previously studied. Second, these animals often exhibited paradoxical or unexpected phenotypes that may be explained by developmental abnormalities or compensatory processes that are inherent to "classical" knockouts. Lastly, these mutant animals demonstrate that different mutations can modulate aggressive behavior in either direction – up or down.

Finally, in Table 1, we present a sample of the knockout strains currently available which may be of interest in understanding behaviors related to psychiatric conditions. Since new knockouts are continually being created, this cannot be thought of as a complete list, but rather as a representative sample of what is currently available. The knockouts involve either receptors for neurotransmitters, enzymes involved in neurotransmitter synthesis, enzymes involved in second messenger pathways, or growth factors. All these classes of compounds will be involved in multiple behaviors and neuronal systems, making simple interpretations of the resulting phenotype difficult. The aggressive mutations have been discussed in some detail. In addition, there are several strains of mice which have been utilized in the study of memory and learning. Interestingly, almost all these mice involve alterations in second messenger pathways or transcription factors activated as a consequence of second messenger activation. The clearest information comes from those knockouts which do not impair learning, such as NOS knockouts, which suggests that these systems are not involved in the learning process. Definitive interpretations will only be possible with conditional site-specific knockouts, but these data provide suggestive clues. The mice which show changes in stress response can be used as models for anxiety and depression and generally involve the catecholamine systems and the hypothalamic–pituitary–adrenal (HPA) axis. Again, interpretation is difficult, but the models in which a substance is overexpressed point to a role for corticotropin-releasing factor (CRF) in anxiety. The catecholamine knockouts have provided less information about mood disorders or schizophrenia than might be expected until we consider the difficulty in defining the phenotype in a mouse model. In general, dopamine changes involve drug effects and motor function, while noradrenergic changes are either lethal or involved the vascular system.

Thus we believe that the transgenic technologies described here will hold promise for helping researchers to dissect the role of different genes in influencing behavior. As the human genome approaches being fully sequenced in the next 5–10 years, the application of these technologies to newly described genes will continue to provide insights into the genetic regulation of behavior.

8 References

Abel T, Nguyen PV, Barad M, Deuel TA, Kandel ER, Bourtchouladze R (1997) Genetic demonstration of a role for PKA in the late phase of LTP and in hippocampus-based long-term memory. Cell 88: 615–626

Abeliovich A, Chen C, Goda Y, Silva AJ, Stevens CF, Tonegawa S (1993a) Modified hippocampal long-term potentiation in PKC gamma-mutant mice. Cell 75: 1253–1262

Abeliovich A, Paylor R, Chen C, Kim JJ, Wehner JM, Tonegawa S (1993b) PKC gamma mutant mice exhibit mild deficits in spatial and contextual learning. Cell 75: 1263–1271

Accili D, Fishburn CS, Drago J, Steiner H, Lachowicz JE, Park BH, Gauda EB, Lee EJ, Cool MH, Sibley DR, Gerfen CR, Westphal H, Fuchs S (1996) A targeted mutation of the D3 dopamine receptor gene is associated with hyperactivity in mice. Proc Natl Acad Sci USA 93: 1945–1949

Aiba A, Chen C, Herrup K, Rosenmund C, Stevens CF, Tonegawa S (1994a) Reduced hippocampal long-term potentiation and context-specific deficit in associative learning in mGluR1 mutant mice. Cell 79: 365–375

Aiba A, Kano M, Chen C, Stanton ME, Fox GD, Herrup K, Zwingman TA, Tonegawa S (1994b) Deficient cerebellar long-term depression and impaired motor learning in mGluR1 mutant mice. Cell 79: 377–388

Babinet C, Morello D, Renard JP (1989) Transgenic mice. Genome 31: 938–949

Baik JH, Picetti R, Saiardi A, Thiriet G, Dierich A, Depaulis A, Le Meur M, Borrelli E (1995) Parkinsonian-like locomotor impairment in mice lacking dopamine D2 receptors. Nature 377: 424–428

Blottner D, Grozdanovic Z, Gossrau R (1995) Histochemistry of nitric oxide synthase in the nervous system. Histochem J 27: 785–811

Bonson KR, Johnson RG, Fiorella D, Rabin RA, Winter JC (1994) Serotonergic control of androgen-induced dominance. Pharmacol Biochem Behav 49: 313–322

Bourtchuladze R, Frenguelli B, Blendy J, Cioffi D, Schutz G, Silva AJ (1994) Deficient long-term memory in mice with a targeted mutation of the cAMP-responsive element-binding protein. Cell 79: 59–68

Brennan PA, Mednick SA, Jacobsen B (1996) Assessing the role of genetics in crime using adoption cohorts. Ciba Found Symp 194: 115–123, 123–128

Bronson SK, Smithies O (1994) Altering mice by homologous recombination using embryonic stem cells. J Biol Chem 269: 27155–27158

*Brunner HG, Nelen M, Breakefield XO, Ropers HH, van Oost BA (1993) Abnormal behavior associated with a point mutation in the structural gene for monoamine oxidase A. Science 262: 578–580

*Cases O, Seif I, Grimsby J, Gaspar P, Chen K, Pournin S, Muller U, Aguet M, Babinet C, Shih JC et al (1995) Aggressive behavior and altered amounts of brain serotonin and norepinephrine in mice lacking MAOA. Science 268: 763–1766

Cases O, Vitalis T, Seif I, De Maeyer E, Sotelo C, Gaspar P (1996) Lack of barrels in the somatosensory cortex of monoamine oxidase A-deficient mice: role of a serotonin excess during the critical period. Neuron 16: 297–307

Chen C, Rainnie DG, Greene RW, Tonegawa S (1994) Abnormal fear response and aggressive behavior in mutant mice deficient for alpha-calcium-calmodulin kinase II. Science 266: 291–294

Coccaro EF, Bergeman CS, Kavoussi RJ, Seroczynski AD (1997) Heritability of aggression and irritability: a twin study of the Buss-Durkee aggression scales in adult male subjects. Biol Psychiatry 41: 273–284

Dahl AA (1993) The personality disorders: a critical review of family, twin, and adoption studies. J Pers Disord Suppl 1: 86–99

De Filipe C, Herrero JF, O'Brien JA, Palmer JA, Doyle CA, Smith AJH, Laird JMA, Belmonte C, Cervero F, Hunt SP (1998) Altered nociception, analgesia, and aggression in mice lacking the receptor for substance P. Nature 392: 394–397

Demas GE, Eliasson MJ, Dawson TM, Dawson VL, Kriegsfeld LJ, Nelson RJ, Snyder SH (1997) Inhibition of neuronal nitric oxide synthase increases aggressive behavior in mice. Mol Med 3: 610–616

DeVries AC, Young WSR, Nelson RJ (1997) Reduced aggressive behaviour in mice with targeted disruption of the oxytocin gene. J Neuroendocrinol 9: 363–368

Fredericson E (1952) Reciprocal fostering of two inbred mouse strains and its effect on the modification of aggressive behavior. Am Psychol 15: 241–242

Ginsburg BE, Allee WC (1942) Some effects of conditioning on social dominance and subordination in inbred strains of mice. Physiol Zool 15: 485–506

*Giros B, Jaber M, Jones SR, Wightman RM, Caron MG (1996) Hyperlocomotion and indifference to cocaine and amphetamine in mice lacking the dopamine transporter. Nature 379: 606–612

Grimsby J, Toth M, Chen K, Kumazawa T, Klaidman L, Adams JD, Karoum F, Gal J, Shih JC (1997) Increased stress response and beta-phenylethylamine in MAOB-deficient mice. Nat Genet 17: 206–210

Grisel JE, Fleshner M, Watkins LR, Maier SF (1993) Opioid and nonopioid interactions in two forms of stress-induced analgesia. Pharmacol Biochem Behav 45: 161–172

Guillot PV, Chapouthier G (1996) Intermale aggression and dark/light preference in ten inbred mouse strains. Behav Brain Res 77: 211–213

Gustavsson JP, Pedersen NL, Asberg M, Schalling D (1996) Exploration into the sources of individual differences in aggression-, hostility- and anger-related (AHA) personality traits. Pers Indiv Diff 21: 1067–1071

Hilakivi-Clarke L, Durcan M, Goldberg R (1993) Effect of alcohol on elevated aggressive behavior in male transgenic TGF alpha mice. Neuroreport 4: 155–158

Hoffmann HJ, Schneider R, Crusio WE (1993) Genetic analysis of isolation-induced aggression. II. Postnatal environmental influences in AB mice. Behav Genet 23: 391–394

Huang YY, Kandel ER, Varshavsky L, Brandon EP, Qi M, Idzerda RL, McKnight GS, Bourtchouladze R (1995) A genetic test of the effects of mutations in PKA on mossy fiber LTP and its relation to spatial and contextual learning. Cell 83: 1211–1222

Konig M, Zimmer AM, Steiner H, Holmes PV, Crawley JN, Brownstein MJ, Zimmer A (1996) Pain responses, anxiety and aggression in mice deficient in pre-proenkephalin. Nature 383: 535–538

Koolhaas JM, Meerlo P, De Boer SF, Strubbe JH, Bohus B (1997) The temporal dynamics of the stress response. Neurosci Biobehav Rev 21: 775–782

Lagerspetz KM, Lagerspetz KY (1971) Changes in the aggressiveness of mice resulting from selective breeding, learning and social isolation. Scand J Psychol 12: 241–248

Lauder JM, Krebs H (1978) Serotonin as a differentiation signal in early neurogenesis. Dev Neurosci 1: 15–30

Ledent C, Vaugeois JM, Schiffmann SN, Pedrazzini T, El Yacoubi M, Vanderhaeghen JJ, Costentin J, Heath JK, Vassart G, Parmentier M (1997) Aggressiveness, hypoalgesia and high blood pressure in mice lacking the adenosine A2a receptor. Nature 388: 674–678

Lenders JW, Eisenhofer G, Abeling NG, Berger W, Murphy DL, Konings CH, Wagemakers LM, Kopin IJ, Karoum F, van Gennip AH, Brunner HG (1996) Specific genetic deficiencies of the A and B isoenzymes of monoamine oxidase are characterized by distinct neurochemical and clinical phenotypes. J Clin Invest 97: 1010–1019

Lenders JW, Brunner HG, Murphy DL, Eisenhofer G (1998) Genetic deficiencies of monoamine oxidase enzymes: a key to understanding the function of the enzymes in humans. Adv Pharmacol 42: 297–301

Link RE, Desai K, Hein L, Stevens ME, Chruscinski A, Bernstein D, Barsh GS, Kobilka BK (1996) Cardiovascular regulation in mice lacking alpha2-adrenergic receptor subtypes b and c. Science 273: 803–805

Link RE, Stevens MS, Kulatunga M, Scheinin M, Barsh GS, Kobilka BK (1995) Targeted inactivation of the gene encoding the mouse alpha 2c-adrenoceptor homolog. Mol Pharmacol 48: 48–55

Linnoila VM, Virkkunen M (1992) Aggression, suicidality, and serotonin. J Clin Psychiatry 53: 46–51

Lisciotto CA, DeBold JF, Haney M, Miczek KA (1990) Implants of testosterone into the septal forebrain activate aggressive behavior in male mice. Aggress Behav 16: 249–258

*Lucas JJ, Hen R (1995) New players in the 5-HT receptor field: genes and knockouts. Trends Pharmacol Sci 16: 246–252

Mann JJ, Arango V, Underwood MD (1990) Serotonin and suicidal behavior. Ann NY Acad Sci 600: 476–484, 484–485

Mayford M, Wang J, Kandel ER, O'Dell TJ (1995) CaMKII regulates the frequency-response function of hippocampal synapses for the production of both LTD and LTP. Cell 81: 891–904

Mayford M, Bach ME, Huang YY, Wang L, Hawkins RD, Kandel ER (1996) Control of memory formation through regulated expression of a CaMKII transgene. Science 274: 1678–1683

McHugh TJ, Blum KI, Tsien JZ, Tonegawa S, Wilson MA (1996) Impaired hippocampal representation of space in CA1-specific NMDAR1 knockout mice. Cell 87: 1339–1349

Mednick SA, Kandel ES (1988) Congenital determinants of violence. 17th Annual Meeting of the American Academy of Psychiatry and the Law (1986, Philadelphia, Pennsylvania). Bull Am Acad Psychiatry Law 16: 101–109

Mehlman PT, Higley JD, Faucher I, Lilly AA, Taub DM, Vickers J, Suomi SJ, Linnoila M (1994) Low CSF 5-HIAA concentrations and severe aggression and impaired impulse control in nonhuman primates. Am J Psychiatry 151: 1485–1491

Miczek KA, Thompson ML, Shuster L (1982) Opioid-like analgesia in defeated mice. Science 215: 1520–1522

Miczek KA, Weerts E, Haney M, Tidey J (1994) Neurobiological mechanisms controlling aggression: preclinical developments for pharmacotherapeutic interventions. Neurosci Biobehav Rev 18: 97–110

Miles DR, Carey G (1997) Genetic and environmental architecture on human aggression. J Pers Soc Psychol 72: 207–217

Moiseiwitsch JR, Lauder JM (1995) Serotonin regulates mouse cranial neural crest migration. Proc Natl Acad Sci USA 92: 7182–7186

Mombaerts P, Wang F, Dulac C, Chao SK, Nemes A, Mendelsohn M, Edmondson J, Axel R (1996) Visualizing an olfactory sensory map. Cell 87: 675–686

Moratalla R, Xu M, Tonegawa S, Graybiel AM (1996) Cellular responses to psychomotor stimulant and neuroleptic drugs are abnormal in mice lacking the D1 dopamine receptor. Proc Natl Acad Sci USA 93: 14928–14933

Nelson RJ, Demas GE, Huang PL, Fishman MC, Dawson VL, Dawson TM, Snyder SH (1995) Behavioural abnormalities in male mice lacking neuronal nitric oxide synthase. Nature 378: 383–386

Nishimori K, Young LJ, Guo Q, Wang Z, Insel TR, Matzuk MM (1996) Oxytocin is required for nursing but is not essential for parturition or reproductive behavior. Proc Natl Acad Sci USA 93: 11699–11704

O'Dell TJ, Huang PL, Dawson TM, Dinerman JL, Snyder SH, Kandel ER, Fishman MC (1994) Endothelial NOS and the blockade of LTP by NOS inhibitors in mice lacking neuronal NOS. Science 265: 542–546

Ogawa S, Taylor JA, Lubahn DB, Korach KS, Pfaff DW (1996) Reversal of sex roles in genetic female mice by disruption of estrogen receptor gene. Neuroendocrinology 64: 467–470

Ogawa S, Lubahn DB, Korach KS, Pfaff DW (1997) Behavioral effects of estrogen receptor gene disruption in male mice. Proc Natl Acad Sci USA 94: 1476–1481

Olivier B, Mos J, van Oorschot R, Hen R (1995) Serotonin receptors and animal models of aggressive behavior. Pharmacopsychiatry 28: 80–90

Pepin MC, Pothier F, Barden N (1992) Antidepressant drug action in a transgenic mouse model of the endocrine changes seen in depression. Mol Pharmacol 42: 991–995

Pine DS, Coplan JD, Wasserman GA, Miller LS, Fried JE, Davies M, Cooper TB, Greenhill L, Shaffer D, Parsons B (1997) Neuroendocrine response to fenfluramine challenge in boys. Associations with aggressive behavior and adverse rearing. Arch Gen Psychiatry 54: 839–846

Popova NK, Nikulina EM, Kulikov AV (1993) Genetic analysis of different kinds of aggressive behavior. Behav Genet 23: 491–497

Raab A, Seizinger BR, Herz A (1985) Continuous social defeat induces an increase of endogenous opioids in discrete brain areas of the mongolian gerbil. Peptides 6: 387–391

Ramboz S, Saudou F, Amara DA, Belzung C, Dierich A, LeMeur M, Segu L, Misslin R, Buhot MC, Hen R (1996) Behavioral characterization of mice lacking the 5-HT1B receptor. NIDA Res Monogr 161: 39–57

Rossant J (1990) Manipulating the mouse genome: implications for neurobiology. Neuron 4: 323–334

Rotenberg A, Mayford M, Hawkins RD, Kandel ER, Muller RU (1996) Mice expressing activated CaMKII lack low frequency LTP and do not form stable place cells in the CA1 region of the hippocampus. Cell 87: 1351–1361

Roubertoux PL, Carlier M, Degrelle H, Haas-Dupertuis MC, Phillips J, Moutier R (1994) Co-segregation of intermale aggression with the pseudoautosomal region of the Y chromosome in mice. Genetics 136: 225–230

Sallinen J, Link RE, Haapalinna A, Viitamaa T, Kulatunga M, Sjoholm B, Macdonald E, Pelto-Huikko M, Leino T, Barsh GS, Kobilka BK, Scheinin M (1997) Genetic alteration of alpha 2C-adrenoceptor expression in mice: influence on locomotor, hypothermic, and neurochemical effects of dexmedetomidine, a subtype-nonselective alpha 2-adrenoceptor agonist. Mol Pharmacol 51: 36–46

Sandnabba NK (1995) Predatory aggression in male mice selectively bred for isolation-induced intermale aggression. Behav Genet 25: 361–366

Sandnabba NK (1996) Selective breeding for isolation-induced intermale aggression in mice: associated responses and environmental influences. Behav Genet 26: 477–488

Saudou F, Hen R (1994) 5-Hydroxytryptamine receptor subtypes: molecular and functional diversity. Adv Pharmacol 30: 327–380

Saudou F, Amara DA, Dierich A, LeMeur M, Ramboz S, Segu L, Buhot MC, Hen R (1994) Enhanced aggressive behavior in mice lacking 5-HT1B receptor. Science 265: 1875–1878

Sheard MH, Davis M (1976) *p*-Chloroamphetamine: short and long term effects upon shock-elicited aggression. Eur J Pharmacol 40: 295–302

Silva A, Paylor R, Wehner J, Tonegawa S (1992a) Impaired spatial learning in alpha-calcium calmodulin kinase II mutant mice. Science 257: 206

Silva AJ, Stevens CF, Tonegawa S, Wang Y (1992b) Deficient hippocampal long-term potentiation in alpha-calcium-calmodulin kinase II mutant mice. Science 257: 201–206

Smithies O, Kim HS (1994) Targeted gene duplication and disruption for analyzing quantitative genetic traits in mice. Proc Natl Acad Sci USA 91: 3612–3615

Son H, Hawkins RD, Martin K, Kiebler M, Huang PL, Fishman MC, Kandel ER (1996) Long-term potentiation is reduced in mice that are doubly mutant in endothelial and neuronal nitric oxide synthase. Cell 87: 1015–1023

Southwick CH (1968) Effect of maternal environment on aggressive behavior of inbred mice. Comm Behav Biol 1: 129–132

Stenzel-Poore MP, Heinrichs SC, Rivest S, Koob GF, Vale WW (1994) Overproduction of corticotropin-releasing factor in transgenic mice: a genetic model of anxiogenic behavior. J Neurosci 14: 2579–2584

Stork O, Welzl H, Cremer H, Schachner M (1997) Increased intermale aggression and neuroendocrine response in mice deficient for the neural cell adhesion molecule (NCAM). Eur J Neurosci 9: 1117–1125

Thomas SA, Matsumoto AM, Palmiter RD (1995) Noradrenaline is essential for mouse fetal development. Nature 374: 643–646

Tsien JZ, Chen DF, Gerber D, Tom C, Mercer EH, Anderson DJ, Mayford M, Kandel ER, Tonegawa S (1996a) Subregion- and cell type-restricted gene knockout in mouse brain. Cell 87: 1317–1326

Tsien JZ, Huerta PT, Tonegawa S (1996b) The essential role of hippocampal CA1 NMDA receptor-dependent synaptic plasticity in spatial memory. Cell 87: 1327–1338

Valzelli L, Bernasconi S, Garattini S (1981) *p*-Chlorophenylalanine-induced muricidal aggression in male and female laboratory rats. Neuropsychobiology 7: 315–320

Vergnes M, Depaulis A, Boehrer A (1986) Parachlorophenylalanine-induced serotonin depletion increases offensive but not defensive aggression in male rats. Physiol Behav 36: 653–658

Wilson MA, Tonegawa S (1997) Synaptic plasticity, place cells and spatial memory: study with second generation knockouts. Trends Neurosci 20: 102–106

Wu ZL, Thomas SA, Villacres EC, Xia Z, Simmons ML, Chavkin C, Palmiter RD, Storm DR (1995) Altered behavior and long-term potentiation in type I adenylyl cyclase mutant mice. Proc Natl Acad Sci USA 92: 220–224

Xu M, Hu XT, Cooper DC, Moratalla R, Graybiel AM, White FJ, Tonegawa S (1994a) Elimination of cocaine-induced hyperactivity and dopamine-mediated neurophysiological effects in dopamine D1 receptor mutant mice. Cell 79: 945–955

Xu M, Moratalla R, Gold LH, Hiroi N, Koob GF, Graybiel AM, Tonegawa S (1994b) Dopamine D1 receptor mutant mice are deficient in striatal expression of dynorphin and in dopamine-mediated behavioral responses. Cell 79: 729–742

Yavarone MS, Shuey DL, Tamir H, Sadler TW, Lauder JM (1993) Serotonin and cardiac morphogenesis in the mouse embryo. Teratology 47: 573–584

Young LJ, Winslow JT, Wang Z, Gingrich B, Guo Q, Matzuk MM, Insel TR (1997) Gene targeting approaches to neuroendocrinology: oxytocin, maternal behavior, and affiliation. Horm Behav 31: 221–231

Zhang J, Snyder SH (1995) Nitric oxide in the nervous system. Annu Rev Pharmacol Toxicol 35: 213–233

Zhou QY, Palmiter RD (1995) Dopamine-deficient mice are severely hypoactive, adipsic, and aphagic. Cell 83: 1197–1209

Zhou QY, Quaife CJ, Palmiter RD (1995) Targeted disruption of the tyrosine hydroxylase gene reveals that catecholamines are required for mouse fetal development. Nature 374: 640–643

CHAPTER

6

F. Henn, R.J. Hitzemann

Neurochemistry: The Basis of Psychopharmacology

1 Introduction

For ages, man has taken psychic comfort in the effects of drugs such as alcohol, morphine and cocaine. However, it was not until the 1950s that drugs became available with a demonstrated efficacy in the treatment of psychiatric disorders such as schizophrenia and depression. Attempts to understand the mechanisms of action for these drugs have been essential in the development of modern concepts in biological psychiatry. For example, the dopamine (DA) hypothesis of schizophrenia largely derives from experiments showing that antipsychotic drugs such as haloperidol and chlorpromazine affect the metabolism of DA but not other neurotransmitters (although at the time this hypothesis was formulated in the early 1960s, there were relatively few "other" neurotransmitters). Roles for norepinephrine (NE) and serotonin (5-HT) in depression developed from the observation that reserpine produces a "depressed" behavior in rodents and humans which is associated with the depletion of these transmitters. Further, it was found that the "depressed" behavior could be reversed by the administration of precursor molecules which elevate NE and serotonin levels. Soon thereafter, it was demonstrated that the first generation of antidepressant drugs block the synaptic reuptake of NE and serotonin, thus securing a prominent role in depression for these transmitters. Benzodiazepines were first synthesized in the late 1950s and introduced into clinical practice as anxiolytics soon thereafter. However, it was not until the 1970s that benzodiazepines were discovered to both potentiate the binding and enhance the neurophysiological effects of the inhibitory neurotransmitter γ-aminobutyric acid (GABA). Drugs which show cross-tolerance to and dependence on the benzodiazepines, e.g. the barbiturates and ethanol, were also found to similarly affect GABA neurotransmission.

Drugs have also been used to induce transient states which model psychiatric disorders. As early as 1845, Moreau suggested that hashish intoxication provided a model psychosis useful in the study of insanity. The use and abuse of cocaine and amphetamine led to the recognition that these drugs could induce a paranoid delusional state (see, e.g. Bell 1973). There was a brief period of scientific interest in the psychotomimetic properties of mescaline in the early 1900s and again in the 1930s. However, the modern era of psychotomimetic drugs begins with the synthesis of lysergic acid diethylamide (LSD) in 1943 by Stoll and Hoffmann (1943) and the discovery of its remarkable potency. Subsequent studies showed that LSD bound to serotonin receptors, and soon thereafter it was recognized that a wide variety of indolealkylamines had psychotomimetic properties. These included dimethyltryptamine (DMT) and psilocin. Even the psychotomimetic phenylalkylamines such as mescaline and dimethoxyamphetamine (DMA) were considered to act via a LSD- or indolealkylamine-like mechanism. Finally, it is important to remember that the rediscovery (from a scientific perspective) of natural psychotomimetics (e.g. mescaline) and the discovery of new agents (e.g. LSD) were important components of a view popular from the 1950s to the early 1970s which held that schizophrenia was due to the production of some abnormal natural product. This line of reasoning reached its zenith with attempts to cure schizophrenia via renal dialysis (Wagemaker and Cade 1978; Emrich et al. 1979). Interestingly, the development of Li^+ for the treatment of mania followed a related path. While investigating psychotoxic nitrogenous substances in the urine of mental patients for testing in guinea pigs, Cade administered lithium salts to the animals in an attempt to increase the solubility of urates. Cade noticed that lithium carbonate made the animals lethargic and decided to try lithium carbonate in manic psychiatric patients (Cade 1949). Fortunately, lithium was effective in the first few patients, and lithium remains the standard against which new drugs for the treatment of mania (e.g. valproic acid) are assessed.

Finally, pharmacological developments in nonpsychiatric areas have contributed significantly to modern concepts in biological psychiatry. For example, the development of 3,4-dihydroxyphenylalanine (dopa) therapy for the treatment of Parkinson's disease solidified the evidence for DA as a neurotransmitter. The finding that some Parkinson's patients on dopa therapy developed psychotic symptoms further strengthened the DA hypothesis of schizophrenia. Phencyclidine was developed as a "dissociative" anesthetic agent, but was abandoned because of the high incidence of postoperative delirium with hallucinations. Nearly 30 years later, it was found that phencyclidine and related compounds (e.g. ketamine and MK-801) bind to the *N*-methyl-D-aspartate (NMDA) glutamate receptors, an observation which in part led to the glutamate hypothesis of schizophrenia (Carlsson and Carlsson 1990).

To understand the significance of these historical observations, the current concepts of psychiatric disorders, and the future directions of psychiatric and psychopharmacological research, one needs to develop a perspective which integrates neuroanatomy, neurobiology, neurochemistry, neuropharmacology, and behavior. We begin with a brief review of cellular elements of the central nervous system

(CNS) that are especially relevant to psychopharmacology.

2 Cells of the Nervous System

An understanding of the action of drugs on the nervous system demands that we look at what is special about the nervous system, and that begins with the cells from which it is constructed, nerve cells or neurons. These cells possess two special properties related to information transfer and processing not found in other cells of the body. First, they can transmit electrical signals over long distances without loss of signal strength, and second, they have special areas for intercellular communication, synapses. There are also other cells in brain called glia. The two main types are the astrocyte and the oligodendrocyte, which latter of which wraps around the long axonal processes forming a mylin sheath to insulate the axon transmitting the electrical signals. In general, there are three types of nerve cells: (1) unipolar cells, such as sensory neurons, which transmit information from a sensory receptor to the dorsal root ganglion, (2) bipolar cells, such as granule cells, and (3) multipolar cells with an efferent axon and a dendritic tree. This latter structure is the receptive end of the cells where messages arrive and often can be a complex network like the branches of a tree.

The special property of neurons relates to electrical excitability. When these cells are at rest, they have an electrical potential across the cell membrane like all cells in the body. This is due to the distribution of ions inside the cell as opposed to outside the cell. The interior is low in sodium and free calcium and high in potassium and compares to the extracellular fluid. The ions move across the membrane through specific ion channels (see Vol. 1, Part 1, Chap. 7), and the membrane potential is maintained through metabolic pumps which maintain the ion gradients across the membrane. In nerve cells, when the membrane is depolarized from its resting potential, a sudden, rapid change in membrane potential is seen in which the potential actually reverses polarity for a short period of time, going from -70 mV to +20 mV, for example. This is termed electrical excitability; it is unique to nerve cells and is due to short-term changes in ion permeability. This is the electrical signal which is propagated down the axon and causes the release of neurotransmitters at the synapse. It is through the action of neurotransmitters at the synapse that most of the drugs which we use in psychiatry act, and we need to consider this area in more detail.

3 Neurotransmitters, Receptors, and Transporters

3.1 What Is a Neurotransmitter?

The role that psychopharmacology has had in the development of biological psychiatry requires an understanding of the neurotransmitters which provide the mechanisms of action for psychoactive substances. Surprisingly, the definition of what makes a substance a neurotransmitter has been changing ever since the original proposals for neurohumoral transmission by Lewandowsky (1898) and Langley (1901). Some of the criteria which have been more or less constant include the following:

1. *The transmitter must be present in the presynaptic terminal of the synapse.* One historical corollary to this criteria was that the ability to synthesize the neurotransmitter must also be present in the presynaptic terminal. However, this criteria would eliminate the neuropeptides, which are synthesized in the cell body and transported to the terminal from consideration as neurotransmitters.
2. *The transmitter must be released from the nerve terminal in response to neuronal activity.* Thus the putative transmitter must be released under conditions (electrical or chemical stimulation) which are known to increase the rate of neuronal firing. Acetylcholine (ACh) was the first neurotransmitter to satisfy this criteria. Loewi (1921) observed that, when he stimulated the vagus nerve of a perfused frog heart, the perfusate caused a second heart to respond (slowing of the heart rate) in the same way as the donor heart. Subsequently, Loewi and Navratil (1926) identified ACh as the substance being released. In vivo microdialysis is a modern adaptation of Loewi's classic experiments, in which a small region (0.5–1 μl) of the brain is perfused and the transmitter present from synaptic overflow is collected and measured. Although an indirect measure, in vivo electrochemical detection accomplishes essentially the same goal. Overall, these and other techniques have provided compelling evidence that all of the transmitter-like substances shown in Table 1 are released by neuronal activity.
3. *Exogenous application of the putative transmitter substance to the target cells must produce the same physiological response as that seen from increasing synaptic release.* Until recent years, meeting this criteria without ambiguity was difficult in brain because of the problem of measuring the transmitter effect on a single cell. However, this problem was largely overcome by the development of whole-cell

patch clamp technology (see Aston-Jones and Siggins 1994).

In general, the original idea of neurotransmitter activity involved point-to-point signaling. However, it is now clear that many substances such as the catecholamines, originally identified as neurotransmitters, may in fact involve modulation of an effect, and the term neuromodulator is often used. Modulation is often effected through actions on second messenger systems such as cyclic adenosine monophosphate (cAMP). A list of transmitters and modulators active in the CNS is presented in Table 1 along with the site of action either on ion channels or second messenger systems.

The discovery of and characterization of neurotransmitters in the brain (including the spinal cord) has developed from several directions, including those mentioned below.

3.1.1 Technology

For the student of the 1990s, it is difficult to recognize that, until the mid-1950s, the only reliable methods for assessing neurotransmitters were bioassays. With the development of the spectrophotofluorometer, it became possible to detect microgram quantities of epinephrine, NE, DA, and serotonin. One of the first pharmacological applications of this technology was the demonstration that reserpine depleted the biogenic amines from the brain; further, the recovery of amine levels was associated with the recovery from reserpine associated "depression." The fluorescence technology was extended to histochemistry, which permitted the visualization of the amines within intact tissue (see Bjorklund et al. 1968). When applied to the brain, this technique revealed the presence of catecholamine and later serotonergic pathways and cell bodies with quite specific regional localizations. The terminology used to describe these groupings, e.g. A9 and A10 to describe the DA neurons of the substantia nigra and ventral tegmental area, respectively, is still found in the current literature. Subsequently, the availability of radioisotopes made it possible to study transmitter metabolism, to visualize the uptake of transmitters into specific brain areas, to study transmitter release, and to characterize the transmitter receptors. The availability of new, highly sensitive detection systems, e.g. gas chromatography/mass spectrometry, made it possible to detect transmitters present in only minute amounts. Advances in molecular biology have made it possible to clone, sequence, and investigate the active sites of neurotransmitter receptors (see Vol. 1, Part 1, Chap. 5).

3.1.2 Brain–Gut Connection

With the advancements in technology, new transmitters, particularly new neuropeptides, were discovered outside the CNS. The intestine was a particularly rich tissue for discovery. A number of the peptides first isolated from the intestine, e.g. cholecystokinin and gastrin, were later found in brain and to be associated with specific transmitter systems. In addition to the intestinal peptides, small peptides first identified either as pituitary peptides (e.g. vasopressin) or as pituitary-releasing factors were subsequently shown to have a transmitter role in the CNS. Corticotropin-releasing factor (CRF) is one such peptide; from the psychiatric perspective, the CRF neurons which project from the central nucleus of the amygdala to the locus ceruleus appear to be important, especially in the regulation of fear, anxiety, and stress.

The neuropeptides are generally colocalized with another transmitter, e.g. DA and neurotensin. The idea that more than one transmitter-like substance is found in a nerve terminal was only grudgingly accepted. Dale (1935) proposed that a neuron releases the same substance at each synapse. However, the recognition that most (if not all) neurons contain more than one transmitter substance has led to a revision of Dale's hypothesis, and it is now generally held that each neuron releases the same set of transmitters at each synapse. However, one wonders whether even this view may be an oversimplification. Consider the DA neurons which release DA from both axonic and dendritic synapses. We really do not know whether the same neuropeptides are released at both types of synapse and in the same proportions.

3.2 Synapse

The general features of the synapse are familiar (Fig. 1). Neurotransmitters, the chemical messengers, are stored in synaptic vesicles; the vesicle membranes contain specific transporters for the neurotransmitters. When the action potential reaches the presynaptic terminal, an influx of Ca^{2+} causes these vesicles to fuse with the presynaptic membrane and release either excitatory or inhibitory transmitters into the synaptic cleft (Fig. 1). Many details of the release process have been characterized. Special proteins, the synapsins and Rab3, control the trafficking and mobilization of vesicles. The docking of the vesicles to release sites on the plasma membrane is controlled by other proteins, synaptagmin and synaptobrevin, in the vesicle membrane and the neurexins and syntaxins

Table 1. Overview of transmitter pharmacology in the central nervous system

Transmitter	Receptor subtype and motif	Effector mechanisms
GABA	$GABA_A$ (ion channel) α, β, γ, δ, σ isoforms $GABA_B$ (G protein)	↑Cl^- conductance ↓cAMP ↑K^+ and Ca^{2+} conductance
Glycine	α and β subunits of (IR)	↑Cl^- conductance
Glutamate	AMPA (IR)	↑Na^+ and K^+ conductance
Aspartate	GLU 1–4 (IR) KA (IR) GLU 5–7; KA 1, 2 (IR) NMDA (IR) NMDA 1, 2_{A-D} (IR)	↑Na^+ and K^+ conductance ↑Na^+ and K^+ conductance ↓cAMP ↑IP_3/DG
Acetylcholine	Nicotinic (IR) α_{2-4} and β_{2-4} isoforms Muscarinic M1–4 (GPCR)	↑Na^+, K^+ and Ca^{2+} conductance M1, M3: ↑IP_3/DG M2, M4: ↓cAMP, ↑K^+ conductance
Dopamine	D1–5 (GPCR)	D1, 5: ↑cAMP D2: ↓cAMP, ↑K^+ and ↓Ca^{2+} conductance D3 + D4: ↓cAMP
Norepinephrine	α_{1A-D} (GPCR) α_{2A-C} (GPCR) β_{1-3} (GPCR)	↑IP_3/DG ↓cAMP ↑K^+ and ↓Ca^{2+} conductance ↑cAMP
Serotonin	5-HT_{1A-F} (GPCR) 5-HT_{2A-C}(GPCR) 5-HT_3 (IR) 5-HT_{4-7} (GPCR)	↓cAMP; ↑K^+ conductance ↑IP_3/DG ↑Na^+ and K^+ conductance 5-HT4, 6, 7: ↑cAMP
Histamine	H_1 (GPCR) H_2 (GPCR) H_3 (?)	↑IP_3/DG ↑cAMP
Vasopressin	$V1_{A,B}$ (GPCR) V2 (GPCR)	↑IP_3/DG ↑cAMP
Oxytocin	(GPCR)	↑IP_3/DG
Tachykinins	NK1 (SP>NKA>NKB) (GPCR) NK2 (NKA>NKB>SP) (GPCR) NK3 (NKB>NKA>SP) (GPCR)	↑IP_3/DG
CCK	CCK_A (GPCR) CCK_B (GPCR)	↑IP_3/DG
NPY	Y1 (GPCR) Y2 (GPCR)	↓cAMP, ↓Ca^{2+} and ↑K^+ conductance
Neurotensin	(GPCR)	↓cAMP ↑IP_3/DG
Opioid peptides	μ (β-endorphin) (GPCR) δ(Met^5-Enk) (GPCR) κ(Dyn A) (GPCR)	↓cAMP ↓Ca^{2+} and ↑K^+ conductance
Somatostatin	$SRIF_{1A-C}$ (GPCR) $SRIF_{2A,\ H}$ (GPCR)	↓cAMP ↓Ca^{2+} and ↑K^+ conductance
Purines	P_1 ($A_{1,\ 2a,\ 2b,\ 3}$) (GPCR) P_{2X}(IR) P_{2Y} (GPCR)	↓cAMP ↓Ca^{2+} and ↑K^+ conductance ↑Ca^{2+}, K^+, and Na^+ conductance ↑IP_3/DG

IR, ionophore receptor; GPCR, G protein-coupled receptor.

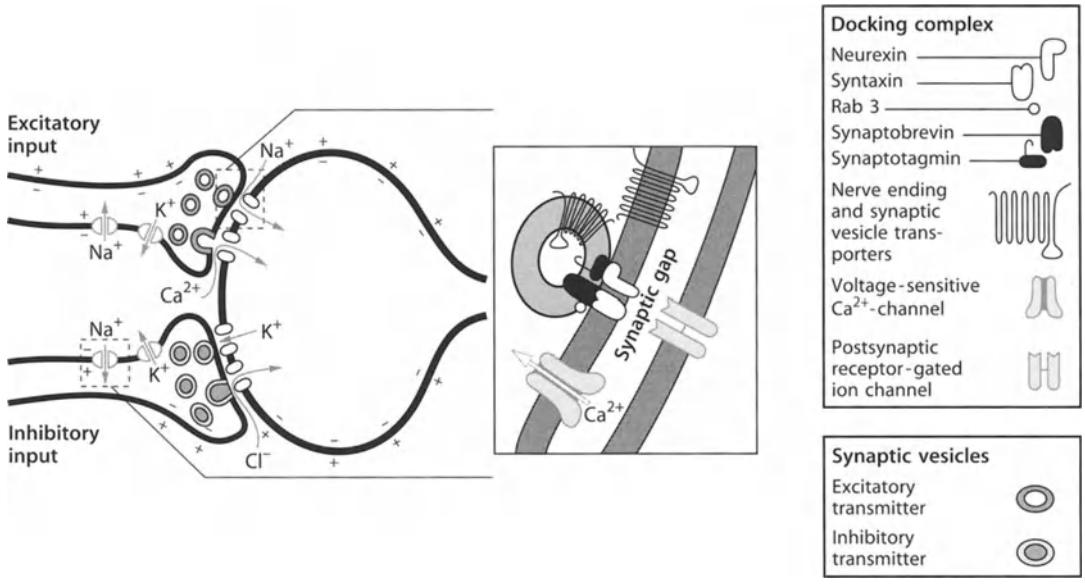


Fig. 1. The synapse. Neurotransmitters are stored in vesicles which fuse to the cell membrane and release transmitters in the postsynaptic cleft, which then interact with post- or presynaptic receptors

found in the plasma membrane. Synaptophysin is important for the formation of the fusion pore.

Once released, some of these transmitter molecules diffuse across the synapse and interact with postsynaptic receptors. The transmitter molecules may also interact with presynaptic receptors, of which there are two types, release-regulating and synthesis-regulating receptors. The binding of the transmitter to these receptors decreases release and inhibits local transmitter synthesis (for the neuropeptides, this mechanism is not operational since the peptides are synthesized in the cell body or soma). When the presynaptic receptor binds the transmitter released from the terminal, it is designated as a presynaptic homoreceptor. When the presynaptic receptor binds a different transmitter, it is termed a heteroreceptor. There are also receptors on the soma and dendrites (somatodendritic receptors); these may be either homo- or heteroreceptors.

The action of the neurotransmitter may be terminated by three processes: metabolism, reuptake, and diffusion. For some transmitters (e.g. the biogenic amines), reuptake into the presynaptic terminal is the most important mechanism of terminating transmitter action. The reuptake, however, may not occur into the terminal from which it was released; in fact, some data suggest that there is considerable diffusion of the transmitter from the synapse of origin before reuptake occurs. The transport of the transmitter need not occur into neurons; glial cells which surround the synapse also contain the transporters for some neurotransmitters (e.g. for glutamate and GABA). For other transmitters, metabolism is the primary mechanism for terminating transmitter action. Acetylcholine esterase (AChE), which is located on the membrane surface, rapidly metabolizes ACh into acetate and choline. The choline is then taken back up by the nerve terminal by a specific transporter for resynthesis into ACh; choline is also provided for ACh biosynthesis from the breakdown of membrane phosphatidylcholine. Metabolism by both specific and nonspecific proteases is important for terminating the action of the neuropeptides.

3.3 Neurotransmitter Receptors

In addition to listing the major CNS neurotransmitters, Table 1 also lists their receptor subtypes and their effector mechanisms. The neurotransmitter receptors come in essentially two motifs: ionophore receptor (IR) and G protein-coupled receptor (GPCR). However, it is important to remember that other receptor mechanisms operate in the CNS just as they do in other tissues. Ionophore receptors, which are also known as ligand-gated ion channels, act when the binding of the neurotransmitter results in the opening of a specific ion channel. These ionophore receptors are usually composed of multiple subunits (usually tetramers or pentamers). Each subunit contains four transmembrane domains, composed of hydrophobic amino acids. The subunits frequently contain one or more sites for phosphorylation by protein kinases and

dephosphorylation by protein phosphatases; in addition, there are frequently sites for voltage gating. Receptors of this type include the nicotinic cholinergic, some of the receptors for GABA and glutamate, the aspartate, glycine, and one subtype of serotonin receptor.

The G protein-coupled receptors may be directly or indirectly linked to ion channels, e.g. activation of a G protein-coupled receptor could activate a protein kinase which phosphorylates the ion channel proteins. G protein-coupled receptors are monomeric and contain seven transmembrane domains. The G proteins are heterotrimeric molecules composed of α-, β-, and γ-subunits; the complex is defined by the α-subunit composition. For example, α_s is associated with stimulation of adenylyl cyclase, while α_i is associated with inhibition of the enzyme. The agonist–receptor interaction facilitates the disassociation of GDP and the binding of GTP. At this point, the equilibrium is strongly in favor of the α-GTP subunit disassociating from the βγ-subunit. The α-GTP subunit can then interact with various effectors, e.g. adenylyl cyclase. The activity of the α-GTP subunit is terminated by the endogenous GTPase activity of the α-subunit.

Two of the principal targets of the G protein-coupled receptors are adenylyl cyclase and phospholipase C (PLC) (Fig. 2). The PLC-mediated metabolism of phosphatidylinositol-4,5 bisphosphate (PIP_2) nicely illustrates the interactions between two second messengers cAMP and Ca^{2+}. The PIP_2 signaling cascade is of particular interest in psychiatry, since Li^+, which is used to treat mania, blocks the reutilization of inositol for PIP_2 synthesis. The metabolism of PIP_2 yields two products, diacylglycerol (DAG) and inositol triphosphate (IP_3). IP_3 releases Ca^{2+} from internal stores, which in turn can interact with protein kinase C (PKC), calmodulin (CaM) kinase and cAMP. DAG is involved in the activation of PKC. Additional interactions are illustrated in the figure, which illustrates the cellular interactions of just two of the many second messengers and the complexity of the relationships.

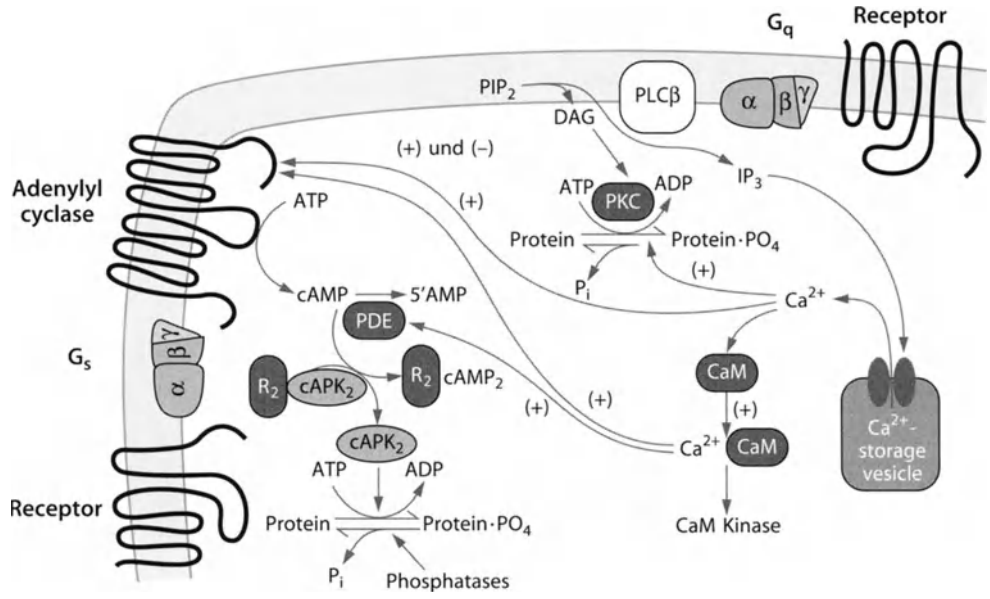
Fig. 2. Interactions between the second messengers cyclic AMP and Ca^{2+}. *CaM*, calmodulin; *DAG*, diacylglycerol; *PDE*, phosphodiester; *PIP_2*, phosphatidylinositol 4,5-bisphosphate; *PKC*, protein kinase C; *PLC*, phospholipase C

4 Specific Neurotransmitters

4.1 Catecholamines

Synthesis

The pathways for the synthesis and metabolism of the catecholamines are shown in Fig. 3. Tyrosine, either formed by the hepatic metabolism of phenylalanine or from dietary sources, is actively transported across the blood–brain barrier (BBB). The first enzyme in catecholamine biosynthesis is tyrosine hydroxylase (TH), which converts tyrosine to dopa. TH is the rate-limiting enzyme in catecholamine biosynthesis and requires both Fe^{2+} and tetrahydrobiopterin. In catecholamine-containing neurons, TH is the main target for the regulation of catecholamine biosynthesis; in particular, the phosphorylation and dephosphorylation



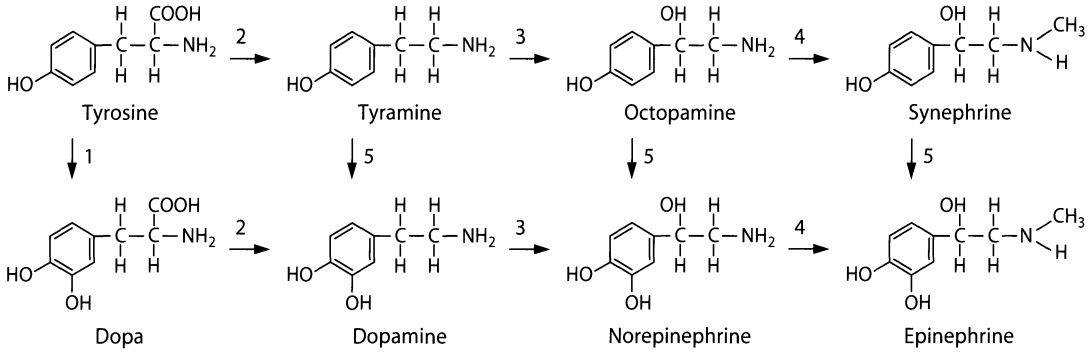


Fig. 3. Pathways in the formation of catecholamines. *1*, tyrosine hydroxylase; *2*, amino acid decarboxylase; *3*, dopamine β-hydroxylase; *4*, phenylethanolamine-*N*-methyltransferase; *5*, catechol-forming enzyme

of TH is used to regulate the enzyme affinity for tetrahydrobiopterin. Normally, the brain concentration of tyrosine is several times greater than the K_m (affinity) of TH for tyrosine, and thus enzyme activity is normally not substrate limited. However, in some pathological conditions, tyrosine becomes scarce (e.g. phenylketonuria, PKU) and an adequate production of dopa is not maintained. TH does not have an absolute specificity for tyrosine; for example, TH will convert α-methyl tyrosine to α-methyl dopa, which is then subsequently decarboxylated to α-methyl DA, a false transmitter. False transmitters can displace the endogenous transmitter. Further, when the false transmitter is released, it frequently has lower efficacy at the relevant receptor sites. Tyrosine analogues such as α-methyl-para-tyrosine (α-MT) reversibly inhibit TH activity; α-MT has been found to block the behavioral effects of some central stimulants, e.g. amphetamine, but not the effects of others, e.g. cocaine and methylphenidate. Such data may suggest that intact catecholamine synthesis is required for the full expression of amphetamine's behavioral effects.

The second step in catecholamine synthesis is catalyzed by aromatic amino acid decarboxylase (AAD) and converts dopa to DA. AAD is ubiquitously distributed in the brain and this step is not rate limiting. AAD also converts 5-hydroxytryptophan (5-HTP) to 5-hydroxytryptamine (5-HT) or serotonin.

The third step in catecholamine biosynthesis occurs only in noradrenergic and adrenergic neurons and catalyzes the conversion of DA to NE. The enzyme responsible is DA-β-hydroxylase (DBH). Thus it is possible to distinguish noradrenergic and adrenergic neurons from dopaminergic neurons on the basis of the presence of DBH (which may be detected by immunocytochemical techniques). A relatively sparse distribution of catecholamine neurons, largely in the hypothalamus and brainstem, synthesize epinephrine and thus require the fourth enzyme associated with catecholamine biosynthesis, phenylethanolamine-*N*-methyltransferase (PNMT). The presence of PNMT can be used to distinguish noradrenergic from adrenergic neurons.

Metabolism

The metabolism of catecholamines is shown in Fig. 3. Both NE and DA can be sequentially metabolized by monoamine oxidase (MAO) and catecholamine-*O*-methyltransferase (COMT) or vice versa. In brain, the primary metabolites detected for DA are dihydroxyphenylacetic acid (DOPAC), 3-methoxytyramine (3-MT), and homovanillic avid (HVA); the primary metabolites detected for NE are normetanephrine (NM) and 3-methoxy-4-hydroxy-phenylglycol (MHPG). There are two forms or isoenzymes of MAO in the brain. MAO-A is found intraneuronally and preferentially metabolizes NE and serotonin. MAO-B is found in glial cells and preferentially metabolizes phenethylamine (PEA).

Like MAO-B, COMT is also localized extraneuronally. Thus the presence of the COMT metabolites NM and 3-MT has been taken as an index of NE and DA release. Although this index has been largely supplanted by direct measures of release, e.g. in vivo microdialysis, measures of NM and 3-MT levels were historically important in understanding the mechanisms of drug action. For example, amphetamines and related compounds release both NE and DA, which is evidenced by an increase in NM and 3-MT levels.

4.1.1 Dopamine

Dopamine Pathways

From a psychiatric perspective, the three DA pathways of importance are the nigrostriatal, the mesolimbic/cortical, and the tuberoinfundibular (Fig. 4). The nigrostriatal projects form the substantia nigra zona reticulata to the striatum. Degeneration of this pathway results in Parkinson's disease; similar symptoms are

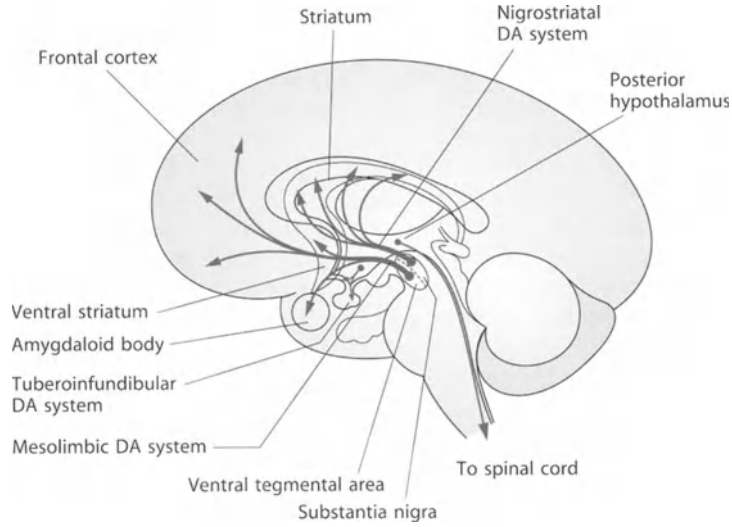


Fig. 4. Dopaminergic (*DA*) pathways. The nigrostriatal DA system originates in the substantia nigra and terminates in the main dorsal part of the striatum. The ventral tegmental area gives rise to the mesolimbic DA system. The tuberoinfundibular system innervates the median eminence and the posterior and intermediate lobes of the pituitary. (From Heimer 1983)

produced by the administration of typical antipsychotic drugs, e.g. haloperidol and chlorpromazine, which block D_2 DA receptors (see below). Activation of these receptors, e.g. after the administration of a direct DA receptor agonist such as apomorphine, produces a pattern of behavior known as stereotyped activity. These highly repetitive behaviors frequently involve cleaning and grooming activities. In the early days of antipsychotic drug development, new drugs were chosen for their ability to block such stereotyped activity in laboratory animals. Not surprisingly, all drugs chosen in this way produced extrapyramidal symptoms (EPS) when used clinically.

The mesolimbic/cortical DA pathway originates from DA cells located more medially to those of the nigrostriatal pathway. It has been useful to collectively name the nuclei which contain these cells the ventral tegmental area (VTA). This collective VTA actually contains the ventral tegmental area, the pigmented parabrachial nuclei, the rostral linear nuclei, and the paranigral nuclei. Historically, it was considered that the DA neurons of the VTA projected largely to the ventral striatum, particularly the shell of the nucleus accumbens, and to the cortex, in particular the orbitofrontal and dorsolateral prefrontal cortices. There are also projections to the amygdala and hippocampus. More recent evidence suggests that the nigrostriatal and mesolimbic pathways are not as compartmentalized as originally thought and some DA neurons of the VTA have targets in motor areas of the striatum, while some DA neurons in the substantia nigra have limbic targets. It is the mesolimbic/cortical pathway which has long been seen as important in a wide variety of behaviors, including drug reinforcement and psychosis; thus blocking the DA receptors of this pathway is seen as the locus for antipsychotic drug action.

The DA cell bodies of the tuberoinfundibular pathway are located in the arcuate and periventricular nuclei of the hypothalamus and project to the anterior pituitary and the infundibulum. DA inhibits the release of prolactin; thus blocking DA receptors in the pituitary increases prolactin release. This increase in prolactin levels can be measured with standard techniques and thus provides some indirect measure of DA receptor blockade.

Dopamine Receptors

For nearly 30 years prior to the actual cloning of DA receptors, it was recognized that the typical antipsychotic drugs must exert their therapeutic actions by blocking some population of these receptors (Carlsson and Lindquist 1963). Ligand-binding studies (Creese et al. 1975; Seeman et al. 1975) revealed that the typical antipsychotic drugs blocked DA receptors in direct relation to their clinical potency. Spano et al. (1978) and Kebabian and Calne (1979) classified the DA receptors into two groups, D_1 and D_2; DA stimulated adenyl cyclase activity at the D_1 receptor and inhibited adenyl cyclase at the D_2 receptor.

The D_2 receptor was cloned in 1988 (Bunzow et al. 1988) and not unexpectedly had the structure of a typical G protein receptor. Five classes of DA receptors have been cloned. The D_1 and D_5 receptors fall into the class of the D_1-like receptors; these receptors are coded by genes on human chromosomes 5 and 4, respectively and are more than 80% homologous. The receptors have similar agonist/antagonist profiles, except that DA appears to be more potent at the D_5 receptor. D_1 receptors are the major DA receptor subtype in the brain and as such play an important role in the regulation of the "direct" output pathway in the basal ganglia (see below).

The D_2, D_3, and D_4 receptors fall into the class of D_2-like receptors, and the genes are found on human chromosomes 11, 3, and 11, respectively. The D_2 receptor has two main variants, D_{2S} and D_{2L}; the long version has an additional 29-amino acid insert in the third cytoplasmic loop. In most brain regions, the long version is present in a higher concentration than the short version. However, no consistent functional difference between the long and short version has been detected. In addition, there has been no demonstration associating a specific psychopathological state with a change in the ratio of the long and short isoforms. Some recent studies have suggested that a polymorphism in the receptor gene is associated with substance abuse and alcoholism, although other studies have not confirmed this relationship. Finally, animal studies have found that a polymorphism in or near the receptor gene is associated with increased receptor density (Kanes et al. 1996).

In comparison to the D_2 receptor, the D_3 and D_4 receptors are present in much lower quantities and appear to have some preferential distribution for limbic areas. Moreover, some data suggest that a relatively higher proportion of D_3 receptors are autoreceptors. Like the D_2 receptor, the D_3 receptor has both a long and short form as well as several nonfunctional variants. D_4 has several variants (Seeman and Van Tol 1994), each of which consists of a multiple of a 16-amino acid insert in the third cytoplasmic loop. Most humans have four repeats, although this varies widely among different races. In addition, there are 19 different types of repeating units; in general, the first and last repeats are the same. No specific psychopathology has been associated with variants in the D_4 receptor.

Dopamine and the Regulation of the Basal Ganglia

More than 95% of the cells within the striatum are inhibitory efferents which receive excitatory glutaminergic input from the cortex. It is convenient to consider that these cells contain either D_1- or D_2-like dopamine receptors, although it is likely that some cells contain both receptor subtypes. (Interestingly many cells in the striatum contain the mRNA for both subtypes, but the expression of the receptor protein appears to be compartmentalized.) The D_1 receptor-containing or -preferring cells are part of the "direct output pathway". The release of DA onto these cells increases GABA (an inhibitory transmitter) release at the main output nuclei – the internal aspect of the globus pallidus and the substantia nigra reticulata – and thus decreases the activity of the inhibitory GABA output neurons in these regions. The D_2 receptor-containing or -preferring cells are part of the "indirect output pathway." The release of DA inhibits the activity of these inhibitory GABA efferents, which send projections to the ventral pallidum; the GABA neurons project to glutaminergic neurons of the subthalamic nucleus, which in turn project to the output nuclei. Enkephalin is colocalized with GABA in the striatal efferents of the indirect pathway. This pathway also has the potential for greater interaction with the limbic system at the level of the ventral pallidum and subthalamic nucleus. Overall, the simple schematic of the direct and indirect pathway would suggest that increasing inhibition of either pathway would have rather similar behavioral consequences. For example, the administration of either a D_1 or a D_2 receptor antagonist should produce a Parkinson-like syndrome (the inability to initiate movement). However, it is important to remember that the regulation of these pathways is complex. D_2 receptors are also autoreceptors, and blockade of these receptors will stimulate the dopaminergic neurons and increase DA release. Thus, depending on the relative difference in pre- and postsynaptic blockade, the behavioral outcome may be quite different.

There is an age-related loss of both DA neurons and D_2 receptors. A substantial loss occurs in the first years of life (>40%). Then there is an age-related loss of approximately 5% per decade beginning in the third decade. For most individuals, this loss of neurons appears to have no behavioral consequences. The remaining neurons synthesize more DA. However, the combination of a more sparse innervation, increased rate of synthesis, and preserved function is more consistent with DA acting as a neurohormone which diffuses a considerable distance from the site of release. When the loss of neurons is greater than approximately 80%, regulation of the system fails, and the symptoms of Parkinson's disease appear. Overall, it is important to recognize that substantial changes in neurotransmitter systems may occur with no apparent loss of function; many systems have remarkable adaptive mechanisms which sustain function. However, once a threshold is reached, deterioration of function is rapid, irreversible, and generally resistant to pharmacological manipulation.

It has long been recognized that there are substantial interactions between brain DA and serotonin systems and that these interactions may be of clinical importance. For example, it has been proposed that the relative ability of a drug to block D_2 and 5-HT_{2a} receptors will determine whether or not a drug is a typical or atypical antipsychotic (atypicals are defined as drugs which will produce little or no EPS). Importantly, it appears that serotonin can regulate DA release by interactions with dendritic, somatic, and terminal serotonin receptors.

Drugs Which Affect Brain Dopamine Systems

The typical antipsychotic drugs, e.g. haloperidol, fluphenazine, thiothixene, preferentially block D_2 DA

receptors, although most of these drugs also block D_1 receptors at higher concentrations.[1] Drugs which block D_2 receptors generally also block D_3 and D_4 receptors, although there are some notable exceptions. The substituted benzamides, such as raclopride and s-sulpiride, have a very low affinity for D_4 receptors; these drugs also bind poorly to D_1 (and D_5) receptors. Interestingly, spiroperidol, a butryphenone-like haloperidol, has more than a 1000-fold selectivity for D_2-like as opposed to D_1-like receptors; spiroperidol and *N*-methyl-spiroperidol are widely used to measure D_2-like receptor densities in both animals and clinical imaging strategies such as positron emission tomography (PET). The atypical antipsychotic clozapine has a significantly higher affinity for D_4 than for D_2 and D_3 receptors. However, olanzepine, which has the pharmacological profile of an atypical drug, i.e. does not produce EPS, binds with relatively equal affinity to D_2 and D_4 receptors. Thus the D_2/D_4 dichotomy cannot explain the unique actions of the atypical antipsychotic drugs. There are no D_1 antagonists approved for clinical use; SCH 23390 has been used in a variety of clinical trials, but the results have been difficult to evaluate because the drug has a short half-life.

The direct DA agonists (drugs which bind directly to DA receptors) include DA, apomorphine, and bromocriptine. DA (like NE and serotonin) is not clinically useful to treat CNS disorders, since it does not penetrate the BBB. It binds with an approximately tenfold higher affinity to D_1-like than to D_2-like receptors. Apomorphine is an emetic agent; the drug binds with relatively equal affinity to the D_1 and D_2 receptors and with tenfold lower affinity to the D_3 and D_4 receptors. When administered at subemetic doses, apomorphine produces marked sleepiness; apomorphine is important in psychiatry since it has been widely used to stimulate the hypothalamic DA systems and affect the release of pituitary hormones, e.g. growth hormone and prolactin. Bromocriptine is still used clinically in the treatment of Parkinson's disease. Bromocriptine has a high affinity for D_2 and D_3 DA receptors and a poor affinity for D_1, D_5, and D_4 receptors.

The indirect DA agonists which increase the synaptic concentration of DA include drugs such as cocaine, methylphenidate, and amphetamine. These drugs act by blocking DA reuptake and/or increasing DA release. The antiviral drug amantidine increases DA release and has been used to treat the dyskinesias induced by antipsychotic drugs.

[1]The law of mass action shows that the percentage of receptors bound is equal to the amount of unbound or free drug divided by the sum of the amount of free drug plus the concentration of drug which just occupies 50% of the receptors. This 50% occupancy concentration is known as the K_D concentration and is frequently discussed as the receptor affinity for compound: the lower the number, the higher the affinity. Many drugs used in psychiatry have affinities of 10^{-7} M or lower. Consider a typical antipsychotic drug which blocks both D_1 and D_2 receptors but with a tenfold greater affinity (lower value of K_D) for the D_2 receptor. Thus, when such a drug occupies 75% of the D_2 receptors, D_1 receptor occupancy will be 23%.However, if 90% receptor occupancy at the D_2 receptor is required for therapeutic effect, D_1 receptor occupancy will be 47%, thus greatly increasing the likelihood of D_1 related side effects.

4.1.2 Norepinephrine

Norepinephrine Pathways

The major noradrenergic pathways in the brain are illustrated in Fig. 5. The highest concentration of noradrenergic cell bodies is found in the locus ceruleus in the pons. It is important to note that these noradrenergic neurons project throughout the cortex, to the hippocampus, the amygdala, and also to the cerebellum. Because of this widespread distribution, drugs which affect the noradrenergic neurons have profound behavioral effects. When examined physiologically, the major effect of activating the locus ceruleus neurons has been a hyperpolarizing response in the target cells, followed by an increase in membrane resistance. Pharmacologically, this response appears to be associated with the activation of β-receptors (see below). Many noradrenergic neurons are found outside the locus ceruleus and are scattered throughout the lateral tegmental area. Many of these neurons project to areas in the basal forebrain such as the amygdala. The adrenergic neurons arise from the same sources, but the terminals appear to be largely localized to the brainstem.

Adrenergic Receptors

The receptors for both the adrenergic and noradrenergic pathways are referred to collectively as the adrenergic receptors. The concept of receptor subtypes began in 1948 with the classification by Ahlquist of the adrenergic receptors into the α- and β-categories based on the responses of tissues to epinephrine, NE, and isoproterenol (Ahlquist 1948). The recognition that a subclass of the α-receptors was acting presynaptically to inhibit NE release led to a further subdivision of the α-receptors into α_1 (postsynaptic and excitatory) and α_2 (presynaptic and inhibitory). Unlike the situation for the DA D_2 receptor, where the same gene encodes the pre- and postsynaptic receptor, the α_1- and α_2-receptors are different gene products. It is now recognized that there are three subtypes of α_1-receptors (A, B, and D) and three subtypes of α_2-receptors (A–C). In addition, there are three subtypes of β-receptors (1–3). A summary of these receptors and

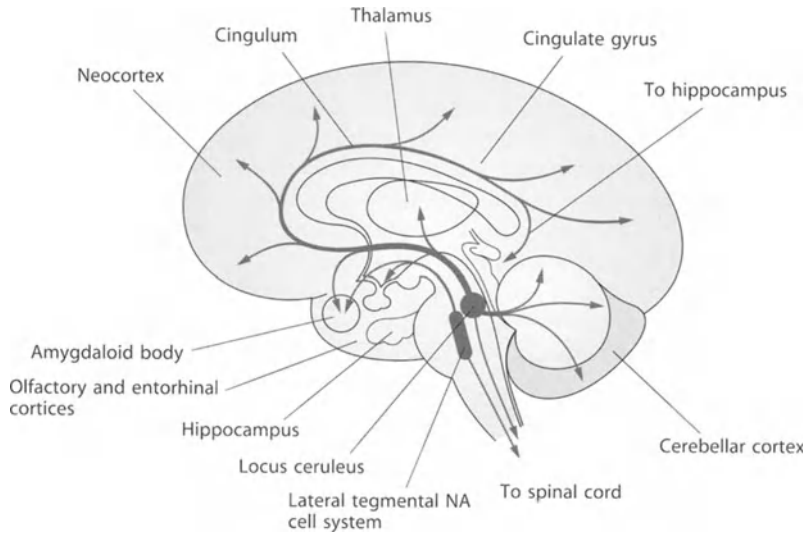


Fig. 5. Noradrenergic pathways. The locus ceruleus projects to many areas in the forebrains, the cerebellum, and the spinal cord. Noradrenergic neurons in the lateral brainstem tegmentum innervate several structures in the basal forebrain, including the hypothalamus and the amygdaloid body. (From Heimer 1983)

their effector systems is found in Table 1. Originally, it was considered that the α_1-receptors were coupled to PIP_2 breakdown, the α_2-receptors to the inhibition of adenylyl cyclase, and the β-receptors to the stimulation of adenylyl cyclase. However, as indicated in Table 1, the effector systems are somewhat more complicated. Furthermore, it appears that the presynaptic/postsynaptic dichotomy between the α_2- and α_1-receptors has some exceptions; it is now clear that some α_2-receptors are postsynaptic.

The adrenergic receptors have been important in understanding at the molecular level the mechanisms which regulate receptor desensitization. When a drug or neurotransmitter is repeatedly applied to the receptor, the physiological effect diminishes over time and more rapidly than could be accounted for by a change in receptor synthesis. For β-receptors, two types of desensitization have been detected, homologous and heterologous. Homologous desensitization is limited to the β-receptors and appears to involve the activity of a receptor-directed protein kinase termed the β-adrenergic receptor kinase (β-ARK). When an agonist (e.g. NE) interacts with the β-receptor, it causes G_S to disassociate into α- and βγ-subunits. The βγ-subunit binds to the plasma membrane, where it appears to stabilize the association between β-ARK and the receptor protein, leading to phosphorylation of multiple sites (serine residues) near the carboxy terminus of the protein. However, this phosphorylation is not sufficient to completely desensitize the receptor. Another protein, β-arrestin, is required, which binds to the phosphorylated protein and sterically inhibits its interaction with G_S. Heterologous desensitization refers to a process whereby stimulation of the β-receptor leads to desensitization of other G protein-coupled receptors. Here, the mechanism appears to involve the activation of the cAMP-dependent protein kinase A (PKA), which in turn phosphorylates the receptor. The phosphorylation sites have been mapped to the third cytoplasmic loop and to a location on the carboxy terminus (but a site different from the sites phosphorylated by β-ARK).

Drugs Which Affect the CNS Noradrenergic Systems

The CNS noradrenergic systems have long been associated with depression and the mechanisms of action of the antidepressant drugs. This association began with the observation that reserpine depleted brain NE and produced behavioral depression in laboratory animals. Further, the depression could be reversed by the administration of the synthetic amino acid dihydroxyphenylserine (DOPS), which is decarboxylated in the brain to NE. The first tricyclic antidepressant was imipramine. Imipramine was synthesized during the 1940s and is structurally similar to the phenothiazines (e.g. chlorpromazine) and quite similar to the substituted benzamide antipsychotics such as raclopride. In 1958, Kuhn began to test imipramine and related phenothiazine analogs for antipsychotic efficacy (Kuhn 1958). Although imipramine was not an effective antipsychotic, it did appear to have some efficacy in depressed patients. This observation led to the discovery and synthesis of a wide variety of tricyclic antidepressants. The first generation of these compounds were found to be potent inhibitors of both NE and serotonin reuptake and strongly anticholinergic (muscarinic receptor antagonists); drugs in this category included imipramine, amitryptyline, and doxepin. The second-generation tricyclic antidepressants were secondary amines and showed a much greater specificity for

blockade of NE as opposed to serotonin reuptake and showed less anticholinergic effects.

MAO inhibitors (MAOI) have been found useful in the treatment of some depressive disorders. Iproniazid was the first clinically useful MAOI. Iproniazid was developed for the treatment of tuberculosis; however, in some patients it was noted that the drug produced euphoria-like symptoms. It was also found that, in animals, iproniazid reversed the "depressive" state produced by reserpine. These observations led Kline (1958) to propose that iproniazid might be useful antidepressant. Because it produces hepatotoxicity, iproniazid is no longer used clinically; however, it is important historically because it helped to establish a relationship between the metabolism of neurotransmitters and psychiatric disorders.

4.2 Serotonin

Serotonergic Pathways in the CNS

The major serotonergic pathways are shown in Fig. 6. The serotonin-containing cells are restricted to clusters of cells lying in or near the midline of the upper pons and midbrain and specifically include the median and dorsal raphe nuclei, the caudal locus ceruleus, the area postrema, and the interpeduncular area. The more caudal cell groups are thought to project largely to the medulla and spinal cord; some of these spinal projections play an important role in the regulation of pain. The more rostral serotonergic groups project throughout the diencephalon and telencephalon, including the basal ganglia, the limbic system, and the association cortex.

Serotonin Synthesis

As with the catecholamines, serotonin is primarily synthesized in the nerve terminal. The first step in serotonin synthesis is the hydroxylation of tryptophan. In contrast to the synthesis of the catecholamines, this step is not limited by the availability of the enzyme tryptophan hydroxylase, but rather by the availability of the substrate. Thus dietary manipulations or the administration of exogenous tryptophan can significantly affect brain serotonin levels. 5-HTP is rapidly decarboxylated to serotonin by the nonspecific aromatic amino acid decarboxylase (Fig. 7). In the nerve, serotonin is metabolized by MAO-A to 5-hydroxyindole acetic acid (5-HIAA). Low levels of 5-HIAA in the cerebrospinal fluid (CSF) have been consistently associated with impulsive behavior and suicide. In the pineal, serotonin is a precursor for the synthesis of melatonin.

Serotonergic Receptors

Gaddum and Picarelli (1957) defined serotonin responses as being mediated by either D or M receptors. The D receptors were found mainly in smooth muscles (e.g. uterus), and their responses were blocked by D-LSD. This observation led investigators to postulate that a similar antagonist mechanism may be operating in the CNS. However, the observation that 2-brom-LSD, which is not a hallucinogen, also blocked the smooth muscle responses shed doubt on this theory. Further, it was found that, at low doses, LSD had agonist-like properties; thus LSD appeared to meet the criteria for a partial agonist. The D receptors of Gaddum and Picarelli are now recognized to be the serotonin receptors, of which there are three subclasses (A–C). These are G protein receptors, are linked to PLC, and largely modulate the breakdown of

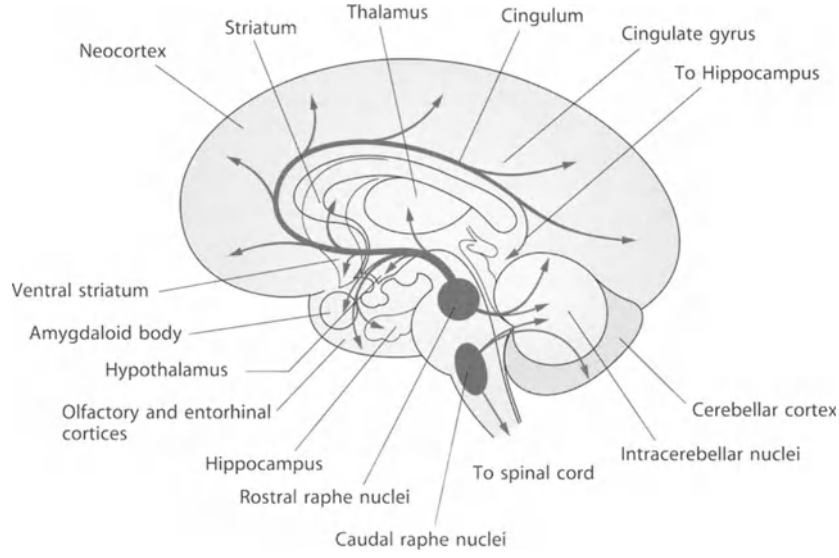


Fig. 6. Serotonergic pathways. The rostral raphe nuclei project laterally through the internal and external capsules to widespread areas of the neocortex (not indicated)

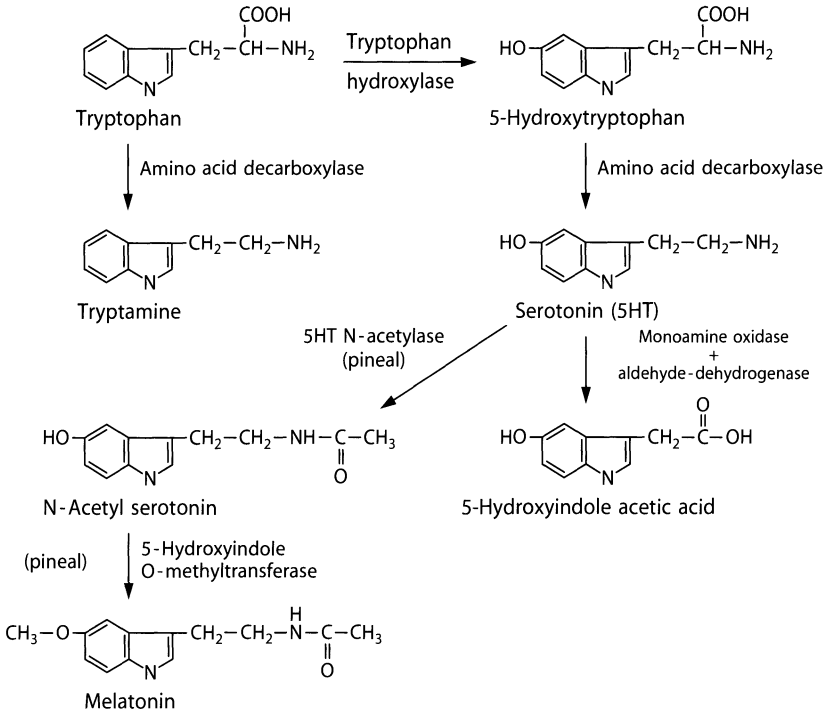


Fig. 7. The metabolic pathways for the synthesis and metabolism of serotonin

PIP_2. 5-HT_{2A} receptors are abundant in the neocortex and the olfactory tubercle, but are also found in some limbic areas and the basal ganglia. 5-HT_{2C} receptors are enriched in the choroid plexus. When applied directly to raphe neurons, LSD inhibits firing, suggesting that it is acting as an agonist at somatic 5-HT_2 autoreceptors. This action of LSD could be consistent with the drug's hallucinogenic activity, since the serotonergic system tonically inhibits visual and other sensory systems; thus LSD would act via *disinhibition*. However, other data suggest that the actions of LSD are more complex and may involve interactions with other serotonergic receptors.

The M receptors of Gaddum and Picarelli were defined by their ability to mediate the release of ACh. These M receptors are now recognized to be 5-HT_3 receptors. The 5-HT_3 receptors are not G protein-linked receptors but rather ligand-gated ion channels and thus are more similar to the $GABA_A$ or nicotinic cholinergic receptors. Activation of the 5-HT_3 receptor leads to increased Na^+ and K^+, but not Ca^{2+} currents. Clinically, the 5-HT_3 receptor antagonists are used to treat chemotherapy-induced emesis. In the laboratory, these antagonists have been found to block both the increase in synaptic DA and some of the behaviors associated with the administration of the central stimulants such as cocaine.

There are 5 subtypes of 5-HT_1 receptors (A, B, D–F). These are all G protein receptors linked to the inhibition of adenylyl cyclase or to the regulation of K^+ and Ca^{2+} channels. 5-HT_{1A} receptors are found on the serotonergic neurons of the raphe nucleus as well as throughout the limbic system, e.g. the hippocampus and amygdala. Activation of these receptors leads to an inwardly rectifying K^+ current and neuronal inhibition. The novel anxiolytic agent buspirone is an agonist at these receptors. $5HT_{1B}$ receptors are presynaptic heteroreceptors, associated with increased transmitter release; in humans, the $5HT_{1B}$ was formerly the 5-$HT_{1D\alpha}$ receptor. No specific agonists or antagonists for these receptors have entered into clinical practice. However, mice with the $5HT_{1B}$ gene "knocked out" show increased aggressive behavior and elevated ethanol consumption (see Vol. 1, Part 1, Chap. 5).

The $5HT_4$, $5HT_6$, and $5HT_7$ receptors are linked to activation of adenylyl cyclase. Two gene products with distinct anatomical distributions, 5-HT_{5A} and 5-HT_{5B}, have been identified, but coupling to specific effectors has not been demonstrated.

Drugs Affecting the CNS Serotonergic Systems

The hallucinogens and buspirone were discussed above. Drugs used in the treatment of depression have long been associated with the serotonergic systems. Thus the first-generation tricyclic antidepressants affected both NE and serotonin transport. The second-generation compounds suggested that blockade of NE transport was sufficient for antidepressant efficacy. The introduction of the serotonin-specific reuptake inhibitors (SSRI) showed that blockade of serotonin reuptake was also sufficient. The first of these compounds was fluoxetine, which was quickly followed by others, e.g. paroxetine and sertraline. Given the

complexity and number of serotonergic receptors, the precise antidepressant mechanism of the SSRI remains unclear. However, some data would suggest that it is the serotonin–NE interaction which is important. The SSRI are also useful in the treatment of obsessive compulsive disorder (OCD), which is thought in part to involve a dysregulation of the limbic loop of the basal ganglia and in particular the orbitofrontal cortex.

4.3 Acetylcholine

Cholinergic Pathways in the CNS

Mesulam and Geula (1994) note that the brain's cholinergic systems appear to be phylogenetically ancient, anatomically ambiguous, and do not respect traditional nuclear boundaries. Thus cholinergic pathways are intermixed with noncholinergic cells. There are eight major cholinergic groups which project to other CNS structures; the Ch1–Ch8 nomenclature was introduced to describe these cell groups. In addition, there are the intrinsic cholinergic neurons of the ventral and dorsal striatum. The location of the various cell groups is as follows: Ch1, medial septal nucleus; Ch2, vertical nucleus of the diagonal band; Ch3, horizontal limb of the diagonal band; Ch4, nucleus basalis of Meynert; Ch5, pedunculopontine nucleus; Ch6, laterodorsal tegmental nucleus; Ch7, medial habenula; and Ch8, parabigeminal nucleus. The major targets of these cell groups may be summarized as follows: Ch1 and Ch2, the hippocampal complex; Ch3, the olfactory bulb; Ch4, the cerebral cortex and amygdala; Ch5 and Ch6, the thalamus; Ch7, the interpeduncular nucleus; and Ch8, the superior colliculus. All regions of the cortex receive intense cholinergic innervation, and the function of these neurons appears to be complex. A recurring theme is the modulatory role of the cholinergic neurons. For example, in the primary visual cortex, Steriade and McCarley (1990) have found that cholinergic stimulation does not affect the orientation specificity of a given neuron, but does increase the likelihood that it will fire in response to its preferred stimulus. The Ch1–Ch4 cell groups are known to have an important role in memory, and in particular lesions of the Ch4 cell groups produce marked memory deficits. The precise mechanisms by which these cholinergic neurons affect memory are not clear, but some data suggest that the cortical cholinergic innervation may help to channel (or gate) sensory information into and out of the limbic system in a way that is sensitive to the behavioral salience of the associated experience (Mesulam and Geula 1994). Thus the Ch1–Ch4 cells are necessary for effective sensory–limbic communication. Cholinergic neurons are also known to have an important role in arousal mechanisms, e.g. the shift from the synchronized electroencephalography (EEG) of deep sleep to the desynchronized EEG of waking and rapid eye movement sleep; in this later respect, the Ch5–Ch6 cell groups appear to be important.

Acetylcholine Synthesis and Metabolism

ACh is synthesized in the cholinergic nerve terminal from acetyl coenzyme A and choline by the enzyme choline acetyltransferease (ChAT). Once released, ACh is metabolized in the synaptic cleft by AChE; thus termination of the activity of ACh does not depend on reuptake. The choline liberated by the metabolism of ACh is taken back up into the nerve terminal for resynthesis to ACh.

Cholinergic Receptors

Dale (1914) originally defined cholinergic receptors on the basis of their responses to certain agonists and antagonists. Muscarinic receptors were activated by the alkaloid muscarine (isolated from the poisonous mushroom *Amanita muscaria*) and antagonized by atropine (isolated from the flowering plant deadly nightshade, *Atropa belladona*). Nicotinic receptors were defined by their response to nicotine and blockade by curarae. Subsequent research has identified a number of snake toxins which block nicotinic receptors, including α-bungarotoxin (α-BGT). Nicotinic cholinergic receptors (nAChR) are multi-subunit (pentameric) ligand-gated ion channels which increase Na^+, K^+, and Ca^{2+} conductances. In muscle, the receptor may contain α-, β-, ε-, and δ-subunits (adult muscle) or α-, β-, γ-, and δ-subunits (embryonic muscle). However, in neurons, the receptor appears to be largely, if not completely, formed from both α- and β-subunits, with one notable exception. In brain, the mRNA for six α-subunits (α_{2-7}) and three β-subunits (β_{2-4}) are routinely detected. However, it is not clear whether all of these subunits are expressed in functional receptors. The most common receptor conformation appears to be $(\alpha_4)_2(\beta_2)_3$. Two types of brain nicotinic receptors can be detected from binding studies, those receptors which recognize α-BGT with a high affinity and those that do not. The former has a low affinity for nicotine, while the latter has a high affinity for nicotine. The brain α-BGT-binding sites, which are enriched in the hippocampus and hypothalamus, appear to be composed of a homopentamer formed from α_7-subunits. Recently, Freedman et al. (1997) have shown that a polymorphism in the α_7-subunit may be linked to schizophrenia and deficits in sensory gating.

Nicotine is known to have diverse and potent effects on CNS function. For example, nicotine interacts with presynaptic nAChR to increase the release of a variety

of neurotransmitters, including DA, NE, serotonin, GABA, and glutamate, all of which are important in regulating behavior. In animal models, nicotine can both enhance learning and memory and reverse deficits in cognitive performance, e.g. those induced by septal lesions. The loss of the Ch1–Ch4 cell groups in Alzheimer's diseases has been associated with the cognitive and attentional deficits of this disorder, and to some extent these deficits (at least early in the disorder) may be reversed by nicotine. This also appears to be the set of action of the newer drugs used to treat Alzheimer's disease, such as rifastigmin and donepezil.

Five muscarinic receptor genes (m_1–m_5) have been cloned and all are G protein-coupled receptors. The m_1, m_3, and m_5 subtypes are coupled to inositol phosphate release, while the m_2 and m_4 subtypes are coupled to inhibition of adenylyl cyclase. There are marked regional differences in the distribution of the various subtypes; for example, in the striatum, the principal subtypes are m_1 and m_4, while in the medulla and pons and cerebellum, the principal subtype is m_2. A moderate level of the m_3 subtype is expressed in most brain regions; m_5 expression is low to negligible in most areas. To date, there are no clinically useful drugs which discriminate among the various subtypes.

Drugs Affecting CNS Cholinergic Systems

The reversible and irreversible AChE inhibitors have marked CNS effects; accidental poisoning from these agents (e.g. from pesticides) frequently causes a syndrome which is not unlike endogenous depression. Muscarinic receptor antagonists such as benztropine are widely used to treat movement disorders, especially the EPS induced by the typical antipsychotic drugs. Dopamine inhibits the release of ACh from the cholinergic interneurons within the striatum; blocking the DA receptors increases ACh release, leading to stimulation of the striatal GABA efferents. A wide variety of drugs have moderate to strong anticholinergic activity; these include some of the typical antipsychotics such as chlorpromazine, the tricyclic antidepressants, and some antihistamines, e.g. diphenhydramine.

4.4 Glycine

Glycine serves two transmitter roles in the CNS. First, it acts on inhibitory glycine receptors which are primarily located in the brainstem and spinal cord (on the basis of observed receptor density). However, from a neurophysiological perspective, nearly all neurons in the brain show an inhibitory response to the application of glycine, suggesting a somewhat more widespread distribution of receptors. The inhibitory glycine receptor is a pentamer formed from homologous 48-kDa and 58-kDa subunits which form an intrinsic Cl^- channel. Four isoforms of the α-subunit (α_1–α_4) and one variant of the β-subunit have been detected. All subunits have four transmembrane domains. Further molecular diversity is generated from alternative splicing of the intracellular loop domains. The α-subunit contains the ligand-binding site, while the β-subunit is a major determinant of channel conductance. Endogenous ligands for the receptor include glycine, β-alanine, and taurine. Strychnine is the best-known glycine antagonist.

Glycine also binds to the NMDA type of glutamate receptor and is required for receptor activation (see below).

4.5 γ-Aminobutyric Acid

GABA is the most abundant inhibitory neurotransmitter in the CNS. However, because of this ubiquitous distribution, it was argued that GABA could not act specifically and the acceptance of GABA as a neurotransmitter was therefore somewhat slowed. In 1950, Roberts and Frankel and Awapara et al. independently discovered GABA in the brain (Awapara et al. 1950; Roberts and Frankel 1950). In 1953, Florey discovered "factor I" in mammalian brain and showed that it inhibited the crayfish stretch receptor neuron (a typical example of the early bioassay approach; Florey 1953). Subsequent work showed that factor I was GABA and that it could duplicate transmitter action at certain crayfish synapses. However, it was not until the late 1960s and early 1970s that GABA was fully accepted as a neurotransmitter in the mammalian brain. Importantly, Roberts and colleagues (1950) were able to show that GABA synthesis is localized to only certain neurons.

As noted in previous sections, some regions, e.g. the striatum, are almost completely composed of GABA neurons. By some estimates, GABA neurons comprise 30%–40% of all neurons in the brain. Further, essentially all neurons in the brain respond to GABA application. GABA is synthesized from glutamate by glutamic acid decarboxylase (GAD); there are two isoforms of the enzyme, both of which require vitamin B_6 as a cofactor. Once released into the synapse, GABA is taken up by specific transporters in both the neurons and glia. GABA is metabolized by GABA-transaminase (GABA-T), a mitochondrial enzyme. The transamination to form succinic semialdehyde can take place only if α-ketoglutarate is the acceptor of the amine group. This transforms the α-ketoglutarate to glutamate, the precursor of GABA.

There are two types of GABA receptors, denoted as $GABA_A$ and $GABA_B$. $GABA_A$ receptors are in the class of ligand-gated ion channels, while the $GABA_B$ recep-

tors are typical G protein-coupled receptors. $GABA_A$ receptors appear to be composed of multiple subunits (generally a pentamer) formed from various combinations of six α-, four β-, four γ-, one δ-, three ρ-, and one ε-subunit. In addition, alternatively spliced versions exist of the α_6-, β_2-, β_4-, and γ_2-subunits. The ρ-subunits appear to be found only in the retina; α- and β-subunits form fully functional GABA receptors. A γ-subunit is required for benzodiazepine potentiation of the GABA facilitated Cl^- flux, although the site of benzodiazepine binding appears to be the α-subunit. Probably the most common form of the $GABA_A$ receptor is $\alpha_1\beta_2\gamma_2$, which has a high affinity for benzodiazepines and related ligands. Other subunit combinations, e.g. those which contain the α_6-subunit, have a negligible affinity for traditional benzodiazepines. In addition to the benzodiazepine-binding site, there are binding sites for the barbiturates and the neuroactive steroids. Progesterone also binds to this steroid site. Other drugs, including the volatile anesthetics and ethanol, potentiate the effects of GABA at the $GABA_A$ receptor, although the mechanism or mechanisms of interaction at the molecular level are poorly understood.

$GABA_B$ receptors were first detected indirectly in the late 1970s from the observation that GABA could block the release of NE but the effect was not blocked by typical GABA antagonists such as bicuculline. Subsequently, it was found that β-*p*-chlorophenyl-GABA (baclofen) was inactive at $GABA_A$ receptors but active at $GABA_B$ receptors. Activation of $GABA_B$ receptors increases K^+ conductance in many regions, leading to membrane hyperpolarization. Since this effect is blocked by pertussis toxin, it is assumed that most $GABA_B$ receptors are coupled to G proteins.

4.6 Glutamate and Aspartate

Glutamate and aspartate are both found widely distributed and in a high concentration in brain. Both amino acids excite nearly all neurons in the CNS, and there is now general acceptance that glutamate and aspartate function as the classical fast-excitatory transmitters. Glutamate can be synthesized from several sources in the CNS, including the metabolism of GABA (see above). After release into the synapse, glutamate can be taken up via specific transporters into either the presynaptic neuron or the adjacent glia in order to terminate the postsynaptic receptor effects. There are two broad classes of glutamate receptors, the ionotropic (or ligand-gated ion channel receptors) and the metabotropic (or G protein-coupled receptors).

The ionotropic receptors are further subdivided into three subclasses on the basis of agonist specificity: these are the *N*-methyl-D-aspartate (NMDA), the α-amino-3-hydroxy-5-methyl-4-isoxazole propionic acid (AMPA), and the kainic acid receptors. The best known of these receptors is the NMDA subtype. The NMDA receptor requires two molecules of glutamate and one molecule of glycine for activation. The glycine-binding site on the NMDA receptor is not strychnine sensitive, which distinguishes this binding site from the site on the inhibitory glycine receptor. Additional modulators of the NMDA receptor include zinc, some neurosteroids, arachidonic acid, redox reagents, and polyamines including spermidine. The NMDA receptor is permeable to both Na^+ and Ca^{2+}; the counterion is K^+. The NMDA receptor is a typical multi-subunit ligand ion channel, formed from Grin1 (glutamate receptor inotropic NMDA) and $Grin2_{a-d}$ subunits. Alternative splicing generates eight isoforms of the Grin1 subunit (one nonfunctional), each with differing pharmacological properties. The $Grin2_d$ exists as two splice variants.

Several lines of evidence point to the importance of the NMDA receptor subtype. First, it has been shown that the chemically related compounds phencylidine ("angel dust"), ketamine, and MK-801 (dizoclipine) are all noncompetitive open-channel NMDA antagonists. Not all the behavioral effects of these compounds can be entirely attributed to NMDA antagonism, e.g. phencylidine is also a sigma opiate receptor antagonist. However, the confusion, disorientation, and psychoses induced by these compounds have contributed to the "glutamate" hypothesis of schizophrenia. Second, exposure of neurons to high concentrations of glutamate for only a few moments can result in cell death; this effect is mediated by excessive stimulation of NMDA receptors and the enhanced influx of Ca^{2+} into the cell. This mechanism has been evoked to explain the cell death which occurs after ischemia or hypoglycemia. Third, long-term potentiation (LTP) is in part associated with activation of NMDA receptors. There are two types of LTP, associative and nonassociative. The associative type of LTP, especially in the CA_1 region of the hippocampus, is perhaps the best understood and will be briefly reviewed here. When a stimulus train of strong intensity brings a relatively large population of presynaptic neurons to threshold, there is a resulting synaptic enhancement which persists for hours or in some cases days, i.e. LTP. As noted by Kandel (1991) in the CA_1 regions of the hippocampus, LTP has three interesting properties: (1) cooperativity (more than one fiber must be activated to produce LTP), (2) associativity (the contributing fibers and the postsynaptic cell need to be active together, in an associative way), and (3) specificity (LTP is specific to the active pathway) (Fig. 8). The non-NMDA glutamate receptors are the AMPA/kainate receptors described below. Some evidence suggests that

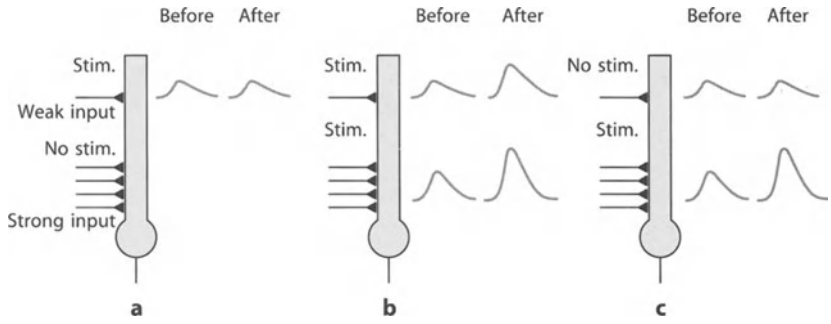


Fig. 8a–c. Single pyramidal cell receiving weak and strong synaptic inputs. Long-term potentiation illustrating **a** cooperativity, **b** associativity, and **c** specificity

the diffusable retrograde messenger is nitric oxide (NO). Is LTP involved in learning and memory, in particular the spatial learning which is so clearly dependent on the hippocampus? Until recently, the data to support such an association was largely pharmacological in nature, i.e. drugs which block the NMDA subtype of glutamate receptors which are required for LTP also block spatial learning. More recently, genetic strategies have been brought to bear on this problem. For example, mice genetically engineered to be deficient in Ca^{2+}/calmodulin-dependent kinase II (so-called knockout mice) show poor LTP and poor spatial learning. However, these mice perform nearly normally on simple, conditioned response paradigms, suggesting the deficit in the knockout mice is selective for certain types of associative learning.

AMPA receptors are formed from Gria1–4 subunits and may exist in homomeric or heteromeric form; each of the polypeptides can exist in two forms generated by alternative splicing known as "flip" and "flop" that differ in their desensitization kinetics. The flip forms give rise to a larger sustained current (slower to desensitize) than the flop forms. Gria1, Gria3, and Gria4 are Ca^{2+} permeable; through RNA editing and the consequent change in composition of a single amino acid, Gria2 is converted from Ca^{2+} permeable to impermeable. Kainate receptors are formed from Grik1–5 subunits. Autoradiographic studies show that both AMPA and kainate receptors are largely localized to telencephalic regions, but each receptor class shows a distinct distribution. Kainate receptors are principally localized in the CA_3 region of the hippocampus, the striatum, the deep cortical layers, the reticular nucleus of the thalamus, and the granule cell layer of the thalamus. AMPA receptors appear to principally localized to the CA_1 region of the hippocampus, the outer cortical layers, the lateral septum, and the molecular layer of the cerebellum. This distribution is similar to that for the NMDA receptors and is consistent with the idea that these receptors act in concert. Overall, the molecular diversity of the glutamate receptors is apparently necessary to maintain fine control.

4.7 Others

4.7.1 Peptides

As noted in previous sections, in most (if not probably all) neurons, neuropeptides are colocalized with the more traditional neurotransmitters. By some estimates, there may be more than 300 neuropeptides in the brain. A partial list of the peptides identified to date is presented below:

- Adrenocorticotropic hormone (ACTH)
- Androgens
- Angiotensin I, II, and III
- Bradykinin
- Calcitonin
- Cardioexcitatory peptide
- Carnosine
- Cholecystokinin
- Corticotropin-releasing hormone (CRH)
- Cortisol
- Endogenous opioids
- Estrogens
- Follicle-stimulating hormone (FSH)
- Gastrin
- Gastrin-inhibiting peptide
- Glucagen
- Gonadotropin-releasing hormone
- Growth hormone
- Growth hormone-releasing factor
- Insulin
- Luteinizing hormone
- Melanocyte-inhibiting factor
- Melanocyte-stimulating hormone (MSH)
- Melatonin
- Motilin
- Neural growth factor
- Neuronal polypeptide
- Neuropeptide Y
- Neurotensin
- Oxytocin
- Progesterone

- Prolactin
- Secretin
- Sleep-inducing peptide
- Somatostatin
- Substance K
- Substance P
- Thyroid hormones
- Thyroid-stimulating hormone
- Thyrotropin-releasing hormone
- Vasoactive intestinal peptide
- Vasopressin

Some well-established examples of coexistence of biogenic amine transmitters and neuropeptides are given in Table 2. It has been postulated that neurotensin and cholecystokinin may be involved in the pathology of schizophrenia because of their colocalization with DA; however, pharmacological data to support this position has been inconsistent. Substance P, a tachykinin, is best known as the neurotransmitter in most primary sensory neurons; substance P is colocalized with DA and serotonin and has been hypothesized to be involved in a variety of psychiatric disturbances.

Of all the neuropeptides, the endogenous opioids have probably received the most attention. There are several reasons for this focus.

1. Morphine and related compounds are widely used in clinical practice for severe pain. Attempts to develop morphine-like compounds which do not produce tolerance and dependence dates back to the synthesis of heroin in the late 1800s and continues to this day. Nearly 50 years of highly sophisticated structure–activity studies led pharmacologists to the conclusion that there were specific CNS opiate receptors. This led to the synthesis of a variety of very powerful synthetic opiates, e.g. fentanyl. Receptor models were proposed that could accommodate both the natural and synthetic opiate agonists. However, until the discovery by Hughes, Kosterlitz, and colleagues of methionine-enkephalin (Hughes et al. 1975), there was little or no interest in the possibility that endogenous opiate-like compounds existed which could interact with these receptors. Subsequently, other peptides with opiate-like activity were identified, including leucineenkephalin, dynorphin, and β-endorphin. Recently, two new endogenous opiate peptides have been identified, endomorphin-1 and endomorphin-2 (Zadina et al. 1997). It has been suggested that these tetrapeptides may be the specific ligands for the μ-opiate (or morphine-like) receptor. Endomorphin-1 displays a 4000- and 15,000-fold selectivity for the μ-opiate receptor as compared to the δ- and κ-opiate receptors (see below). Overall, the initial discovery of the endogenous opioid peptides was significant in stimulating research to find endogenous ligands for receptors (e.g. the benzodiazepine receptors) which were known to be present in the CNS.
2. Morphine and related compounds have marked effects on mood; interestingly, there is a substantial body of older literature suggesting that morphine is an effective antidepressant.
3. The abuse of opiates is a significant contemporary problem, complicated by the favored route of administration (i.v.) and the associated spread of certain infectious diseases, e.g. hepatitis and human immunodeficiency virus (HIV).
4. Specific opiate antagonists have been available for more than 40 years; this greatly facilitated advances in opiate pharmacology and also led directly to the discovery of the opiate peptides.
5. The opiate receptors were among the first of the brain receptors to be characterized (early 1970s) through simple binding assays.

Table 2. Examples of biogenic amine and peptide coexistence in neurons

Acetylcholine	Vasoactive intestinal peptide Substance P
Dopamine	Cholecystokinin Neurotensins
GABA	Somatostatin Cholecystokinin
Norepinephrine	Somatostatin Enkephalin Neuropeptide Y Neurostensin
Serotonin	Substance P Enkephalin

Three distinct families of opiate peptides have been identified – the enkephalins, the endorphins, and the dynorphins, each of which is generated from a distinct precursor polypeptide. The endorphins and enkephalins are widely distributed throughout the CNS; the distribution of the enkephalins is especially enriched in areas which might be considered the extended limbic system. β-Endorphin has a more limited distribution. It is derived from proopiomelanocortin (POMC) (Fig. 9). POMC is processed to produce γ-MSH, ACTH, and β-lipotropin (β-LPH); within β-LPH is β-endorphin and β-MSH. In addition to a high concentration in the pituitary, the POMC peptides are found in a high concentration within the arcuate nucleus, which projects both to the limbic system and spinal cord; this distribution of the POMC peptides corresponds to the areas of the human brain where electrical stimulation can relieve pain.

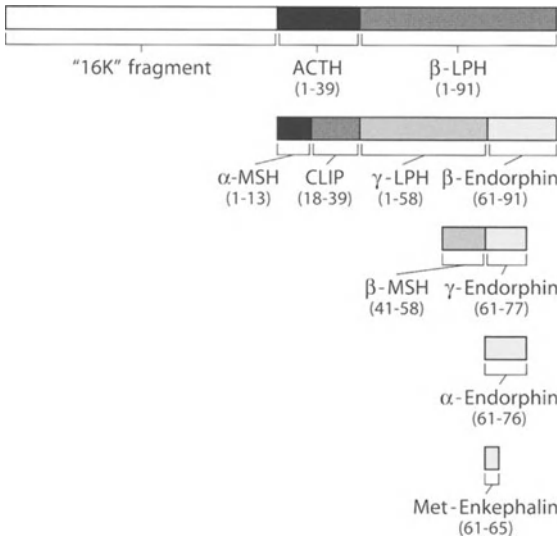


Fig. 9. Bovine precursor molecule, indicating an NH_2-terminal fragment and the following adrenocorticotropic hormone (*ACTH*) and β-lipotropic hormone (*LPH*) sequence. *CLIP*, corticotropin-like intermediate lobe peptide; *MSH*, melanocyte-stimulating hormone

The POMC protein illustrates at the molecular level a mechanism for the coordinate regulation of response to stress and pain. In addition to their effects as neuropeptides, ACTH increases adrenal function, β-LPH increases the mobilization of fat as source of energy, and β-endorphin can have "morphine-like" effects on a variety of organ systems, e.g. the intestine. It is also important to note that the release of the POMC peptides is synergistic with the stress effects mediated by CRF. CRF is not only required for the release of ACTH and β-endorphin from the pituitary, but, as noted previously, also serves as an important neuromodulator of the stress responses. Importantly, when injected into the brain, the CRF antagonist α-helical CRF blocks the pituitary and nonpituitary responses to stress.

There are three main classes of opiate receptors, μ, δ, and κ. On the basis of pharmacological profiles and receptor cloning, each of these classes have two or more subtypes. All of the cloned opiate receptors are of the G protein-coupled type. Most of the clinically used opiates exert their pharmacological effects through the μ-receptor, reflecting their similarity to morphine. The enkephalins and β-endorphin bind with a relatively high affinity to the μ- and δ-receptors, but with a fairly low affinity to the κ-receptors. The dynorphins bind with a high affinity to the κ-receptors and moderate affinity to μ- and δ-receptors. Some of the physiological functions affected by the various receptors include pain, appetite control, sedation, and lymphocyte mitosis.

4.7.2 Purines

It is now recognized that adenosine monophosphate (AMP), adenosoine triphosphate (ATP), and free adenosine can act as neuromodulators. Adenosine can act presynaptically, especially in the cortex and hippocampus, to inhibit the release of amino acid and amine neurotransmitters. Adenosine also inhibits the release of mesopontine ACh through the activation of an inwardly rectifying K^+ current and inhibition of hyperpolarization-activated conductance. Since these effects are blocked by the methylxanthines (caffeine, theophylline, and theobromine), it may account for the stimulant effects of these drugs. Two broad classes of purine receptors have been identified. The P1 class is G protein coupled and has four subtypes (A1–A4); only the A1 and A2 subtypes are inhibited by the xanthines. The P2 class refers to receptors for ATP and uridine triphosphate (UTP); the $P2_x$ subtype is a ligand-gated ion channel, and the $P2_y$ subtype is a G protein-coupled receptor.

4.7.3 Histamine

It has long been recognized that antihistamines have marked behavioral effects. Since many of these compounds were also strongly anticholinergic (e.g. diphenhydramine), it was easy to associate the behavioral effects with muscarinic receptor blockade. However, a well-established network of histamine-containing neurons has now been described; most of these neurons are located in the posterior hypothalamus, but project widely in a manner reminiscent of the other amine transmitters. Three subtypes of histamine receptors have been described; the H_1 and H_2 subtypes are G protein-coupled receptors. The H_1 subtype acts to mobilize intracellular Ca^{2+}, while the H_2 subtype activates adenylyl cyclase. The H_3 subtype is very sensitive to the effects of histamine and is enriched in the basal ganglia and olfactory regions of the brain. Interestingly, there appears to be no active reuptake mechanism for histamine, and the release of histamine from neurons has never been described.

4.7.4 Nitric Oxide

NO is thought to be an important modulator of neuronal function. Four isoforms of the enzyme which synthesizes NO (nitric acid synthase) have been described. NO has been proposed to have a role in LTP, transmitter release, and NMDA receptor-mediated toxicity. Some data also suggest that CO may be an important neuroregulator, especially of guanyl cyclase activity.

4.7.5 Cannabinoids

The pharmacological effects of marijuana include changes in mood, perception, and motivation; in addition, marijuana reduces nausea, is an anticonvulsant, and reduces intraocular pressure. These effects are related to the cannibinoids found in marijuana; Δ-9-tetrahydrocannabinol is responsible for most of the pharmacological effects. Two cannabinoid receptors have been cloned, CB1 and CB2; these are G protein-coupled receptors where the effector is $G_{i/o}$ for both receptors. A number of potential endogenous ligands for these receptors have been identified; the most widely studied is the arachidonic acid derivative arachidonoylethanolamide (anandamide). Both the receptors and the endogenous ligands are widely distributed in the brain, with the highest concentrations being in the cerebral cortex, cerebellum, striatum, and hippocampus.

5 Conclusion

This overview of the major neurochemical components of synaptic transmission has provided the building blocks on which almost all psychopharmacologically active substances act. However, as we have seen, neurotransmission is only the cell-to-cell signaling mechanism, and what is signaled within the cell is central to changing the output of the system. We have mentioned several of the second messenger cascades which can be activated or inhibited by cell-to-cell signaling. Sites within these cascades could clearly be critical for the development of psychiatric diseases, and sites beyond the second messenger systems, involving the regulation of specific gene expression within nerve cells, will almost certainly play a role in psychiatric disease.

As an example of how programmed gene expression can regulate behavior, consider the case of egg-laying behavior of aplysia. Here stimulation leads to a set of stereotypical behaviors culminating in egg-laying. Two sets of cells are involved, the bag cells and the atrial gland, and once the system is set in motion subsequent behavior appears to be rigidly programmed. Genes are activated which express two hormones in the bag cells leading to both long-term and short-term effects. The hormones arise from a common precursor, and through postprocessing the timing of short- versus long-term effects is regulated. The atrial gland expresses two closely related peptides which appear to come from closely related genes and direct the CNS output which leads to egg-laying behavior. Thus, in this case, behavior is rigidly programmed through the influence of specific hormones which appear through the activation of specified genes; it really is all in the hormones, and if appropriate signals arrive intracellularly, the subsequent behavior is programmed. Similar activation of specific hormones and neuromodulators may play a role in mammalian systems, although perhaps with less specified programming. There are, for example, several experiments which suggest that cellular connections can be changed due to the activation of genes. One example which may play a role in psychiatry involves NE activation of β-receptors leading to the activation of the gene for brain-derived nerve growth factor (BDNF), which has been shown to alter LTP and perhaps learning ability. These examples suggest that the next stage in understanding the neurobiology of mental illness may involve understanding how changes in signal transduction result in specific gene activation and changes in CNS function.

6 References

*Ahlquist RP (1948) A study of the adrentropic receptors. Am J Physiol 153: 586–600

**Aston-Jones G, Siggins GK (1994) Electrophysiology. In: Bloom FE (ed) Psychopharmacology: fourth generation of progress. Raven, New York, pp 95–110

Awapara J, Landua A, Fuerst R, Seale B (1950) Free gamma-aminobutyric acid in brain. J Biol Chem 187: 35–39

Bell DS (1973) The experimental reproduction of amphetamine psychosis. Arch Gen Psych 29: 35–40

Bjorklund A, Ehinger B, Falck B (1968) A method for differentiating dopamine from norepinephrine in tissue sections by microspectrofluorometry. J Histochem Cytochem 16: 243–257

Bunzow JR, Van Tol HHM, Grandy DK et al (1988) Cloning and expression of a rat D2 dopamine receptor cDNA. Nature 336: 783–787

Cade JFJ (1949) Lithiurn salts in the treatrnent of psychotic excitement. Med J Aust 2: 349–352

*Carlsson A, Lindqvist M (1963) Effect of chlorpromazine or haloperidol on formation of 3-methoxytyramine and normetaonephrine in mouse brain. Acta Pharmacol Toxicol 20: 140–144

*Carlsson M, Carlsson A (1990) Schizophrenia: a subcortical neurotlransmitter imbalance syndrome. Schiz Bull 16: 425–432

Creese I, Burt DR, Snyder SH (1975) Dopamine receptor binding: differentiation of agonist and antagonist states with 3H-dopamine and 3H-haloperidol. Life Sci 17: 993–1002

Dale HH (1914) The action of certain esters and ethers of choline, and their relation to muscarine. J Pharmacol Exp Ther 6: 147–190

Dale HH (1935) Pharmacology and nerve endings. Proc R Soc Med 28: 319–332

Emrich HM, Kissling W, Fischler M, Zerssen DV, Riedhammer H, Edel HH (1979) Hemodialysis in schizophrenia: three failures with chronic patients. Am J Psychiatry 136: 1095

Florey E (1953) Über einen nervösen Hemmungsfaktor in Gehirn und Rückenmark. Naturwissenschaften 40: 295–296

Freedman R, Coon H, Myles-Worsley M et al (1997) Linkage of a neuropsychological deficit in schizophrenia to a chromosome 15 locus. Proc Natl Acad Sci USA 94: 587–592

Gaddum JH, Picarelli ZP (1957) Two kinds of tryptamine receptors. Br J Pharmacol 12: 323–328

Heimer L (1983) The human brain and spinal cord. Springer, Berlin Heidelberg New York

Hoffmann BB, Lefkowitz RJ, Taylor P (1996) Neurotransmission: the autonomic and somatic nervous systems. In: Hardman JG, Limbird LE (eds) The pharmacological basis of therapeutics. McGraw-Hill, New York, pp 105–140

Hughes J, Smith W, Kosterlitz HW, Fothergill LH, Morgan GH, Morris HR (1975) Identification of two related pentapeptides from brain with potent opiate agonist activity. Nature 258: 577

**Kandel ER (1991) Cellular mechanisms of learning and the zbiological basis of individuality. In: Kandel ER, Schwartz JH, Jessell TM (eds) Principles of neural science, 3rd edn. Appleton and Lange, New York, pp 1009–1040

Kanes S, Dains K, Cipp L, Gattey J, Hitzemann B, Rasmussen E, Larderson S, Silverman M, Hitzemann R (1996) Mapping the genes for halperidol induced catalepsy. J Pharmacol Exp Ther 277: 1016–1025

Kaplan HI et al (1991) Synopsis of psychiatry. Williams & Wilkins, New York

*Kebabian JW, Calne DB (1979) Multiple receptor for dopamine. Nature 277: 93–96

Kline N (1958) Clinical experience with iproniozid. J Clin Exp Psychopathol 19[Suppl 1]:72–78

Kuhn R (1958) The treatment of depressive states with G22355 (imipramine hydrochloride). Am J Psychiatry 115: 459–464

Langley JN (1901) Observations on the physiological action of extracts of the suprarenal bodies. J Physiol (Lond) 27: 237–256

Lewandowsky M (1898) Über eine Wirkung des Nebennierenextractes auf das Auge. Zentralbl Physiol 12: 599–600

Loewi O (1921) Über humorale Übertragbarkeit der Herznervenwirkung. Pflugers Arch 189: 239–242

Loewi O, Navratil E (1926) Über humorale Übertragbarkeit der Herznervenwirkung. Mitteilung. Über das Schicksal des Vagusstoff. Pflugers Arch Ges Physiol 214: 678–688

Mesulam MM, Geula C (1994) Chemoarchitectonics of axonal and perikaryal acetylcholinesterase along information processing systems of the human cerebral cortex. Brain Res Bull 33: 137–153

Mesulam MM, Mufson EJ, Wainer BH, Levey AI (1983) Central cholinergic pathways in the rat: an overview based on an alternative nomenclature (Ch1–Ch6). Neuroscience 10: 1185–1201

Moreau de Tours JJ (1845) Du haschisch et de l'aberration mentale: études psychologiques. Masson, Paris, p 428

Roberts E, Frankel S (1950) Gamma-aminobutyric acid in brain: its formation from glutamic acid. J Biol Chem 187: 55–63

Seeman P, Van Tol HHM (1994) Dopamine receptor pharmacology. Trends Pharmacol Sci 15: 264–274

**Seeman P, Chau-Wong M, Tedesco J, Wong K (1975) Brain receptors for antipsychotic drugs and dopamine: direct binding assays. Proc Natl Acad Sci USA 72: 4376–4380

Spano PF, Govoni S, Trabucchi M (1978) Similarities and dissimilarities between dopamine and neuroleptic receptors: further evidence for type 1 and type 2 dopamine receptors in the CNS. Adv Biochem Psychopharmacol 19: 155–165

Steriade M, McCarley RW (1990) Brainstem control of wakefulness and sleep. Plenum, New York

Stoll A, Hoffmann A (1943) Helv Chem Acta 26: 944

Wagemaker H, Cade R (1978) Hemodialysis in chronic schizophrenic patients. South Med J 71: 1463–1465

Zadina JE, Hackler L, Ge LJ, Kastin AJ (1997) A potent and selective endogenous agonist for the mu-opiate receptor. Nature 386: 499–502

CHAPTER

7

J. Aldenhoff

Fundamental Cellular Principles of Psychological Disturbances

Translator:
S. Coupland-Heimann

1 Introduction

Virtually no other field of modern science has been subject to such rapid development in knowledge as seen in neurobiology in the last decade. Many of the known functions of nerve cells represent targets for pathophysiological changes. However, we are still very far from the development of pathogenetic models which could be useful in determining the prognosis of diseases or in providing causal-based treatment.

The discussion of possible pathophysiological mechanisms is based upon normal neurobiology. As this is so complex, the reader is referred to a good and readable standard work (Kandel et al. 1995).

2 Basic Neurobiology

2.1 Physiology of Nerve Cells

Our current understanding of neurons is mainly based upon extensive biochemical investigations. These investigations chemically extracted and measured the concentrations of certain substances and their molecules from cells, which were rendered non-functional in the process. Important functional characteristics of the cell are lost in such experiments, in particular the time-related functions of the cell. Time-dependent electrical discharge patterns of neurons are essential information. It now has to be assumed that the discharge pattern of nerve cells is just as important for their function, and in their disturbances, as changes in the amount of neurotransmitters, receptors or subcellular factors. These functional aspects can only be determined with a few techniques such as extra- or intracellular recording or the patch clamp technique (Neher and Sakmann 1992). Only by combining these results with those of biochemical studies is it possible to achieve a relatively complete picture of the functions of nerve cells.

The complex functions of the central nervous system can be summarised as follows:

- The processing of registered information
- The programming of motor and emotional responses
- Learning and memory

Each of these functions can be traced to the level of the nerve cells and their interactions. It can be assumed that the complex disturbances which have also been described in psychiatry are due, at least theoretically, to disturbances of such functions and their cellular correlates.

Nerve cells can be understood as biochemically and electrophysiologically interactive units which are modulated by a continuous exchange of acquired and genetic transferred information.

Neurons express larger amounts of genetic information than cells of other organs, with approximately 200,000 differing sequences of messenger RNA (Spudich 1993). The proteins which are subsequently synthesised remain either in the cytoplasm or are integrated into the cell nucleus, the mitochondria or the cell membrane system.

The cytoplasmic proteins consist of the fibrillary elements, which form the cellular skeleton, responsible not only for the structural consistency of the cell but also for the functioning of axonal transport. Other cytoplasmic proteins include multiple enzymes required for the metabolic functioning of the cells. The proteins of the cell nucleus regulate genetic expression. Membrane proteins contribute to a large extent to the structure of the cell membrane in the form of both ion channels and cellular receptors. They are also involved in the production and release of neurotransmitters and neuropeptides. With the opening and closing of the ion channels, the proteins undergo a change in their conformation. They therefore represent a connection between the electrical and biochemical functions of the cell (Table 1).

The cell membrane is closed to ions in the resting phase. Through the activity of various ion pumps, a disparity or potential is created across the cell membrane, which underlies the excitability of the nerve cell. In the resting phase, the concentration of Na^+ is high externally, and that of K^+ is high internally. This results in an intracellular and extracellular electrochemical balance, creating a negative resting phase potential (Hille 1992). On stimulation of the neuron, the ion channels open and a current of

Table 1. Ion currents of nerve cells

Signal	Characteristics	Ion currents	Mechanism
Resting potential	Stabile, –45 to –90 mV	K^+, Cl^-	Leaky currents, voltage-dependent K^+ currents
Action potential	"All or nothing"; amplitude approx. 100 mV	Na^+, K^+	Voltage-dependent channels

positively laden Na^+ passes into the cell, causing a reverse of the transmembrane potential. This is followed by an outward current of K^+, leading to membrane repolarisation (Hodgkin 1964).

A special role is played by calcium ions (Ca^{2+}); in the resting phase, a concentration gradient of approximately 10^{-6} *M* exists between the extracellular and intracellular spaces. Thus only a few Ca^{2+} ions are required in the inner part of the cell before an enormous signalling effect occurs. This takes place via a cascade of calcium-binding structures, including calmodulin and calbindin. Ca^{2+} is thus one of the most important second messengers in addition to cyclic adenosine monophosphate (cAMP).

Since an increase in the internal concentration of Ca^{2+} represents a "risk" to the cell – as a concentration of 10^{-5} *M* calcium phosphate can be detected (Kretsinger 1979) – a number of mechanisms exist to prevent this. The Ca^{2+}-dependent K^+ current is particularly interesting in neurophysiology; when the intracellular concentration of free ionised calcium ($[Ca^{2+}]_i$) increases above a critical value during the stimulation phase, the Ca^{2+}-dependent K^+ current ($I_{K(Ca^{2+})}$) is activated and results in the hyperpolarisation of the cell. Consequently, a membrane potential is produced, which prevents penetration of the Ca^{2+} ions into the inner part of the cell and thereby prevents neuronal stimulation. $I_{K(Ca^{2+})}$ is therefore fundamental for phasic cell discharge. $I_{K(Ca^{2+})}$ occurs in neuronal structures such as the hippocampus, the locus coeruleus and the cell nuclei of the thalamus and the hypothalamus and therefore plays a decisive role in determining the spectrum of activity in the central nervous system (CNS). The metabolic performance of a cell is also influenced by the $I_{K(Ca^{2+})}$-determined phasic activation pattern; since cell metabolism is triggered by an increase in $[Ca^{2+}]_i$, it must be assumed that cell metabolism is activated and inactivated in a phasic manner.

Ion channels consist of characteristic units. The so-called α-unit makes up the central part of the voltage-dependent canal for Na^+, K^+ and Ca^{2+}. It consists of a chain-like polypeptide structure in which a six-transmembrane α-helix is repeated four times. The fourth α-helix is most likely to be the voltage detector. Channel structures for the other ions consist of other units, e.g. β and γ (Catterall 1995, 1996a,b). The various ion channels consist of these basic elements.

The channel types are not homogeneously distributed across the nerve cell; instead, they are found grouped in areas which correspond to their function.

The post-synaptic part of the dendritic tree collects and transmits impulses arising from other nerve cells and transfers them to the body or soma of the nerve cell. Both stimulatory and inhibitory potentials are integrated into the existing activation phase of the cell. The dendrite and soma do not react passively to the input, but integrate the information received via the synapses into the cells' spontaneous activity (Llinas 1988).

Again, the K^+ currents, particularly $I_{K(Ca^{2+})}$, play an essential role here. The state of the K^+ currents determines what type of potential (stimulatory or inhibitory) is ultimately generated and transmitted along the axon; if $I_{K(Ca^{2+})}$ is strong, the transmission is blocked; if it is moderate in strength, phasic activation is demonstrated; and in the case of weak $I_{K(Ca^{2+})}$, the signal is directly transmitted. It is apparent that the inhibitory capacity of the K^+ currents represent a protective mechanism against overexcitability of the CNS.

The signal is not changed by the axon, which transmits all threshold depolarisations in the form of action potentials. In the synapsis, the action potentials produced by Na^+ and K^+ release Ca^{2+} currents of the N-type, which in turn lead to the opening of synaptic vesicles and the release of the neurotransmitters (Neher 1992).

All neurons have 1000 synapses and receive information from approximately 10,000 connections. Synapses have always been considered to be stable structures. As a matter of fact, electron microscopy examinations have demonstrated that the "spines", "spikes" or sack-like processes of the dendrite tree, which are probably responsible for cell communication, are continuously being dismantled and rebuilt (Murphy and Segal 1997; Rusakov et al. 1996). Important influences on the size of the inflows include the activation of the synapses and hormones such as oestrogen (Lewis et al. 1995).

The electrical transmission at the so-called gap junctions, channels through which a direct ion transfer can take place between cells (Bennett 1997), is essential for the synchronisation of neighbouring cells. The chemical transmission at the synapses involves the release of transmitter substance by the presynaptic cell and its binding on the postsynaptic cell. Upon binding to the postsynaptic cell, there is either a direct or indirect interaction with ion channels and the chemical signal is therefore transformed into an electrical one again.

Ion channels are controlled through various stimuli:

- Ligand-controlled channels open when a ligand combines with its receptor.
- Second messenger-dependent channels are opened and closed via protein phosphorylation.
- Potential-dependent channels open or close in response to changes in the membrane potential.

Ion channels can be grouped into families according to their common genetic origins (Stevens 1987). Thus potential-dependent channels, which are selective for

Ca^{2+}, Na^{+} or K^{+}, belong to one family. A second gene family is responsible for the ligand-controlled channels, which are activated via acetylcholine (ACh), gamino butyric acid (GABA), glycine or glutamate. A third family codes for the gap junctions.

2.2 Neurotransmitters

Neurotransmitters either bind directly to a receptor which is bound to an ion channel and thereby activate or end an ion current, or they bind to a receptor which activates a G protein, resulting in the activation of the second messenger cascade. This ultimately causes the opening of the ion channel via its phosphorylation.

In order to define a substance as a "neurotransmitter", certain criteria have to be fulfilled (Cooper et al. 1996). These include the following:

- Its synthesis by a neuron
- Its storage and sufficient release from a presynaptic terminal in order to produce a defined postsynaptic effect
- The reproducibility of this effect by exogenous application
- A specific mechanism which removes or inactivates the neurotransmitter in the synaptic space

There is a definite grey zone between the "classical" neurotransmitters and other substances, such as peptides, which do not fulfil all these criteria.

"Classic" neurotransmitters and neuropeptides are often located together in one cell and are released simultaneously. The functional significance of this is not known.

The various mechanisms which remove a neurotransmitter from the synaptic space are interesting. The smaller molecules are for the most part degraded enzymatically, whereas the larger peptides remain in the synapsis and can only be removed via diffusion and proteolysis. For several neurotransmitters, there are many high-affinity transporters which take up the neurotransmitter in the terminals or in the glial cells. The majority of the transporters belong to the so-called superfamilies. These are proteins which pass through the membrane 12 times, bind the neurotransmitter and divert them into the internal parts of the cell (Amara and Kuhar 1993).

The border between "neurotransmitters" and "neuromodulators" is also not clear, since many substances can take on the functions of both. Whereas neurotransmitters are important in various theories regarding the development of depression or schizophrenia, neuromodulators perhaps arouse more interest in connection with psychiatric illnesses. In the case of neurotransmission, a substance functions as a transmitter between two nerve cells. In the case of neuromodulation, on the other hand, the existing activity pattern of a nerve cell is altered in that $I_{K(Ca^{2+})}$ is reduced. For example, a phasic activation pattern in a cell can be modulated to a regular one. The reasons for such neuromodulation cannot be explained at the cellular level and may only be clarified in the context of a neuronal network or the homeostasis of the organism. Two examples of differing activating neuromodulation are given in detail below:

Noradrenaline (NA) is a biogenic amine; NA released in the CNS arises to a large extent from a syncytium of nerve cells in the upper brain stem, namely the locus coeruleus (Foote et al. 1983). These cells demonstrate a synchronous phasic activation pattern, which is determined mainly by $I_{K(Ca^{2+})}$. Their discharge occurs when an external, often strong stimulation is to be processed. NA projections occur in the cortex, hippocampus, thalamus, cerebellum and other important areas of the brain (Nieuwenhuys 1987). Correspondingly, following such a stimulus, there is a simultaneous release of NA in these areas. NA increases the effect of the local transmitter; in other words, the incoming stimulation is increased at the point of the target neuron ("enabling") (Bloom 1984). According to the receptor involved ($\alpha 1$, $\alpha 2$, β), NA increases excitatory or inhibitory potentials.

Corticotropin-releasing hormone (CRH) is a peptide. The distribution of CRH reflects its neuroendocrine function; CRH-containing interneurons and projections are not only found in the hypothalamus and locus coeruleus but also in the higher brain regions such as the hippocampus and cortex. Via a mechanism which was determined in vitro, a reduction of the $I_{K(Ca^{2+})}$ (Aldenhoff et al. 1983a) results in an increase in the phasic activity pattern found predominantly in these neurons. These cellular mechanisms might correspond to the CNS "alarm" or stress reactions postulated by Selye which allow the CNS to respond more rapidly to life-threatening situations. The above-described excitatory effect of NA is driven further by the activation of the locus coeruleus, for example. It is apparent from behavioural experiments with CRH (Koob et al. 1984) that the effectiveness of this "stress" reaction decreases with increasing dosage of CRH. A hyperexcited state no longer improves the reactive ability of the organism to the threat; in contrast, it worsens it. It is currently assumed that a CRH overdrive such as this could be the cause of neuroendocrine and vegetative symptoms seen in patients with depression (Holsboer 1989; see also below).

The receptors for glutamate represent a special situation, as glutamate controls two types of ionotrope receptors. The *N*-methyl-D-aspartate (NMDA) receptor controls one channel which allows the passage of Ca^{2+},

K^+ and Na^+. The non-NMDA receptor, on the other hand, binds the glutamate agonists kainat, quisqualat and aminomethyl propionic acid AMPA and regulates a channel which allows the passage of Na^+ and K^+. Further, glutamate functions via second messenger activation; when glutamate binds to a metabolically effective receptor, it activates phospholipase C (PLC), inositol phosphate (IP) and diaglycerol (DAG).

The standard model in which long-term potentiation (LTP) (Nicoll et al. 1988) is experimentally examined is the binding between CA3 and CA1 neurons in the hippocampus, the so-called Schaffer collaterals. With a strong stimulus which activates many fibres simultaneously, it is possible to produce persisting changes in the excitability of a chosen neuron. In this way, certain events occur which are reminiscent of conditioning processes in psychology; should both mild and strong excitatory stimuli activate a pyramid cell, the weaker stimulus can be permanently amplified only if it arrives in close association with the stronger stimulus. The permanent amplification of stimuli is restricted to the associated synaptic structures, which are correspondingly activated.

The decisive prerequisite for the occurrence of LTP is sufficient depolarisation of the postsynaptic cell. A simultaneous discharge of pre- and postsynaptic cells occurs in this way.

There are both NMDA and non-NMDA receptors at glutamate synapses; only the non-NMDA receptors play a role during normal synaptic activation, because the NMDA receptors are blocked by magnesium ions. With a stronger depolarisation of the postsynaptic cell, the positively charged magnesium ions are released from the NMDA receptor channel and the channel thus becomes permeable to Ca^{2+}. It is presumed that the Ca^{2+} current is the decisive factor for the triggering of LTP (Schneggenburger et al. 1993).

The ion channel of the NMDA receptor represents a unique form of channel, namely one with double activation. The channel only becomes permeable to Ca^{2+} when a simultaneous depolarisation of the cell and the binding of glutamate on the receptor occurs. The permanent increase in the postsynaptic excitability is probably transmitted by two Ca^{2+}-dependent protein kinases, protein kinase C and Ca^{2+} calmodulin kinase. The activation of gene expression also plays an essential role in LTP (Tully 1997; see also below).

LTP can only be maintained when an increase in the excitability of the postsynaptic neuron occurs together with an increase in the presynaptic transmitter release. It must be assumed that a retrograde transfer of information occurs from postsynaptic to presynaptic neurons. Nitrous oxide (NO) is considered to be a promising candidate for this function (Holscher 1997).

LTP is a physiological description of a process which was described 50 years ago by Hebbs Law (Bonhoeffer et al. 1989).

LTP is of great significance for complex neurological functions. Its most obvious role is probably memory, but it can be generally assumed that all processes may involve the NMDA receptors, which require the coupling of various experience-dependent neuronal systems. Further, one of the most interesting theories for the explanation of schizophrenia includes the NMDA receptor as an essential structure (see below).

2.3 Receptors and Second Messenger Systems

Various receptors are associated with differing second messenger systems. They transfer the signal carried by the neurotransmitter or neuromodulator to the centre of the cell.

The NA-binding β-adrenergic receptor functions in this way via adenylate cyclase and a G protein, which produces cAMP (Lefkowitz and Caron 1988). ACh binds to the muscarinic ACh receptor, which in turn activates a G protein and PLC. This leads to an increase in inositol triphosphate (IP_3), Ca^{2+} and diaglycerol (Nathanson and Harden 1990). Histamine, together with phospholipase A_2 (PLA_2), produces arachnidonic acid.

Here, again, there is a close connection between biochemical and electrical functions: cAMP activates a biphasic current response, consisting firstly of a faster inward current of Ca^{2+} and Na^+, followed by a slower K^+ outward current. The production of cAMP below the cell membrane results in a shorter excitatory and a longer inhibitory period for the individual cell. These currents might contribute to the phasic activation pattern in the neurons of the locus coeruleus (Wang and Aghajanian 1987), which are characteristic of this syncytium's function.

Ion channels and receptors occur not only in the outer cell membrane, but also within the internal parts of the cell. Ca^{2+} is therefore released from the intracellular stores via the IP_3 receptor in order to achieve optimal concentrations for Ca^{2+}-activated structures.

As an intracellular "excitable medium" (Berridge 1997), it contributes to the amplification and transmission of Ca^{2+} signals in cells (Jacobs and Meyer 1997). Further, it appears that intracellular stores of Ca^{2+} contribute to a high concentration of calcium in the nuclei of hippocampus neurons following glutamate application, suggesting a transcriptional significance of Ca^{2+} (Korkotian et al. 1996). The induction of LTP in the hippocampus requires the release of calcium from the endoplasmic reticulum (ER) (Reyes and Stanton 1996).

A subsequent oscillating uptake and release of Ca^{2+} ions occurs until this process comes to rest (Berridge 1993). The calcium response correlates to the various parts of the signal transduction. In neurons and lymphocytes, for example, the non-receptor kinases of the *src* subfamily, in particular the *fyn* kinases, play an important role in the commencement and arrest of the calcium response (Grant et al. 1992). In all cell types, the Ca^{2+} response is dependent upon the metabolism of phosphatidylinositol. The IP3-sensitive Ca^{2+} channel of the ER is a point of attack for the dismantling of IP3 produced by phosphatidylinositol 4,5-bisphosphate (PIP2). In both neurons and lymphocytes, the IP3-sensitive intracellular Ca^{2+} stores are important. It has also been demonstrated that the so-called Ca^{2+}-induced Ca^{2+} release (CICR) is transmitted by the ryanodine-sensitive receptors. Not only can Ca^{2+} increase its own release from the intracellular stores, but it can also inhibit it depending on the concentration present.

As a result of these complex interactions, a Ca^{2+} response occurs not only as a simple increase in the $[Ca^{2+}]_i$, but mainly in the form of oscillations. Recent results demonstrate that the frequency, amplitude and length of the signal has functional consequences for the cell. B cells, for example, demonstrate differing activations of transcription factors according to the amplitude and length of the calcium signal (Dolmetsch and Lewis 1997). It has also been shown in T cells that, although a sufficient stimulus was required for the cells to respond, there was no difference in the Ca^{2+} response following various stimuli if they exceeded the threshold for a response.

The structures required for the development of Ca^{2+} oscillation vary from cell to cell. Those cells with so-called membrane oscillation have a periodic opening and closing of the potential-dependent Ca^{2+} channels. In contrast, the development of Ca^{2+} oscillation by so-called cytosolic oscillators is dependent upon the release of ions from the ER. Cytosolic oscillators include nerve cells and hormone-releasing cells of the pituitary gland.

Activation of cellular mechanisms via a transmitter may have effects of varying duration. For example, the binding of NA to the α_2-receptor leads to the phosphorylation of a Ca^{2+} channel via the cAMP system and in turn to an outward K^+ current, which inhibits the excitability of the cell for several minutes. Should this process be repeated or should binding of NA to the receptor be amplified, an activation of gene expression follows through the phosphorylation of transcription-regulatory proteins. Such proteins can influence the channel directly and may produce changes in the neuronal excitability over a period of days or weeks. This mechanism, which influences the rate of mRNA transcription, has largely been clarified (Strachan and Read 1996). It is particularly interesting, as it appears to be modified in psychiatric diseases and therefore may play a role in their pathogenesis.

Normally, certain genes are expressed, whereas others are not in a transcriptable form. This situation can be changed by proteins which bind to the regulatory parts of the gene. They are controlled via receptors, which bind substances such as steroid hormones, growth factors and neurotransmitters. The regulatory parts of a DNA strand contain two control elements, a promoter and an enhancing region. The promoter region has many genes with adenine and thymine and is therefore called the TATA box. It brings the RNA polymerase to the region where transcription should commence.

The structures of the enhancer region are the binding points for the proteins which decide whether a gene will be transcribed or not. They are called response elements. Examples include the cAMP response element (CRE), which recognises the CRE-binding (CREB) proteins (these proteins are activated under the control of the cAMP-dependent protein kinases); the phorbolester response element and the glucocorticoid (GC) response element. Some enhancer regions continually activate, whereas others are only active when gene expression occurs as a result of hormonal influence or learning processes (Yin et al. 1995).

It was considered for many years that the subcellular functions described above could not contribute to understanding complex brain functions. This opinion only changed with the help of molecular biology techniques which managed to turn off single receptors or kinases by manipulating their RNA.[1] CREB-knock-out mice, for example, show a definite disturbance of long-term memory following adverse stimulation (Bourtchuladze et al. 1994). Similarly, the turning off of Ca^{2+} calmodulin kinase α*CaMKII* or the *fyn* tyrosine kinase, both of which are involved in Ca^{2+} regulation, leads to disruption of LTP and memory processes (Grant et al. 1992; Rotenberg et al. 1996).

These results converge to a uniform mechanism of the storage of information; the decisive mechanism occurs when a particularly large inward Ca^{2+} current is allowed into the cell. This can occur via a non-specific cation channel with activation of the NMDA receptor or via another non-specific channel leading to an intracellular increase of cAMP. The usually effective controls for an increase in $[Ca^{2+}]_i$ are bypassed via the use of these non-specific channels. The $[Ca^{2+}]_i$ increase

[1] Such techniques include the development of transgenic animals or the antisense knockout technique (Montkowski and Holsboer 1996).

favours enzymatic changes, the time period of which can be influenced by gene expression.

2.4 Summary

The cellular apparatus, which integrates both biochemical and electrical functions, has a relatively high threshold of functional stability in both the resting and activation phases. Modulating influences use the second messenger systems of cAMP and Ca^{2+} via biogenic amines, peptides and hormones. These systems allow medium- to long-term modification of the expression of genes which are not usually expressed. A further possibility for modification is the cellular storage of information.[2]

Neuronal activity can thus be altered at several cellular levels:

- Through the balance between excitatory and inhibitory ion channels
- Through short-term involvement of second messenger systems of cAMP and Ca^{2+}
- Through the expression of existing but "silent" genetic information
- Through new information via processes of cellular information storage

In terms of the pathophysiology of psychological disease, these alterations also represent points at which disturbances may develop. The various therapeutic options to date can also be placed in a pharmacological hierarchy. Furthermore, the description of new mechanisms such as second messenger-activated gene expression should allow for new therapy options.

3 Pathophysiological Mechanisms

It will probably never be possible to undertake a direct examination of the neurons of psychiatric patients. Indirect conclusions, however, can be made in a number of ways:

- The "pharmacological bridge": the effect of a substance on behaviour in an animal experiment is compared with the clinical effect of the substance and a mechanism of action is then postulated (Aldenhoff 1989). Examples of this process include many therapeutic agents, including the pharmacological hierarchy of new substances, which support the developmental history of CRH, etc.
- Cellular research models in humans: certain cellular functions can be investigated on peripheral human cells, e.g. K^+ channels in fibroblasts or lymphocytes (Cohen 1996), $[Ca^{2+}]_i$ in lymphocytes (Aldenhoff et al. 1997) and PLA in thrombocytes (Gattaz et al. 1996). The advantages of this model lie in the availability of the human cell and its identical genome. A disadvantage is the differing functions of the cells.

3.1 Disturbed Balance Between the Excitatory and Inhibitory Ion Channels

A typical example can be considered to be the disturbance of interactions between the hormones of the hypothalamus, the cortex of the adrenal glands and their peripheral target nerve cells. These are presumed to be associated with depression (Holsboer 1989).

Following environmental stress stimuli, a release of CRH occurs from the neurons of the hypothalamus, the central cortical CRH neurons and the neurons of the hippocampus. The above-mentioned neuronal activation begins as a result of an inhibition of $I_{K(Ca^{2+})}$. When the levels of corticosteroids released by the endocrinal effect of CRH are sufficiently high, they have an inhibitory effect on and, ultimately, arrest neuronal activity (Vidal et al. 1986). This neuroendocrine feedback loop is turned off when a CRH overdrive occurs as a result of moderate (but not overly strong) external stresses. Following such stimuli, the $I_{K(Ca^{2+})}$ is further inhibited via a continuous dominant excitatory influence. In addition, as a result of the continuously high levels of cortisol, there is a down-regulation of the GC receptors, which, in contrast to the mineralocorticoid (MC) receptors, have a inhibitory effect on corticosteroids (de Kloet et al. 1990) and, consequently, lead to a weakening of the effect of cortisol. The normal feedback loop is thus hypersensitised. This is accompanied by a CRH-induced noradrenergic stimulation. Under high doses of CRH, the nerve cells of the locus coeruleus lose their phasic activation pattern and discharge in an increased and continuous way (Valentino et al. 1987).

The target neurons of the locus coeruleus also lead to an increased NA effect, which is no longer sensitive to environmental stimuli but increases in excitability independently of this.

This could be an explanation for a decrease in ability to concentrate and in perception in patients with depression. A further explanation could be the

[2] It does not seem appropriate to use the terms "learning" and "memory", as these describe complex abilities which differ with time and objects. Even if new molecular biological techniques support the dependence of differentiated and complex abilities of an individual on subcellular mechanisms, the relationship of storage of information is meant here.

downregulation of GC receptors. Mice that have been treated with antisense to the GC receptor have a definite loss of cognitive capabilities (Montkowski et al. 1995). Interestingly, the impaired capabilities improved after treatment with anti-depressants.

A further example of the disturbed balance between the excitatory and inhibitory ion channels are the cellular changes in addiction and tolerance development. Benzodiazepines increase the physiological effect of the transmitter GABA via the benzodiazepine receptor complex-coupled chloride channel. In other words, they increase the inhibitory effect of the chloride current. With regular use of benzodiazepines, this effect decreases until the appearance of a rebound phenomenon.

This can be explained electrophysiologically by the weakening of the inhibiting GABA effect, leading to a predominant excitatory influence on the membrane potential. From the clinical viewpoint, it is interesting that, during a withdrawal syndrome, anxiety can be initiated by other cellular excitatory substances, such as yohimbine, coffee or intrinsically effective benzodiazepine antagonists.

Similar to alcohol dependence, there is an adjustment of the excitatory balance; on the one hand, there is a decrease in the inhibitory mechanisms of the cell with the regular supply of alcohol. On the other hand, new Ca^{2+} channels are produced (Dolin et al. 1987). Both effects lead to an adjustment of the excitability balance, which can explain most of the clinically observed phenomena of alcohol withdrawal, such as increased transfer in all sensory channels and increased vegetative and epileptic fits.

3.2 Disturbances of the Second Messenger Systems of Cyclic Adenosine Monophosphate and Ca^{2+}

The influence of various anti-depressive medications and treatments, such as imipramine, monoamine oxidase (MAO) inhibitors and electroconvulsive therapy, on the β-receptor of activated cAMP-dependent protein kinase A, has been investigated in detail in clinical trials (Nestler et al. 1990). It was demonstrated that these therapeutic measures, which have completely differing mechanisms of actions from the pharmacological viewpoint, all lead to a translocation of kinase from the cytosol to the cell nucleus, suggesting a possible change in gene expression. A significant decrease in the activity of kinase was demonstrated in depressive patients (Shelton et al. 1996).

Direct investigations of $[Ca^{2+}]_i$ on human peripheral nerve cells do not exist, but there is indirect evidence that a second messenger might be involved in the development of psychiatric disturbances.

One of the most attractive models for the development of bipolar affective illness involves the phenomenon of kindling (Post et al. 1986). Following repeated stimulation of the amygdala, there is a permanent increase in the excitability level ("sensitisation"). Although occurring in a different location, kindling is similar to LTP. This observation led Post et al. to the conclusion that a similar process might be responsible for the development and acceleration of the phases of affective illnesses. According to this hypothesis, it is possible that spontaneously occurring affective episodes lead to biological scars in the form of kindling, which favour the further development of the illness phases. Each of these phases collides with the pathological process and enhances it, so that the time difference between each phase decreases. In the worst case, rapid cycling occurs. Supporting this hypothesis are the effects of the main therapeutic agents used in bipolar illnesses – lithium and carbamazepine – i.e. the regulation of $[Ca^{2+}]_i$ (see below).

The observation that an increase in $[Ca^{2+}]_i$ above a critical value leads to cell death resulted in a hypothesis that age and, in particular, dementia are associated with an increased level of intracellular $[Ca^{2+}]_i$ (Landfield 1987). Supporting this theory is the observation that the $I_{K(Ca^{2+})}$ increases in the hippocampus of old rats (Landfield and Pitler 1984). However, the difficulty of fitting such acute effects into a gradual process such as aging speaks against this model. By directly measuring $[Ca^{2+}]_i$ in single T lymphocytes, we found that the Ca^{2+} signals in the cells of older patients did not differ significantly from – and indeed were less than those – of younger patients (Sulger et al., in press). In T lymphocytes of patients with Alzheimer's disease, we also demonstrated "above normal" Ca^{2+} signals in normal resting phase concentrations. These results cannot support the original thesis and instead suggest a decrease in Ca^{2+}-binding proteins such as calbindin in dementia.

The introduction of antisense technology in neurobiology has recently brought better understanding of the effects of $[Ca^{2+}]_i$-regulating kinases.

On turning off the *fyn* tyrosine kinase, a "syndrome" of effects occur which obviously have an intimate causal relationship with one another (Grant et al. 1992); spatial memory in the water maze becomes considerably worse; the LTP effect is electrophysiologically weakened; and in anatomical studies of the hippocampus, which is considered to be responsible for spatial memory, there is a disintegration of granula cells and of pyramid cells. The *fyn* tyrosine kinase appears to attack at several points in the memory process, as this is essential for the development of kindling (Cain et al. 1995). Similarly, disturbances of *CaMKII* lead to damage in spatial but not contextual memory (Bach et al. 1995).

Olney and Farber (1995) postulate that subfunctioning of the NMDA receptor is a central mechanism which can explain the pathophysiological and clinical aspects of schizophrenia. This proposal is based upon the clinical effect of the substances such as phencyclidine and ketamine which block the NMDA receptor-channel complex. Phencyclidine induces schizophrenia-like psychoses in normal subjects and results in relapses in stable schizophrenic patients. The anaesthetic ketamine also induces psychoses in adults, but not in the young. Olney and Farber support their hypothesis with neuropathological findings which demonstrate that NMDA antagonists result in neurodegenerative changes in corticolimbic regions.

Glutamate functions via the NMDA receptors on the GABA and noradrenergic nerve cells and via a tonic inhibitory control of multiple stimulating afferents to the posterior cingulum and to the retrosplenial cortex (PC/RS cortex). The PC/RS neurons regulate their activity through recurrent collaterals, the inhibitory influence of which is in turn determined by the effect of glutamate on the GABA neurons. In such a network, NMDA hypofunctioning leads to a loss of inhibitory controls in the region of the PC/RS cortex. Consequently, a massive disruption of the second messenger systems occurs. These are subsequently activated and lead to severe neuropsychological disturbances. This model represents a fascinating connection between physiological, pharmacological, neuroanatomical and clinical findings of the cell, which integrate many observations and findings of research in schizophrenia.

3.3 Genetics of Psychiatric Illnesses

The discovery of enhancer proteins – and thereby of the function-dependent expression of genes that were previously not transcriptable – demonstrated many new possibilities for interactions between function and the genome. Long-term application of anti-depressants leads to an increased expression of CREB in the hippocampus of the rat (Nibuya et al. 1996). This effect is specific to anti-depressants. We are sure to see very interesting results from these pharmacological measures in the near future.

To date, there are no reports which describe a psychiatric disturbance as a result of a gene mutation. As far as we know, the main psychiatric illnesses appear to be polygenetically determined (Propping et al. 1994). This may have several reasons: the diagnostic entities of the psychiatric disturbances, as they are defined in the DSM-III classification, may group together multiple neuropsychological subsyndromes. The latter may develop incompletely or subclinically following the appearance of pathogenetic mechanisms and only develop into the complete disturbance, as defined in the classification, following a long latency period. It must be assumed that a genetic mutation occurring in a cell will be initially compensated for by multiple mechanisms. Only when these compensatory mechanisms no longer function effectively do the symptoms become clinically apparent. Individual genetic defects that lead to striking behavioural changes in antisense experiments appear to be compensated for and therefore remain subclinical in humans for a longer period.

One way to clarify the role of mutations in psychiatric disease would be to take the evidence obtained from research, e.g. from experiments with transgenic animals and the antisense technique, and apply it in clinical practice. The absence of the serotonin ($5\text{-}HT_2$) C receptors, for example, leads to eating problems and cerebral fits (Tecott et al. 1995). Further, the loss of α-calcium calmodulin kinase II leads to an abnormal angst response and aggressive behaviour in addition to the above-mentioned memory disturbances (Chen et al. 1994). When the corresponding behavioural disturbances occur in patients, we could narrow our search to the respective genes rather than conducting large population genetic studies, as done previously.

Disturbances of the genetic codes of the ion channels, which could also be relevant in psychiatry, were recently published. These disturbances occur predominantly in the region of the Ca^{2+} and K^+ channels. Missense mutations in the region of the gene for an α_1-subunit of a Ca^{2+} channel of the P7Q type were found in patients with familial migraines and episodic ataxia (Ophoff et al. 1996). Similarly, a false genetic coding for a K^+ channel was demonstrated in patients with episodic ataxia (Browne et al. 1994). A detailed discussion is being conducted presently in the literature about the *weaver* gene, whose product blocks G protein-coupled K^+ channels (Liao et al. 1996). Apart from functional and anatomical peculiarities in the cerebellum, a specific and progressive loss of dopamine neurons has been observed. This may be of possible relevance to Parkinson's disease (Oo et al. 1996).

The ionic mechanisms can be explained as follows: in the case of a wild type, as a result of the binding of ligand, a G protein-coupled K^+ channel, which contributes to the resting phase of the cell, is opened, and a gene transcription is activated, which has an influence on the further differentiation of the cell. In the case of the mutant GIRK2, a large amino acid blocks the K^+ channel, resulting in depolarisation, excitotopic cell death and a blockade of the genetic transcription results (Goldowitz and Smeyne 1995). Interestingly, the consequences of such mutations are not constantly observed, but only occur episodically; this is of particular relevance to psychiatric illnesses. Such

evidence could promote the search for the subtypes of otherwise genetically but not clinically obvious depression. Absence seizures are similarly determined by a mutation in the potential dependent Ca^{2+} channel (Fletcher et al. 1996).

3.4 Cellular Targets for Pharmacological Effects

A typical example of a cellular clarification model is the synaptic model for the effect of anti-depressants. At the time of the production of these medications, the actual function of synapses was not really known. This led to a one-sided overinterpretation of the re-uptake mechanism as the principle mechanism of action of these medications. The effect of the anti-depressants can now be explained via a functioning neuroendrocrine model at the cellular level (Barden et al. 1997; Karanth et al. 1997; Modell et al. 1997). Other pharmacological mechanisms can be explained via the neuromodulatory effects of lithium.

This example clearly demonstrates the interaction between electrical and biochemical effects. In the nerve cells of the snail *Helix pomatia* – some (Burster cells, D cells) of which regulate their activation pattern predominantly via $I_{K(Ca^{2+})}$ – a disintegration of the phasic activation pattern is observed under lithium treatment. The measurement of $[Ca^{2+}]_i$ demonstrates that lithium clearly slows down the intracellular feedback of $[Ca^{2+}]_i$ following stimulation (Aldenhoff and Lux 1985). It can be postulated that this effect must vary according to the excitability of the cell; if the excitability is low, the Ca^{2+} signal is likely to be increased, whereas in the case of a cell with high rate of discharge, the Ca^{2+} transients overlap and there is a gradual increase in $[Ca^{2+}]_i$; the latter can be demonstrated experimentally. As a strong activator of $I_{K(Ca^{2+})}$, this increase of $[Ca^{2+}]_i$ should lead to a more rapid commencement of neuronal inhibition, i.e. it should have a "dampening" effect on neuronal activity. These alterations in function were confirmed years later when the biochemical mechanism of the action of lithium on IP was demonstrated (van Calker et al. 1987).

If the physiological model is discarded, the effect of lithium can be summarised as a calcium antagonism. Such considerations have led in the past to attempts to treat affective illnesses with calcium antagonists, e.g. verapamil. Whereas the effect of this treatment was not clear in older studies (Aldenhoff et al. 1986), recent investigations with elaborate study designs clearly support this hypothesis (Soares and Mallinger 1997). Carbamazepine, which has a similar clinical effect to lithium, also appears to function via the phosphoinositol system (Biber et al. 1996) and further to reduce Ca^{2+} currents (Schirrmacher et al. 1995).

One research path which has been pursued for many years is the functional effect of second messenger systems in relation to the effects of anti-depressants (Shelton et al. 1996). This has recently come into the limelight as a result of the discovery of enhancement factors (see above).

4 Relationship Between Cell Physiology and Psychiatry: The Compensatory Potential of the Brain

Circumscribed defects in the regions of genes or receptors do not necessarily lead to observable changes in behaviour, as they can be compensated for over a long period of time. Important cell functions are often protected by many mechanisms; if one cellular or intercellular mechanism does not work, it will therefore be compensated for by another. Although this leads to the survival of the individual, it makes it more difficult to interpret the role of gene mutations from the phenotypical or clinical viewpoint. Similarly, the application of substances against theoretical interventional targets, i.e. using the antisense knockout techniques and transgenic animals, produces no obvious effect on behaviour. This is particularly true when certain defects are apparent in embryonal and early development. Under such conditions, the pressure for the application of compensatory mechanisms appears to be very high.

A reasonable alternative experiment would be to construct certain genetic defects in an inactive form, which can then be activated at the planned time of investigation. This is possible in some cases via a sensitivity for substances such as tetracyclines (Gossen et al. 1995); certain constructed genetic defects can be activated via tetracycline in adults and completely behaviourally competent animals.

On the other hand, it has to be presumed that the manifestation of a disturbance or a disease requires the dysfunctioning of multiple cellular functions. On construction of a research model, it is often overlooked that most of these models are very simple and are not multifactorial. The majority are not suitable for the application of the complex theories, such as the chaos theory, which is still very much in the early stages of clinical observation.

Even now, the fundamental cellular principles of classical psychopathology are not clearly understood, e.g. ego disturbances, affect, impulse. The relationship between global terms such as psychosis or depression and the cellular level can be best understood via the so-called pharmacological bridge (Aldenhoff 1989). It

becomes more difficult when we try to go into more concrete detail, but the interaction of neuropsychological and neuroanatomic methods may provide assistance.

Discrete new treatment strategies are being defined with newer techniques of modern neuropsychology, such as animal experiments where the respective lesions are demonstrated. This has been shown very clearly and impressively with the combination of the results of animal experiments and neuroanatomical and neuropsychological work on the clinical symptoms of schizophrenia.

Older neuromorphological findings have shown that schizophrenic patients have a higher cell density and a disturbed cell architecture in the prefrontal cortex (Beckmann 1992). In an impressive series of animal experiments (Friedman and Goldman-Rakic 1994) and magnetic resonance imaging (MRI) examinations (McCarthy et al. 1996), Goldman-Rakic's research group was able to show that the dorsolateral prefrontal cortex plays an important role in activities in which the individual has to maintain a purpose despite many intermediary steps. According to recent neuropsychological investigations, this mechanism appears to be one which may help explain the typical characteristics of schizophrenia. In schizophrenia, there are defects in working memory, namely the inability to allow behaviour to be guided by representation and to differentiate between inner representations and external stimuli. These neuropsychological deficits lead to the clinical symptoms, i.e. the inability to keep to a conscious plan and to monitor it during its progression, which can lead to disorganised speech and thoughts; the inability to maintain a purpose, which can lead to negative symptoms; and the inability to differentiate between specific external or internal experiences and associative memory, which in turn can lead to madness and hallucinations (Andreasen et al. 1997).

Thus it does seem possible to demonstrate a relationship between clinical phenomenology and the cellular substrate.

5 References

Aldenhoff JB (1989) Imbalance of neuronal excitability as a possible cause of psychic disorder. Pharmacopsychiatry 22: 227–240

*Aldenhoff JB (1997) Reflections on the psychobiology of depression. Nervenarzt 68/5: 379–389

Aldenhoff JB, Lux HD (1985) Lithium slows neuronal calcium regulation in the snail Helix pomatia. Neurosci Lett 54(1): 103–108

Aldenhoff JB, Gruol DL, Rivier J, Vale W, Siggins GR (1983a) Corticotropin releasing factor decreases postburst hyperpolarizations and excites hippocampal neurons. Science 221(4613): 875–877

Aldenhoff JB, Hofmeier G, Lux HD, Swandulla D (1983b) Stimulation of a sodium influx by cAMP in Helix neurons. Brain Res 276(2): 289–296

Aldenhoff JB, Schlegel S, Heuser I, Wetzel H (1986) Antimanic effects of the calcium-antagonist D600. A double-blind placebo-controlled study. Clin Neuropharmacol 9[Suppl 4]: 553–555

Aldenhoff JB, Dumais-Huber C, Fritzsche M, Sulger J, Vollmayr B (1997) Altered Ca(2+)-homeostasis in single T-lymphocytes of depressed patients. J Psychiatr Res 31(3): 315–322

Amara SG, Kuhar MJ (1993) Neurotransmitter transporters: recent progress. Annu Rev Neurosci 16: 73–93

Andreasen NC, O'Leary DS, Flaum M, Nopoulos P, Watkins GL, Boles Ponto LL, Hichwa RD (1997) Hypofrontality in schizophrenia: distributed dysfunctional circuits in neuroleptic-naive patients. Lancet 349(9067): 1730–1734

Bach ME, Hawkins RD, Osman M, Kandel ER, Mayford M (1995) Impairment of spatial but not contextual memory in CaMKII mutant mice with a selective loss of hippocampal LTP in the range of the theta frequency. Cell 881: 905–915

*Barden N, Stec IS, Montkowski A, Holsboer F, Reul JM (1997) Endocrine profile and neuroendocrine challenge tests in transgenic mice expressing antisense RNA against the glucocorticoid receptor. Neuroendocrinology 66: 212–220

Beckmann H (1992) 'Temporal lobe cytoarchitectural neuropathology in schizophrenia. Clin Neuropharmacol 15[Suppl 1/A]: 493A–494A

Bennett MV (1997) Gap junctions as electrical synapses. J Neurocytol 26(6): 349–366

**Berridge MJ (1993) Inositol trisphosphate and calcium signalling. Nature 361(6410): 315–325

Berridge MJ (1997) Elementary and global aspects of calcium signalling. Physiol Lond 499: 290–306

Biber K, Walden J, Gebicke Harter P, Berger M, van Calker D (1996) Carbamazepine inhibits the potentiation by adenosine analogues of agonist induced inositolphosphate formation in hippocampal astrocyte cultures. Biol Psychiatry 40(7): 563–567

Bloom FE (1984) The functional significance of neurotransmitter diversity. Am J Physiol 246(3/1): C184–C194

Bonhoeffer T, Staiger V, Aertsen A (1989) Synaptic plasticity in rat hippocampal slice cultures: local "Hebbian" conjunction of pre- and postsynaptic stimulation leads to distributed synaptic enhancement. Proc Natl Acad Sci USA 86(20): 8113–8117

Bourtchuladze R, Frenguelli B, Blendy J, Cioffi D, Schutz G, Silva AJ (1994) Deficient long-term memory in mice with a targeted mutation of the cAMP-responsive element-binding protein. Cell 79(1): 59–68

Browne DL, Gancher ST, Nutt JG, Brunt ER, Smith EA, Kramer P, Litt M (1994) Episodic ataxia/myokymia syndrome is associated with point mutations in the human potassium channel gene, KCNA1. Nat Genet 8(2): 136–140

Cain DP, Grant SG, Saucier D, Hargreaves EL, Kandel ER (1995) Fyn tyrosine kinase is required for normal amygdala kindling. Epilepsy Res 22(2): 107–114

Catterall WA (1995) Structure and function of voltage-gated ion channels. Annu Rev Biochem 64: 493–531

*Catterall WA (1996a) Ion channels in plasma membrane signal transduction. J Bioenerg Biomembr 28(3): 217–218

*Catterall WA (1996b) Molecular properties of sodium and calcium channels. J Bioenerg Biomembr 28(3): 219–230

Chen C, Rainnie DG, Greene RW, Tonegawa S (1994) Abnormal fear response and aggressive behavior in mutant mice deficient for alpha-calcium-calmodulin kinase II. Science 266(5183): 291–294

Cohen C, Vollmayr B, Aldenhoff JB (1996) K^+ currents of human T-lymphocytes are unaffected by Alzheimer's disease and amyloid beta protein. Neurosci Lett 202(3): 177–180

**Cooper JR, Bloom FE, Roth RH (1996) The biochemical basis of neuropharmacology. Oxford University Press, Oxford

de Kloet ER, Reul JM, Sutanto W (1990) Corticosteroids and the brain. J Steroid Biochem Mol Biol 37(3): 387–394

Dolin S, Little H, Hudspith M, Pagonis C, Littleton J (1987) Increased dihydropyridine-sensitive calcium channels in rat brain may underlie ethanol physical dependence. Neuropharmacology 26(2–3): 275–279

Dolmetsch RE, Lewis RS (1997) Signaling between intracellular Ca^{2+} stores and depletion-activated Ca^{2+} channels generates $[Ca^{2+}]i$ oscillations in T lymphocytes. J Gen Physiol 103(3): 365–388

Fletcher CF, Lutz CM, O'Sullivan TN, Shaughnessy JD Jr, Hawkes R, Frankel WN, Copeland NG, Jenkins NA (1996) Absence epilepsy in tottering mutant mice is associated with calcium channel defects. Cell 87(4): 607–617

*Foote SL, Bloom FE, Aston-Jones G (1983) Nucleus locus ceruleus: new evidence of anatomical and physiological specificity. Physiol Rev 63(3): 844–914

Friedman HR, Goldman-Rakic PS (1994) Coactivation of prefrontal cortex and inferior parietal cortex in working memory tasks revealed by 2DG functional mapping in the rhesus monkey. J Neurosci 14(5/1): 2775–2788

Gattaz WF, Cairns NJ, Levy R, Forstl H, Braus DF, Maras A (1996) Decreased phospholipase A2 activity in the brain and in platelets of patients with Alzheimer's disease. Eur Arch Psychiatry Clin Neurosci 246(3): 129–131

Goldowitz D, Smeyne RJ (1995) Tune into the weaver channel. Nat Genet 11(2): 107–109

*Gossen M, Freundlieb S, Bender G, Müller G, Hillen W, Bujard H (1995) Transcriptional activation by tetracyclines in mammalian cells. Science 268: 1766–1769

Grant SG, O'Dell TJ, Karl KA, Stein PL, Soriano P, Kandel ER (1992) Impaired long-term potentiation, spatial learning, and hippocampal development in fyn mutant mice. Science 258: 1903–1910

Hille B (1992) Ionic channels of excitable membranes. Sinauer, Sunderland, MA

Hodgkin AD (1964) The conduction of the nervous impulse. Thomas, Springfield

**Holsboer F (1989) Psychiatric implications of altered limbic-hypothalamic-pituitary-adrenocortical activity. Eur Arch Psychiatry Neurol Sci 238(5–6): 302–322

Holscher C (1997) Nitric oxide, the enigmatic neuronal messenger: its role in synaptic plasticity. Trends Neurosci 20(7): 298–303

Jacobs JM, Meyer T (1997) Control of action potential-induced Ca^{2+} signaling in the soma of hippocampal neurons by Ca^{2+} release from intracellular stores. J Neurosci 17: 4129–4135

**Kandel ER, Schwartz JH, Jessell TM (1995) Essentials of neural science and behavior. Appleton and Lange

Karanth S, Linthorst AC, Stalla GK, Barden N, Holsboer F, Reul JM (1997) Hypothalamic-pituitary-adrenocortical axis changes in a transgenic mouse with impaired glucocorticoid receptor function. Endocrinology 138: 3476–3485

Koob GF, Swerdlow N, Seeligson M, Eaves M, Sutton R, Rivier J, Vale W (1984) Effects of alpha-flupenthixol and naloxone on CRF-induced locomotor activation. Neuroendocrinology 39(5): 459–464

Korkotian E, Segal M (1996) Lasting effects of glutamate on nuclear calcium concentration in cultured rat hippocampal neurons: regulation by calcium stores. J Physiol (Lond) 496(1): 39–48

Kretsinger RH (1979) The informational role of calcium in the cytosol. Adv Cyclic Nucleotide Res 11: 1–2610

*Landfield PW (1987) 'Increased calcium current' hypothesis of brain aging. Neurobiol Aging 8: 346–347

Landfield PW, Pitler TA (1984) Prolonged Ca^{2+}-dependent after hyperpolarizations in hippocampal neurons of aged rats. Science 226(4678): 1089–1092

Lefkowitz RJ, Caron MG (1988) Adrenergic receptors. Models for the study of receptors coupled to guanine nucleotide regulatory proteins. J Biol Chem 263(11): 4993–4996

Lewis C, McEwen BS, Frankfurt M (1995) Estrogen-induction of dendritic spines in ventromedial hypothalamus and hippocampus: effects of neonatal aromatase blockade and adult GDX. Brain Res Dev Brain Res 87(1): 91–95

Liao YJ, Jan YN, Jan LY (1996) Heteromultimerization of G-proteingated inwardly rectifying K^+ channel proteins GIRK1 and GIRK2 and their altered expression in weaver brain. Neuroscience 16(22): 7137–7150

**Llinas RR (1988) The intrinsic electrophysiological properties of mammalian neurons: insights into central nervous system function. Science 242(4886): 1654–1664

McCarthy G, Puce A, Constable RT, Krystal JH, Gore JC, Goldman-Rakic P (1996) Activation of human prefrontal cortex during spatial and nonspatial working memory tasks measured by functional MRI. Cereb Cortex 6(4): 600–611

*Modell S, Yassouridis A, Huber J, Holsboer F (1997) Corticosteroid receptor function is decreased in depressed patients. Neuroendocrinology 65: 216–222

Montkowski A, Holsboer F (1996) Transgene Tiermodelle in der psychiatrischen Forschung. Neuroforum 1: 25–32

*Montkowski A, Barden N, Wotjak C, Stec I, Ganster J, Meaney M, Engelmann M, Reul JM, Landgraf R, Holsboer F (1995) Long-term antidepressant treatment reduces behavioural deficits in transgenic mice with impaired glucocorticoid receptor function. J Neuroendocrinol 7(11): 841–845

Murphy DD, Segal M (1997) Morphological plasticity of dendritic spines in central neurons is mediated by activation of cAMP response element binding protein. Proc Natl Acad Sci USA 94(4): 1482–1487

Nathanson NM, Harden TK (1990) G-proteins and signal transduction. Rockefeller University Press, New York (Societies of General Physiologists Series, no. 45)

Neher E (1992) Ion channels for communication between and within cells. Science 256(5056): 498–502

*Neher E, Sakmann B (1992) The patch clamp technique. Sci Am 266(3): 28–35

Nestler EJ, McMahon A, Sabban EL, Tallman JF, Duman RS (1990) Chronic antidepressant administration decreases the expression of tyrosine hydroxylase in the rat locus coeruleus. Proc Natl Acad Sci USA 87(19): 7522–7526

*Nibuya M, Nestler EJ, Duman RS (1996) Chronic antidepressant administration increases the expression of cAMP response

element binding protein (CREB) in rat hippocampus. J Neurosci 16(7): 2365–2372

Nicoll RA, Kauer JA, Malenka RC, Nicoll RA et al (1988) The current excitement in long-term potentiation. Neuron 1(2): 97–103

*Nieuwenhuys R (1987) Chemoarchitecture of the brain. Springer, Berlin Heidelberg New York

Olney JW, Farber NB (1995) Glutamate receptor dysfunction and schizophrenia. Arch Gen Psychiatry 52(12): 998–1007

Oo TF, Blazeski R, Harrison SM, Henchcliffe C, Mason CA, Roffler-Tarlov SK, Burke RE (1996) Neuron death in the substantia nigra of weaver mouse occurs late in development and is not apoptotic. J Neurosci 16(19): 6134–6145

Ophoff RA, Terwindt GM, Vergouwe MN, van Eijk R, Oefner PJ, Hoffman SM, Lamerdin JE, Mohrenweiser HW, Bulman DE, Ferrari M, Haan J, Lindhout D, van Ommen GJ, Hofker MH, Ferrari MD, Frants RR (1996) Familial hemiplegic migraine and episodic ataxia type-2 are caused by mutations in the Ca2+ channel gene CACNL1A4. Cell 87(3): 543–552

Post RM, Rubinow DR, Ballenger JC (1986) Conditioning and sensitisation in the longitudinal course of affective illness. Br J Psychiatry 149: 191–201

*Propping P, Nothen MM, Korner J, Rietschel M, Maier W (1994) Genetic association in psychiatric diseases. Concepts and findings. Nervenarzt 65(11): 725–740

Reyes M, Stanton PK (1996) Induction of hippocampal long-term depression requires release of Ca^{2+} from separate presynaptic and postsynaptic intracellular stores. J Neurosci 16: 5951–5960

Rotenberg A, Mayford M, Hawkins RD, Kandel ER, Muller RU (1996) Mice expressing activated CaMKII lack low frequency LTP and do not form stable place cells in the CA1 region of the hippocampus. Cell 87(7): 1351–1361

Rusakov DA, Stewart MG, Korogod SM (1996) Branching of active dendritic spines as a mechanism for controlling synaptic efficacy. Neuroscience 75(1): 315–323

Schirrmacher K, Mayer A, Walden J, Dusing R, Bingmann D (1995) Effects of carbamazepine on membrane properties of rat sensory spinal ganglion cells in vitro. Eur Neuropsychopharmacol 5(4): 501–507

Schneggenburger R, Zhou Z, Konnerth A, Neher E (1993) Fractional contribution of calcium to the cation current through glutamate receptor channels. Neuron 11(1): 133–143

Shelton RC, Mainer DH, Sulser F (1996) cAMP-dependent protein kinase activity in major depression. Am J Psychiatry 153(8): 1037–1042

Soares JC, Mallinger AG (1997) Intracellular phosphatidylinosital pathway abnormalities in bipolar disorder patients. Psychopharmcol Bull 33: 685–691

Spudich JA (ed) (1993) Molecular genetic approaches to protein structure and function: applications to cell and developmental biology. Liss, New York

Stevens CF (1987) Molecular neurobiology. Channel families in the brain. Nature 328(6127): 198–199

Strachan T, Read AP (1996) Human molecular genetics. BIOS

Sulger J, Dumais-Huber C, Zerfass R, Henn FA, Aldenhoff J (1999) The calcium response of human T-lymphocytes is decreased in aging but increased in Alzheimer's dementia. Biol Psychiatry 45: 737–742

Tecott LH, Sun LM, Akana SF, Strack AM, Lowenstein DH, Dallman MF, Julius D (1995) Eating disorder and epilepsy in mice lacking 5-HT2c serotonin receptors. Nature 374(6522): 542–546

Tully T (1997) Regulation of gene expression and its role in long-term memory and synaptic plasticity. Proc Natl Acad Sci USA 94: 4239–4241

Valentino RJ, Foote SL, Aston-Jones G (1983) Corticotropin-releasing factor activates noradrenergic neurons of the locus coeruleus. Brain Res 270(2): 363–367

van Calker D, Forstner U, Bohus M, Gebicke-Harter P, Hecht H, Wark HJ, Berger M (1993) Increased sensitivity to agonist stimulation of the Ca^{2+} response in neutrophils of manic-depressive patients: effect of lithium therapy. Neuropsychobiology 27(3): 180–183

Vidal C, Jordan W, Zieglgansberger W (1986) Corticosterone reduces the excitability of hippocampal pyramidal cells in vitro. Brain Res 383(1–2): 54–59

Wang YY, Aghajanian GK (1987) Excitation of locus coeruleus neurons by an adenosine 3′,5′-cyclicmonophosphate-activated inward current: extracellular and intracellular studies in rat brain slices. Synapse 1(5): 481–487

Yin JC, Del Vecchio M, Zhou H, Tully T (1995) CREB as a memory modulator: induced expression of a dCREB2 activator isoform enhances long-term memory in Drosophila. Cell 81(1): 107–115

CHAPTER

8

I. Heuser

Psychoneuroendocrinology

Translator: W. VanSyckel

1 Historical Overview

References to biological disturbances as the basis of specific psychiatric disorders can be found in works from as early as the *Corpus Hippocratum* dating back to the fifth century B.C. Within this context and in keeping with humoral theory, psychological disorders were also ascribed to an imbalance (dyscrasia) of black and yellow bile, mucus and blood, with melancholy being attributed to a predominance of black bile, pointing, once again, to a humoral interpretation of psychiatric abnormalities. In contrast to this scholastic approach to medicine, Paracelsus (1491–1541), who also laid the foundations for the development of iatrochemistry (which gained particular importance during the seventeenth century), saw the human body as a chemical system whose ailments could therefore also be cured using chemical compounds. His findings concerning the relationship between struma and cretinism first pointed to endocrinological changes in association with psychiatric disorders.

The term "endocrinological psychiatry" first came in use in connection with a conference of psychiatrists in Dijon in 1908, during which Laignel-Lavastine encouraged his audience to intensify research efforts exploring the interaction between personality and endocrine systems. However, it was the work of Manfred Bleuler which first provided a detailed account of the relationship between endocrine disturbances (e.g. hypothyroidism, acromegalia, Cushing's syndrome, and Addison's disease) and psychiatric disorders or "defects." Based on his observation that the onset of psychoses often coincided with phases of endocrinological change (pregnancy or postpartum), Bleuler (1948) posted the hypothesis that "psychiatric and endocrine control mechanisms not only influence each other but are also highly integrated." He further noted that all endocrine diseases, even those with a rather mild and nonacute course, are generally accompanied by disturbances in mood, vitality, and motivation (Bleuler 1948).

2 The Brain: Endocrine Target Organ and Endocrine Gland

Present theories assume a model of bidirectionality involving the brain and endocrine systems; a change in peripheral hormone activity may not only be the result of disturbances in the central regulatory function, but may also be the cause of a change in brain function, thus modifying behavior. During the last few decades, modern neurobiological methods and biotechnology have broadened this conceptual idea in as much as the brain is no longer simply seen as an endocrine target organ, but also as an endocrine gland; in addition to housing the receptors of so-called peripheral hormones (e.g. somatostatin, vasoactive intestinal polypeptide, cholecystokinin), the brain itself is also the site at which various hormones such as prolactin are synthesized, governed by the so-called classical neurotransmitters, e.g. dopamine, or other central peptides.

Thus a vast number of neurons in the hypothalamus synthesize and secrete peptides, while individual neurons also express more than one neuropeptidergic hormone. While some of these neurons release their peptides into the synaptic cleft, where they act as neurotransmitters, others deliver neuropeptidergic hormones directly into the circulatory system, where they either act as true hormones on a number of tissues and/or as neuromodulators on more remote cells.

Within this context it is important to understand that hormones serving as neurotransmitters also have behavioral effects. Thus thyrotropin-releasing hormone (TRH), for example, causes increased motoric activity; corticotropin-releasing hormone (CRH) lowers food intake and decreases libido and slow-wave sleep; and cholecystokinin (CCK) is known to exert anxiogenic effects. The aforementioned are only a few examples illustrating the psychotropic effects of different hormones in order to underscore the importance that neuroendocrinology holds for psychiatry.

These findings have also resulted in the development of therapeutic strategies. Observations of the anxiolytic effects of CCK have, for example, led to the development of a neurokinin-3 receptor antagonist presently undergoing clinical studies. This is based on the assumption that mesolimbic dopaminergic neurons are involved in the development of anxiety, as are noradrenergic neurons located in the locus ceruleus. Administration of this functional tachykinin antagonist is aimed at inhibiting these neuron clusters, thereby reducing the level of anxiety.

According to their mechanisms of action, hormones may roughly be divided into two large groups:

1. Aminergic and peptide hormones act through receptors located on the cell surface: ligand binding to the receptor causes a change in the permeability of ion channels or the modulation of second-messenger systems (e.g. adenylcyclase, phosphorylase).
2. Steroid and thyroid hormones have an intracellular effect as they readily pass through cell membranes due to their lipophilic properties. Binding of these

hormones to their respective receptor proteins within the cytoplasm causes a conformational change in the receptor protein. This allows the hormone receptor complex to bind to specific regions, so-called hormone responsive elements, within the promotor region, thereby either stimulating or inhibiting the production of the gene products. However, only recently a second, so-called nongenomic, rapid steroid effect has been described; this effect is mediated by steroid receptors located on the cell surface (Wehling et al. 1998).

As a characteristic feature, hormonal secretion follows a specific temporal pattern. Accordingly, many hormones, particularly those of the pituitary gland such as luteinizing hormone (LH) and follicle-stimulating hormone (FSH), are released in a pulsatile manner. This basal hormone release follows a rhythm which can either be ultradian (shorter than 1 day), circadian (a period of approximately 24 h), or infradian (a period of longer than 24 h). In determining the exact rhythm, the suprachiasmatic nucleus plays an important role as pacemaker.

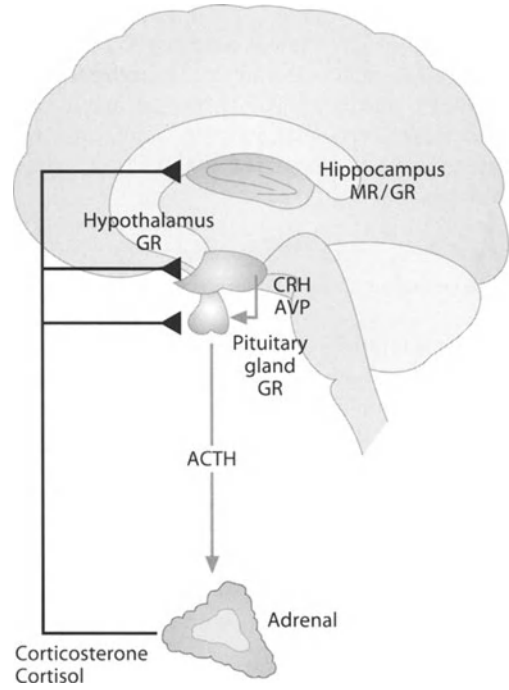


Fig. 1. Hypothalamic–pituitary–adrenal system. *MR*, mineralocorticoid receptor; *GR*, glucocorticoid receptor; *CRH*, corticotropin-releasing hormone; *AVP*, arginine–vasopressin; *ACTH*, adrenocorticotropin-releasing hormone

3
Endocrine Systems and Psychiatric Disorders

Kraepelin was the first to note that the onset of the first clinically relevant episode of a psychiatric illness can be triggered by adverse life events. Early research from the 1960s and 1970s was then able to show that depressed patients display similar endocrine changes as mentally healthy individuals subjected to stress. These findings proved to be tremendously stimulating for neurobiological research. Using depression as an example, the following will provide a more detailed illustration of this fact.

3.1
The Hypothalamic–Pituitary–Adrenal System

The hypothalamic–pituitary–adrenal system is the most significant system of hormonal stress regulation (Fig. 1). It is governed by a complex regulatory mechanism which is equally influenced by both peripheral factors and aspects of the central nervous system (CNS). Within this mechanism, the hypothalamus with its afferent and efferent connections represents a central regulatory structure. While receiving neuronal afferents from the hippocampus, the amygdala and septum, the cortex, thalamus, reticular formation, and autonomous centripedal nerve fibers of the spinal cord, it also receives direct input from the eyes via retinohypothalamic connections.

Neuroendocrine changes are aspects of a stress reaction aimed at enabling the organism to adapt to stressful situations; thus activation of sympathetic portions of the CNS leads to heightened vigilance and attention, while specific vegetative functions such as libido, hunger, and need for sleep decrease ("flight response").

3.1.1 The Stress Cascade

The transduction of "psychological" stress into "physical" events (reactions) takes place in the hypothalamus. Whenever the organism is exposed to stressors, this results in the increased hypothalamic release of CRH and also, in specific cases, arginine–vasopressin. Both neuropeptides are synthesized and stored in the paraventricular nucleus of the hypothalamus and augment each other's effect on the pituitary release of adrenocorticotropin hormone (ACTH) and β-lipocortin from their common precursor pro-opiomelanocortin (POMC). Independent of this mechanism, epinephrine and norepinephrine, as well as other peptides (e.g. angiotensin II), also stimulate the secretion of ACTH from the pituitary gland, while the atrial natriuretic peptide (ANP) inhibits ACTH release (Fink et al. 1991).

As a tropic hormone, ACTH increases both the biosynthesis and release of glucocorticoids from the adrenal cortex, while catecholamines are released from the adrenal medulla. The hypothalamic–pituitary–adrenal system regulates its own activity by way of negative feedback loops at the adrenal, pituitary, hypothalamic, and hippocampal level.

3.1.2 Hippocampal Glucocorticoid Receptors

The hippocampus is thought to be of particular importance for both the inhibition and activation of the hypothalamic–pituitary–adrenal system. Two different receptors, the mineralocorticoid receptor and the glucocorticoid receptor, are responsible for binding glucocorticoids in the hippocampus. Although they have different functions, both receptors are involved in the homeostatic control of the hypothalamic–pituitary–adrenal system (Funder 1994).

Through their research on agonists and antagonists, the research group of de Kloet was able to demonstrate that, in rats, the mineralocorticoid receptors in the brain are almost exclusively expressed within the hippocampus, while glucocorticoid receptors are found throughout the entire brain. Later studies then described the hippocampal mineralocorticoid receptor as having a tenfold greater affinity for glucocorticoids than the central glucocorticoid receptor with generally more than 90% of its binding sites occupied by glucocorticoids. The glucocorticoid receptor, on the other hand, is only occupied by its ligand as the result of increasing glucocorticoid concentrations, e.g. in the course of diurnal fluctuations or following stress (de Kloet 1991).

In summary, this indicates that, compared with mineralocorticoid receptors, glucocorticoid receptors possess a significantly lower ligand affinity, which, as already mentioned, acts in the hippocampus as a receptor with a clear preference for glucocorticoids. By means of animal experiments, it has been shown that mineralocorticoid receptors are mainly responsible for the regulation of basal rhythms such as sleep–wake cycles and food intake, while glucocorticoid receptors appear to play a role in terminating stress-responsive glucocorticoid secretion (McEwen 1998). As a result of repeated severe stress or chronic glucocorticoid treatment, hippocampal glucocorticoid receptors are down-regulated.

Furthermore, it has been demonstrated that hypersecretion of glucocorticoids in the presence of other noxious stimuli, such as hypoxemia, may lead to cell death within the CA_3 region of the hippocampus. This observation led Sapolsky to formulate the so-called cascade theory. Heightened glucocorticoid exposition of the brain may result in the disturbed arborization of dendrites, inhibition of glucose-3 transporters (which are mainly expressed in hippocampal neurons), and increased excitatory amino acid function (Sapolsky 1996). Accordingly, the impact of glucocorticoids may either cause increased vulnerability, temporary damage, or even neuronal death.

3.1.3 Clinical Results of Neuroendocrine Research on Depression

Patients with hypercortisolemia are known to frequently suffer from depressive syndromes (Wybrow and Horwitz 1976). In a follow-up study on depressed patients, Gibbons (1964) was also able to demonstrate that cortisol concentrations in remitted patients were significantly lower than during preliminary examinations, i.e during their acute episode. Recent findings were able to confirm this observation (Deuschle et al. 1997). Furthermore, patients suffering from depression typically display a number of symptoms pointing to hypothalamic malfunctioning, such as sleep disorders, loss of appetite and libido, circadian abnormalities, and autonomous dysregulation.

In 50%–70% of all depressed patients, it was noted that, in contrast to healthy controls, the administration of small doses of the synthetic glucocorticoid dexamethasone did not lead to a decrease in cortisol concentrations (dexamethasone suppression test; Carroll et al. 1981). Together with the findings of increased basal cortisol levels in depressed patients, this points to an apparent dysregulation in the hypothalamic–pituitary–adrenal system in these patients.

In endocrinology, the CRH test has come to provide a reliable differential diagnosis for Cushing's syndrome (for an overview, see Müller 1987). In addition, the CRH test has been carried out with patients suffering from depression. Within this context, it was demonstrated that hypercortisolemic patients exhibit a reduced ACTH response, while their cortisol response follows the same course as in healthy controls. This reduction in ACTH response has been termed ACTH blunting (Holsboer et al. 1984, 1987; Gold et al. 1986).

Following pretreatment with metyrapone, a 11-β-steroid hydroxylase inhibitor which inhibits the biosynthesis of cortisol from its precursor, followed by stimulation with CRH, depressed patients show both a normalized ACTH response and a normal cortisol response. These findings seem to suggest that the increased plasma cortisol concentrations found in depressed patients have a particular braking effect on CRH-induced ACTH release, i.e the cortisol feedback loop between the adrenal cortex and pituitary gland in depressed patients appears to be intact.

In order to further investigate this inverse relationship between basal plasma cortisol concentrations and

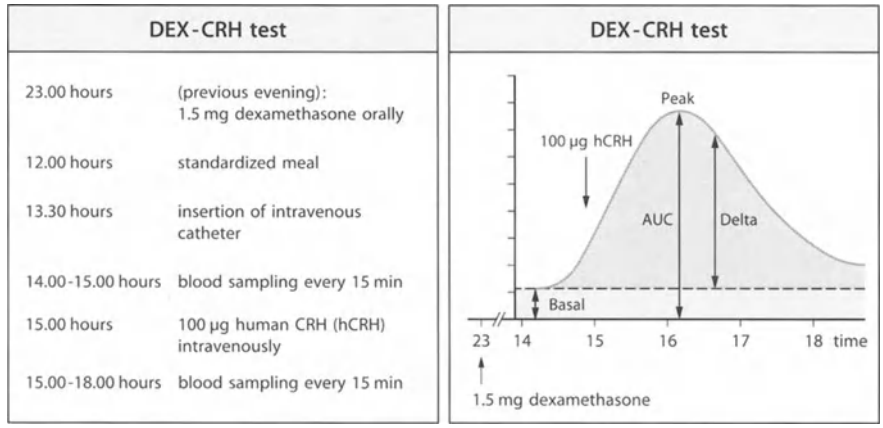


Fig. 2. Dexamethasone suppression/corticotropin-releasing hormone (CRH) stimulation test. *AUC*, area under the curve

CRH-induced ACTH release, hypercortisolemic patients suffering from depression and healthy normocortisolemic controls were both given dexamethasone (at 23:00 on day 1), followed by CRH stimulation 16 h later (at 15:00 the following day), a procedure known as the combined dexamethasone suppression/CRH stimulation test.

It had been assumed that in depressed patients the increased overall cortisol level (endogenous and exogenous) would result in significantly diminished ACTH stimulation through CRH. However, the complete opposite was true: in hypercortisolemic, depressed patients, the additional dose of dexamethasone and subsequent CRH injection led to both an excessive ACTH and an excessive cortisol response. In healthy controls, however, a decrease in both the cortisol and the ACTH response was noted, as expected, with increasing dexamethasone doses (Heuser et al. 1994; Fig. 2).

To date an explanation for this apparently paradox phenomenon is still lacking, particularly in light of the fact that "standard" CRH testing (without prior dexamethasone treatment) in depressed patients seems to point to regular functioning of pituitary–glucocorticoid receptor-mediated feedback. To explain this phenomenon of the dexamethasone suppression/CRH stimulation test, we might either look to an increase in the number of CRH receptors, and/or a decrease in glucocorticoid receptors, and/or a decrease in sensitivity on the different levels of the stress-responsive hypothalamus–pituitary–adrenal systems.

3.1.4 The Corticotropin-Releasing Hormone-Overdrive Hypothesis of Depression

Findings in depressed patients point to suprapituitary changes as the cause of the hypercortisolemia found in these patients. In all likelihood, increased hypothalamic release of CRH plays a major role but is not the only factor involved, as changes in glucocorticoid receptors with associated feedback disturbances also have to be taken into consideration. In order to test this so-called CRH-overdrive hypothesis of depression-associated hypercortisolemia, a number of studies have recently been performed (Heuser 1998; Heuser et al. 1998). Most but not all of these cerebrospinal fluid (CSF) studies showed that, compared to healthy controls, depressed patients have significantly increased cerebrospinal CRH concentrations, which return to normal levels following successful antidepressive treatment. Within this context, it was further shown that patients in remittance whose test results continued to be pathological in either the dexamethasone suppression or the dexamethasone suppression/CRH stimulation test are at an increased risk for suffering a relapse and yet another depressive episode (Zobel et al. 1999).

In summary, it can be noted that both increased activity and abnormal regulation of the hypothalamic–pituitary–adrenal system represent a frequent endocrine symptom of depression. It is assumed that the noted increase in hypothalamic–pituitary–adrenal system activity associated with depression may, among other factors, be caused by increased hypothalamic release of CRH, since a number of studies were able to illustrate increased CRH concentrations in the CSF of depressed patients, reduced CRH receptor density in the frontal cortex of suicide victims, and the activation of CRH neurons in the hypothalamus of depressed patients (Nemeroff et al. 1984, 1988).

Two distinct CRH receptors, CRH-1 and CRH-2, have been identified in rodents. CRH-1 is largely found within the adenohypophysis, neocortex, hippocampus, amygdala, and cerebellum. Recent findings have been able to demonstrate that transgenic mice who do not

express this CRH-1 receptor not only show an atrophy of the adrenal cortex, but also a diminished stress-induced release of ACTH and corticosterone. Furthermore, these mice also displayed less "anxiety" and a reduced "stress response" (Timpl et al. 1998).

Based on such animal experiments, a number of laboratories are presently developing human CRH receptor antagonists as antidepressant and/or anxiolytics for the treatment of affective disorders. However, other neuropeptide hormones, such as substance P, which was recently shown to have an antidepressive effect similar to that of paroxetin, also appear to be promising in this context (Kramer et al. 1998).

Thus far, this chapter has provided a detailed account of the hypothalamic–pituitary–adrenal system with all its connections and effects during stress and depression. This account also provided an exemplary illustration of the "cross-fertilization" between clinical observations and preclinical animal experiments within the field of psychoneuroendocrinological research and the resulting developments for new therapeutic approaches. In the following, we will outline the somatotropic system (hypothalamic–pituitary–growth hormone system) and the hypothalamic–pituitary–gonadal system.

3.2 Hypothalamic–Pituitary–Growth Hormone System

During the course of a lifetime, the secretion of growth-stimulating hormones differs depending on age. Low secretion rates after birth are followed by increased rates during childhood growth, which then decrease with advancing age. Regulation of the somatotropic effect occurs by way of peripherally active insulin-like growth factors (IGF), which, as the name indicates, exert an insulin-like effect and are involved in protein synthesis, lipolysis, and carbohydrate metabolism.

IGF represent a structurally related family of growth factors. IGF-1 and IGF-2 are responsible for transforming the growth-stimulating effect of the growth hormone on the peripheral reacting organs. In addition to these circulating IGF with an endocrine action, their local production in different tissues, such as the bones, is also important for an autocrine or paracrine effect.

The formation and secretion of IGF-1 in the liver, the main site of its synthesis, is stimulated by the secretion of growth hormone from the pituitary. The release of GH again is stimulated by the hypothalamic growth hormone-releasing hormone (GHRH) and inhibited by somatostatin. Along with its neurons, GHRH is located in the arcuate nucleus and is regulated by numerous transmitter systems. Special mention should be made of the stimulatory effect of norepinephrine via α_2 receptors and its inhibitory effect via β_2 receptors.

The hypothalamic concentration and the release of somatostatin are increased by growth hormones, while ACTH leads to a reduction. Somatostatin itself mainly exerts inhibitory effects on the secretion of neuropeptide hormones. It is generally assumed that somatostatin inhibits its own release through autoreceptors located on the cell body or dendrites of somatostatin-containing neurons (Epelbaum 1986). In addition to GHRH, naturally occurring molecules, such as the synthetic growth hormone-releasing peptide, an encephalin-derived hexapeptide, may function as ligands for receptors increasing growth hormone release. At this point we should also briefly mention the pituitary adenylate cyclase peptide (PACAP), which stimulates somatotropic activity and, differentially activating two distinct second-messenger systems, appears to be a hypophysiotropic factor (Spengler et al. 1993).

Depressed patients frequently display a weakened growth hormone response to GHRH (GHRH test). In addition, they have also been shown to have reduced nocturnal growth hormone secretion amplitude and frequency. It has been suggested that this might be the result of increased depression-associated CRH release, as CRH is a functional antagonist of GHRH. According to the hypothesis postulated by Ehlers and Kupfer (1987), this reciprocal interaction between GHRH and CRH plays a particularly important role in sleep regulation. In keeping with this hypothesis, Steiger et al. (1994) were able to show that in depressed patients, who typically suffer from sleep disturbances (particularly difficulty falling asleep, disrupted sleep, and waking up early), the administration of GHRH led to a marked improvement in sleep architecture. The same holds true for older healthy probands, who typically display diminished GHRH system activity in association with reduced growth hormone release and who complain about "poor sleep."

Within this context, Rudman et al. (1990) reported that older healthy probands described an increased sense of well-being following a 12-week regimen of subcutaneous growth hormone injections. At the same time, they also showed an increase in muscle tone, physical activity, and energy. These findings suggest that growth hormone substitution in elderly, otherwise healthy probands might represent a "psychotropic" strategy within the field of geriatric medicine; clinical studies on this topic are, however, yet to be carried out.

3.3 Hypothalamic–Pituitary–Gonadal System

The hypothalamic–pituitary–gonadal system regulates the production of sexual steroids and gametogenesis. The hypothalamic gonadotropin-releasing hormone

(GnRH), a decapeptide, stimulates the release of LH and FSH from the pituitary gland. GnRH neurons are located in the preoptic area of the hypothalamus, with their axons projecting mainly to the median eminence. In addition to its effect on LH and FSH secretion, GnRH also exerts behavioral effects such as intensifying an existing lordosis in rats.

The pituitary glucoprotein hormones LH and FSH are made up of α- and β-subunits, are secreted by the same cell, and bind to receptors within the ovaries and testicles. In males, LH stimulates testosterone production in the Leydig cells of the testes, while FSH stimulates testicle growth; both hormones must be present for spermatogenesis.

In females, LH stimulates ovarian estrogen and progesterone production. A rapid increase in LH in the middle of the menstrual cycle causes ovulation, with the continued influx of LH stimulating the corpus luteum to produce progesterone. While maturation of the follicle is mainly controlled by FSH, the secretion of estrogen from the follicle requires both FSH and LH. Finally, both FSH and LH are released episodically depending on the sleep–wake pattern.

Behavioral effects associated with the hormones of the hypothalamic–pituitary–gonadal system have been demonstrated within the framework of animal studies. Their findings suggest that the two gonadal steroids testosterone and estradiol play a modulating role with regard to aggression, motivation, and "emotionality" in general. Overall, however, these animal findings must be considered preliminary, and applications to humans are not yet permissible.

The effect of sexual hormones on cognitive processes, on the other hand, are well documented; here, estradiol appears to be responsible for the consistently noted female superiority in the areas of verbal fluency, verbal comprehension, and analogous thinking, while male superiority in solving visuospatial problems can be attributed to testosterone (Sherwin 1998a,b). Based on very recent epidemiological and practical animal experiments, it can be concluded that estrogens have a neuroprotective effect on the particularly sensitive neurons of the hippocampus.

It has long been a well-documented fact that chronic stress or severe acute stressors can impair the functioning of the hypothalamic–pituitary–gonadal system, a phenomenon which finds expression in the term "distress amenorrhea." As was shown, stress-induced CRH hypersecretion inhibits GnRH, thus leading to reduced pituitary LH and FSH secretion. In addition, high plasma glucocorticoid concentrations reduce ovarian and testicular sensitivity to LH, again resulting in diminished estradiol, progesterone, and testosterone secretion.

The relationship observed between the stress-induced increase in activity in the hypothalamic–pituitary–adrenal system and the associated inhibition of activity in the hypothalamic–pituitary–gonadal system has given rise to a number of studies on gonadal functions in depressed patients. The range of typical symptoms of moderate to severe depression includes loss of libido, erectile dysfunction, impotence, anovulatory menstrual cycles, and/or amenorrhea. Within this context, a recent study on severely depressed men showed that, compared with a healthy, age-matched control group, these patients have significantly lower testosterone concentrations, which negatively correlate with the level of their plasma cortisol concentration. At the same time, these men also showed signs of lowered LH pulse frequency (Schweiger et al. 1999).

4 Neurosteroids

Within the CNS, the de novo synthesized steroids are termed neurosteroids (Banlieu and Robel 1990). Based on the fact that substantial amounts of pregnenolone, dehydroepiandrosterone (DHEA), and their metabolites were found in the brain of mice, rats, pigs, and primates, Mathur et al. (1993) formulated the hypothesis that the CNS consists of steroid-producing tissue. Here, it is important to note that the presence of neurosteroids within the CNS appears to be independent of gonadal and adrenal synthesis, as they could be detected there even following adrenalectomy and gonadectomy (Corpéchot et al. 1993).

In contrast to the classical, genome-mediated steroid effects, neurosteroids interact with γ-aminobutyric acid $(GABA)_A$ and *N*-methyl-D-aspartate (NMDA) receptors located on the cell membrane. Furthermore, incubation of glial cell cultures of rats with a cholesterol precursor led to the production of cholesterin, pregnenolone, 20-OH-pregnendone, and progesterone (Jung-Testas et al. 1989). These results of animal experiments were then substantiated by postmortem studies in humans, which also showed many times higher concentrations of neurosteroids in the CNS compared to the amounts found in plasma (Lanthier and Patwardhan 1986).

Independent of their location of synthesis, glucocorticoids, mineralocorticoids, androgens, estrogens, and progestins exert their effects in the CNS by means of intracellular steroid hormone receptors, with their expression being influenced by specific genes (McEwen 1991). In addition, it has been demonstrated that neuronal functions may also be regulated by modulating gene expression, whereby intracellular progesterone receptors are known to play a crucial role (Rupprecht et al. 1993).

Neurosteroids act as allosteric agonists on $GABA_A$ receptors by increasing the frequency and duration of

chloride channel permeability and by potentiation of the inhibitory effect of GABA (Majewska 1992). Sulfated neurosteroids, e.g. pregnenolone sulfate and DHEA sulfate, act as noncompetitive antagonists on $GABA_A$ receptors and inhibit GABA-induced ion transportation by reducing the frequency of ion channel permeability (Demirgoren et al. 1991).

While benzodiazepines, barbiturates, anticonvulsants, and GABA all bind to subunits of $GABA_A$ receptors, neurosteroids in all probability bind at a different location of the GABA receptor (Costa et al. 1994). Due to their effect on GABA and NMDA receptor systems, neurosteroids might be of clinical relevance in the treatment of memory deficits, pain syndromes, and affective and panic disorders. This assumed potential of neurosteroids suggests their use in pharmacotherapy and is presently the topic of intense study efforts (Rupprecht 1997).

5 Outlook

The admittedly arbitrary insights into the field of neuroendocrinology given in this chapter were intended to show that our understanding concerning the relevance and effect of hormones has considerably increased; hormones not only act as the effectors and regulators of metabolism and are not restricted to controlling endocrinological functions in the narrower sense, but also have distinct behavioral, i.e psychotropic effects. On the other hand, the discussion of affective disorders, in particular depression, was aimed at illustrating that so-called "mental" disorders have both "somatic" causes and consequences, a fact which needs to be recognized, described, and taken into account in order for treatment to be successful.

6 References

Baulieu EE, Robel P (1990) Neurosteroids: a new brain function? J Steroid Biochem 37: 395–403

Bleuler E (1948) Untersuchungen aus dem Grenzgebiet zwischen Psychopathologie und Endokrinologie. Arch Psychiatr Nervenkr 180: 271–528

Carroll BJ, Feinberg M, Greden JF et al (1981) A specific laboratory test for the diagnosis of melancholia. Standardization, validation, and clinical utility. Arch Gen Psychiatry 38(1): 15–22

Corpéchot C, Young J, Calvel M et al (1993) Neurosteroids: 3-alpha-hydroxy-5 alpha-pregnan-20-one and its precursors in the brain, plasma, and steroidogenic glands of male and female rats. Endocrinology 133(3): 1003–1009

Costa E, Auta J, Guidotti A, Korneyev A, Romeo E (1994) The pharmacology of neurosteroidogenesis. J Steroid Biochem Mol Biol 49: 385–389

de Kloet ER (1991) Brain corticosteroid receptor balance and homeostatic control. Front Neuroendocrinol 12: 95–164

Demirgoren S, Majewska MD, Spivak CE, London ED (1991) Receptor binding and electrophysiological effects of dehydroepiandrosterone sulfate, an antagonist of the GABAA receptor. Neuroscience 45(1): 127–135

Deuschle M, Schweiger U, Weber B et al (1997) Diurnal activity and pulsatility of the hypothalamus-pituitary-adrenal system in male depressed patients and healthy controls. J Clin Endocrinol Metab 82(1): 234–238

Dorn LD et al (1997) The longitudinal course of psychopathology in Cushing's syndrome after correction of hypercorticolism. J Clin Endocrinol Metab 82/3: 912–919

Ehlers CL, Kupfer DJ (1987) Hypothalamic peptide modulation of EEG sleep in depression: a further application of the S-process hypothesis. Biol Psychiatry 22(4): 513–517

Epelbaum J (1986) Somatostatin in the central nervous system: physiology and pathological modifications. Prog Neurobiol 27: 63–100

Fink G, Dow RC, Casley D et al (1991) Atrial natriuretic peptide is a physiological inhibitor of ACTH release: evidence from immuno-neutralization in vivo. J Endocrinol 131: 9–12

Funder JW (1994) Corticosteroid receptors and the central nervous system. J Steroid Biochem Mol Biol 49: 381–384

Gibbons JL (1964) Cortisol secretion rate in depressive illness. Arch Gen Psychiatry 10: 572–575

Gold PW, Loriaux DL, Roy A et al (1986) Responses to corticotropin-releasing hormone in the hypercortisolism of depression and Cushing's disease. N Engl J Med 314: 1329–1335

Heuser I (1998) The hypothalamic–pituitary–adrenal system in depression. Pharmacopsychiatry 31: 10–13

Heuser I, Yassouridis A, Holsboer F (1994) The combined dexamethasone/CRH-test: a refined laboratory test for psychiatric disorders. J Psychiatr Res 28: 341–356

Heuser I, Bissette G, Dettling M et al (1998) Cerebrospinal fluid concentrations of corticotropin-releasing hormone, vasopressin, and somatostatin in depressed patients and healthy controls: response to amitriptyline treatment. Depress Anxiety 8(2): 71–79

Holsboer F, Doerr HG, Gerken A, Muller OA, Sippell WG (1984) Cortisol, 11-deoxycortisol, and ACTH concentrations after dexamethasone in depressed patients and healthy volunteers. Psychiatry Res 11(1): 15–23

Holsboer F, Gerken A, Stalla GK, Muller OA (1987) Blunted aldosterone and ACTH release after human CRH administration in depressed patients. Am J Psychiatry 144(2): 229–231

Jung-Testas I, Hu ZY, Baulieu EE, Robel P (1989) Neurosteroids: biosynthesis of pregnenolone and progesterone in primary cultures of rat glial cells. Endocrinology 125(4): 2083–2091

Kramer MS, Cutler N, Feighner J et al (1998) Distinct mechanism for antidepressant activity by blockade of central substance P receptors. Science 281(5383): 1640–1645

Lanthier A, Patwardhan VV (1986) Sex steroids and 5-en-3 beta-hydroxysteroids in specific regions of the human brain and cranial nerves. J Steroid Biochem 25(3): 445–449

Majewska MD (1992) Neurosteroids: endogenous bimodal modulators of the GABAA receptor. Mechanism of action and physiological significance. Prog Neurobiol 38: 379–395

Mathur C, Prasad VV, Raju VS, Welch M, Lieberman S (1993) Steroids and their conjugates in the mammalian brain. Proc Natl Acad Sci USA 90(1): 85–88

McEwen BS (1991) Non-genomic and genomic effects of steroids on neural activity. Trends Pharmacol Sci 12(4): 141–147

McEwen BS (1998) Protective and damaging effects of stress mediators. N Engl J Med 338/3: 171–179

Müller OA (ed) (1987) Corticotropin releasing hormone. Thieme, Stuttgart (Hormone metabolic research supplement series, vol 16)

Nemeroff CB, Widerlov E, Bissette G et al (1984) Elevated concentrations of CSF corticotropin-releasing factor-like immunoreactivity in depressed patients. Science 226(4680): 1342–1344

Nemeroff CB, Owens MJ, Bissette G, Andorn AC, Stanley M (1988) Reduced corticotropin releasing factor binding sites in the frontal cortex of suicide victims. Arch Gen Psychiatry 45(6): 577–579

Rudman D, Feller AG, Nagraj HS et al (1990) Effects of human growth hormone in men over 60 years old. N Engl J Med 323(1): 1–6

Rupprecht R (1997) The neuropsychological potential of neuroactive stroieds. J Psychiatr Res 31(3): 297–314

Rupprecht R, Reul JM, Trapp T et al (1993) Progesterone receptor-mediated effects of neuroactive steroids. Neuron 11(3): 523–530

Sapolsky RM (1996) Stress, glucocorticoids, and damage to the nervous system: the current state of confusion. Stress 1: 1–19

Schweiger U, Deuschle M, Weber B et al (1999) Testosterone, gonadotropin and cortisol secretion in male patients with major depression. Psychosomatic Med 61: 292–296

Sherwin BB (1998a) Use of combined estrogen-androgen preparations in the postmenopause – evidence from clinical studies. Int J Fertil Menopaus Stud 43(2): 98–103

Sherwin BB (1998b) Estrogen and cognitive functioning in women. Proc Soc Exp Biol Med 217(1): 17–22

Spengler D, Waeber C, Pantaloni C, Holsboer F, Bockaert J, Seeburg PH, Journot L (1993) Differential signal transduction by five splice variants of the PACAP receptor. Nature 365(6)442: 170–175

Steiger A, Guldner J, Colla-Muller M, Friess E, Sonntag A, Schier T (1994) Growth hormone-releasing hormone (GHRH)-induced effects on sleep EEG and nocturnal secretion of growth hormone, cortisol and ACTH in patients with major depression. J Psychiatr Res 28(3): 225–238

Timpl P, Spanagel R, Sillaber I et al (1998) Impaired stress response and reduced anxiety in mice lacking a functional corticotropin-releasing hormone receptor 1. Nat Genet 19(2): 162–166

Wehling M, Spes CH, Win N, Janson CP, Schmidt BMW, Theisen K, Christ M (1998) Rapid cardiovascular action of aldosteronein man. J Clin Endocrin Metabol 83(10): 3517–3522

Wybrow TC, Horwitz T (1976) Psychological disturbances associated with endocrine disease and hormone therapy. In: Sachar EJ (ed) Hormones, behavior and psychopathology. Raven, New York, pp 125–144

Zobel A, Yassouridis A, Frieboes R-M, Holsboer F (1999) Prediction of medium-term outcome by cortisol response to the combined dexamethasone-CRH test in patients with remitted depression. Am J Psychiatry 156: 949–951

CHAPTER 9

W. Strik

Psychiatric Neurophysiology

Translator: E. Taub

1 Relevance of Neurophysiology to Psychiatry

Electroencephalography (EEG) is a highly sensitive method of detecting alterations of brain function. It was, in fact, the first technique of biological measurement with which conscious human thought processes, such as mental calculation, could be detected. EEG became well established in clinical use shortly after its introduction, primarily as a diagnostic tool in neurology, but the early hope that it would provide a better knowledge of the physiology and pathophysiology of mental processes remained unfulfilled for many years. Only later, with the advent of computer technology, could event-related potentials (ERP), frequency spectra, and reconstructions of cerebral electric fields begin to be studied. The use of these neurophysiologic methods has yielded major insights into higher brain function, e.g. regarding the chronology of sensory perception and of decision processes. The physiologic background of these modern methods of analysis in psychiatric neurophysiology will be briefly discussed in what follows.

The cerebral cortex is organized for the processing of a very large amount of information concerning the structure of incoming sensory stimuli. In a system of sufficient complexity, it would be conceivable that a single neuron, at the end of a long chain of transformations, might represent all of the information originally contained in the pattern of activation of the receptor cells. This model, though pleasing in its simplicity, cannot be fully realized because of the unrealistically high number of "grandmother cells" that would be needed to represent the experiences of an entire lifetime. Neuroanatomical considerations and evidence from cell physiology have lead to the current assumption that the higher integrative functions of the brain, such as conscious perception, ideation, and the planning and execution of voluntary movements, depend on the simultaneous, concerted activity of spatially distributed neurons on all relevant hierarchical levels, including the primary brain areas. The flexibility and capabilities of the system derive from the high degree of organization of these patterns of spatial activation.

From this perspective, consciousness corresponds to the totality of the cortical electrochemical pattern of activation, which is subjectively experienced either as conscious perception or as ideation according to the differential weighting of its exogenous and endogenous portions. In fact, neuropsychological and neurophysiological studies imply that such a pattern of activation, in the form of a subliminally activated search pattern, can assume a filtering function by admitting only those sensory stimuli that match it, while rejecting others (Freeman 1983; Wilson and McNaughton 1993; Eichenbaum 1993).

These background considerations explain why cerebral electrical activity and its spatial configuration play such an important role in the study of higher brain function. The spatial resolution of the EEG is more than adequate for the detection of macroscopic changes of cerebral electrical activity: varying degrees of activation of different cortical areas are reflected in topographical variations of the cerebral electric field. The direct measurement of cerebral electric fields has yielded valuable insights into the characteristics and chronology of cortical activation associated with mental processes in terms of its intensity (amplitude), time point (latency, with a resolution of up to approximately 1 ms), spatial pattern (topography), and oscillating changes (frequency). This method has not yet been supplanted even by modern imaging techniques such as functional magnetic resonance imaging (fMRI) or proton emission tomography (PET), as the latter are currently directed at the measurement of slower, secondary metabolic processes.

Psychiatric disorders affect the highest, and often specifically human, functions of the brain, such as language, thinking, voluntary movement, emotion, and mood. Thus psychiatric neurophysiology, unlike its neurological counterpart, is not concerned with local, isolated manifestations of neural activity such as epileptic foci, early evoked potentials, or the excitation of individual nerve pathways. Of interest here are events associated with cognitive processes, such as ERP and spontaneous cerebral activity.

2 Measurement of Cerebral Electrical Activity

The difference in electrical potential between two electrodes is the net result of all electrical activity occurring between them. The EEG is presumably generated by the activity of large aggregates of neurons, primarily through the summation of excitatory and inhibitory postsynaptic potentials. These potentials diminish as the inverse square of the distance from their source; hence superficial activity rather than deeper activity is predominantly measured. The surrounding tissues – cerebrospinal fluid, bone, meninges, and scalp – alter the fields to some extent.

Cerebral electrical activity is measured as the potential difference between two electrodes. It is customary in clinical EEG to measure either the potential difference between scalp and reference electrodes (so-called monopolar recording) or a potential

gradient (the potential difference between neighboring electrodes, so-called bipolar recording). As both types of measurement concern the potential difference between two electrodes, the distinction between mono- and bipolar recording, like that between "active" and "inactive" electrodes, is misleading and should be avoided.

Every electric dipole generates a magnetic field that is perpendicular to its electric field. Magnetoencephalography (MEG), unlike EEG, detects only dipole components that are tangential to the cranial surface and thus primarily measures the neural activity within the cortical sulci because of the way these axons are oriented. Unlike the electric field, the magnetic field is not distorted by the surrounding tissues, but its field strength diminishes more rapidly with distance from the source (inverse-cube law). Sources deep to the surface are thus more difficult to detect. Compared to EEG, MEG has the practical advantage that a rigorous placement of scalp electrodes is unnecessary. Nonetheless, the head must be kept fixed in relation to the sensors, and MEG may therefore be difficult to perform in psychiatric patients.

By virtue of the above-described properties, MEG is best considered a complementary technique to EEG that may be used to greatest advantage in combination with it. The joint application of these two techniques in psychiatric patients has, however, not yet been systematically studied.

Cerebral magnetic fields are analyzed with essentially the same methods as cerebral electric fields. MEG has a somewhat better spatial resolution for localization of the intracerebral sources than EEG (Cohen et al. 1990; Wikswo et al. 1993).

3 Quantitative Analysis of Cerebral Electrical Activity

3.1 Electroencephalography in the Frequency Domain

In the electroencephalographic tradition, information about cerebral processes is derived from the wave patterns of the EEG or ERP. A description of wave patterns was developed, primarily by neurologists, and it still possesses clinical validity for the detection of localized neurologic processes or of severe, generalized functional disturbances. In contrast, modern analytical methods enable quantification of the EEG in the frequency domain for descriptive and statistical purposes.

The digitized signal of each EEG channel can be converted by a fast Fourier transformation (FFT) into a set of power values (in μV^2) of its individual frequency components. The frequency resolution depends on the frequency at which the incoming signal is sampled. For example, with a sampling frequency of 128 MHz, power values can be obtained at 64 frequency points spaced 0.5 Hz apart, from 0 Hz to 32 Hz. For further analysis, these power values are summarized by average values in each of the classical frequency bands – delta, theta, alpha, and beta. The results of the FFT may be given either as absolute or as relative power values. The disadvantage of frequency or spectral analysis is its low temporal resolution, as it generally requires a data series that is at least 1 s long. Power measurement in the four frequency bands of the EEG has been used to investigate differences between patient groups, the effects of psychoactive drugs, and changes in brain activity induced by sensory stimuli. It is thus a valuable tool in the study of normal and pathological information processing as well as their hemispheric lateralization (Flor-Henry and Gruzelier 1983; Koukkou-Lehmann 1987).

Coherence analysis is used to investigate the relationship of electrical signals that are recorded simultaneously in different channels. This method involves a statistical estimation of the correlation of two signals in the frequency domain and is carried out as a pairwise comparison between two different channels at each frequency point or in each frequency band. Coherence is mathematically independent of the amplitude of the signal and is, in principle, inversely proportional to the distance between the two electrodes (French and Beaumont 1984). If one assumes that an increase in coherence in the frequency domain is a manifestation of coupled neural activity, this parameter may be interpreted as a measure of the functional relationship between two areas of the brain.

Digital filtering is used to enable measurement of the electrical activity in narrow frequency bands, which may be difficult or impossible with the unfiltered EEG. The major application of this method is in the study of the EEG in the so-called gamma band, which contains frequencies in the neighborhood of 40 Hz. Cognitive stimuli have been reported to evoke activity in the gamma band, and the relationship of this gamma response to evoked potentials has been studied. A distributed gamma system of functional relevance to cognitive processes has been postulated, by analogy to the 40-Hz action potentials of certain neurons as well as for other reasons (Basar-Eroglu et al. 1996).

The frequency characteristics of the resting EEG are not stationary (Lopes Da Silva et al. 1974). Changes can be measured quantitatively with temporal resolution in the 1-s range by segmentation of the spontaneous EEG activity, followed by analysis in the frequency domain. Studies involving autocorrelational analysis of the signals of individual channels (Barlow

et al. 1981) have not yet yielded findings of relevance to psychiatry.

Working in the frequency domain, it must be borne in mind that waveforms are not unique when measured with respect to reference electrodes. If recordings are taken from N channels, there are $N \times (N-1)/2$ possible distinct waveforms – thus, for example, 21 channels generate 210 possible waveforms. Conversion to a reference-independent mode (e.g. average reference) is to be preferred in all topographical analyses.

3.2 Spatial Analysis of Cerebral Electric Fields

In the study of widely distributed aggregate neural activity with unknown source localization and, from the neurologic viewpoint, only mild functional abnormality (as generally found in psychiatric disease), the analysis of the spatial configuration of the cerebral electric field has a particular importance. Its topographical alterations allow study of the chronology of the activation of spatially distributed groups of neurons, which cannot be performed with any other currently available imaging technique.

The information contained in a multichannel EEG or ERP recording can be viewed as a spatiotemporal series of instantaneous potential distributions from which the cerebral electric field can be reconstructed. The potential distributions on the scalp are represented in a two-dimensional, map-like projection of the electrode array, including peaks (positive values) and valleys (negative values), isopotential lines, and often color coding of amplitude ("brain mapping"; Lehmann 1971). The landscape map allows a clear visualization of topographical relationships. A change of reference changes only the baseline value (offset), while the use of an average reference removes spatial offset from the map without changing the landscape in any other way. Baseline corrections are not allowed, however, as they will distort the electric fields.

Not only potential distributions, but also FFT values (see above), the potential gradient (the first derivative of the distribution, i.e. the local steepness of the field), and current source density (CSD, the second derivative of the distribution) can be represented as landscape maps. Potential gradients and current source density both have the effect of a spatial high-pass filter, in which values obtained from electrodes at the edges of the array are lost. Potential differences between widely separated electrodes are suppressed, while those between neighboring electrodes are highlighted. The resulting maps are reference independent and give prominence to superficial cortical sources. Current source density, in particular, has been widely used, and various processing algorithms have been proposed (Nunez 1989).

A prerequisite for the valid interpretation of any landscape map, whether it depicts the distribution of the EEG, ERP, potential differences, potential gradients, or CSD or FFT values, is that the raw signal must be thoroughly checked for artifacts. Furthermore, it must be ensured that the anatomic–topological interpretation of the map is not merely trivial. In particular, the common practice of assuming that the neural sources lie directly under the maxima and minima of the landscape, on lines projected perpendicularly inward, is not correct.

The landscape maps of the spontaneous EEG and of cognitive ERP are usually simply configured, with one or at most two peaks and one valley. The isopotential lines generally form concentric circles around these extrema. For descriptive and statistical purposes, therefore, the field strength and topography of these landscapes may reasonably be summarized by a few local and global descriptors. The global field power (GFP) corresponds to the standard deviation of the measured potentials and is an index of the strength of the field as a whole. It can be used for the reference-independent determination of ERP components. A further, nonparametric index of field strength is the reference-independent amplitude, which is calculated as the difference between the highest and lowest values on the map. A major advantage of these reference-independent, global indices of field strength, compared to the traditional measures of amplitude at given electrode positions, is that they are largely independent of the topography of the field (Strik et al. 1994a).

The calculation of field topography descriptors is performed by projection of the electrode array onto a two-dimensional coordinate system. The coordinates of the extrema (maxima and minima) of the map, or of the mathematical centers of gravity of the positive and negative regions, constitute a quantitative measure of the field configuration that is independent of both amplitude and reference. Simple statistical tests involving these values may be used for the determination of topographical differences between maps. The "dissimilarity" of two maps is a global parameter of the difference between them and is mathematically related to a correlation coefficient (Lehmann 1987). This method allows amplitude-independent topographical comparisons and has proved to be robust, as well as being superior to multivariate tests at multiple electrode positions (Strik et al. 1994b).

In the topographical representation, the landscape maps of the spontaneous EEG typically show a periodic phase reversal with opposite polarity in the occipital and frontal regions. Earlier theories of "wave fronts" and "traveling waves" spreading across the cranial

surface could not be confirmed. The periodic background activity of the EEG may be regarded as the overall, tonic state (macrostate) in which smaller, more rapid alterations of the topographic field configuration (microstates) are embedded (Lehmann 1995). Macrostates are relatively stable for each individual and are partly genetically determined (Buchsbaum and Gershon 1984), while microstates vary over intervals of a fraction of a second (Lehmann et al. 1987).

A topographically oriented segmentation method for the study of these microstates was proposed by Lehmann and coworkers (1987). It was shown that microstates remain as stable field configurations for a duration of 50 ms to more than 1 s (mean, 144 ms; Strik and Lehmann 1993) and are distinct from rapid shifts of the field landscape. Durations in this range correspond well to those of the individual building blocks of cognitive processes that have been revealed by psychological experiments. Indeed, newer studies support the hypothesis of a relationship between microstates and conscious thought processes, as visual elements and linguistic/abstract thoughts were associated with differently configured microstates in a relaxed state (Lehmann 1995; Lehmann et al. 1998). Spatially oriented segmentation methods have also been applied to the determination of ERP components (Koenig and Lehmann 1996).

3.3 Localization of Neural Generators

Neural sources cannot be localized merely by projection from the recording electrode to the underlying cortical area, because the implicit assumption of a dipole lying on a vertical line between the recording and reference electrodes is not justified for any of the types of field that are relevant to psychiatry. From the physical viewpoint it is, of course, possible to calculate the field of an electric dipole; however, if we try to proceed in the other direction and rederive the localization and number of the electric sources from the field configuration, we find that they are not uniquely determined. This unavoidable mathematical limitation is known as the inversion problem.

The computation of a model dipole that accounts for the measured field up to a given accuracy (explained variance) must therefore be performed by means of iterative algorithms. Such a computation requires a number of assumptions about the shape of the skull, the surrounding media (skin, bone, meninges, cerebrospinal fluid), and the localization and number of the generators.

The natural dipoles of the brain are its neurons. Each computed source represents a center of gravity of neural activity and thus may well turn out to be located in an electrically inactive area, e.g. in the ventricular system. The greater the number of definitely valid constraints that can be placed on dipole source localization, symmetry, number, and so forth, the more reliable the dipole solution becomes from the physiologic viewpoint. Conversely, speculative constraints on the dipole model only increase the arbitrariness of the result (Fender 1987).

Thus only physiologically valid assumptions concerning the number, symmetry, and localization of dipole sources are permitted. Very limited information of this type is available for cerebral electric fields that are relevant to cognitive processes; consequently, the computed dipoles must be regarded as mathematical models, and their localizations cannot simply be equated with anatomical sites. New developments point to interesting ways of solving the source problem, in which sharpness of localization is sacrificed in favor of a data-driven solution that requires no prior assumptions regarding the number of dipoles (low-resolution electromagnetic tomography, LORETA; Pascual-Marqui et al. 1994).

In summary, the computation of dipole sources is a useful method of obtaining a quantitative description of a cerebral electric field in terms of six parameters (three spatial coordinates, two directional angles, and dipole strength). Dipole sources can be calculated for the spontaneous EEG, for ERP, and in the frequency domain (FFT approximation; Lehmann and Michel 1990). This method has found application in psychiatry (Michel et al. 1993; Dierks et al. 1995).

3.4 Other Modern Analytical Methods

3.4.1 *Dimensional Complexity*

Algorithms for computing the complexity of biological signals have been borrowed from the theoretical physics of nonlinear systems (chaos theory). Such procedures are based on the possibility of projecting the dynamics of the signal of a deterministic system onto a static attractor. The attractor is described by several parameters, including its dimension; a higher dimension indicates a higher complexity of the system. Dimension is thus also referred to as dimensional complexity and is computed as a correlational dimension (Grassberger and Procaccia 1983).

As a rule, the determination of this parameter requires a signal with a minimum number of data points, and it therefore cannot be used to characterize very short intervals of cerebral electric activity. In general, high-frequency patterns such as rapid eye movement (REM) sleep, waking EEG with the eyes open, and elevated arousal are associated with a

high-dimensional complexity, while complexity is lower for epileptic 3-Hz discharges, with the eyes closed, and in deep sleep (Röschke and Aldenhoff 1992; Babloyantz 1989).

For multichannel recordings, it is recommended that the number of channels be used as the "embedding dimension." In this way, the global correlational dimension of a series of cerebral electric maps can be calculated (Wackermann et al. 1993).

3.4.2 Multivariate Statistics and Neural Networks

With the aid of multivariate statistical methods, one can attempt to identify clinical subgroups on the basis of the parameters of the multichannel EEG. The neurometrics system (John et al. 1988) is the best-known example. In this type of procedure, discriminance analysis or similar methods are used to identify characteristic constellations of parameters in the EEG of patients with known diagnoses. The validity of the resulting patterns is then tested against the clinical diagnoses of a further, independent patient cohort. This is a matter of optimization of an expert system, which, however, cannot be definitively validated, because there is as yet no etiologic–pathogenetic classification of psychiatric diagnoses. This method is thus intrinsically unable to furnish new information regarding natural disease entities. Furthermore, its practical application is limited, because every change in diagnostic criteria requires a corresponding adaptation of the system, and its use as a simple means of diagnosis in the hands of nonpsychiatrists cannot be envisioned.

A further interesting application is the attempt to identify new patient subgroups on the basis of constellations of EEG parameters derived from principal component analysis (PCA) or cluster analysis. Although some of these subgroups include a high percentage of patients treated with neuroleptics, no clinical features relevant to psychiatric classification have been identified to date (John et al. 1994).

Neural networks are control or decision systems capable of learning. They have found application in statistics, machine control, adaptive signal processing, and pattern recognition (e.g. language). The name is taken from the neurophysiological tradition (Hebb 1949). Recently, it has sometimes been implied that these modern mathematical methods have made the biological foundations of conscious information processing "calculable." In fact, however, theories based on analogies between biological and mathematical neural networks are highly speculative, because realistic simulations of higher brain functions remain impossible. In practical terms, neural networks constitute a statistical method which, like PCA or discriminance analysis, can be used to optimize – in this case, by training – the separation of a multidimensional data set into previously specified groups (Arbib 1995).

4 Electroencephalography and Sleep

EEG remains the most sensitive method for the noninvasive measurement of transient changes in wakefulness. It plays an important role in the current understanding of sleep physiology and is the basis of classification systems for the stages of sleep. It can be used not only to document the presence of hypnagogic and hypnotic disorders, but also to obtain parameters of relevance to psychiatric disease, such as the time from falling asleep to the onset of the first dreaming stage (REM latency) or the absolute or proportional duration of the various sleep stages (e.g. deep sleep duration, REM density). The separation into stages is carried out on the basis of the wave pattern, either by visual inspection or by computer algorithm. The reader is referred to specialized texts for the classification and description of sleep stages (Rechtschaffen and Kales 1968; Schulz 1997).

All of the quantitative methods described in general in Sect. 3 above can be used to study the sleep EEG. The most important findings of sleep EEG studies involving psychiatric patients are described below in Sect. 7.

5 Event-Related Potentials and Their Components

ERP are obtained from the averaging of segments of the EEG that are temporally coupled to a specific event. Periodic activity that is not coupled to the event is eliminated by the averaging process, while the specific cerebral electric activity in response to the event, or in preparation for it, remains. This section contains a few examples with particular relevance to psychiatric disease, selected from among the many paradigms and components described to date. An overview of early evoked potentials can be found in Maurer (1993), and a thorough classification and description of ERP in Olbrich (1989). The nomenclature of ERP generally consists of an abbreviation for polarity according to the traditional system of referential recording (P, positive; N, negative) followed by the typical latency in milliseconds.

Among ERP of intermediate latency, P50 is of special psychiatric interest. This component diminishes in

healthy individuals upon repetition of the provoking auditory stimulus (click). Maximal suppression is reached by repeating the stimulus at an interval of 500 ms; with intervals of 8 s or more, the effect is lost. P50 is regarded as a reflection of the activity of an early sensory filtering channel that inhibits the neural response to irrelevant stimuli and thus restricts the incoming flood of sensory information. Evidence suggests that P50 is generated in the medial temporal lobe (Freedman et al. 1991).

Acoustic stimuli evoke an N100 component whose amplitude is proportional to the attention paid to the stimulus. This component is thus interpreted as a reflection of automatic stimulus processing and selective attention; it is regarded as an expression of the activity of the auditory cortex (Näätänen 1990). Increasing the stimulus intensity may result in an increase or a decrease of the N100 amplitude, depending on the subject. This intensity dependence is thought to be related to central serotonergic activity, as a high level of serotonergic activation is associated with a low N100 amplitude (Hegerl et al. 1996).

A small change of the duration, pitch, or intensity of an acoustic stimulus in the midst of an otherwise unvarying series leads to an increase in negativity at a latency of approximately 200 ms. This effect is known as mismatch negativity (MMN). In this paradigm, unlike the P300 paradigm, the subject does not direct his or her attention to the stimuli, but rather is given a distracting task to perform. This component is interpreted as a reflection of automatic perception and discrimination against the background of an "echoic" acoustic memory task (Näätänen 1990). The amplitude of the MMN increases with age (Woods 1992).

The best-known and best-studied late ERP component is P300, which occurs only when the stimulation involves a conscious mental process. For example, a typical P300 experiment concerns the recognition of a rare target stimulus against a background of meaningless stimuli ("oddball paradigms"). The amplitude increases with increasing rarity of the target stimulus. P300 can also be evoked by the omission of an expected stimulus.

In the acoustic modality, normal individuals generally display a center of gravity of the P300 field in the left hemisphere (Morstyn et al. 1983; Strik et al. 1993a), while in the visual modality there is a mild right hemispheric asymmetry (Alexander et al. 1995). The positive center of gravity lies farther forward when a motor response is suppressed than when it is carried out (Fallgatter and Strik 1996); in LORETA, this is explained by inhibitory functions of the frontal lobe (Strik et al. 1998).

The neural generators of P300 are still not known for certain. It is assumed that the temporal, parietal, and frontal lobes all contribute to this summed potential (Halgren 1986). Temporal lobe activity alone does not account for the potential (Johnson 1988). P300 latency increases after the 18th year of life (Goodin et al. 1978).

The N400 component, which arises only when a sentence ends with an inappropriate word (semantic incongruence), is of interest to cognitive science. This component is not present when sentence construction or spelling are faulty. It is a central negativity arising at a latency of approximately 400 ms. The location of its neural generators is not known with certainty.

Post-imperative negative variation (PINV) is a late ERP component that typically arises in normal individuals when the expected effect of a motor action fails to occur. It is regarded as a reflection of continuous cognitive processing (Pritchard 1986).

Contingent negative variation (CNV) arises before an expected sensory stimulus. In a typical experiment, two stimuli are presented with a fixed delay, usually of less than 2 s. The CNV occurs before the second stimulus in the form of a central negative potential lasting from 100 to 500 or more ms. This potential reflects the preparation of the brain for the incoming stimulus and is interpreted as an expression of the neural resources made available in this process (Rockstroh et al. 1989).

6 Psychopharmacology of Electroencephalography and Event-Related Potentials

The EEG reacts sensitively to psychotropic substances. Their mechanism of effect, dose–response and temporal response profiles, brain penetration, and bioequivalent dosing can all be investigated with quantitative EEG methods (Herrmann 1982; Saletu 1989). Because the resulting EEG changes show considerable interindividual differences, expert systems have been developed to distinguish psychoactive substances with antidepressant, neuroleptic, anxiolytic, and neurotropic effects (Itil et al. 1979).

In particular, benzodiazepines induce a characteristic beta activation with spindles at 14–16 Hz, as well as slow waves at higher doses (Friedman et al. 1992). In normal subjects, neuroleptic medications induce an increase of delta and theta activity, while alpha activity remains unchanged or is mildly reduced, and beta activity decreases occipitally and increases frontally (Fink 1974). Clozapine induces severe generalized slowing and, occasionally, spikes and spike-wave complexes, a finding that invites caution in raising the dose, but is not in itself a reason to discontinue the medication (Koukkou et al. 1979). Tricylic agents elevate delta and beta activity. The EEG changes produced by both neuroleptic agents and tricyclic

agents are dose dependent (Czobor and Volavka 1992). Mind-altering drugs generally induce an increase of high-frequency components in association with an increase in wakefulness (Fischhof et al. 1992), though not in a clearly dose-dependent fashion. Lithium treatment is often associated with major EEG changes, including slowing and increased amplitude of background activity, paroxysmal dysrhythmias, and occasionally hypersynchronous discharges (Helmchen and Kanowski 1971).

The influence of psychotropic substances on ERP components is best documented for P300. Centrally active sedatives such as alcohol, benzodiazepines, flupentixol, and anticholinergics lower the P300 amplitude. Cholinergics such as physostigmine, however, raise the P300 amplitude in normal subjects (Rösler et al. 1985; Picton 1992; Maurer et al. 1990). An elevation of amplitude without a change of latency was also seen in normal elderly subjects after the administration of the cholinergic substance pyritinol (Dierks et al. 1994). There is no single explanation for the effects of the dopamine agonist methylphenhydate, which raises the P300 amplitude in hyperactive children, but not in normal individuals (Klorman and Brumaghim 1991). If we consider these specific changes of P300 amplitude along with simultaneously measured cerebrospinal fluid levels of 3-methoxy-4-hydroxyphenylglycol (MHPG; the major metabolite of norepinephrine) and with psychometrically quantified changes of attention, we arrive at a unified interpretation of the P300 amplitude as a reflection of cortical arousal.

The P300 latency is increased in normal subjects by anticholinergics and by haloperidol. Haloperidol has no effect on P300 amplitude (Stanzione et al. 1990). In contrast, dopaminergic agonists such as L-dopa normalize the prolonged P300 latencies of patients with Parkinson's disease, but have no effect on P300 parameters in normal subjects.

7
Neurophysiologic Manifestations of Psychiatric Disease

7.1
Schizophrenias

Schizophrenic patients typically show an increased amplitude of both slow (delta, theta) and fast (beta) frequency components. In contrast, background alpha activity is reduced and slowed (Shagass 1987). Ocular artifact has been discussed as a possible explanation for these findings (Guenther et al. 1988), but several studies have failed to demonstrate the frontal accentuation of delta activity that is characteristic of ocular artifacts. An effect of neuroleptics cannot account for these findings either, because slow-wave activity is actually more pronounced in nontreated than in treated patients (Galderisi et al. 1991). The EEG of schizophrenic patients shows a lesser degree of modulation than that of normal controls; this property holds both for the spontaneous EEG and in the presence of visual stimuli and has been referred to as the "rigidity" of the schizophrenic EEG (Shagass 1987). The reactivity of the EEG normalizes upon remission from an acute schizophrenic episode, but to different extents in different frequency bands (Koukkou-Lehmann 1987). At the onset of disease, untreated patients have diminished reactivity, primarily of frequency and power in the alpha band. Alpha-band frequency reactivity normalizes in patients in stable remission who have been off medication for 3 months, but alpha-band power reactivity remains rigid. Alpha-band power reactivity has thus been proposed as a "trait correlate" of schizophrenia. The reactivity of the EEG has also been interpreted as an electrophysiologic manifestation of the reorganization of working memory in psychotic patients (Koukkou et al. 1995).

Single doses of neuroleptics change the EEG independently in a manner that correlates with their later therapeutic effect. Responders to pharmacotherapy show an increase in slow-frequency components (Herrmann and Winterer 1996; Gaebel et al. 1988). Seventeen of 18 responders and eight of ten nonresponders were correctly classified on the basis of their $alpha_1$ reactivity (7.7–9.5 Hz); responders reacted with an increase of $alpha_1$ activity, and nonresponders with a decrease (Galderisi et al. 1994). These findings support the hypothesis that pharmacotherapy is effective only in those schizophrenics whose EEG on neuroleptic medication reacts similarly to that of normal subjects.

The application of newer methods has revealed differences in the field configurations of the spontaneous EEG of normal subjects and schizophrenics. Patients who have never been treated manifest an increase of correlational dimension in anterior regions of their EEG recordings (Koukkou et al. 1993). The related finding of diminished coherence between prefrontal cortical areas under conditions of activation implies a functional uncoupling of these areas (Hoffman et al. 1991). Changes in orientation of the EEG microstates have been described on the basis of topographically oriented segmentation. The microstate configuration corresponds to that of normal subjects during concrete pictorial thought (Strik et al. 1995).

The deeper stages of sleep (i.e. non-REM or slow-wave stages) are shortened by approximately 50% in approximately 60% of schizophrenic patients, whether treated or untreated. This phenomenon is

not specific, however, as it is also found in healthy elderly persons as well as in depressed or demented patients (Keshavan et al. 1990; Fleming 1994). More recent studies of correlational dimension have shown a diminished complexity of the EEG signal during stage II and REM sleep. There is as yet no plausible explanation of these findings (Röschke and Aldenhoff 1993).

Evidence of changes in perceptual modalities can be seen as early as the intermediate ERP components. Repeated presentation of a simple acoustic stimulus leads to a diminution of the evoked P50 component in normal subjects, but not in schizophrenics (Adler et al. 1982; Judd et al. 1992). This has been interpreted as a manifestation of "sensory gating," i.e. regulation of the flow of sensory information through a hypothetical channel of limited capacity. The persistence of P50 in schizophrenics is regarded as a sign of hypervigilance, leading to a loss of the ability to suppress irrelevant external stimuli, and is said to be relatively specific to schizophrenia. Neuroleptic medications do, in fact, raise P50 amplitude generally, but do not alter the ratio of the responses to the warning and test stimuli. Only manic patients manifest a similar abnormality in the acute stage, which then disappears upon pharmacologic treatment and remission. This abnormality is said to be more common among relatives of schizophrenics than in the normal population (Freedman et al. 1991). The P50 response of schizophrenics normalizes after a brief period of deep sleep or after smoking a cigarette (Griffith et al. 1993; Adler et al. 1993).

The findings of recent studies of P300 in schizophrenics are among the more interesting developments in psychiatric neurophysiology. P300 amplitude has been known for many years to be lower in schizophrenics than in normal subjects, while a prolongation of P300 latency has not been consistently demonstrated. The reduction of amplitude cannot, however, be used as a diagnostic marker, because of its lack of specificity and a considerable overlap of values with those of normal subjects. An interesting new aspect came about with the description of topographic alterations in the form of an asymmetric left-hemispheric reduction of amplitude, with a maximum in the right hemisphere. Normal control groups showed the reverse (physiologic) asymmetry, with a maximum in the left hemisphere (Morstyn et al. 1983). This finding has not been described in other types of patients and is thus more specific than the reduction of amplitude.

Because of the inversion problem (see the above discussion of dipole analysis), the finding of a maximum in the right hemisphere cannot be taken as proof of a functional deficit in left temporal cortical areas. Indirect evidence from neuropsychologic and MRI studies is, however, consistent with this hypothesis (Heidrich and Strik 1997; McCarley et al. 1993). Consistent results have been obtained with visual hemifield stimulation according to an oddball paradigm: stimulation of the right hemifield is associated with prolonged P300 latency and diminished P300 amplitude (Galderisi et al. 1988).

Recent clinical studies indicate that these P300 changes have prognostic significance, as well as a relationship to diagnostic subcategories. Low amplitudes are significantly correlated with the social impairment caused by negative symptoms (Strik et al. 1993b). A cross-sectional study by Hergerl and coworkers revealed a higher frequency of tardive dyskinesia and a less favorable clinical course in patients with lower P300 amplitudes (Hegerl et al. 1995). In fact, a longitudinal study showed an association of low P300 amplitudes with poor social integration after a mean follow-up period of 2.4 years, which implies a possible prognostic significance for this parameter (Strik et al. 1996). In view of the known relationship of the P300 amplitude to attention (Heidrich and Strik 1997) and to the cerebrospinal fluid level of the major norepinephrine metabolite MHPG (Ford et al. 1994), it may be assumed that the P300 amplitude is modulated by the state of arousal of the central nervous system. If so, it would then be nonspecifically associated with the subject's readiness to perform a cognitive task. There is evidence that a corresponding phenomenon is at work in normal subjects as well: the P300 amplitude was found to be negatively correlated with a psychometric index of anhedonia (Simons 1982).

One successful application of P300 measurement has been found in the subclassification of the schizophrenias into cycloid psychoses and (proper) schizophrenias, according to the scheme of Leonhard (1986). It has been demonstrated, and replicated in several independent patient cohorts, that the P300 maximum lies in the right hemisphere only in patients who are schizophrenic in Leonhard's sense. P300 topography has been found to be normal in cycloid psychoses, even when the DSM-III-R criteria for schizophrenia are fulfilled. P300 amplitude, however, was recently found to be elevated in a cohort of psychiatric patients with cycloid psychoses (Strik et al. 1993a; Strik et al. 1997). It should be noted that Leonhard's distinction between cycloid psychoses and schizophrenias does not simply correspond to a grouping of schizophrenias into positive and negative types, as many subtypes of schizophrenia in the Leonhard scheme in fact have predominantly positive symptoms.

It is not yet known whether the elevation of P300 amplitude seen in cycloid psychoses regresses after complete remission and reintegration into everyday life. Other open questions concern the psychophysiological implications of P300 asymmetry

for the pathogenesis of schizophrenic symptoms. Language functions, which are highly lateralized to the left hemisphere in normal subjects, are of particular interest for the development of hypotheses for further study, as many central schizophrenic symptoms reflect a disturbance of particular aspects of cerebral language function, including expression (circumstantiality), conceptualization (disordered thinking), and perception (hallucinations of speech). Pathologic asymmetries have been described only when silent counting, i.e. the construction of imagined speech, was used as a control for attention, and not when a motor reaction was carried out; this provides a further indication of the important role of linguistic processes in P300 asymmetry.

A diminished amplitude of mismatch negativity has been found repeatedly in both treated and untreated schizophrenic patients. The magnitude of this diminution correlates, albeit weakly ($p = 0.4$), with the extent of negative symptoms. Mismatch negativity was normal, however, in a control group of patients with bipolar affective disorder (Catts et al. 1995). It should be added that practically all of the traditionally analyzed, cognitively influenced ERP components are diminished in amplitude among schizophrenics, so that the operation of a general modulating factor, rather than a disturbance of specific functions, must be postulated.

Similar limitations necessarily apply to the interpretation of findings regarding the N400 component, although the latter is, theoretically, a potential of great relevance to schizophrenia. The traditional waveform analyses performed to date have revealed inconsistent reductions of amplitude and/or prolongations of latency, analogous to the results obtained with other ERP components. There is evidence that these abnormalities are present only in a subgroup of schizophrenics and that the N400 amplitude is reduced only when the task does not require a decision (Andrews et al. 1993). There has been, as yet, no comprehensive topographic description of the N400 of schizophrenic patients. One may expect, in analogy to the recent findings of P300 studies, that the topography of N400 landscapes will turn out to be highly relevant to schizophrenic disorders.

CNV, like other ERP components, is reduced in amplitude in schizophrenic patients. There are contradictory reports of associations with particular symptoms. The reported finding of low amplitudes in floridly psychotic patients and normalization after clinical remission (Pritchard 1986) could not be replicated by other investigators (van den Bosch et al. 1988). A reduction of amplitude affecting predominantly the early components of the CNV over the vertex area has been described as typical for schizophrenic patients (Rockstroh et al. 1989).

Schizophrenics given a variety of stimulus–response tasks often spontaneously display a late negativity (PINV) that is seen in normal subjects only under certain experimental conditions. The topographical aspects of the distribution of electrical potential are correlated with negative symptoms (Eikmeier et al. 1993).

7.2 Affective Disorders

Even though there is no characteristic abnormality of the frequency pattern of the EEG in patients with affective disorders, studies of depressed patients have yielded evidence of an increased rigidity of alpha activity in psychotic depressions and an increased lability of alpha activity in neurotic depressions (Herrmann and Winterer 1996). The findings in psychotic depressions are analogous to the abnormalities typically seen in schizophrenic patients (see above). With spatially oriented segmentation of the spontaneous EEG, an elevated spatial variability was found, i.e. the cerebral electric field changed its configuration both more frequently per unit time and more extensively than in normal subjects. This finding has been thought to bear a relation to psychopathologically described properties of the cognitive strategies of depressed patients. Microstate topography in schizophrenics is, however, no different from that of normal control subjects (Strik et al. 1995).

A shortening of REM latency and elevation of REM density has been found in the sleep EEG of depressed patients, i.e. these patients fall significantly earlier into the first dreaming stage of sleep than control subjects, and they have briefer periods of deep sleep. This finding is quite robust but, unfortunately, nonspecific, as it is present in other psychiatric illnesses as well. Antidepressants suppress REM sleep, and the parameters of sleep normalize in many patients upon clinical remission. Surprisingly, no relationship was found between REM latency and the therapeutic efficacy of sleep deprivation treatment (Fleming 1994). Relatives of patients with sleep stage anomalies manifest this abnormality with a concordance of approximately 70%; REM latency has therefore been proposed as a marker of vulnerability to affective disorders. The association of short REM latency with an elevated rate of relapse, and the finding of early normalization among responders to amitriptyline, raised interest in the possible predictive value of this parameter, but these findings have not yet been sufficiently refined and validated to allow an application in clinical practice. Furthermore, the dependence of REM latency on the clinical state of affected patients seems to

contradict the notion that it is a useful marker of vulnerability to the disease.

A diminution of P300 amplitude less severe than that seen in schizophrenics has been described in severely depressed patients. In the present state of research, no clinical significance can be attached to this abnormality, which is assumed to be due to a lack of motivation and, indeed, disappears upon remission (Picton 1992). Systematic studies of P300 in manic patients are still rare. In a recently completed study, evidence for a frontal disinhibition has been found in manic patients, in contrast to the results in cycloid psychoses, which indicate a cerebral overreaction (Strik et al. 1998). The normal reduction of P50 amplitude upon repetition of the stimulus is absent in acutely manic patients, as in schizophrenics, but returns upon clinical remission. The CNV of depressed patients tends to show a diminution of amplitude that is negatively correlated with the severity of depression (Ashton et al. 1988). Timsit-Berthier, however, found a subgroup of depressed patients with elevated rather than diminished amplitudes. This subgroup also had a higher dopamine reactivity on apomorphine testing than the remainder of the cohort of depressed patients (Timsit-Berthier 1986).

7.3 Dementia

Features typical of dementia can be seen even on qualitative assessment of the EEG. An increase in theta and delta activity occurs in a generalized fashion in Alzheimer-type dementia, but focally in cerebrovascular dementias. Furthermore, a relative increase of alpha activity is found in frontal tracings, and there is a decreased reactivity of alpha activity upon eye opening or photic stimulation. These findings have been documented statistically with the aid of quantitative topographic FFT analysis (Dierks et al. 1991). Recent studies have shown a correlation between the increase in slow-wave activity and the severity of dementia. A frontal displacement of the center of gravity of alpha activity can be seen even in the early stages of Alzheimer-type dementia, and the extent of this abnormality correlates with the severity of dementia. There is evidence that functional assessment by means of EEG parameters correlates more closely with cognitive impairment than does regional cerebral blood flow (Müller et al. 1997). The dementias are thus a disease category in which EEG can supply clinically relevant information about the severity and nature of the disorder (Alzheimer-type versus multi-infarct dementia). EEG mapping allows a perspicuous representation of the abnormalities described (frontal displacement, focal versus global increase of slow-wave activity), provided that an experienced interpreter has ruled out the presence of significant artifact.

A diminution of P300 amplitude and a prolongation of P300 latency are typically found in patients with dementing illnesses. While a diminution of amplitude is nonspecific, a prolongation of latency is regarded as relatively characteristic of these processes of cognitive deterioration. In a meta-analysis, the sensitivity of this abnormality was found to vary from 13% to 80%, but its specificity (against a background of patients with general psychiatric illnesses) was nearly 90% (Goodin 1990). The sensitivity has been found to be higher in studies of more severe dementias. Of particular interest is the reported ability to distinguish the early stages of Alzheimer-type dementia from depressive pseudodementias, although the use of the P300 latency as an aid to differential diagnosis remains to be validated. P300 testing is of less value in the later stages of the disease, as many severely demented patients have a severe impairment of attention that makes such testing impossible.

7.4 Obsessive–Compulsive Disorder

Obsessive-compulsive patients given an acoustic oddball task were found to have an elevated N200 amplitude and a shortened P300 latency, but a normal P300 amplitude. While normal subjects had a prolongation of P300 latency that increased with the difficulty of the task, the patients did not show this effect. The elevation of N200 amplitude has been interpreted as a sign of elevated selective attention, and the shortening of P300 latency as a sign of accelerated information processing (Towey et al. 1990). In a more recent study, elevations of both N100 and P300 amplitudes were described, and the latter elevation was correlated with the severity of symptoms (Olbrich et al. 1996). There is as yet no explanation for these inconsistent results.

7.5 Anxiety Disorders

A reduction of alpha activity and a simultaneous increase of activity in both lower and higher frequency ranges have been described as characteristic of the EEG of patients with anxiety disorders. In anxious individuals or patients with chronic anxiety disorders, unlike normal subjects, alpha activity increases with cognitive activation, while the alpha blockade response is diminished, as are amplitude and frequency variability. At the same time, fast- and slow-wave activity decreases. Epileptic discharges have occasionally been

described as occurring during panic attacks, although it is not clear whether these may have been cases of comorbidity with two different diseases (Herrmann and Winterer 1996).

7.6 Personality Disorders

Individuals with personality disorders (Raine 1989) and normal extroverts (Cahill and Polich 1992) have an elevated P300 amplitude compared to control subjects, while normal adults with high psychometric values for anhedonia have a diminished amplitude (Simons 1982). A higher amplitude is interpreted as a reflection of elevated cortical arousal, of which risk-positive behavior (sensation seeking) is an epiphenomenon.

The N100 augmenting-reducing phenomenon in personality disorders has been extensively studied. In the visual modality, it has been shown that the N100 amplitude can be augmented by increasing the light intensity and that this augmentation is correlated with risk-positive behavior, extroversion, and impulsiveness; likewise, the opposite personality traits can be brought about by reducing the N100 amplitude (Pritchard 1986). In a manner analogous to the P300 findings, this phenomenon has been interpreted as the expression of a psychophysiologically based tendency to seek or to avoid excitatory stimuli.

7.7 Alcoholism

The consumption of moderate amounts of alcohol by healthy young men leads to a slowing of the EEG frequency spectrum (Ehlers et al. 1989). This effect was less marked in subjects who were alcohol dependent in a 10-year follow-up period than in subjects who did not become dependent on alcohol (Volavka et al. 1996). This result is in good accordance with the finding that a small intoxicating effect of alcohol is a predictor of later alcoholism (Schuckit 1994).

The P300 amplitude is diminished in alcoholics, and the P300 latency prolonged. This abnormality is more pronounced with acoustic than with visual stimulation. The reported diminution of P300 amplitude in relatives of alcoholics is of particular clinical and theoretical interest, although replication studies have not uniformly confirmed this finding (Picton 1992).

In both normal subjects and alcoholics, alcohol reduces the intensity dependence of the acoustic N1/P2 components. In consideration of the known effects of other drugs, this finding has been attributed to the serotonergic effect of alcohol (Hegerl et al. 1996).

8 Conclusions and Overview

Neurophysiologic testing is of particular importance in psychiatry because it allows a biological measurement of the activity of the intact brain on a timescale appropriate to the study of conscious mental processes. Nonetheless, the problem of anatomical-topological coordination of the events that are measured has yet to be satisfactorily solved. Furthermore, the sensitivity of these methods is still insufficient for the study of mental processes such as mental calculation, conscious stimulus discrimination, or semantic language processing, and their temporal resolution is not high enough to allow a description of the time course of perceptual and decision processes in the millisecond range.

Modern analytical methods and highly refined paradigms that activate well-defined brain functions while providing reasonable reference conditions will likely lead to a further expansion of our understanding of the concerted neuronal activity that is the basis of the highest functions of the human brain. The accumulated knowledge of the conditions and chronology of cognitive activity, the opportunities for practical application, and the high sensitivity to functional abnormalities will be invaluable for the development and testing of suitable activating conditions that will also meet the needs of modern, high-resolution functional imaging methods such as PET and fMRI.

9 References

Adler LE, Patchman E, Franks, RD, Pechevich M, Waldo MC, Frzeedman R (1982) Neurophysiological evidence for a defect in neural mechanisms involved in sensory gating in schizophrenia. Biol Psychiatry 17: 639–654

Adler LE, Hoffer LD, Wiser A, Freedman R (1993) Normalization of auditory physiology by cigarette smoking in schizophrenic patients. Am J Psychiatry 150: 1856–1861

Alexander JE, Porjesz B, Bauer LO, Kuperman S, Morzorati S, O'Connor SJ, Rohrbaugh J, Begleiter H, Polich J (1995) P300 hemispheric amplitude asymmetries from a visual oddball task. Psychophysiology 32: 467–475

Andrews S, Shelley AM, Ward PB, Fox A, Catts SV, McConaghy N (1993) Event-related potential indices of semantic processing in schizophrenia. Biol Psychiatry 34: 443–458

Arbib MA (1995) The handbook of brain theory and neural networks. MIT, Cambridge, MA

Ashton H, Golding JF, Marsh VR, Thompson JW, Hassanyeh F, Tyrer SP (1988) Cortical evoked potentials and clinical rating scales as measures of depressive illness. Psychol Med 18: 305–317

Babloyantz A (1989) Estimation of correlation dimensions from single and multiple recordings – a critical review. In: Basar E,

Bullock TH (eds) Brain dynamics. Springer, Berlin Heidelberg New York, pp 122–130

Barlow JS, Creutzfeld OD, Michael D, Houchin J, Epelbaum H (1981) Automatic adaptive segmentation of clinical EEGs. Electroencephal Clin Neurophysiol 51: 512–525

*Basar-Eroglu C, Strüber D, Schürmann M, Stadler M, Basar E (1996) Gamma-band responses in the brain: a short review of psychophysiological correlates and functional significance. Int J Psychophysiol 24: 101–124

Buchsbaum MS, Gershon ES (1984) Genetic factors in EEG. In: Davidson J, Davidson RJ, Schwartz GE (eds) Human consciousness and its transformations. Plenum, New York

Cahill JM, Polich J (1992) P300, probability, and introverted/extroverted personality types. Biol Psychol 33: 23–35

Catts SV, Shelley AM, Ward PB, Liebert B, McConaghy N, Andrews S, Michie T (1995) Brain potential evidence for an auditory sensory memory deficit in schizophrenia. Am J Psychiatry 152: 213–219

Cohen D, Cuffin BN, Yunokuchi K, Maniewski R, Purcell C, Cosgrove GR, Ives J, Kennedy JG, Schomer DL (1990) MEG versus EEG localization test using implanted sources in the human brain. Ann Neurol 28: 811–817

Czobor P, Volavka J (1992) Level of haloperidol in plasma is related to electroencephalographic findings in patients who improve. Psychiatry Res 42: 129–144

Dierks T, Perisic I, Froelich L, Ihl R, Maurer K (1991) Topography of the quantitative electroencephalogram in dementia of the Alzheimer type: relation to severity of dementia. Psych Res Neuroimaging 40: 181–194

Dierks T, Frölich L, Ihl R, Maurer K (1994) Event-related potentials and psychopharmacology. Cholinergic modulation of P300. Pharmacopsychiatry 27: 72–74

*Dierks T, Strik WK, Maurer K (1995) Electrical brain activity in schizophrenia, measured by center of gravity dipoles of FFT-data. Schizophr Res 14: 145–154

Ehlers CL, Wall TL, Schuckit MA (1989) EEG spectral characteristics following ethanol administration in young men. Electroencephalogr Clin Neurophysiol 73: 179–187

Eichenbaum H (1993) Thinking about brain cell assemblies. Science 261: 993–994

Eikmeier G, Lodemann E, Olbrich HM, Pach J, Zerbin D, Gastpar M (1993) Altered fronto-central PINV topography and the primary negative syndrome in schizophrenia. Schizophr Res 8: 251–256

Fallgatter AJ, Strik WK (1996) Topographische Veränderungen der hirnelektrischen Felder während des Continuous Performance Test (CPT). Fortschr Neurol Psychiatr 64[Suppl]: 85

*Fender DH (1987) Source localization of brain electrical activity. In: Gevins AS, Rémond A (eds) Methods of analysis of brain electrical and magnetic signals. Handbook of electroencephalography and clinical neurophysiology, vol 1 (revised). Elsevier, Amsterdam, pp 355–403

Fink M (1974) EEG profiles and bioavailability measure of psychoactive drugs. In: Itil TM (ed) Psychotropic drugs and the human EEG. Modern problems of pharmacopsychiatry, vol 8. Karger, Basel, pp 76–98

Fischhof PK, Saletu B, Ruther E, Litschauer G, Moslinger R, Herrmann WM (1992) Therapeutic efficacy of pyritinol in patients with senile dementia of the Alzheimer type (SDAT) and multi-infarct dementia (MID). Neuropsychobiology 26: 65–70

Fleming JA (1994) REM sleep abnormalities and psychiatry. J Psychiatry Neurosci 19: 335–344

*Flor-Henry P, Gruzelier J (1983) Lateralization and psychopathology. Elsevier, Amsterdam

Ford JM, White PM, Csernansky JG, Faustman WO, Roth WT, Pfefferbaum A (1994) ERPs in schizophrenia: effects of antipsychotic medication. Biol Psychiatry 36: 153–170

Freedman R, Waldo M, Bickford-Wimer P, Nagamoto H (1991) Elementary neuronal dysfunctions in schizophrenia. Schizophr Res 4: 233–243

**Freeman WJ (1983) The physiological basis of mental images. Biol Psychiatry 18: 1107–1125

French CC, Beaumont JG (1984) A critical review of EEG coherence studies of hemisphere function. Int J Psychophysiol 1: 241–254

Friedman H, Greenblatt DJ, Peters GR, Metzler CM, Charlton MD, Harmatz JS, Antal EJ, Sanborn EC, Francom SF (1992) Pharmacokinetics and pharmacodynamics of oral diazepam: effect of dose, plasma concentration, and time. Clin Pharmacol Ther 52: 139–150

Gaebel W, Ulrich G, Pietzker A, Müller-Oerlinghausen B (1988) Elektroenzephalographische Indikatoren der neuroleptischen Akutresponse. In: Beckmann H, Laux G (eds) Biologische Psychiatrie, Synopsis 1986/87. Springer, Berlin Heidelberg New York, pp 303–306

Galderisi S, Maj M, Mucci A, Monteleone P, Kemali D (1988) Lateralization patterns of verbal stimuli processing assessed by reaction time and event-related potentials in schizophrenic patients. Int J Psychophysiol 6: 167–176

Galderisi S, Mucci A, Mignone ML, Maj M, Kemali D (1991) CEEG mapping in drug-free schizophrenics. Differences from healthy subjects and changes induced by haloperidol treatment. Schizophr Res 6: 15–23

*Galderisi S, Maj M, Mucci A, Bucci P, Kemali D (1994) QEEG alpha1 changes after a single dose of high-potency neuroleptics as a predictor of short-term response to treatment in schizophrenic patients. Biol Psychiatry 35: 367–374

Goodin DS (1990) Clinical utility of long latency 'cognitive' event-related potentials (P3): the pros. Electroencephalogr Clin Neurophysiol 76: 2–5

Goodin DS, Squires KC, Hendersen BH, Starr A (1978) Age-related variations in evoked potentials to auditory stimuli in normal human subjects. Electroencephalogr Clin Neurophysiol 44: 447–458

Grassberger P, Procaccia I (1983) Measuring the strangeness of strange attractors. Physica 9: 189–208

Griffith JM, Waldo M, Adler LE, Freedman R (1993) Normalization of auditory sensory gating in schizophrenic patients after a brief period for sleep. Psychiatry Res 49: 29–39

Guenther W, Davous P, Godet JL, Guillibert E, Breitling D, Rondot P (1988) Bilateral brain dysfunction during motor activation in type II schizophrenia measured by EEG mapping. Biol Psychiatry 23: 295–311

Halgren E, Stapleton JM, Smith M, Altafullah I (1986) Generators of the human scalp P3(s). In: Cracco Q, Bodis-Wollner I (eds) Evoked potentials. Liss, New York, pp 269–284

Hebb D (1949) Organization of behavior. Wiley, New York

Hegerl U, Juckel G, Müller-Schubert A, Petzcker A, Gaebel W (1995) Schizophrenics wit small P300: a subgroup with a neurodevelopmental disturbance and a high risk for tardive dyskinesia? Acta Psychiatr Scand 91: 120–125

Hegerl U, Juckel G, Möller HJ (1996) Ereigniskorrelierte Hirnpotentiale als Indikatoren neurochemischer Dysfunktionen bei psychiatrischen Patienten. Nervenarzt 67: 360–368

*Heidrich A, Strik WK (1997) Auditory P300 topography and neuropsychological test performance: evidence for left hemispheric dysfunction in schizophrenia. Biol Psychiatry 41: 327–345

Helmchen H, Kanowski S (1971) EEG changes under lithium (Li) treatment. Electroencephalogr Clin Neurophysiol 30: 269

Herrmann WM (1982) Development and critical evaluation of an objective procedure for the electroencephalographic classification of psychotropic drugs. In: Herrmann WM (ed) EEG in drug research. Fischer, Stuttgart, pp 249–351

*Herrmann WM, Winterer G (1996) Über die Elektroenzephalographie in der Psychiatrie – gegenwärtiger Stand und Ausblick. Nervenarzt 67: 348–359

Hoffman RE, Buchsbaum MS, Escobar MD, Makuch RW, Nuechterlein KH, Guich SM (1991) EEG coherence of prefrontal areas in normal and schizophrenic males during perceptual activation. J Neuropsychiatry Clin Neurosci 3: 169–175

Itil TM, Shapiro DM, Herrmann WM, Schulz W, Morgan V (1979) HZI Systems for EEG parametrization and classification of psychotropic drugs. Pharmacopsychiatria 12: 4–19

John ER, Prichep LS, Easton P (1988) Neurometrics: computer assisted differential diagnosis of brain dysfunction. Science 293: 162–169

John ER, Prichep LS, Alper KR, Mas FG, Cancro R, Easton P, Sverdlov L (1994) Quantitative electrophysiological characteristics and subtyping of schizophrenia. Biol Psychiatry 36: 801–826

Johnson R Jr (1988) Scalp-recorded P300 activity in patients following unilateral temporal lobectomy. Brain 111: 1517–1529

Judd LL, McAdams L, Budnick B, Braff DL (1992) Sensory gating deficits in schizophrenia: new results. Am J Psychiatry 149: 488–493

Keshavan MS, Reynolds CF, Kupfer DJ (1990) Electroencephalographic sleep in schizophrenia: a critical review. Comp Psychiatry 30: 34–47

Klorman R, Brumaghim JT (1991) Stimulant drugs and ERPs. Electroencephalogr Clin Neurophysiol 42[Suppl]: 135–141

Koenig T, Lehmann D (1996) Microstates in language-related brain potential maps show noun-verb differences. Brain Language 53: 169–182

Koukkou M, Angst J, Zimmer D (1979) Paroxysmal EEG activity and psychopathology during treatment with clozapine. Pharmacopsychiatria 12: 173–183

Koukkou M, Lehmann D, Wackermann J, Dvorak I, Henggeler B (1993) Dimensional complexity of EEG brain mechanisms in untreated schizophrenia. Biol Psychiatry 33: 397–407

Koukkou M, Lehmann D, Federspiel A, Merlo MCG (1995) EEG reactivity and EEG activity in never-treated acute schizophrenics, measured with spectral parameters and dimensional complexity. J Neural Transm Gen Sect 99: 89–102

*Koukkou-Lehmann M (1987) Hirnmechanismen normalen und schizophrenen Denkens. Springer, Berlin Heidelberg New York

Lehmann D (1971) Multichannel topography of human alpha EEG fields. Electroencephalography Clin Neurophysiol 31: 439–449

**Lehmann D (1987) Principles of spatial analysis. In: Gevins AS, Rémond A (eds) Methods of analysis of brain electrical and magnetic signals. Handbook of electroencephalography and clinical neurophysiology, vol I (revised). Elsevier, Amsterdam, pp 309–354

Lehmann D (1995) Brain electrical microstates, and cognitive and perceptual states. In: Kruse P, Stadler M (eds) Ambiguity in mind and nature: multistable cognitive phenomena. Springer, Berlin Heidelberg New York

Lehmann D, Michel CM (1990) Intracerebral dipole source localization for FFT power maps. Electroencephalogr Clin Neurophysiol 76: 271–276

Lehmann D, Ozaki H, Pal I (1987) EEG alpha map series: brain micro-states by space-oriented adaptive segmentation. Electroenceph Clin Neurophysiol 48: 609–621

Lehmann D, Wackermann J, Michel CM, Koenig T (1993) Space-oriented EEG segmentation reveals changes in brain electric field maps under the influence of a nootropic drug. Psychiatry Res Neuroimaging 50: 275–282

Leonhard K (1986) Aufteilung der endogenen Psychosen und ihre differenzierte Aetiologie. Akademie, Berlin

Lopes Da Silva FH, Hoeks A, Smits H, Zetterberg LH (1974) Model of brain rhythmic activity. Kybernetik 15: 27–37

*Maurer K (1993) Akustisch evozierte Potentiale (AEP) und ereigniskorrelierte Potentiale (P300). In: Lowitsch K, Maurer K, Hopf HC, Tackmann W, Claus D (eds) Evozierte Potentiale bei Erwachsenen und Kindern. Thieme, Stuttgart, pp 142–212

Maurer K, Dierks T, Strik WK, Frölich L (1990) P3 topography in psychiatry and psychopharmacology. Brain Topogr 3: 79–84

*McCarley RW, Shenton ME, O'Donnel BF, et al (1993) Auditory P300 abnormalities and left posterior superior temporal gyrus volume reduction in schizophrenia. Arch Gen Psychiatry 50: 190–197

Michel CM, Koukkou M, Lehmann D (1993) EEG reactivity in high and low symptomatic schizophrenics, using source modelling in the frequency domain. Brain Topogr 5: 389–394

Morstyn R, Duffy FH, McCarley RW (1983) Altered P300 topography in schizophrenia. Arch Gen Psychiatry 40: 729–734

Müller TJ, Thome J, Chiaramonti R, Dierks T, Frölich L, Scheubeck M, Strik WK (1997) A comparison of qEEG and HMPAO-SPECT in relation to the clinical severity of Alzheimer's disease. Eur J Psychiatr Neurol Sci 247: 259–263

Näätänen R (1990) The role of attention in auditory information processing as revealed by event-related potentials and other brain measures of cognitive function. Behav Brain Sci 13: 201–288

Nunez PL (1989) Estimation of large scale neocortical source activity with EEG surface Laplacians. Brain Topogr 2: 141–154

*Olbrich HM (1989) Ereigniskorrelierte Potentiale (EKP). In: Stöhr M, Dichgans J, Diener HC, Buettner UW (eds) Evozierte Potentiale. Springer, Berlin Heidelberg New York, pp 513–587

Olbrich HM, Hohagen F, Lis S, Krieger S (1996) Ereigniskorrelierte Potentiale als Korrelate kognitiver Störungen bei Wahnstörungen und Schizophrenie. Nervenarzt 67[Suppl]: 92

*Pascual-Marqui RD, Michel CM, Lehmann D (1994) Low resolution electromagnetic tomography: a new method for localizing electrical activity in the brain. Int J Psychophysiol 18: 49–65

Picton TW (1992) The P300 wave of the human event-related potential. J Clin Neurophysiol 9: 456–479

Pritchard WS (1986) Cognitive event-related potential correlates of schizophrenia. Psychol Bull 100: 43–66

Raine A (1989) Evoked potentials and psychopathy. Int J Psychophysiol 8: 1–16

Rechtschaffen A, Kales A (eds) (1968) A manual of standardized terminology, techniques and scoring systems for sleep stages of human sleep. Public Health Service, Washington, DC

Rockstroh B, Elbert T, Canavan A, Lutzenberger W, Birbaumer N (1989) Slow cortical potentials and behaviour, 2nd edn. Urban and Schwarzenberg, Baltimore

Röschke J, Aldenhoff JB (1992) A nonlinear approach to brain function: deterministic chaos and sleep EEG. Sleep 15: 95–101

*Röschke J, Aldenhoff JB (1993) Estimation of the dimensionality of sleep-EEG data in schizophrenics. Arch Psychiatry Clin Neurosci 242: 191–196

Rösler, F, Manzey D, Sojka B (1985) Delineation of pharmacopsychological effects by means of endogeneous event-related brain potentials: an exemplification with flupentixol. Neuropsychobiology 13: 81–92

*Saletu B (1989) EEG imaging of brain activity in clinical psychopharmacology. In: Maurer K (ed) Topographic brain mapping of EG and evoked potentials. Springer, Berlin Heidelberg New York, pp 482–506

Schuckit MA (1994) Low level of response to alcohol as a predictor of future alcoholism. Am J Psychiatry 151: 184–189

Schulz H (ed) (1997) Kompendium der Schlafmedizin. Ecomed, Landsberg

Shagass C (1987) Deviant cerebral functional topography as revealed by electrophysiology In: Helmchen H, Henn FA (eds) Biological perspectives of schizophrenia. Wiley, New York, pp 237–253

Simons RF (1982) Physical anhedonia and future pathology: an electrocortical continuity? Psychophysiology 19: 433–441

Stanzione P, Fattapposta F, Tagliati M, D'Alessio C, Marciani MG, Foti A, Amabile G (1990) Dopaminergic pharmacological manipulations in normal humans confirm the specificity of the visual (PERG-VEP) and cognitive (P300) electrophysiological alterations in Parkinson's disease. Electroencephalogr Clin Neurophysiol 41[Suppl]: 216–220

Strik WK, Lehmann D (1993) Data-determined window size for space oriented segmentation of EEG-map series. Electroencephal Clin Neurophysiol 87: 169–174

Strik WK, Dierks T, Franzek E, Maurer K, Beckmann H (1993a) Differences in P300 amplitudes and topography between cycloid psychosis and schizophrenia in Leonhard's classification. Acta Psychiatr Scand 87: 179–183

Strik WK, Dierks T, Maurer K (1993b) Amplitudes of auditory P300 in remitted and residual schizophrenics: correlations with clinical features. Neuropsychobiology 27: 54–60

Strik WK, Dierks T, Franzek E, Stöber G, Maurer K (1994a) P300 in schizophrenia: interactions between amplitudes and topography. Biol Psychiatry 35: 850–856

Strik WK, Dierks T, Franzek E, Stöber G, Maurer K (1994b) P300 asymmetries in schizophrenia revisited with reference-independent methods. Psychiatry Res Neuroimaging 55: 153–166

Strik WK, Lehmann D, Dierks T, Becker T (1995) Larger topographical variance and decreased duration of brain electric microstates in depression. J Neural Transm Gen Sect 99: 213–222

Strik WK, Dierks T, Kulke H, Maurer K, Fallgatter AJ (1996) The predictive value of auditory P300 on the course of schizophrenia. J Neural Transm 103: 1351–1359

Strik WK, Fallgatter AJ, Stöber G, Franzek E, Beckmann H (1997) Specific features of auditory P300 in cycloid psychosis. Acta Psychiatr Scand 95: 67–72

*Strik WK, Fallgatter AJ, Brandeis D, Pascual-Marqui RD (1998) Three dimensional tomography of event-related potentials during response inhibition: evidence for phasic frontal lobe inhibition. Electroencephalogr Clin Neurophysiol 108: 406–413

**Strik WK, Ruchsow M, Fallgatter AJ, Mueller TJ (1998) Distinct neurophysiological mechanisms for manic and cycloid psychoses: evidence from a P300 study on manic patients. Acta Psychiatr Scand 98: 459-466

Timsit-Berthier M (1986) Contingent negative variation (CNV) in psychiatry. In: McCallum WC, Zappoli R, Denoth F (eds) Cerebral psychophysiology: studies in event-related potentials. Elsevier, Amsterdam, pp 429–438

Towey J, Bruder G, Hollander E, Friedman D, Erhan H, Liebowitz M, Sutton S (1990) Endogenous event-related potentials in obsessive-compulsive disorder. Biol Psychiatry 28: 92–98

van den Bosch RJ, Rozendaal N, Mol JM (1988) Slow potential correlates of frontal function, psychosis, and negative symptoms. Psychiatry Res 23: 201–208

Volavka J, Czobor P, Goodwin DW, Gabrielli WF Jr, Penick EC, Mednick SA, Jensen P, Knop J (1996) The electroencephalogram after alcohol administration in high-risk men and the development of alcohol use disorders 10 years later. Arch Gen Psychiatry 53: 258–263

Wackermann J, Lehmann D, Dvorak I, Michel CM (1993) Global dimensional complexity of multichannel EEG indicates change of human brain functional state after a single dose of a nootropic drug. Electroencephalogr Clin Neurophysiol 86: 193–198

Wikswo JP, Gevins A, Williamson SJ (1993) The future of the EEG and MEG. Electroencephalogr Clin Neurophysiol 87: 1–9

Wilson MA, McNaughton BL (1993) Dynamics of the hippocampal ensemble code for space. Science 261: 1055–1058

Woods DL (1992) Auditory selective attention in middle-aged and elderly subjects: an event-related potential study. Electroencephalogr Clin Neurophysiol 84: 456–468

CHAPTER

10

B. Bogerts, P. Falkai

Neuroanatomical and Neuropathological Basis of Mental Illness

Translator: E. Taub

1 Study of Brain Structure in Psychiatry

The role of neuropathology in the study of classical psychiatric diseases such as the schizophrenias, affective disorders, personality disorders, anxiety disorders, and obsessive-compulsive disorder was until recently highly controversial. Neuropathological findings have been largely impossible to demonstrate in psychiatric disease, except in the case of organic mental syndromes resulting from focal brain lesions or degenerative processes. A few structural abnormalities have been detected to date, but only with difficulty, with the aid of highly sophisticated techniques.

Neuropathological substrates for the schizophrenias and affective psychoses were long thought not to exist. Recent research, however, benefiting from statistical morphometric techniques and new histochemical techniques for postmortem analysis, as well as from in vivo techniques of structural and functional imaging, has revealed the presence of moderate abnormalities in certain brain areas in many patients. These macroscopic and histological abnormalities are not as extensive as those found in brain illnesses that are known to be organic; they are relatively inhomogeneous, both in type and in localization, and are found in only a subset of patients carrying the diagnosis of an "endogenous psychosis." Sophisticated statistical methods are often needed to demonstrate a difference between a group of psychiatric patients and a normal control group. There is generally considerable overlap between the two groups, and the morphologic parameters of many patients lie within the normal range. No pathognomonic abnormalities of brain tissue have yet been found that might be used for a reliable differential diagnosis of schizophrenic and affective disorders, as these are defined in the current systems of disease classification. The pathological findings in this area are thus of an entirely different nature from those familiar to us from the realm of neurology and cerebral degenerative disease.

Many studies of brain structure in schizophrenic patients have been performed to date, but the affective disorders have been the subject of almost no neuropathological studies, and of relatively few structural imaging studies. There have been no neuropathological studies at all to date of the so-called neuroses or personality disorders, which is not surprising, as it is commonly assumed that these result from psychosocial factors or variations of normal personality traits (although genetic factors may also be important). It is quite conceivable that some structural features of the brain may represent "vulnerability factors" for the latter class of disorders; no postmortem studies have yet been performed to provide evidence for or against this hypothesis.

There are several reasons why discrete, homogeneous neuropathological substrates for the named psychiatric disorders are so difficult to find. The first reason is that the schizophrenias and affective psychoses, unlike the classical organic mental disorders, are not unitary diseases but, rather, conventional diagnostic constructs, whose neurobiological causes may be as diverse as their clinical manifestations. In psychiatry as in internal medicine, a given constellation of signs and symptoms may result from one of multiple biological substrates that exert their effects through a final common pathway (as is true of fever or hypertension, for example).

It is also possible that the brains of many of these patients are entirely normal in structure, and that their mental illnesses are caused by reversible neurochemical disturbances or by stress-induced changes in neurotransmitter or neurohormonal systems. Another reason is that the histopathological substrates of the typical psychiatric diseases may be too subtle to be detected by the traditional qualitative methods of neuropathological research or that they may lie in brain areas or cell types that have not yet been investigated. The functional importance of the limbic system, for example, was not fully recognized until the 1950s (McLean 1952), and it has only recently become a major object of investigation in the neuropathology and pathophysiology of mental disorders. The neurotransmitter systems of the brain were discovered in the 1960s (Dahlström and Fuxe 1964), and the cell types containing neuropeptides 10 years later (for a review, see Nieuwenhuis 1985). There has been no systematic histopathological study of these cell types to date, despite their major relevance to the theory and pharmacology of mental diseases.

One possible technical reason for the slowness of progress in this area was the lack, until recently, of adequate neuropathological techniques. It was only after the introduction of immunohistochemistry and in situ hybridization that neuroglial subpopulations could be conveniently studied and observations regarding their function could be made (Table 1). These techniques were the first to add a functional dimension to the study of structural abnormalities, yielding important clues both to disease etiology (e.g. the expression of gene products) and to normal function (e.g. the characterization of inhibitory interneurons).

Another reason for the lagging interest in the neuropathology of mental illness is the conspicuous lack of success of neuropathological research on schizophrenia in the first half of the twentieth century. At that time, a dualistic attitude toward mind and

Table 1. New research techniques for postmortem study of the brain

Technique	Object under study
Stereology	Absolute cell number in a brain structure
Immunohistochemistry	Protein expression in brain sections
Western blot	Protein expression in brain homogenate
In situ hybridization	mRNA expression in brain sections
RT-PCR	mRNA expression in brain homogenate

RT-PCR, reverse transcription polymerase chain reaction.

brain was widespread among psychiatrists, in view of the prevailing lack of knowledge of many aspects of the central nervous system. Since then, however, modern brain research has yielded a great deal of insight into the functional architecture of the brain (Fig. 1).

A historical overview of neuropathological research in mental illness, documenting the gradual abandonment of the dualistic viewpoint, is provided in Tables 2 and 3.

The following discoveries stand out among all those that have led to an improved understanding of the neurobiology of psychotic syndromes:

- Hess, in 1949, discovered that elementary drives and emotions can be evoked by direct electrical stimulation of the diencephalon.
- McLean, in 1952, recognized the importance of the limbic system in the neural modulation of feelings and emotions.
- Dahlström and Fuxe, in 1964, localized the intracerebral neurotransmitters dopamine, norepinephrine, and serotonin.
- Jones and Powel, in 1970, described the neuroanatomical basis of the cortical integration and association of sensory input.
- The first integrated, multidimensional model of a mental symptom taking neurophysiological, psychological, ethological, and pharmacological aspects into account was proposed by Gray (for anxiety).

Finally, it must be noted that the introduction of computed tomography (CT) and magnetic resonance imaging (MRI), and the subsequent discoveries of structural abnormalities in psychiatric diseases (e.g. Johnstone et al. 1976), were a major stimulus to the performance of further studies of brain structure in this area. Since then, MRI has developed further and given rise to the two related techniques of magnetic resonance spectroscopy (MRS) and functional magnetic resonance imaging (fMRI), which yield information on brain biochemistry and functional activity, respectively, and help put the structural findings into a functional context.

2 Functional Neuroanatomical Concepts

2.1 The Limbic System and Mental Illness

The limbic system plays a central role in the neuropathology and pathophysiology of the schizophrenias, affective disorders, secondary psychoses, anxiety disorders, Alzheimer's disease, and the amnestic syndrome. The anatomy of the most important components of the limbic system is shown in Fig. 2. Information from the entire association cortex converges upon the limbic system, where it is then processed in coordination with other signals arising

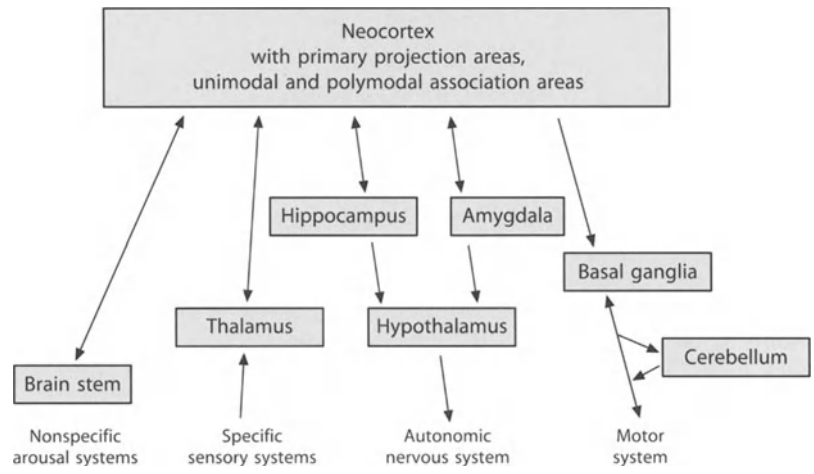


Fig. 1. Functional systems of the human brain

Table 2. Dates in the history of neuroanatomy and psychiatry

Year	Event
Sixth to fourth century B.C.	Presocratics: Democritus: thinking and perception have a material basis Hippocrates: epilepsy and psychoses are produced by irritation of the brain
Fourth century B.C.	Aristotle: the heart is the seat of reason and perception; the brain functions as a cooling organ and elaborates nasal mucus
Eighth century A.D.	Galen: discovery of the cerebral ventricles
1662	Descartes: the pineal gland as nexus of body and soul
Late eighteenth century	Gall: phrenology and cranioscopy Flourens: holistic vs. regionalistic views of the brain Psychiatry: somatic vs. physical concepts
1822	Bayle: description of progressive paralysis
1861	Broca: the case of Leborgne (motor aphasia)
1868	Harlow: the case of Phineas Gage (frontal brain syndrome)
1867	Griesinger: mental illnesses are brain illnesses
1874	Darwin: body and mind are both products of biological evolution
Approx. 1900	Cajal, Golgi, Nissl: the study of neurons (Nobel Prize 1906) Sherrington: the concept of the synapse Alzheimer: description of the neuropathological basis of the dementias, epilepsies, progressive paralysis, and psychoses Kraepelin: systematic classification of psychiatric disease Freud: psychoanalysis Pavlov, Thorndike: classical and operant conditioning, foundations of ethology and behavioral therapy
1949	Hess: intrahypothalamic stimulation (release of elementary drives and emotions by electrical stimulation of the brain)
1952	McLean: concept of the limbic system First International Congress of Neuropathology Introduction of antipsychotic neuroleptic drugs
1965	Dahlström and Fuxe: discovery of intracerebral neurotransmitter systems
1970	Jones and Powel: the neuroanatomical basis of cortical information processing
1975	Hughes and Kosterlitz: discovery of the endorphins (neuropeptide hormones)
Since approx. 1980	Structural and functional imaging of the brain by computed tomography (CT), magnetic resonance imaging (MRI), positron emission tomography (PET), and single photon emission computed tomography (SPECT)
Since approx. 1990	Functional magnetic resonance imaging (fMRI) and magnetoencephalography (MEG)

from the brain stem. The central limbic structures form a ring ("limbus") around the corpus callosum on the medial surface of the cerebral hemisphere. The key limbic structures of the medial temporal lobe are sites of convergence for information from the entire higher association cortex. The hippocampus and amygdala project to the phylogenetically oldest parts of the brain stem and the septal nuclei. All of these pathways are bidirectional, i.e. the temporal limbic areas project back to the association cortex, and the septal nuclei and diencephalon project back to the hippocampus and amygdala. Structural and functional lesions at these sites may give rise to psychiatric syndromes.

The basic principles of cortical information processing will be discussed next, as a knowledge of these is indispensable for an understanding of the neurophysiological basis of psychiatric disease (see also Fig. 3).

All sensory information originates in the peripheral sensory organs, passes through specific sensory pathways (such as the visual, auditory, and somatosensory pathways), and then continues, through a thalamic relay, into the primary sensory cortical areas (visual cortex, auditory cortex, somatosensory cortex; see Fig. 3, *1a–c*). Sensory information then passes into the secondary, unimodal sensory association areas, which

Table 3. Historical aspects of neuropathological research on schizophrenia

Year	Research
Approx. 1900	Cortex (Kraepelin, Alzheimer, Nissl)
1930–1960	Thalamus, basal ganglia (C. and O. Vogt, Fünfgeld, Hopf, Hempel, Treff)
1927–1969	Pneumencephalography, ventricles (Jacobi and Winkler, Huber)
1952	First International Congress of Neuropathology in Rome "Schizophrenia is the graveyard of neuropathology" (Plum)
Approx. 1960–1976	Abandonment of neuropathological research on schizophrenia, dominance of psychoanalytic, biochemical, and social theories, antipsychiatry
1976–	Ventricular dilatation discovered by CT (e.g., Johnstone, Weinberger)
Since 1983	Morphometric postmortem studies of the dopaminergic system, limbic system, basal ganglia, thalamus (e.g., Bogerts, Pakkenberg, Benes)
Since 1984	Cytoarchitectural studies (e.g., Kovelman and Scheibel, Benes, Jakob and Beckmann, Falkai, Arnold)
Since 1989	MRI volumetry of the ventricular system, limbic system, cortex, thalamus (e.g., Suddath, Andreasen, Bogerts, Degreef)
Since 1990	Studies of cerebral structural asymmetry (e.g., Crow, Falkai, Bilder)
Since 1995	Immunohistological studies of cell types and cellular components

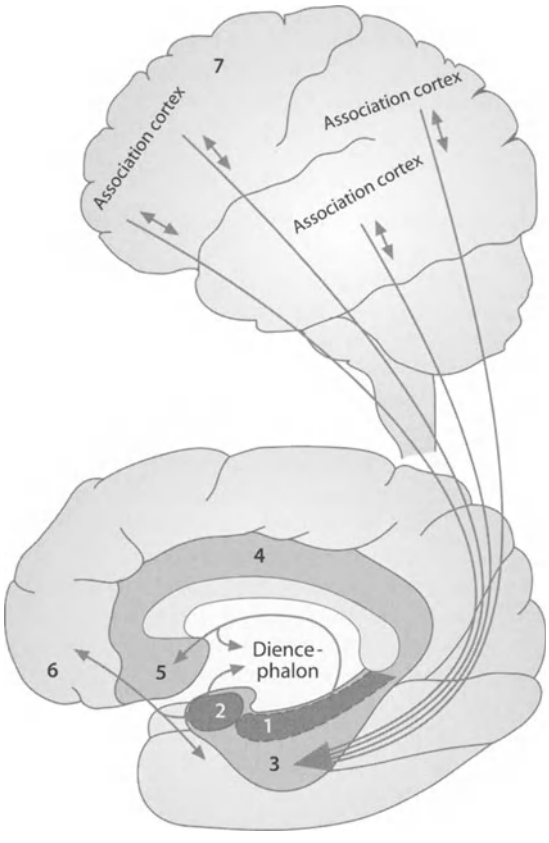


Fig. 2. Anatomical relationships of the limbic telencephalic structures and their most important fiber pathways. *1*, hippocampus; *2*, amygdala; *3*, parahippocampal and endorhinal cortex; *4*; cingulate gyrus; *5*, septum and posterior orbital cortex; *6*, mediobasal frontal lobe; 7, dorsolateral frontal cortex

are adjacent to the primary cortex (*2a–c*). In these unimodal association areas, incoming sense impressions are processed at a higher level of abstraction, so that their underlying properties become evident – this is the level at which pattern recognition occurs, so that words or spatial patterns, for example, can be discerned. "Associated information" of this type, generated in the unimodal association areas, then passes to the tertiary, polymodal cortical association areas of the frontal, parietal, and temporal lobes. In these neocortical areas, information derived from multiple sensory modalities (e.g. visual and auditory) converges onto single neurons. This fact is the neuroanatomical basis of the unitary nature of sensory perception, which comes about despite the segregation of peripheral sensory channels into different modalities. The tertiary association cortex projects further to supramodal cortical association regions in the prefrontal and temporal cortex, in which, on a still higher level, the incoming information is integrated and associated with previously acquired information (*3*).

After this cascade-like traversal of the unimodal, polymodal, and supramodal association cortices, all information finally converges upon the key limbic structures of the medial temporal lobe, the hippocampus and amygdala (*4*). The latter structures influence the activity of the hypothalamus (*5*), which controls the autonomic (sympathetic and parasympathetic) centers of the medulla (*6*). The relevant fiber pathways are shown in Fig. 3.

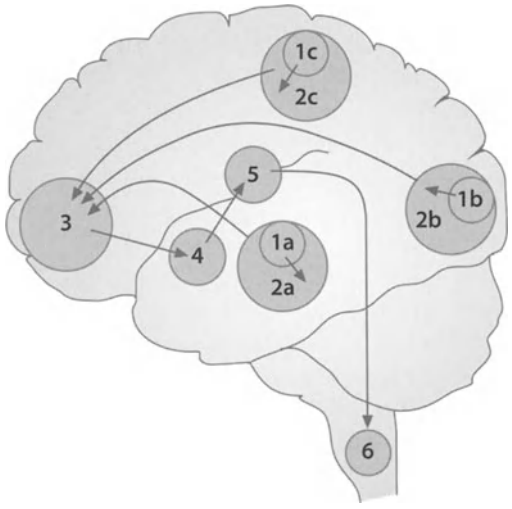


Fig. 3. Anatomic sites of cerebral information processing. *1a–c*, primary sensory cortical area: *a* primary auditory cortex, *b* primary visual cortex, *c* somatosensory cortex; *2a–c*, unimodal sensory cortical association areas: *a* auditory, *b* visual, *c* somatosensory; *3*, polymodal cortical association areas; *4*, temporal limbic structures (supramodal association areas); *5*, hypothalamus; *6*, autonomic brain stem centers (parasympathetic, sympathetic); *arrows* in cortex, short and long association fibers; *pathways* marked *4* and *5*, fornix, stria terminalis, ventral amygdalofugal pathway; *pathways* marked *5* and *6*, dorsal longitudinal fasciculus)

The limbic structures play a crucial role in the filtering out of irrelevant information, or "sensory gating." They work in a coordinated fashion with the cortical sensory areas that lie proximal to them along the afferent pathway, comparing past and present incoming sensory data and evaluating these data for emotional relevance. Whatever information is emotionally relevant goes on to influence further brain activity, while superfluous information is deleted (Mesulam 1986; van Hoesen 1982; Swanson 1983; Gray 1982; Millner 1992).

In addition, the same central limbic structures of the temporal lobe stimulate or inhibit the activity of the hypothalamus through several pathways. The hypothalamus contains the neural generators of the drives and emotions, which constitute a phylogenetically ancient sphere of brain function (Palkovits and Zaborski 1979). Elementary drives such as aggression, flight, and sexuality, which may indeed be evoked by direct electrical stimulation in the area of the hypothalamus and septum (Hess 1949), are mediated by input to the hypothalamus arising from the amygdala and hippocampus and thus ultimately depend on information that the latter structures receive from the sensory association cortex.

The hypothalamus, in turn, exerts its effects on both parasympathetic and sympathetic functions through a long, descending autonomic pathway, the Edorsal longitudinal fasciculus. The latter delivers afferent input to the dorsal nucleus of the vagus, which provides parasympathetic innervation to almost the entire body, as well as to medullary sympathetic centers. The hypothalamus thus controls all aspects of peripheral autonomic function (Palkovits and Zaborski 1979; Nieuwenhuis 1985).

This completes the chain of sensory information processing, beginning at the sense organs and proceeding, by way of cortical association and integration areas, to further processing by the limbic system, activation or inhibition of the hypothalamus, and generation of peripheral sympathetic or parasympathetic reactions.

The dorsal longitudinal fasciculus, defined above as a fiber bundle connecting the limbic–hypothalamic system, which subserves emotion, to the lower medullary nuclei, which give rise to the parasympathetic and sympathetic outflow, is thus not only an important part of the regulating mechanism for endocrine secretion, but also, in a way, a locus of coupling between the psyche of the individual and his or her peripheral somatic responses. This tract may be considered a "psychosomatic pathway" that provides the anatomical substrate for the regulation of sympathetic and parasympathetic responses by higher cerebral processes.

2.2
The Environment and Brain Structure: Neuroplasticity

The neuropathological approach to the study of psychiatric disease makes exclusive use of reductionist strategies, through which the clinical manifestations of disease may be traced back to primary neurobiological disturbances. Nevertheless, it must be borne in mind that even an initially healthy brain may develop pathological neuroanatomical abnormalities as a result of exposure to a pathogenic psychosocial environment. Studies of brain plasticity have conclusively revealed that an abnormal constellation of environmental stimuli may lead to a lasting impairment of brain development, in both functional and structural terms. The appearance of irreversible changes of brain structure and function after sensory deprivation, particularly in the early phases of postnatal development, is an illustration of this phenomenon (Braun 1996).

In numerous deprivation studies carried out over the last 20 years, it has been shown that the development of cerebral structure and function depends not only on genetic factors, but also, to a large extent, on environmental influences (Singer 1991; Roth 1991; Fields and Nelson 1992). These findings apply to other brain systems as well. The following principles of

cerebral plasticity can be inferred from the results of such studies:

- The functional inactivation of brain systems by sensory deprivation in early postnatal critical developmental phases leads to irreversible, or only partly reversible, structural and functional damage.
- Brain systems that are underdeveloped as a result of previous deprivation and are then reactivated after the end of the critical phase can recover normal function only to a minimal extent or not at all. Conversely, brain systems that have developed normally in the presence of adequate stimulation and are then inactivated by deprivation after the end of the critical phase suffer only mild and reversible functional damage.
- The critical period, i.e. the period in which a cerebral system is vulnerable to irreversible damage as a result of functional deprivation, is different for each individual system. It is determined by the stage of postnatal development in which that system undergoes its genetically determined maturation and the corresponding functional abilities are acquired (e.g. the third to sixth postnatal week for binocular vision in primates).

Similar effects after early deprivation have been demonstrated at the receptor level for the corticosteroid and corticotropin-releasing hormone (CRH) system (for a review, see Aldenhoff 1997).

Deprivation studies have shown that complete recovery of a damaged cerebral system can no longer be achieved after the end of the critical period, when plasticity is severely limited. This explains the lack of effective therapies for mental disorders caused by massive early deprivation (Bogerts 1996). The subtle structural neuroanatomical defects often found in mentally ill patients may be thought of, in relation to brain plasticity, as the possible result of inadequate early sensory stimulation.

Neural plasticity is relevant to neuropsychiatric disease not only in the context of deprivation in early childhood, but also in the response to injury of the mature, normally developed brain. Both neuronal and oligodendroglial mechanisms are involved in the coordination of synaptic plasticity and axon regrowth that lead, if successful, to the restoration of normal function. A typical marker for neural plasticity is growth-associated protein 43 (GAP-43), a substance found in high concentration in synapses that has been shown to be important both in brain development and in the response to brain injury (Benowitz and Perrone-Bizzozero 1991). An autopsy study of schizophrenic patients revealed elevated GAP-43 immunoreactivity in the frontal cortex as well as a decreased concentration of the synaptic marker synaptophysin (Perrone-Bizzozero et al. 1996). This finding might be interpreted as reflecting an attempt by the brain to activate its reserve of plasticity in response to a previous down-regulation of synaptic activity.

3 Organic Mental Syndromes

3.1 Psychoses Resulting from Focal Brain Lesions

Lesions of the primary motor or sensory brain areas, or of the extrapyramidal motor system, cause neurological signs and symptoms, while disturbances of the higher association cortex or of the limbic system cause psychiatric signs and symptoms. In contrast, lesions of the simple unimodal cortical association areas, which occupy an intermediate anatomical and functional position between the primary sensory and motor cortices and the higher association cortex, are clinically manifest in the area of overlap between neurology and psychiatry (e.g. aphasia, apraxia, agnosia).

The frontal and temporal lobes consist predominantly of polymodal and supramodal association cortex. This explains why even large lesions in these areas, such as tumors, infections, trauma, infarcts, and degenerative changes, may not be associated with any motor or sensory neurological deficits (with the exception of anosmia caused by frontal tumors or trauma), while frontal and temporal syndromes typically do include a disturbance of higher cortical or limbic functioning.

Damage to the frontal lobe of any etiology may cause personality changes including loss of motivation, apathy, impairment of judgment, loss of anticipatory ability, social disinhibition or withdrawal, and psychomotor retardation.

Destruction of temporal limbic areas causes amnesia or psychotic manifestations. Total, bilateral hippocampal dysfunction renders the patient incapable of remembering new events, because the transition from short- to long-term memory is no longer possible. Bilateral lesions of the amygdala are clinically manifest as the Klüver-Bucy syndrome, which is characterized by the inability to assign an emotional value to sensory percepts, often in association with hypersexuality and diminished aggressiveness (Mesulam 1986).

Lesions of temporal limbic structures of lesser severity, such as the early stages of tumors and infections, are often associated with manifestations resembling those of schizophrenia. Viral infections with a high affinity for the medial temporal lobe, such as herpes simplex encephalitis or rabies, cause (in their early stages) severe emotional changes including anxiety, fearfulness, overreactions, aggressiveness or apathy, abnormal sexual behavior, delusions, and

Table 4. Distribution of brain lesions in schizophreniform psychoses

Localization	Brain injuries (%) (Hillbom 1951)	Brain tumors (%) (Davison and Bagley 1969)
Temporal lobe	40	35
Frontal lobe	23	19
Parietal lobe	14	–
Occipital lobe	8	2
Diencephalon	–	19
Cerebellum	–	6
Basal ganglia	–	1
Brain stem	–	3
Other areas	15	15

hallucinations (Greenwood et al. 1983). The same phenomena may be produced by trauma, tumors, or ischemia of the medial temporal lobe (Davison and Bagley 1969; Hillbom 1951) or by temporal lobe epilepsy (Slater et al. 1963), particularly when the focus is on the left side and the causative lesion is congenital (Flor-Henry 1969; Perez et al. 1984). Such affections of the limbic or paralimbic regions of the temporal or frontal lobes are often misdiagnosed, in their early stages, as schizophrenia or affective psychosis.

Table 4 provides an overview of the distribution into various brain areas of tumors and injuries that may give rise to schizophreniform psychoses. The frontal and temporal lobes are most often involved, i.e. those areas of the brain in which most of the limbic and paralimbic structures lie (including the hippocampus, parahippocampal cortex, amygdala, temporal pole, cingulate gyrus, and orbital cortex).

3.2 Alcoholic Brain Injury

The injurious effect of chronic alcoholism on brain tissue was described by Wernicke as early as 1881. The best known neuropathological sequela of alcoholism is Wernicke's encephalopathy, which is characterized by degenerative changes, gliosis, and small areas of hemorrhage, mainly in the diencephalon and midbrain. The diencephalic limbic structures bordering on the third ventricle and the cerebral aqueduct, which bear a close relation to the limbic telencephalon, are prominently involved, as is the cerebellum.

The clinical picture of the Wernicke-Korsakoff syndrome, consisting of amnesia, oculomotor dysfunction, ataxia, emotional blunting, and disorientation, is the result of alcoholic damage to the diencephalon, midbrain, and cerebellum. Other neuropathological changes occurring in chronic alcoholism include central pontine myelinolysis, Marchiafava syndrome, and fetal alcohol syndrome (for an overview, see Victor et al. 1989; Phillips et al. 1987; Mann and Widmann 1995).

Recent morphologic studies of the cerebellum in alcoholics revealed a significant diminution of the number of Purkinje cells and thinning of the molecular and granular cell layers. Cerebellar atrophy was seen in approximately half of the alcoholics examined. Alcohol alone can cause cerebellar damage; Wernicke's encephalopathy arises from the combination of alcohol toxicity with thiamine deficiency, a result of poor nutrition (Victor et al. 1989).

The cortical sulci and the cerebral ventricles are often found to be widened on CT and MRI scans of alcoholics (Mann et al. 1989). At the microscopic level, a diminished cell density in the frontal cortex and a diminished single-cell volume in the cingulate, motor, and temporal cortex can be observed. The cortical atrophy is reversible to some extent by prolonged abstinence (Mann and Widmann 1995), especially in younger individuals and those in whom the dependency was of brief duration. Nevertheless, the cerebellar and cerebral cortical cell loss, and the structural changes seen in the diencephalon and midbrain in Wernicke's encephalopathy, are irreversible.

The question of whether regular, but moderate alcohol intake causes brain injury is of great interest. The definition of moderate alcohol intake is largely relative, depending on sociocultural factors, but is usually taken, for scientific purposes, as 40–80 g of ethanol daily. An autopsy study of moderate drinkers revealed an insignificant trend toward cerebral atrophy and a significant degree of dendritic retraction of cortical neurons (Harper et al. 1988). The finding of a focal reduction of volume of the prefrontal cortex in alcoholics and in multiple-substance abusers is of interest (Liu et al. 1998) and has since been replicated. Volume loss appears to be restricted to this region and may have been present before the onset of alcohol abuse, i.e. it may not be an effect of alcohol toxicity.

3.3 Alzheimer's Disease

In view of the extensive literature on the neuropathological features of Alzheimer's disease, including several excellent reviews (Henderson and Henderson 1988; Maurer et al. 1990; Bauer 1994; Kurz 1995), we will not provide a detailed discussion here (see also Vol. 2, Part 2, Chaps. 5–7). We will, however, mention some recent advances that allow neuropathological staging of the disease and improved diagnosis with imaging techniques.

The two neuropathological hallmarks of Alzheimer's disease are the amyloid-containing senile plaques, which lie in the intracellular space, and the intraneuronal neurofibrillary tangles. Plaques and tangles are distributed differently in the brain. While plaques are ubiquitous in the cerebral cortex, tangles are found almost exclusively in the parahippocampal region in early stages of the disease. They then gradually spread to the hippocampus, the rest of the limbic and paralimbic cortex, and, finally, the entire brain, while remaining most numerous in the mesiotemporal limbic area (Braak and Braak 1991).

Plaques and tangles differ at the molecular level as well. The central, amyloid portion of plaques consists of aggregated β-A4 protein, which is formed by pathological cleavage of the amyloid precursor protein (APP), a protein encoded by a gene located on chromosome 21. The formation of plaques takes place over decades and precedes the clinical onset of disease by a long interval.

The major component of neurofibrillary tangles is τ protein (tau protein), whose normal function is thought to be in the stabilization of microtubules, an important substrate for intracellular transport within neurons. The pathological phosphorylation of τ protein results in the formation of neurofibrillary tangles and the disruption of neuronal intracellular transport.

The severity of dementia is more closely correlated with the spatial extent of neurofibrillary tangles than with the number of plaques.

The major site of pathologic changes in Alzheimer's disease is the medial temporal lobe, according to the findings of both MRI studies and histopathological studies (see also this volume, Part 1, Chap. 1). MRI revealed a reduction in overall brain volume by approximately 10% and a generalized widening of the cerebrospinal fluid spaces; the hippocampus–amygdala complex underwent the most extreme reduction in volume, by 30%–40% (Pantel et al. 1997). Several authors have estimated the sensitivity of hippocampal volume reduction as a diagnostic criterion for Alzheimer's disease at between 80% and 90% (Hampel et al. 1997).

Morphometric MRI techniques are thus useful tools for the differential diagnosis of Alzheimer's disease. Nuclear medical techniques are also highly useful for this purpose. Single photon emission computed tomography (SPECT) was initially performed mainly on patients with severe clinical manifestations, and these patients had correspondingly severe abnormalities on SPECT. More recently, however, patients with "milder" disease manifestations have been scanned, including patients who have just made their first visit to a psychiatrist or neurologist because of a newly evident memory deficit. Many of these patients, too, are found to have regional disturbances of cerebral blood flow and glucose utilization.

Extensive, symmetrical temporoparietal hypoperfusion may be considered a nearly pathognomonic finding of Alzheimer's disease and is seen in about one third of patients (the percentage depends on the patient group studied; Holman et al. 1992). Magnetic resonance spectroscopy seems to provide better differential diagnosis, allowing the provision of a diagnosis in between 70% and 80% of patients (see, e.g. Heun et al. 1997).

The fact that Alzheimer's disease originates in the limbic system accounts for the frequent presence not only of memory deficits, but also of emotional changes and delusions among its initial manifestations. These features may make the differential diagnosis of Alzheimer's disease from senile depressive syndromes, paranoid states, and late schizophrenias particularly difficult.

4 The Schizophrenias

4.1 Neuroanatomy: Present State of Knowledge

At the First International Congress of Neuropathology, which took place in Rome in 1952, the view prevailed that no specific neuropathological abnormalities were present in schizophrenic patients. Subsequently, research in this area remained at a standstill until the early 1980s (see Table 3), while psychodynamic and neurochemical theories of etiology attracted the most attention.

The introduction of CT in psychiatry led to a revival of interest in possible cerebral structural abnormalities in schizophrenic patients. All aspects of imaging studies in the schizophrenias are discussed in detail in Chap. 11 (this volume, Part 1). In 1976, Johnstone and colleagues were the first to show by CT that schizophrenics have, on average, wider cerebral ventricles than normal control subjects, thus confirming the findings of earlier, pneumoencephalographic studies (Huber 1957). The earlier studies had not gained full acceptance, because the risks associated with pneumoencephalography had made it impossible to study an adequate number of normal control subjects.

After these initial CT findings were reported (Johnstone et al. 1976), there followed numerous CT and, later, MRI studies of the internal and external cerebrospinal fluid spaces, and of other subcortical structures, in patients with schizophrenic psychosis. As reported in several meta-analyses of these studies, CT reveals a dilatation of the cerebral sulci and cerebral ventricles (mean third ventricular enlargement, approximately 30%) in 30%–50% of schizophrenics

(Raz 1993; Lewis 1990), while MRI reveals a reduction of volume of the entire brain (3%), of both temporal lobes (left 6%, right 9.5%), and of the amygdala–hippocampus complex on both sides (approximately 6% bilaterally) (Lawrie and Abukmeil 1998). The dilatation of the lateral ventricles seen on CT has been confirmed by MRI (Degreef et al. 1992b); according to the most recent meta-analysis, the magnitude of this dilatation is 44% on the left side and 36% on the right side (Lawrie and Abukmeil 1998).

Another meta-analysis yielded a value of 7% for volume reduction in the hippocampus, but no evidence of volume reduction in the amygdala (Nelson et al. 1998). Anatomical differentiation between the amygdala and the hippocampus is difficult in both CT and MRI; this fact may account for the divergent findings regarding amygdalar volume loss.

Nonetheless, there is now a consensus that the sulci of higher integration and association areas in the frontal and temporal lobes are preferentially dilated in schizophrenics. Even qualitative assessment of CT or MRI images of schizophrenic patients allows recognition of ventricular or sulcal dilatation in about half of the patients (Lieberman et al. 1992; Lawrie et al. 1997; Figs. 4, 5).

The findings of recent studies of brain structure in schizophrenia may be summarized as follows:

- Inhomogeneous abnormalities of brain structure
- Dilatation of the lateral ventricles and of the third ventricle
- Mild loss of brain tissue in limbic structures and in the thalamus
- Subtle abnormalities of higher cortical association areas
- Loss of cortical structural asymmetry
- Neuroleptic-induced enlargement of the basal ganglia
- Lack of progression of tissue loss in the limbic system
- Correlation between the extent of ventricular dilatation and an unfavorable disease course

The use of quantitative rating scales (such as that of Smith et al. 1997) is helpful in the assessment of large numbers of CT or MRI images within reasonable time constraints. An especially economical feature of such scales is that they limit the necessary number of measurements to just a few important locations.

4.2 Evidence for Possible Etiologies

In addition to these abundantly replicated findings of dilatation of the internal and external cerebrospinal fluid spaces, major focuses of current interest include studies of limbic structures, the thalamus, the cerebral cortical architecture, and the basal ganglia. A further current topic is the finding of an abnormal alteration of the cortical asymmetry pattern in schizophrenia.

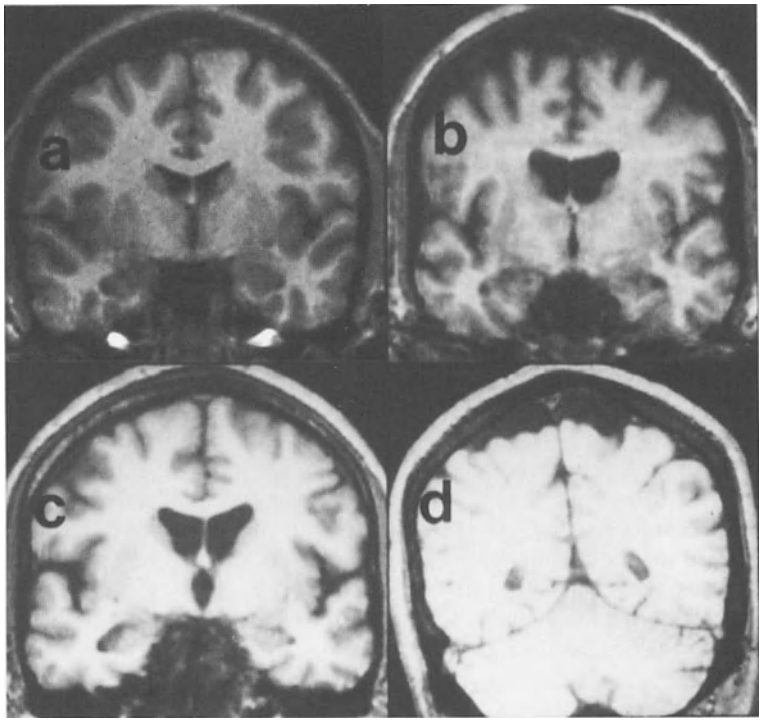


Fig. 4. Coronal magnetic resonance imaging (MRI) sections of **a** a mentally healthy individual, **b,c** two schizophrenic patients of similar ages, with lateral ventricular dilatation (**b**) and a combined dilatation of the lateral ventricles and the third ventricle (c), and **d** a schizophrenic patient with dilatation of the parietal cerebrospinal fluid space

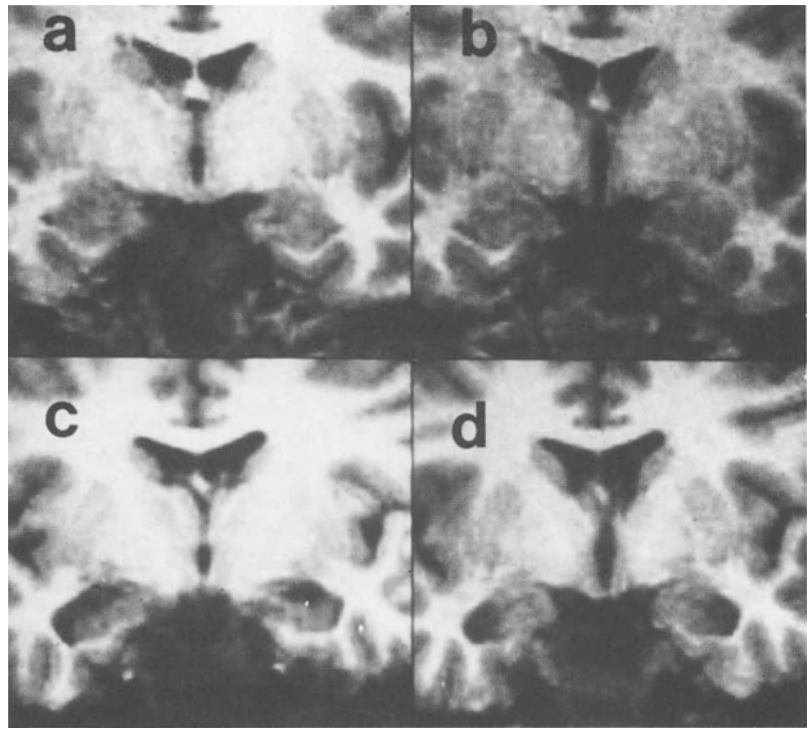


Fig. 5a–d. Coronal, T2-weighted magnetic resonance imaging (MRI) sections through the anterior hippocampus. **a,b** Two consecutive sections showing normal anatomy. **c,d** The same anatomical planes in an age-matched schizophrenic patient with dilatation of the temporal horns and hippocampal hypoplasia

Approximately 50 neuropathological or MRI studies of the limbic structures have been published in the last 10 years (for reviews, see Bachus and Kleinman 1996; Bogerts 1993, 1995; Bogerts and Lieberman 1993; Shapiro 1993; Travis and Kerwin 1997; Nelson et al. 1998; see also Vol. 3, Part 1, Chap. 6). Most of these studies revealed subtle structural defects in limbic areas, including volume loss, cell loss, and cytoarchitectural changes or abnormal cellular configurations in the hippocampus, parahippocampal cortex, amygdala, cingulate gyrus, septum, and orbital frontal cortex. Pathological abnormalities have also been reported in the thalamus and cerebellum, which are closely related to the limbic system (Pakkenberg 1990; Katsetos et al. 1997).

The anterior nucleus of the thalamus (which is a component of the limbic system) and the medial dorsal nucleus of the thalamus appear to be particularly affected by the disease process. In the anterior nucleus, a selective reduction of the number of parvalbumin-containing neurons (inhibitory projection neurons) by approximately 40% has been demonstrated (Danos et al. 1998). Significant cellular losses were also found previously in the medial dorsal nucleus (Pakkenberg 1990). Tissue loss was demonstrated in the thalamic periventricular gray area (Lesch and Bogerts 1984). The finding of reduction in the overall thalamic volume has since been confirmed by MRI (Andreasen et al. 1994).

The presence of schizophreniform symptoms in patients with organic lesions of the limbic system has long led to the suspicion that some of the functional disturbances of schizophrenic patients may be able to be localized to the limbic system, particularly to the medial temporal lobe (Bogerts 1997). As discussed above, the key limbic structures of this area are central sites of convergence for information from the higher cortical association areas of the frontal, temporal, and parietal lobes (see Figs. 1, 2). They play a crucial role in the analysis of situational context, in the filtering of stimuli, and in the comparison of past with present experience. These structures may be regarded as the highest cortical integration and association areas of the brain, and simultaneously as mediators between neocortical cognitive activity and the phylogenetically older neural responses of the septal–hypothalamic–brain stem axis.

It is thus entirely plausible that structural and functional disturbances in temporal limbic areas may lead to a dissociation of higher cognitive processes from elementary emotional modes of response. Bleuler regarded this uncoupling of cognition and emotion as the basic functional disturbance in the schizophrenias.

In addition to limbic structural and functional disturbances, there are many indications of prefrontal cortical disturbance in schizophrenics (Weinberger 1987; Goldman-Rakic 1994). As there is a close connection between the medial temporal lobe and the prefrontal cortex, it is not surprising that prefrontal dysfunction and hippocampal pathology may be described in the same patient. In pairs of monozygotic

twins of whom only one had been diagnosed as having schizophrenia, it was shown that the schizophrenic twins had a significant inverse correlation between hippocampal volume and prefrontal activation by the Wisconsin Card-Sorting Test (Weinberger et al. 1994). The conclusion was drawn that prefrontal cortical function is controlled, at least in part, by the hippocampus.

There is increasing evidence that more widespread loss of cortical volume occurs in schizophrenics, beyond the changes seen in temporal limbic areas. Several CT and MRI studies (Harvey et al. 1993; Raz 1993; Schlaepfer et al. 1994; Ross and Pearlson 1996) have revealed volume loss in polymodal association cortex (dorsolateral prefrontal cortex, inferior parietal lobule, and superior temporal gyrus), but not in occipital or sensorimotor cortex. This volume loss in the polymodal association cortex appears to be specific to schizophrenia, as it has not been found in patients with affective psychoses. This finding supports the hypothesis that not only limbic functioning, but also the functioning of higher cortical association areas is disturbed in schizophrenia.

Several other studies have shown that macroscopic parameters of the brain are also reduced in schizophrenia to a modest extent. A reduction in brain volume by approximately 3%, a reduction in anteroposterior diameter, and a generalized dilatation of the external cerebrospinal fluid space have repeatedly been described (Bogerts 1995).

The pattern of loss of brain substance among schizophrenics is inhomogeneous: some patients have pathological abnormalities mainly in the limbic system and others in the association cortex, while in other patients ventricular dilatation is the most prominent finding. In any case, structural changes in the temporal limbic system and the closely related polymodal association cortex seem to play a major role in the pathophysiology of this disease, as well as of secondary schizophreniform psychoses in patients with organic brain diseases (Ross and Pearlson 1996).

Cerebral structural abnormalities in schizophrenia are summarized at the end of Sect. 4.1. The following findings may be considered supportive evidence for the hypothesis of a disturbance of early brain development in schizophrenics:

- Lack of gliosis in limbic and cortical structures
- Cytoarchitectural changes in temporal and frontal cortex
- Frequent occurrence of cavum septi pellucidi
- Absence of the normal cortical structural asymmetry
- Lack of correlation between structural abnormalities and duration of illness
- Lack of progression of structural abnormalities in follow-up studies

4.3 Clinical Correlates of Structural Abnormalities

Most authors agree that the ventricular dilatation and hippocampal volume loss seen in schizophrenia are not progressive; their severity does not correlate with the duration of the disease, and they do not change over time in follow-up studies, except for the normal effect of aging. These findings seem inconsistent with a progressive, degenerative disease of the brain, but are compatible with hypoplasia acquired early in life (Weinberger 1987; Bogerts 1991; McCarley et al. 1996). On the other hand, there is increasing evidence that cortical pathology in schizophrenia may be progressive (Waddington 1993; Zipurski et al. 1994; DeLisi et al. 1997) (Table 5).

Cytoarchitectural changes in limbic and prefrontal cortex are also important evidence for a disorder of

Table 5. Computed tomography (CT) and magnetic resonance imaging (MRI) follow-up studies in schizophrenic psychoses

Reference	Method	Parameters	Progression
Nasrallah et al. (1986)	CT	VBR	No
Bogerts et al. (1987)	CT	Superior portion of left insular cistern	Yes
Illowsky et al. (1988)	CT	VBR	No
Vita et al. (1988)	CT	VBR	No
Sponheim et al. (1991)	CT	VBR, frontal horns of the lateral ventricles, third ventricle	No
Kemali et al. (1989)	CT	VBR	Yes
Woods et al. (1990)	CT	VBR	Yes
DeLisi et al. (1997)	MRI	Temporal lobes, corpus callosum, lateral ventricles	Yes
Zipurski et al. (1994)	MRI	Temporal lobes	Yes

The parameters were specifically chosen for investigation at the outset of the study or were found to be significantly altered. VBR, ventricle to brain ratio.

early brain development, even though the initial reports of aberrant cell migration in the parahippocampal region (Jakob and Beckmann 1986; Arnold et al. 1991) were criticized on methodological grounds (Falkai et al. 1988a,b; Heinsen et al. 1996; Akil and Lewis 1997; Krimer et al. 1997; Bernstein et al. 1998a). Reports of an abnormal configuration and distribution of neurons in the hippocampus and in cingulate, frontal, and temporal cortex, and in the neighboring subcortical white matter zones (Benes and Bird 1987; Akbarian et al. 1996; Jönsson et al. 1997; Luts et al. 1998), reinforce previous evidence of abnormal cytoarchitecture in schizophrenics and provide further support for the hypothesis of a disturbance of brain development; yet another piece of evidence is the reportedly increased prevalence of cavum septi pellucidi among schizophrenics (Degreef et al. 1992a; Kwon et al. 1998).

Most controlled autopsy studies revealed no evidence of significant gliosis (Falkai et al. 1999). An elevated glial cell density (i.e. gliosis) is present in progressive brain diseases such as Alzheimer's disease and is easily demonstrated by neuropathological techniques. Schizophrenic patients certainly do not have this type of abnormality, but slowly progressive encephalopathies must be mentioned here in the differential diagnosis. Taken together, the histological findings in schizophrenia suggest a disorder of brain development, which possibly triggers another, secondary process in cortical structures (Woods et al. 1996).

Genetic factors play an important role in the etiology and pathogenesis of schizophrenic psychoses (see, e.g. Maier and Schwab 1998), while concordance rates of schizophrenic illnesses in monozygotic twins are no higher than 50%, which implies that environmental factors are important as well. The risk of developing schizophrenia is twice as high in the presence of perinatal complications, according to a recent meta-analysis (Geddes and Lawrie 1995). A prospective study of children at risk revealed that the extent of abnormalities on CT was significantly correlated with genetic risk factors and the presence of perinatal complications (Cannon et al. 1993). Interestingly, perinatal complications were found to be the single important determinant of ventriculomegaly, while cortical sulcal dilatation was a product of both perinatal complications and genetic risk factors (Cannon et al. 1993).

Ventriculomegaly and limbic atrophy cannot be regarded as sequelae of neuroleptic treatment or as secondary effects of the disease. No CT, MRI, or autopsy study to date has shown an association between the duration of therapy and the structural abnormalities found in these brain regions. Despite the sometimes very pronounced clinical manifestations of neuroleptic-induced tardive dyskinesias, no study to date, including well-controlled autopsy studies, has revealed an association of these with localized histopathological changes, e.g. in the basal ganglia (Bogerts 1995; Marsh et al. 1994).

Increased volume of the basal ganglia in chronic schizophrenia is a quite remarkable and repeatedly replicated finding (Chakos et al. 1994; Keshavan et al. 1994; Meshul et al. 1996), surprising because neuropathological processes do not, as a rule, lead to an enlargement of neural tissue. A study in which MRI scanning was performed on schizophrenics at the onset of the illness and 18 months later provides a possible explanation for this finding: the caudate nucleus was significantly larger on the follow-up scans, and the extent of enlargement was correlated with the cumulative dose of neuroleptic medication.

The increased size of the caudate nucleus was attributed to the disinhibitory effect of neuroleptics on the basal ganglia (blockage of the inhibitory effect of dopamine), leading to an overuse hypertrophy of striatal neurons (Chakos et al. 1994). A recent longitudinal MRI study revealed that the superior temporal gyrus is of lower volume in untreated schizophrenics than in normal controls, but that this difference disappeared completely after 1 year of neuroleptic treatment (Keshavan et al. 1998). This finding is clearly in need of replication. The authors explain it by postulating that prolonged overactivity of the dopaminergic system leads to a reversible injury of cortical cells that regresses under neuroleptic treatment. This argument seems plausible, as neuroleptics are known to block apoptosis.

4.4 Use of Structural Findings in Differential Diagnosis and Assessment of the Response to Therapy

Three to 10% of unselected patients with psychotic disorders, in whom the general physical and neurological examinations are unremarkable, are found to have brain lesions on structural imaging (Lewis 1995; Falkai 1996). These are usually clinically irrelevant incidental findings, such as arachnoid cysts (Fig. 6) or septal cysts (Fig. 7), although tumors or subdural hematomata are also occasionally found.

Clear criteria for when structural imaging is indicated are sparse in the literature (e.g. Weinberger 1984). Thorough structural imaging should certainly not be omitted in patients with atypical disease manifestations or atypical progression.

There is some controversy in the literature over the usefulness of structural imaging for differential diagnosis and the prediction of therapeutic response and disease course. A meta-analysis of 33 CT and MRI studies was performed to determine how well structural parameters such as the "ventricle to brain ratio"

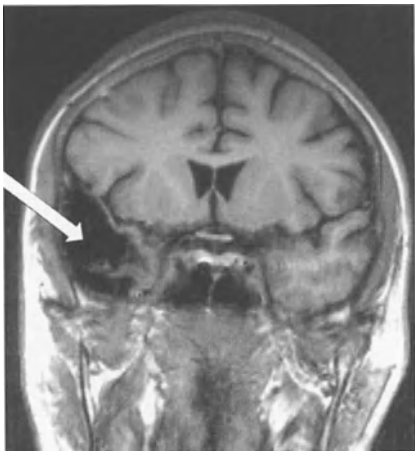


Fig. 6. A cyst in the right temporal lobe in a 33-year-old man

(VBR, an index of ventriculomegaly), the width of the sulci, third ventricular width, and others could be used to predict the short-term response to neuroleptic therapy (Friedman et al. 1992). Although the findings were remarkably heterogeneous, no single parameter with predictive value could be identified. Our own study confirmed this finding with respect to the cerebral ventricles. We did find a significant difference between responders and nonresponders in the mean value of other parameters relating to the frontal interhemispheric fissure, the temporobasal portion of the insular cistern, the inferior horn, and frontal asymmetry, but these parameters could not be used to obtain a reliable differential diagnosis in the individual case (Falkai et al. 1993; Falkai and Bogerts 1994). This finding points to a dilatation of the cerebrospinal fluid space that is most marked in the area of the polymodal association cortex.

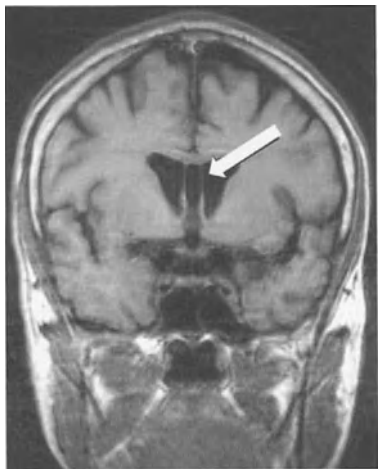


Fig. 7. Verga's ventricle in a 47-year-old man

The predictive value of structural parameters for long-term prognosis is different from that for short-term therapeutic response. Ventricular dilatation is a good predictor of long-term therapeutic success and of disease outcome (Lieberman et al. 1992; Raz 1993; van Os et al. 1995). There is also a highly significant correlation between ventricular dilatation and premorbid social adaptation (van Os et al. 1995): the wider the ventricles, the worse the patient's premorbid social adaptation, and the less favorable the long-term course of the illness.

4.5 Interpretation of Structural Findings in the Brain

The question of whether the brains of schizophrenics are more or less symmetrical than normal has attracted increasing attention recently (Crow 1990, 1993). It has been found repeatedly that the normal asymmetry of the frontal lobes (right larger than left) and of the occipital lobes (left larger than right) is not present in schizophrenics (Bilder et al. 1994; Falkai et al. 1995). This finding may be specific to schizophrenia: Falkai et al. (1995) found the normal asymmetries in patients with affective disorders and neuroses.

In addition to these frontal and occipital asymmetries, the human brain has still more marked asymmetry of other structures, including the temporal plane and the sylvian fissure. Human anatomic variability probably accounts for the contradictory results of recent studies (Rossi et al. 1992; Kleinschmidt et al. 1994; Frangou et al. 1997; for a detailed discussion, see Vol. 3, Part 1, Chap. 5).

Thus far, we have discussed a subtle dilatation of the internal and external cerebrospinal fluid spaces, structural defects in the temporal limbic area, enlargement of the basal ganglia, and absence of the normal asymmetry of the frontal and occipital lobes as recognized categories of cerebral structural abnormality in chronic schizophrenics. Because cerebral asymmetry begins to develop even before birth, the latter finding strongly supports the hypothesis of a prenatal disorder of cerebral development (as do the cytoarchitectural changes and the lack of correlation between structural abnormalities and the duration of disease).

The heterogeneity of these cerebral structural abnormalities among schizophrenics reflects the biological and clinical heterogeneity of the disease. No clear and replicable link has been found to date between the individual types of structural anomaly and the subtypes of clinical psychopathology. One possible exception is the correlation between dilatation of the cerebrospinal fluid space in the left temporal region and positive schizophrenic manifestations (Degreef et al. 1992b).

Dilatation of the lateral ventricles and of the third ventricle, dilatation of the cortical sulci, and abnormal structural findings in the limbic system are the best documented morphological abnormalities in schizophrenics. Furthermore, losses of brain substance in cortical association areas have been described repeatedly. Structural abnormalities in polymodal cortical association areas and the absence of the normal asymmetry of the frontal and occipital lobes have been demonstrated to date only in schizophrenics, but not in patients with affective psychoses or neuroses; these abnormalities thus possess a degree of specificity for schizophrenia.

The results of recent structural studies imply that the schizophrenias are diseases of the entire brain in which the limbic regions, and the cortical association areas most closely related to them in function, are especially severely affected.

The majority of the neuropathological findings, as well as the findings of CT and MRI studies, imply that the ventricular dilatation and limbic structural defects of schizophrenics are static, i.e. not progressive. Studies revealing an association between cortical volume loss and the duration of illness suggest that a second disease process may be at work. As the subcortical abnormalities, at least, seem to be static, they can be used to explain neither the variable course of illness, nor the typical onset of disease manifestations in early adulthood. These abnormalities should thus be regarded as vulnerability factors, which, along with other factors (psychosocial, biochemical, and nonspecific stressors), render the individual more likely to develop the disease.

5 Affective Illnesses

5.1 Histopathological Studies

In contrast to the abundance of neuropathological studies of schizophrenia that have been published since the end of the nineteenth century, there has been relatively little interest in the neuropathological investigation of patients with affective psychoses, and this remains so today. Perhaps the prevailing dualistic view of mind and body made investigators less likely to consider the possibility of organic processes operating in the emotional sphere than in the realm of cognition and perception, which is so prominently affected by schizophrenia (see also Vol. 3, Part 1, Chap. 6).

Two extensive reviews of the world literature (Jeste 1988; Bogerts and Lieberman 1993) turned up only 15 neuropathological studies of the temporal limbic areas, cerebellum, and basal ganglia in patients with affective disorders. In most of these studies, the brains of depressed or cyclothymic patients were used as control cases for the investigation of the neuropathology of schizophrenia. The number of patients in all of these studies was small (five or less), and the qualitative methods used were generally inadequate to yield significant findings in the absence of quantitative, statistical analysis. These methodologically flawed studies yielded no replicable conclusions.

The neurotransmitter-containing cell groups of the brain stem are known to be important in the pharmacology of the affective psychoses, but their histopathology has nevertheless not been investigated in any study of depressive or cyclothymic patients published to date. The situation is similar for the septal–hypothalamic area as well as for the amygdala and other limbic structures, even though the role of these structures in the modulation of elementary drives and emotions, as well as circadian rhythms and endocrine function, has long been recognized (Ketter et al. 1996; Grasby and Bench 1997). It was recently shown, with the aid of immunohistochemical methods, that depressed patients have an increased number of CRF-containing neurons in the paraventricular nucleus of the hypothalamus; this finding is clearly related to the hypercortisolism found in this patient group (Raadsheer et al. 1994). Immunohistochemical methods further revealed a reduced number of nitrous oxide synthase-containing neurons in the hypothalamus, which is perhaps related to an abnormality of endocrine regulation associated with depression (Bernstein et al. 1998b).

5.2 Structural Imaging Techniques

Structural imaging techniques have revealed several abnormalities in the brains of patients suffering from affective disorders (see also this volume, Part 1, Chap. 11). A considerable number of CT and MRI studies have documented the width of the internal and external cerebrospinal fluid spaces as well as the volumes of cortical and subcortical structures. Subtle widening of the cerebral ventricles seems to be present not only in patients with senile depression, but also in some patients with uni- and bipolar affective disorders, even though intergroup comparisons reveal less marked differences than in the case of schizophrenia (for reviews, see Wurthmann et al. 1995; Soares and Mann 1997; Grasby and Bench 1997). Most studies have revealed no dilatation of the cortical sulci in cyclothymic patients. Direct comparisons of the external cerebrospinal fluid spaces of schizophrenics and patients with affective disorders revealed that the

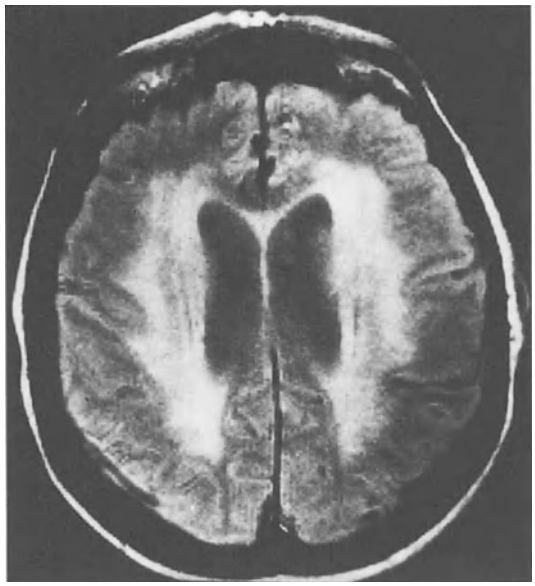


Fig. 8. White-matter lesions in a patient with affective psychosis

supramodal cortical association areas are affected in the former, but not in the latter (Harvey et al. 1993; Schlaepfer et al. 1994). This may explain why patients with affective disorders are less likely to have profound cognitive disturbances than schizophrenics.

The most frequently replicated structural finding in the brains of patients with both uni- and bipolar affective disorders is the increased prevalence of signal-intense areas in the cerebral white matter (Fig. 8).

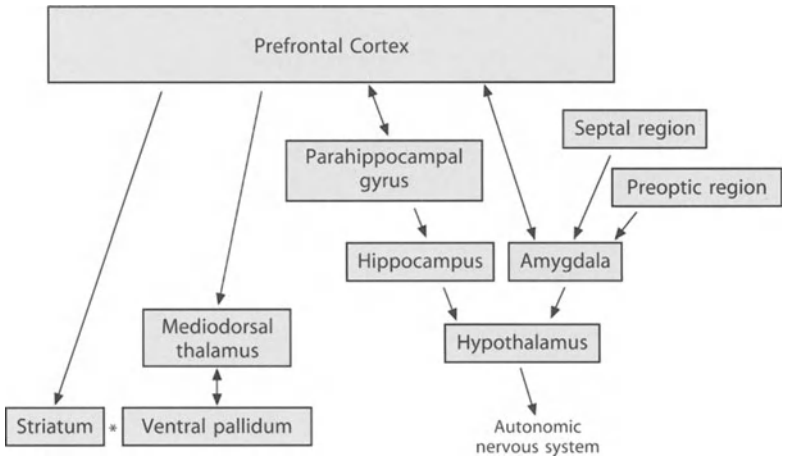
White matter lesions are particularly common in elderly depressed patients compared to age-matched controls, and this is true to a statistically significant extent even after possible vascular causes have been excluded (Soares and Mann 1997; Grasby and Bench 1997). The histological nature of these lesions and their pathophysiological significance are still unknown.

In many other areas of the brain, however, there is a dichotomy between the structural findings of unipolar depression and of bipolar affective disorder. Patients with unipolar depression have diminished volume of the frontal lobe, the cerebellum, and the striatum, while patients with bipolar affective disorder have dilatation of the third ventricle and diminished volume of the cerebellum and temporal lobes. Furthermore, a diminished volume of the basal ganglia of unipolar depressives (caudate, putamen, and nucleus accumbens) has been found in several studies (Soares and Mann 1997).

Among all components of the basal ganglia, its limbic component, the nucleus accumbens, appears to be particularly affected (Baumann et al. 1999). The observation that cerebrovascular accidents involving the basal ganglia often give rise to depressive manifestations is further evidence that the basal ganglia are functionally important not only in motor control, but also in the neural modulation of affect (Herrmann et al. 1993). The key structures of the limbic system involved in the processes of emotion and memory are depicted in Fig. 9.

The dichotomous MRI findings in the two types of affective disorder imply that they have different neurobiological causes. A further finding of interest is that the third ventricle and the frontal and temporal cerebrospinal fluid spaces are widened in patients carrying the diagnosis of endogenous depression (according to ICD-9), but not in neurotic depressives, who have a lifelong susceptibility to depression (Baumann et al. 1997). This finding is an argument against removing the distinction between neurotic and endogenous depressive syndromes, as was done in DSM-IV and ICD-10, as well as a demonstration, on the level of brain structure, of the variable contribu-

Fig. 9. Key structures involved in emotion and memor



tions of psychogenesis and organogenesis to the etiology of affective disorders.

6 Outlook

Recent discoveries in the functional neuroanatomy of the brain, particularly concerning the stages of cortical information processing and the role of the limbic system, have added considerably to our understanding of the neurobiology of psychiatric disease. Especially important findings have been made with respect to the telencephalic limbic structures of the medial temporal lobe, which, by virtue of their anatomic connections, play a mediating role between the neocortex and the brain stem, and thus also between higher cognitive processes and phylogenetically ancient emotions.

If we simplistically divide the brain into primary and higher areas of sensory information processing, then we may conclude from the classic neuropathological studies and the more recent studies of brain structure in psychotic patients that injuries of the primary sensory and motor cortical areas cause neurological manifestations, while disturbances of the higher cortical integration and association areas of the frontal lobes, temporal lobes, and limbic system cause psychiatric syndromes. Pathological abnormalities of the latter brain areas play a central role in organic mental syndromes (including Alzheimer's disease and symptomatic psychoses) and also in those forms of psychoses that were, until recently, termed "endogenous." In Wernicke-Korsakow syndrome, structures of the limbic diencephalon and midbrain are particularly affected, while in affective psychoses the limbic portions of the basal ganglia also appear to be disturbed. Disturbances of memory and cognition, abnormal interpretation of reality, and the dissociation of cognition and emotion that typifies psychosis are all produced by dysfunction of higher cortical areas and of the limbic system.

The neuropathological study of the schizophrenias is still in an early phase, and that of the affective psychoses still more so; nonetheless, considerable progress has been made in recent years, a happy contrast to the failure of neurobiological research on psychosis in the first half of the twentieth century. Most of the currently available studies have revealed moderately extensive, yet statistically significant and replicable losses of volume in areas subjected to gross measurement. Future investigators will have to perform more detailed histological and cytological analyses of these areas, and of neural systems and neuronal cell types that have so far been studied inadequately, or not at all, despite their theoretical relevance to the genesis of psychotic syndromes.

The neuropathological findings in schizophrenia, and probably also in the affective disorders, should be viewed as vulnerability factors for psychiatric disease. It remains entirely unknown what other, nonstructural biological factors (perhaps such as age and stress) might determine the course of psychiatric illness in a patient made vulnerable by a previous structural injury to the brain. If this question were answered, the prospects for effective therapy would be dramatically improved.

7 References

Akbarian S, Kim JJ, Potkin SG, Hetrick WP, Bunney WE, Jones EG (1996) Maldistribution of interstitial neurons in the prefrontal white matter of the brains of schizophrenics. Arch Gen Psychiatry 53: 425–436

Akil M, Lewis DA (1997) Cytoarchitecture of the entorhinal cortex in schizophrenia. Am J Psychiatry 154: 1010–1012

Aldenhoff J (1997) Überlegungen zur Psychobiologie der Depression. Nervenarzt 68/5: 379–389

Andreasen NC, Arndt S, Swayze V et al (1994) Thalamic abnormalities in schizophrenia visualized through magnetic resonance image averaging. Science 266: 294–298

Arnold SE, Hyman BT, van Hoesen GW, Damasio AR (1991) Some cytoarchitectural abnormalities of the entorhinal cortex in schizophrenia. Arch Gen Psychiatry 48: 625–632

Bachus SE, Kleinman JE (1996) The neuropathology of schizophrenia. J Clin Psychiatry 57[Suppl 11]: 72–83

Bauer J (1994) Die Alzheimer-Krankheit. Schattauer, Stuttgart

Baumann B, Bornschlegl C, Krell D, Bogerts B (1997) Changes in CSF spaces differ in endogenous and neurotic depression. A planimetric CT scan study. J Affect Dis 45: 179–188

Baumann B, Danos P, Krell D et al (1999) Reduced volume of limbic system-affiliated basal ganglia in mood disorders: preliminary data from a post mortem study. J Neuropsychiatry Clin Neurosci 11: 71–78

Benes FM, Bird ED (1987) An analysis of the arrangement of neurons in the cingulate cortex of schizophrenic patients. Arch Gen Psychiatry 44: 608–616

Benowitz LI, Perrone-Bizzozero NI (1991) The expression of GAP-43 in relation to neuronal growth and plasticity: when, where, how, and why? Prog Brain Res 89: 69–87

Bernstein HG, Krell D, Baumann B et al (1998a) Morphometric studies of the entorhinal cortex in neuropsychiatric patients and controls: clusters of heterotopically displaced lamina II neurons are not indicative of schizophrenia. Schizophr Res 33: 125–132

Bernstein HG, Stanarius A, Baumann B et al (1998b) Nitric oxide synthase-containing neurons in the human hypothalamus: reduced number of immunoreactive cells in the paraventricular nucleus of depressive patients and schizophrenics. Neuroscience 83: 867–875

Bilder RM, Wu H, Bogerts B et al (1994) Absence of regional hemispheric volume asymmetries in first episode schizophrenia. Am J Psychiatry 151: 1437–1447

Bogerts B (1991) The neuropathology of schizophrenia: pathophysiological and neurodevelopmental implications. In: Mednick SA, Cannon TD, Barr CE (eds) Fetal neural development

and adult schizophrenia. Cambridge University Press, Cambridge, pp 153–173

Bogerts B (1993) Recent advances in the neuropathology of schizophrenia. Schizophr Bull 19: 431–445

Bogerts B (1995) Hirnstrukturelle Untersuchungen an schizophrenen Patienten. In: Lieb K, Riemann D, Berger M (eds) Biologisch-psychiatrische Forschung – Ein Überblick. Fischer, Stuttgart, pp 123–144

Bogerts B (1996) Plastizität von Hirnstruktur und -funktion als neurobiologische Grundlage der Psychotherapie. Z Klin Psychol Psychiatr Psychother 44: 243–252

Bogerts B (1997) The temporolimbic system theory of positive schizophrenic symptoms. Schizophr Bull 23: 423–435

Bogerts B, Lieberman J (1993) Neuropathology in the study of psychiatric disease. In: Costa e Silva ACJ, Nadelson CC (eds) International review of psychiatry, vol 1. American Psychiatric Press, Washington, pp 515–555

Bogerts B, Wurthmann C, Piroth HD (1987) Hirnsubstanzdefizit mit paralimbischen und limbischem Schwerpunkt im CT Schizophrener. Nervenarzt 58(2): 97–106

Braak H, Braak E (1991) Neuropathological stageing of Alzheimer-related changes. Acta Neuropathol 82: 239–259

Braun K (1996) Synaptische Reorganisation bei frühkindlichen Erfahrungs- und Lernprozessen: Relevanz für die Entstehung psychischer Erkrankungen. Z Klin Psychol Psychiatr Psychother 44: 231–242

Cannon TD, Mednick SA, Parnas J, Schulsinger F, Praestholm J, Vestergaard A (1993) Developmental brain abnormalities in the offspring of schizophrenic mothers. I. Contribution of genetic and perinatal factors. Arch Gen Psychiatry 50: 551–564

Chakos MH, Lieberman JA, Bilder RM et al (1994) Increase in caudate nuclei volumes of first-episode schizophrenic patients taking antipsychotic drugs. Am J Psychiatry 151: 1430–1436

Crow TJ (1990) Temporal lobe asymmetries as the key to the etiology of schizophrenia. Schizophr Bull 16(3): 434–443

Crow TJ (1993) Schizophrenia as an anomaly of cerebral asymmetry. In: Maurer K (ed) Imaging of the brain in psychiatry and related fields. Springer, Berlin Heidelberg New York, pp 2–17

Dahlström A, Fuxe K (1964) Evidence for the existence of monoamine-containing neurons in the central nervous system. I. Demonstration of monoamines in the cell bodies of the brain stem neurons. Acta Physiol Scand 62[Suppl 232]: 1–55

Danos P, Baumann B, Bernstein HG et al (1998) Schizophrenia and anteroventral nucleus: selective decrease of parvalbumin-immunoreactive thalamocortical projection neurons. Psychiatry Res Neuroimaging 82: 1–10

Davison K, Bagley CR (1969) Schizophrenia-like psychosis associated with organic disorders of the central nervous system. A review of the literature. In: Hertington RN (ed) Current problems in neuropsychiatry. Br J Psychiatry (special publication) 4: 113–187

Degreef G, Bogerts B, Falkai P, Greve B, Lantos G, Ashtari M, Lieberman J (1992a) Increased prevalence of the cavum septum pellucidum in MRI scans and postmortem brains of schizophrenic patients. Psychiatry Res Neuroimaging 45: 1–13

Degreef G, Ashtari M, Bogerts B, Bilder RM, Jody DN, Alvir JMJ, Lieberman JA (1992b) Volumes of ventricular system subdivisions measured from magnetic resonance images in first episode schizophrenic patients. Arch Gen Psychiatry 49: 531–537

DeLisi LE, Sakuma M, Tew W, Kushner M, Hoff AL, Grimson R (1997) Schizophrenia as a chronic active brain process: a study of progressive brain structural change subsequent to the onset of schizophrenia. Psychiatry Res 74(3): 129–140

Falkai P (1996) Differential diagnosis in acute psychotic episode. Int Clin Psychopharmacol 11[Suppl 2]: 13–17

Falkai P, Bogerts B (1994) Brain morphology and prediction of neuroleptic treatment response in schizophrenia. In: Gaebel W, Awad AG (eds) Prediction of neuroleptic treatment outcome in schizophrenia. Springer, Berlin Heidelberg New York, pp 135–147

Falkai P, Bogerts B, Rozumek M (1988a) Cell loss and volume reduction in the entorhinal cortex of schizophrenics. Biol Psychiatry 24: 515–521

Falkai P, Bogerts B, Roberts GW, Crow TJ (1988b) Measurement of the alpha-cell-migration in the entorhinal region: a marker for developmental disturbances in schizophrenia? Schizophr Res 1: 157–158

Falkai P, Bogerts B, Klieser E, Waters U, Schlüter U, Mooren I (1993) Quantitativ-morphometrische Befunde im CT bei Neuroleptika-Non-Respondern. In: Müller HJ (ed) Therapieresistenz bei Neuroleptikabehandlung. Thieme, Stuttgart, pp 37–48

Falkai P, Schneider T, Greve B, Klieser E, Bogerts B (1995) Reduced frontal and occipital lobe asymmetry on CT-scans of schizophrenic patiens. Its specificity and clinical significance. J Neural Transm (Gen Sect) 99: 63–77

Falkai P, Honert WG, David S, Bogerts B, Maijtenyi C, Bayer TA (1999) No evidence for astrogliosis in brains of schizophrenic patients. A post mortem study. Neuropathol Appl Neurobiol 25: 48–53

Fields DR, Nelson PG (1992) Activity dependent development of the vertebrate nervous system. Int Rev Neurobiol 43: 133–214

Flor-Henry P (1969) Psychosis and temporal lobe epilepsy: a controlled investigation. Epilepsia 10: 363–395

Frangou S, Sharma T, Sigmudsson T, Barta P, Pearlson G, Murray RM (1997) The Maudsley Family Study. 4. Normal planum temporale asymmetry in familial schizophrenia. A volumetric MRI study. Br J Psychiatry 170: 328–333

Friedman L, Lys C, Schulz SC (1992) The relationship of structural brain imaging parameters to antipsychotic treatment response: a review. J Psychiatry Neurosci 17(2): 42–54

Geddes JR, Lawrie S (1995) Obstetric complications and schizophrenia. A meta-analysis. Br J Psychiatry 167: 786–793

Goldman-Rakic P (1994) Cerebral cortical mechanisms in schizophrenia. Neuropsychopharmacology 10[Suppl 3]: 22–27

Grasby PM, Bench C (1997) Neuroimaging of mood disorders. Curr Opin Psychiatry 10: 73–78

Gray JA (1982) The neuropsychology of anxiety: an enquiry into the function of the septo-hippocampal system. Oxford University Press, Oxford

Greenwood R, Bhalla A, Gordon A, Roberts J (1983) Behavior disturbances during recovery from herpes simplex encephalitis. J Neurol Neurosurg Psychiatry 46: 809–817

Hampel H, Teipel SJ, Kötter HU et al (1997) Strukturelle Magnetresonanztomographie in der Diagnose und Erforschung der Demenz vom Alzheimer-Typ. Nervenarzt 68: 365–378

Harper CG, Kril JJ, Daly J (1988) Does 'moderate' alcohol intake damage the brain? J Neurol Neurosurg Psychiatry 51: 909–913

Harvey I, Ron MA, Du Boulay G, Wicks SW, Lewis SW, Murray RM (1993) Reduction of cortical volume in schizophrenia on magnetic resonance imaging. Psychol Med 23: 591–604

Heinsen H, Gössmann E, Rüb U et al (1996) Variability in the human entorhinal region may confound neuropsychiatric diagnoses. Acta Anat 157: 226–237

Henderson AS, Henderson JH (eds) (1988) Etiology of dementias of Alzheimer's type. Wiley, New York

Herrmann M, Bartels C, Wallesch CW (1993) Depression in acute and chronic aphasia: symptoms, pathoanatomical-clinical correlations and functional implications. J Neurol Neurosurg Psychiatry 56: 672–678

Hess WR (1949) Das Zwischenhirn. Schwabe, Basel

Heun R, Schlegel S, Graf-Morgenstern M, Tintera J, Gawehn J, Stoeter P (1997) Proton magnetic resonance spectroscopy in dementia of Alzheimer type. Int J Geriatr Psychiatry 12(3): 349–358

Hillbom E (1951) Schizophrenia-like psychoses after brain trauma. Acta Psychiatr Neurol Scand 60: 36–47

Holman BL, Johnson KA, Gerada B, Carvalho PA, Satlin A (1992) The scintigraphic appearance of Alzheimer's disease: a prospective study using technetium-99m-HMPAO SPECT. Nucl Med 33(2): 181–185

Huber G (1957) Pneumencephalographische und psychopathologische Bilder bei endogenen Psychosen. Springer, Berlin Göttingen Heidelberg

Illowsky BP, Juliano DM, Bigelow LB, Weinberger DR (1988) Stability of CT scan findings in schizophrenia: results of an 8 year follow-up study. J Neurol Neurosurg Psychiatry 51: 209–213

Jakob J, Beckmann H (1986) Prenatal developmental disturbances in the limbic allocortex in schizophrenics. J Neural Transm 65: 303–326

Jeste D, Lohr JB, Goodwin FK (1988) Neuroanatomical studies of major affective disorders. Br J Psychiatry 153: 444–459

Johnstone EC, Crow TJ, Frith CD, Husband J, Kreel L (1976) Cerebral ventricular size and cognitive impairment in chronic schizophrenia. Lancet 2: 924–926

Jones EG, Powell TPS (1970) An anatomical study of converging sensory pathways within the cerebral cortex of the monkey. Brain 93: 793–820

Jönsson SA, Luts A, Guldberg-Kjaer N, Brun A (1997) Hippocampal pyramidal cell disarray correlates negatively to cell number: implications for the pathogenesis of schizophrenia. Eur Arch Psychiatry Clin Neurosci 247:120–127

Katsetos CD, Hyde TM, Herman MM (1997) Neuropathology of the cerebellum in schizophrenia – an update: 1996 and future directions. Biol Psychiatry 42: 213–224

Kemali D, Maj M, Galderisi S, Milici N, Salvati A (1989) Ventricle-to-brain ratio in schizophrenia: a controlled follow-up study. Biol Psychiatry 26: 756–759

Keshavan MS, Bagwell WW, Haas GL, Sweeney JA, Schooler NR, Pettegrew JW (1994) Changes in caudate volume with neuroleptic treatment. Lancet 344(8934): 1434

Keshavan MS, Haas GL, Kahn CE et al (1998) Superior temporal gyrus and the course of early schizophrenia: progressive, static, or reversible? J Psychiatr Res 32: 161–167

Ketter TA, George MS, Kimbrell TA, Benson BE, Post RM (1996) Functional brain imaging, limbic function and affective disorders. Neuroscientist 2: 55–65

Kleinschmidt A, Falkai P, Huang Y, Schneider T, Furst G, Steinmetz H (1994) In vivo morphometry of planum temporale asymmetry in first-episode schizophrenia. Schizophr Res 12: 9–18

Krimer LS, Herman MM, Saunders RC et al (1997) A qualitative and quantitative analysis of the entorhinal cortex in schizophrenia. Cereb Cortex 7: 732–739

Kurz A, Egensperger R, Lautenschlager N, Haupt M, Altland K, Graeber MB, Muller U (1995) Das Apolipoprotein-E-Gen und der Phenotyp der Alzheimer-Krankheit. Z Gerontol Geriatr 28(3): 195–199

Kwon JS, Shenton ME, Hirayasu Y et al (1998) MRI study of cavum septi pellucidi in schizophrenia, affective disorder, and schizotypal personality disorder. Am J Psychiatry 155: 509–515

**Lawrie SM, Abukmeil SS (1998) Brain abnormality in schizophrenia. Br J Psychiatry 172: 110–120

Lawrie SM, Abukmeil SS, Chiswick A, Egan V, Santosh CG, Best JJ (1997) Qualitative cerebral morphology in schizophrenia: a magnetic resonance imaging study and systematic literature review. Schizophr Res 25: 155–166

Lesch A, Bogerts B (1984) The diencephalon in schizophrenia: evidence for reduced thickness of the periventricular grey matter. Eur Arch Psychiatry Neurol Sci 234: 212–219

Lewis SW (1990) Computed tomography in schizophrenia, 15 years on. Br J Psychiatry 157[Suppl 9]: 16–24

Lewis SW (1995) The secondary schizophrenias. In: Hirsch S, Weinberger DR (eds) Schizophrenia. Blackwell, Oxford, pp 324–340

Lieberman J, Bogerts B, Degreef G, Ashtari M, Alvir J (1992) Qualitative assessment of brain morphology in acute and chronic schizophrenia. Am J Psychiatry 149: 784–791

Liu X, Matochik JA, Cadet JL, London ED (1998) Smaller volume of prefrontal lobe in polysubstance abusers: a magnetic resonance imaging study. Neuropsychopharmacology 18(4): 243–252

Luts A, Jonsson SA, Guldberg-Kjaer N, Brun A (1998) Uniform abnormalities in the hippocampus of five chronic schizophrenic men compared with age-matched controls. Acta Psychiatr Scand 98(1): 60–64

**Maier W, Schwab S (1998) Molecular genetics of schizophrenia. Curr Opin Psychiatry 11: 19–25

**Mann K, Widmann U (1995) Zur Neurobiologie der Alkoholabhängigkeit. Fortschr Neurol Psychiatr 63: 238–247

Mann K, Opitz H, Petersen D, Schroth G, Heimann H (1989) Intracranial CSF volumetry in alcoholics: studies with MRI and CT. Psychiatry Res 29: 277–279

Marsh L, Suddath RL, Higgins N, Weinberger DR (1994) Medial temporal lobe structures in schizophrenia: relationship of size to duration of illness. Schizophr Res 11: 225–238

**Maurer K, Riederer P, Beckmann H (eds) (1990) Alzheimer's disease. Epidemiology, neuropathology, neurochemistry, and clinics. Springer, Berlin Heidelberg New York

McCarley RW, Hsiao JK, Freedman R, Pfefferbaum A, Donchin E (1996) Neuroimaging and the cognitive neuroscience of schizophrenia. Schizophr Bull 22: 703–725

McLean PD (1952) Some psychiatric implications of physiological studies on frontotemporal portion of limbic system (visceral brain). Electroencephalogr Clin Neurophysiol 4: 407–418

Meshul CK, Buckman JF, Allen C, Riggan JP, Feller DJ (1996) Activation of corticostriatal pathway leads to similar morphological changes observed following haloperidol treatment. Synapse 22(4): 350–361

Mesulam MM (1986) Patterns in behavioral neuroanatomy: association areas, the limbic system, and hemispheric

specialization. In: Mesulam MM (ed) Principles of behavioral neurology. Davis, Philadelphia, pp 1–70

Millner R (1992) Cortico-hippocampal interplay and the representation of contexts in the brain. Springer, Berlin Heidelberg New York

Nasrallah HA, Olson SC, McCalley-Witters M, Chapman S, Jacoby CG (1986) Cerebral ventricular enlargement in schizophrenia: a preliminary follow-up study. Arch Gen Psychiatry 43: 157–159

Nelson MD, Saykin AJ, Flashman LA, Riordan HJ (1998) Hippocampal volume reduction in schizophrenia as assessed by magnetic resonance imaging. Arch Gen Psychiatry 55: 433–440

Nieuwenhuys R (1985) Chemoarchitecture of the brain. Springer, Berlin Heidelberg New York

Pakkenberg B (1990) Pronounced reduction of total neuron number in mediodorsal thalamic nucleus and nucleus accumbens in schizophrenics. Arch Gen Psychiatry 47: 1023–1028

Palkovits M, Zaborski L (1979) Neural connections of the hypothalamus. In Morgane PJ (ed) Anatomy of the hypothalamus. Decker, New York, pp 379–509

Pantel J, Schröder J, Schad LR et al (1997) Quantitative magnetic resonance imaging and neuropsychological functions in dementia of the Alzheimer type. Psychol Med 27: 221–229

Perez MM, Trimble MR, Reider I, Murray M (1984) Epileptic psychosis, a further evaluation of PSE profiles. Br J Psychiatry 146: 155–163

Perrone-Bizzozero NI, Sower AC, Bird ED, Benowitz LI, Ivins KJ, Neve RL (1996) Levels of the growth-associated protein GAP-43 are selectively increased in association cortices in schizophrenia. Proc Natl Acad Sci USA 93: 14182–14187

Phillips SC, Harper CG Kril J (1987) A quantitative histological study of cerebellar vermis in alcoholic patients. Brain 110: 301–314

Raadsheer FC, Hoogendijk WJ, Stam FC, Tilders FJ, Swaab DF (1994) Increased numbers of corticotropin-releasing hormone expressing neurons in the hypothalamic paraventricular nucleus of depressed patients. Neuroendocrinology 60: 436–444

Raz S (1993) Structural cerebral pathology in schizophrenia: regional or diffuse? J Abnorm Psychol 102: 445–452

Ross CA, Pearlson GD (1996) Schizophrenia, the heteromodal association neocortex and development: potential for a neurogenetic approach. Trends Neurosci 19(5): 171–176

Rossi A, Stratta P, Mattei P, Cupillari M, Bozzao A, Gallucci M, Casacchia M (1992) Planum temporale in schizophrenia: a magnetic resonance study. Schizophr Res 7: 19–22

Roth G (1991) Neuronale Grundlagen des Lernens und des Gedächtnisses. In: Schmidt JS (ed) Gedächtnis: Probleme und Perspektiven der interdisziplinären Gedächtnisforschung. Suhrkamp, Frankfurt, pp 127–158

Schlaepfer TE, Harris GJ, Tien AY et al (1994) Decreased regional cortical gray matter volume in schizophrenia. Am J Psychiatry 151: 842–848

Shapiro RM (1993) Regional neuropathology in schizophrenia: where are we? Where are we going? Schizophr Res 10: 187–239

Singer W (1991) Die Entwicklung kognitiver Strukturen – ein selbstreferentieller Lernprozeß. In: Schmidt JS (ed) Gedächtnis: Probleme und Perspektiven der interdisziplinären Gedächtnisforschung. Suhrkamp, Frankfurt, pp 96–126

Slater E, Beard AW, Glithero E (1963) The schizophrenia-like psychosis of epilepsy. Br J Psychiatry 109: 95–150

Smith GN, Flynn SW, Kopala LC, Bassett AS, Lapointe, JS, Falkai P, Honer WG (1997) A comprehensive method of assessing routine CT scans in schizophrenia. Acta Psychiatr Scand 96: 395–401

**Soares JC, Mann JJ (1997) The anatomy of mood disorders. Biol Psychiatry 41: 86–106

Sponheim SR, Iacono WG, Beiser M (1991) Stability of ventricular size after the onset of psychosis in schizophrenia. Psychiatry Res 40(1): 21–29

Swanson LW (1983) The hippocampus and the concept of limbic system. In: Seifert W (ed) Neurobiology of the hippocampus. Academic, London, pp 3–19

Travis MJ, Kerwin R (1997) Schizophrenia – neuroimaging. Curr Opin Psychiatry 10: 16–25

van Hoesen GW (1982) The parahippocampal gyrus. New observations regarding its cortical connections in the monkey. Trends Neurosci 5: 345–350

van Os J, Fahy A, Jones P et al (1995) Increased intracerebral cerebrospinal fluid spaces predict unemployment and negative symptoms in psychotic illness – a prospective study. Br J Psychiatry 166: 750–758

Victor M, Adams RD, Collins G (1989) The Wernicke-Korsakow syndrome and related neurologic disorders due to alcoholism and malnutrition. Davis, Philadelphia

Vita A, Saccetti G, Cazullo CL (1988) Brain morphology in schizophrenia: A 2- to 5-year CT scan follow-up study. Acta Psychiatr Scand 78: 618–621

Waddington JL (1993) Neurodynamics of abnormalities in cerebral metabolism and structure in schizophrenia. Schizophr Bull 19: 55–69

Weinberger DR (1984) Brain disease and psychiatric illness: when should a psychiatrist order a CAT scan? Am J Psychiatry 141: 1521–1527

**Weinberger DR (1987) Implications of normal brain development for the pathogenesis of schizophrenia. Arch Gen Psychiatry 44: 660–669

Weinberger DR, Aloia MS, Goldberg TE, Berman KF (1994) The frontal lobes and schizophrenia. J Neuropsychiatry Clin Neurosci 6: 419–427

Wernicke C (1881) Lehrbuch der Gehirnkrankheiten für Ärzte und Studierende, vol 2. Fischer, Kassel, pp 229–242

Woods BT, Yurgelun-Todd D, Goldstein JM, Seidman LJ, Tsuang MT (1996) MRI brain abnormalities in chronic schizophrenia: one process or more? Biol Psychiatry 40: 585–596

Wurthmann C, Bogerts B, Falkai P (1995) Brain morphology assessed by computed tomography in patients with geriatric depression, patients with degenerative dementia, and normal control subjects. Psychiatry Res Neuroimaging 61: 103–111

Zipurski RB, Marsh L, Lim KO et al (1994) Volumetric assessment of temporal lobe structures in schizophrenia. Biol Psychiatr 35: 501–516

CHAPTER 11

R. Schlösser, J.D. Brodie

Brain Imaging in Psychiatry

Translator: E. Taub

1 Introduction

Modern in vivo brain imaging, made possible by the major advances in radiology and nuclear medicine over the course of the twentieth century, has led to the formulation of new concepts of brain function and enabled the identification of specific pathological changes in diseases of the central nervous system. For much of its history, the study of brain function depended exclusively on postmortem examination of brain lesions and on the findings of electrophysiological tests. Early maps of brain function were often speculative and based on highly mechanistic assumptions. This approach could yield no more than a rudimentary conception of the modular organization of the brain.

The earliest means of imaging brain structures in vivo were the plain skull X-ray and, somewhat later, pneumoencephalography. The latter was the first technique used to study the relationship between brain parenchyma and ventricular size in various psychiatric disorders.

In the second half of the twentieth century, a first step toward in vivo functional imaging was taken with the introduction of particle emission tomography, which enabled the study of blood flow, metabolism, and specific neurochemical aspects of brain function. The decisive breakthrough for modern functional imaging came with the development of computerized reconstruction algorithms for radiologic and emission-tomographic data. Magnetic resonance tomography, introduced shortly thereafter, was a further addition to the wide range of structural and functional imaging modalities available to neuroscientists and clinicians.

This chapter will provide an overview of computed tomography (CT), magnetic resonance imaging (MRI), single photon emission computed tomography (SPECT), positron emission tomography (PET), functional magnetic resonance imaging (fMRI), and magnetic resonance spectroscopy (MRS). Its primary objective is to describe the major current research strategies in this field. The basic capabilities of each imaging modality and some representative results will be presented and discussed.

An imaging technique, in the broadest possible sense, is any technique that enables a spatial representation of structural or functional data. As virtually any data set obtained from the brain can be displayed as an image, such methods as quantitative electroencephalography (qEEG), event-related potentials (ERP), and magnetoencephalography (MEG) might also be referred to as imaging techniques. These will not be included in the present chapter; the reader is referred to Chap. 9 (this volume, Part 1) for a discussion. Brain imaging studies relating to substance abuse will also be discussed in a separate chapter (Vol. 3, Part 2, Chap. 22).

2 Structural Imaging: Computed Tomography and Magnetic Resonance Imaging

2.1 Technical Principles

Before modern radiologic techniques were available, pneumoencephalography provided no more than an indirect means of estimating the size of some brain structures in vivo. In this technique, air was introduced into the ventricular system by way of a lumbar puncture. This was both invasive and very unpleasant for the patient. CT was one of the first radiologic techniques to provide exact quantitative information about brain structures. It is noninvasive and, aside from radiation exposure, carries no risk to the patient.

X-rays are emitted by a point source rotating around the head, traverse the brain, and are received by a detector on the other side. The attenuation of the beam by the tissue is measured for a large number of trajectories through the brain. These data are then used to calculate the spatial distribution of the attenuating structures, which is usually displayed as a two-dimensional, gray-scaled image.

MRI is based on the fact that atomic nuclei with an odd number of nucleons (i.e., protons and neutrons) act as electrically charged spinning tops, thereby generating a magnetic field. These atomic nuclei align themselves either parallel or antiparallel to a strong, externally applied magnetic field, while simultaneously rotating about their axes ("precessing") at a specific frequency independent of field strength – the Larmor frequency. An externally applied radiofrequency (RF) pulse causes the nuclei to deviate from their original orientation. When the pulse is removed, the nuclei return to their original orientation. This so-called relaxation is accompanied by the emission of an RF signal, which can be detected by appropriately placed detector coils. In order to localize the source of the emitted signal, the RF pulse is applied repeatedly with systematic variation of frequency- and phase-encoding gradients. The data obtained are then converted by Fourier transformation to an image data matrix, which can be represented in the form of two-dimensional, gray-scaled MRI images.

This technique returns three elements of data for each location in imaging space: the two relaxation

times, T_1 and T_2, and the proton density. T_1 is the time constant for the return of the nuclei to their original orientation, while T_2 is the time constant for the loss of phase coherence in the transverse direction.

Gray matter and white matter are easily distinguished in T_1-weighted images, while T_2-weighted images provide a sharper contrast between brain parenchyma and cerebrospinal fluid (Buschong 1996).

2.2 Findings

Schizophrenia

The association of schizophrenia with ventricular dilatation was first suggested by pneumoencephalographic studies and later firmly established using CT (Andreasen et al. 1990b; Jones et al. 1994). This finding implied that at least some psychiatric disorders may be the result of structural abnormalities. The detected abnormalities were small, however, and there was considerable overlap between the schizophrenic patients and the normal control subjects.

MRI has recently enabled the performance of more sensitive and more detailed studies, which will largely form the topic of the present discussion.

The finding of lateral ventricular dilatation was replicated in many MRI studies (Chua and McKenna 1995). The quantitative measure commonly used, the ventricle–brain ratio (VBR), has been found to be elevated in schizophrenics, reflecting an increase in ventricular size at the expense of brain parenchyma. Ventricular dilatation is reportedly associated with both negative (Andreasen et al. 1990a; Gur et al. 1994) and positive (A.H. Young et al. 1991) symptoms of schizophrenia.

Ventricular dilatation and diffuse cerebral atrophy were found to be related to the severity of symptoms, as measured on the Brief Psychiatric Rating Scale (BPRS) (A.H. Young et al. 1991), but the extent of ventricular dilatation seems not to be correlated with the duration of illness (Hoffman et al. 1991; A.H. Young et al. 1991). It may be concluded that brain atrophy is not merely a consequence of the established disease process, but is rather the result of pathogenetic factors that are present at the onset of the disease or even before it. Later prospective studies did show, however, that progressive structural changes occur in the brain over the course of the illness. Progressive volume loss of both cerebral hemispheres, of the right cerebellar hemisphere, and of portions of the corpus callosum have been described (DeLisi et al. 1997a).

Direct measurements have revealed that the normally present asymmetry of the temporal plane (the left side is larger than the right) is less marked in schizophrenic patients than in normal controls (Petty et al. 1995). In pairs of schizophrenic twins, horizontal sylvian fissure asymmetry was found to be more highly correlated within pairs than across pairs, which implies that this structural brain abnormality has a genetic basis (DeLisi et al. 1997b).

Other structural imaging studies have documented an extensive loss of gray matter volume in schizophrenics which is not merely confined to the temporal lobes (Schlaepfer et al. 1994). Schizophrenics have been found to have significantly smaller frontal lobes (Andreasen et al. 1986) as well as a smaller corpus callosum (Woodruff et al. 1993) and an enlargement of the basal ganglia (Hokama et al. 1995).

Normal aging and a variety of demographic variables, including social class, intelligence, and education, are known to be correlated to varying degrees with overall brain size and with the size of individual brain structures (Pearlson et al. 1989; Andreasen et al. 1993). Thus the proper selection of control groups for patient studies must take these variables into account.

Structural abnormalities may be directly related to neuropsychological deficits. A significant correlation has been found between cognitive measures of prefrontal function and the size of the dorsolateral prefrontal cortex (DLPFC), especially in the left hemisphere (Seidman et al. 1994). The results of imaging studies suggest that structural damage to the frontal lobes may account, at least in part, for the cognitive deficits seen in schizophrenia. Studies performed to date also indicate that atrophy may be more marked in patients resistant to neuroleptic therapy (Lawrie et al. 1997).

Affective Disorders

Structural abnormalities of the brain have also been described in patients with both unipolar and bipolar depression. An increased VBR and other manifestations of cerebral atrophy have been found (Videbech 1997). A reduction of the volume of the basal ganglia was found among patients with unipolar, but not bipolar depression. The VBR and the width of the third ventricle were significantly correlated with scores on various rating scales for the severity of depression (Schlegel et al. 1989b; Krishnan et al. 1992). In addition to these findings, patients with affective disorders were found to have subcortical hyperintense signal abnormalities on MRI more frequently than controls (Dupont et al. 1990). In general, findings of cerebral structural abnormalities are less consistent in the affective disorders than in schizophrenia. Measures of volume loss, such as the VBR, are diagnostically nonspecific, as they are found both in schizophrenic patients and in patients with affective disorders.

Dementia

CT and MRI have revealed structural abnormalities in Alzheimer's disease (AD) including cerebral atrophy with widening of the cortical sulci, ventricular dilatation, and deep white-matter lesions with periventricular distribution (Faulstich 1991). MRI has further revealed a particularly severe atrophy of mesial temporal areas (Murphy et al. 1993), including the amygdala–hippocampus complex (Heun et al. 1997a). Brain atrophy is found early in the course of the disease process, and its extent is correlated with the severity of cognitive impairment (Kesslak et al. 1991; Pearlson et al. 1992).

The presence of a large number of hyperintense white-matter lesions, particularly in periventricular areas, is characteristic of dementia of vascular origin, rather than Alzheimer's disease (Erkinjuntti 1987). In this regard, however, there is considerable overlap between demented patients and normal controls, as also between patients with different subtypes of dementia, so that a reliable differential diagnosis generally cannot be made in this way in individual cases.

Other Diseases

A significant difference between the VBR of patients with obsessive-compulsive disorder (OCD) and that of normal controls was reported in CT studies (Behar et al. 1984). Subsequent MRI studies provided evidence of lateral asymmetry, with a larger-sized caudate nucleus on the right side (Calabrese et al. 1993), but failed to confirm a difference in VBR between OCD patients and normal controls (Kellner et al. 1991). The application of structural imaging to OCD is thus a good illustration of the fact that early findings may fail to be replicated when newer, more advanced imaging techniques become available.

There are similar problems regarding the findings of structural abnormalities in schizophrenia, particularly when the differences between groups are small and there is considerable overlap with the normal population. It was noted in a review of the literature (Van Horn and McManus 1992) that the year of publication of studies in this field is significantly negatively correlated with the strength of the effects reported in them. It may be inferred that methodological difficulties account in no small part for this unfortunate situation.

Nonetheless, the advent of modern tomographic techniques has brought about a major resurgence of interest in structural imaging as a tool in psychiatric research. These new methods were the first to enable a correlation of structural changes with psychopathological descriptors, neuropsychological performance parameters, and genetic aspects of various psychiatric disorders.

3
Functional Imaging

The use of short-lived radionuclides to create radiolabeled tracer substances was the key factor in the development of functional imaging. A tracer is a compound that has been "labeled" by the replacement of one or more atoms with a radioactive isotope. The tracer is normally injected into a peripheral blood vessel and transported by the bloodstream into the brain. The tracer has the same biologic activity as the original unlabeled substance; once in the brain, it either binds to specific receptors or becomes concentrated at specific sites through other neurochemical processes. The quantity and distribution of radioactivity in the brain are then measured by the methods described below and displayed as an image.

Although the initial experience of in vivo brain imaging with radiotracers was gained as long ago as the 1950s, the major technical breakthrough in this field – the use of spatial reconstruction algorithms demanding large amounts of computing power – necessarily awaited the development of sufficiently fast computers decades later.

3.1
Positron Emission Tomography

3.1.1 Technical Principles

In PET, unstable positron emitters are used to mark various chemical compounds. When a positron is emitted by one of these radionuclides, it travels a short distance and then collides with an electron, whereupon both the positron and the electron are annihilated and two photons (gamma rays), each with 511 keV of energy, are emitted in opposite directions (i.e., 180° apart). Because the two photons are emitted simultaneously, the site of their production (very near the site of the positron emitter) can be determined by means of a coincidence detector. The output of the coincidence detector is picked up by scintillation crystals and then digitized, and a computerized reconstruction algorithm is used to generate an image.

Positron emitters are normally characterized by short half-lives, and they must therefore be produced in a cyclotron in the immediate vicinity of the PET scanner. As practically all biologically important molecules contain oxygen, carbon, or nitrogen, the short-lived radioisotopes ^{15}O, ^{11}C, and ^{13}N can be used to mark molecules without altering their chemical structure. Because they are short-lived, the patients are exposed to only small amounts of radiation, and studies can be conveniently repeated. Other

Table 1. Commonly used positron emitters for positron emission tomography (PET) studies in psychiatry

	Half-life (min)
^{11}C	20.4
^{13}N	10.0
^{15}O	2.1
^{18}F	109.8
^{75}Br	99.0
^{76}Br	972.0

radioisotopes in use, such as ^{18}F, ^{75}Br, and ^{76}Br, have longer half-lives (Table 1).

Positron emitters and detectors were used to study brain tumors as early as the 1950s. One of the first positron cameras, developed at the Brookhaven National Laboratory in New York, contained an array of ring detectors (Robertson et al. 1973). The steady improvement of PET technology in recent years has led to spatial resolutions as low as 4 mm, depending on the specific tracers and devices used. The theoretical limit of spatial resolution for PET is given by the distance traveled by positrons before their annihilation. This distance is approximately 2 mm for ^{11}C.

Two basic strategies are used in PET functional imaging studies. The first involves the measurement of cerebral blood flow (CBF) or metabolism. The second involves the use of a radiotracer with known specificity for a particular neurotransmitter system. The tracer binds to a site that may be a receptor, a transporter, a neurotransmitter, or a neurotransmitter precursor. PET may be used to measure the in vivo distribution and affinity constants of both pre- and postsynaptic neurotransmitter receptors.

[^{18}F]Fluorodeoxyglucose ([^{18}F]FDG) is the preferred tracer substance for the measurement of cerebral energy metabolism, expressed as the regional cerebral metabolic rate for glucose (rCMRglc). This method has been shown to have a high intrasubject reproducibility (Bartlett et al. 1991a). [^{18}F]FDG is transported across the blood–brain barrier in both directions by the same transporter as unmarked glucose. [^{18}F]FDG also undergoes the same initial metabolic step as glucose, namely 6-phosphorylation. Unlike glucose metabolism, however, the metabolism of [^{18}F]FDG stops at this stage (i.e., as [^{18}F]FDG-6-phosphate).

After 35 min, most of the radioactivity is trapped within the cell in the form of [^{18}F]FDG-6-phosphate, and the amount of radioactivity remains roughly constant for a further 30 min. Tomographic images are obtained during this period to determine the distribution of the positron emitter in the brain. Mathematical algorithms incorporating various correction factors, including the plasma concentration of [^{18}F]FDG, are used to generate a pictorial representation of metabolic activity (Sokoloff et al. 1977; Huang et al. 1980).

When interpreting these data, it should be borne in mind that a given value of regional glucose metabolism does not directly imply any specific functional state of the neurons in the vicinity. Metabolic rates determined with [^{18}F]FDG primarily represent the metabolic activity in nerve terminals, which are rich in mitochondria. Since neurons and neurotransmitters may have either excitatory or inhibitory functions, heightened metabolic activity in a given area may be consistent with either an increase or a decrease of overall information processing in that area.

[^{15}O]H_2O is the preferred tracer substance for the measurement of regional cerebral blood flow (rCBF) with PET. Because the half-life of ^{15}O is much shorter than that of ^{18}F, use of this tracer allows the repetition of studies at short intervals to demonstrate cognitive activation or short-acting pharmacologic effects. ^{18}F, in contrast, is more suitable for the study of relatively stable mental states and longer-lasting clinical pharmacologic effects.

PET and SPECT also have an important application in the study of neuroreceptor pharmacology, including the study of interactions between neurotransmitter systems. In recent years, many radioligands have been developed that are specific for individual receptor types. Radioligands for use in PET receptor studies vary in a number of important technical properties, including tracer pharmacokinetics, reversibility of binding, selectivity, nonspecific binding, and metabolic stability.

Imaging techniques for the study of neuroreceptor systems were first used primarily to study the dopaminergic system. In recent years, ligands have been developed that allow the study of many aspects of this system. One of the first ligands used to study the D_2 receptor system was [^{11}C]*N*-methylspiperone (NMSP) (Wagner et al. 1983). This ligand binds preferentially to D_2 receptors, but also has a high affinity for 5-HT_2 serotonergic receptors. Further ligands for the study of D_2 receptors include [^{76}Br]bromospiperone (Maziere et al. 1985) and [^{11}C]raclopride (Farde et al. 1985), as well as [^{18}F]NMSP.

Ligands successfully used to image the dopamine transporter include [^{11}C]nomifensine (Aquilonius et al. 1987), [^{11}C]WIN35,428 (Bennett et al. 1995), [^{11}C]cocaine (Volkow et al. 1992; Bennett et al. 1995), [^{11}C]methylphenidate (Volkow et al. 1995), and [^{11}C]β-CIT (Farde et al. 1994). The concentration and distribution of the dopamine-metabolizing enzyme monoamine oxidase B (MAO-B) have been studied with the ^{11}C-labeled irreversible enzyme inhibitors

[^{11}C]clorgyline and [^{11}C]L-deprenyl (Bench et al. 1991). Different aspects of presynaptic dopaminergic function have been studied with the radiotracer [^{18}F]fluorodopa (Leenders et al. 1986; Kuwabara et al. 1993).

A number of other neurotransmitter and receptor systems have also been studied by PET. 5-HT_2 receptors may be studied with the radiotracers [^{11}C]setoperone (Blin et al. 1990) and [^{11}C]altanserin (Sadzot et al. 1995). Most of the radiotracers currently available for the study of the serotonergic system have low affinity and low selectivity and are thus of limited usefulness. Newer serotonergic radiotracers, such as [^{11}C]MDL100907 (Lundkvist et al. 1996), may help resolve this problem. For the study of 5-HT_{1A} receptors, ligands such as WAY-100635 appear promising (Pike et al. 1996). The radiotracer [^{11}C]5-hydroxytryptophan has been used to study the function of presynaptic serotonergic nerve terminals (Agren et al. 1991). The radiolabeled benzodiazepine antagonist [^{11}C]flumazenil has been used for quantitative study of benzodiazepine receptors (Frey et al. 1991).

Functional imaging may also be used to study the influence of drugs or neuropsychological tasks on the metabolic and neurochemical processes of the brain. The dynamic response to such challenges may be studied by measuring the regional metabolic rate (rCMRglc) or the regional blood flow (rCBF), as well as by measuring changes in receptor occupancy by means of endogenous ligands (Schlösser et al. 1996).

PET studies incorporating various neuropsychological tasks have enabled the construction of a map of the brain with regard to cognitive function, both in normal subjects and in various patient populations. The typical experimental design of a [^{15}O]H_2O study of regional blood flow involves several alternating epochs of baseline and activation scanning. Scans acquired at baseline and during activation are separately averaged, and a *t* test is performed at each volume element (voxel) to determine whether statistically significant regional activation has occurred. The *t* values are then color coded, and the resulting functional image can be superimposed on a corresponding anatomical image (MRI or CT).

In a further extension of this approach, the predicted response over the time of the activation study is used as a reference function. The observed activation in each voxel over time is correlated with the predicted response function, and the resulting correlation coefficients are displayed as a statistical map. Modern statistical approaches, such as statistical parametric mapping (SPM), involve statistical assessment of signal strength and of the spatial extent of activation, while appropriate corrections are made for multiple comparisons (Friston et al. 1991, 1996). Extensive reviews of the methodological aspects and findings of activation studies with PET and fMRI have been provided by Roland (1993) and by Frackowiak et al. (1997).

3.1.2 Imaging of Metabolism and Blood Flow

A few representative findings of PET studies will be discussed in this and the following section. The distinction between so-called baseline studies and studies performed under the influence of various pharmacologic and neuropsychological manipulations will be kept in mind.

In metabolic PET studies, cognitive tasks are often used to ensure a constant mental state during scanning (Buchsbaum et al. 1992a). The naturally occurring fluctuations of alertness and general mental state are eliminated by the specification of a given state (i.e., the state of performing the given task). Cognitive tasks used in this way serve a normalizing function rather than providing a specific activation of cognitive processes. They are, as a rule, not part of the object under study in such experiments and are applied not only during "activation" but also during the "baseline" component of the study. The term "baseline" should not, therefore, be considered synonymous with "resting state."

Schizophrenia

Many PET studies have been performed concerning the metabolic activity or regional blood flow of the frontal cortex of schizophrenic patients, but their results are, in part, inconsistent. A "hypofrontal" pattern was found by a number of investigators (e.g., Williamson 1987; Andreasen 1988; Buchsbaum et al. 1992a), but this finding could not be replicated by others (Gur and Gur 1995). It should be noted that these studies differ widely in their design, case selection, medication status of patients, and definition of the cerebral region of interest and are thus not easy to compare with one another.

In general, however, it seems unlikely that pharmacotherapy is the sole cause of hypofrontality, as PET studies have shown that drug-free or drug-naive patients also have reduced activity in the prefrontal cortex (Andreasen et al. 1997).

Hypofrontality also seems to be characteristic of chronically ill patients and patients with predominantly negative symptoms (Volkow et al. 1987). Further discoveries concerning metabolic abnormalities in psychotic patients were made with the aid of pharmacologic model substances, as will be discussed further below.

In a study of neuroleptic-naive patients, Andreasen et al. (1997) found a reduction of rCBF not only in the prefrontal cortex (lateral, orbital, and medial), but also in inferior temporal and parietal cortical areas.

This finding suggests that the neural basis for schizophrenia involves widespread dysfunction of cortical systems.

These and similar findings imply that an analysis of functional connectivity should be more profitable than mere isolated analysis of regional changes in rCBF and rCMRglc. The view gradually took hold that an alteration of the overall metabolic pattern, rather than locally restricted alterations, was likely to be the pathophysiological correlate of psychiatric diseases and subsyndromes. It was demonstrated, for example, that various subsyndromes of schizophrenia, including psychomotor slowing, disorganization, and reality distortion, could be distinguished by their characteristic metabolic patterns (Liddle et al. 1992).

As mentioned above, in studies involving the performance of a neuropsychological task, a defined state of activation is compared with the baseline state within a single experiment. It is assumed that the difference between the two states reflects the performance of a specific function by the brain, such as word-finding or working memory (Friston 1992). A word-finding task (i.e., a test of "verbal fluency") has been found to be associated with a comparable activation of the left prefrontal cortex (among other areas) in both schizophrenic and normal control subjects. The schizophrenics, however, did not have a reduction of blood flow in the left superior temporal cortex, as was found in the normal controls. According to the authors of this study, this finding indicates a disturbance of the functional connectivity between the frontal and temporal cortex of schizophrenic patients (Frith et al. 1995).

In addition to the various types of activation paradigm we have discussed, PET may also be used to study the patterns of activation associated with spontaneously occurring psychopathological symptoms. It has been shown, by means of an event-related study design, that active auditory hallucinations in schizophrenic patients are associated with an activation of the subcortical nuclei, limbic structures (particularly the hippocampus), and paralimbic regions as well as of the orbitofrontal cortex. These findings are consistent with the speculation that auditory hallucinations may be generated, or at least modulated, by deep-seated brain structures, while the specific content of a hallucination may depend on the particular neocortical region activated during it (Silbersweig et al. 1995).

Studies of alterations in cerebral metabolism caused by treatment with various neuroleptics have yielded partly inconsistent results. This may be due to differences in treatment regimens or other factors causing heterogeneity in the patient groups studied. The finding of elevated metabolic activity in the striatum during treatment with typical neuroleptics, such as haloperidol, has been replicated repeatedly (Volkow et al. 1986; Bartlett et al. 1991b; Holcomb et al. 1996).

An elevation of metabolic activity in the basal ganglia was also demonstrated after subchronic treatment with the atypical neuroleptic clozapine (Buchsbaum et al. 1992b). On the other hand, reduced neocortical activity, particularly in the frontal lobes, was found after treatment with haloperidol (Bartlett et al. 1991b; Holcomb et al. 1996). In a direct comparison of patients taking either fluphenazine or clozapine, both groups had a reduction of overall cortical glucose metabolism. The use of either medication resulted in a reduction of rCMRglc in the superior prefrontal cortex and an elevation of rCMRglc in the limbic cortex. Fluphenazine use simultaneously increased the rCMRglc subcortically and in the lateral portions of the temporal lobes, while clozapine use reduced metabolic activity in the inferior prefrontal cortex (Cohen et al. 1997).

These results suggest that typical and atypical neuroleptics exert their effects in overlapping brain areas. Both substance classes seem to be associated with an elevation of subcortical metabolism, particularly in the basal ganglia, and with varying degrees of reduction of neocortical metabolism, predominantly in the frontal lobes. With regard to treatment response, it has been shown that an increased metabolic rate in the right putamen and a decreased metabolic rate in the caudate nucleus are associated with clinical improvement after neuroleptic treatment (Buchsbaum et al. 1987). Responders to haloperidol treatment were found to have had a lower metabolic rate in the striatum before beginning treatment (Buchsbaum et al. 1992c); the same was found for patients responding to clozapine (Buchsbaum et al. 1992b). An acute haloperidol challenge was found to cause a significant drop in absolute glucose metabolism in cortical areas, thalamus, and cerebellum in schizophrenics who responded to neuroleptic treatment as compared to nonresponders. This finding implies that pharmacologic challenge paradigms might be used to predict the response to therapy (Bartlett et al. 1998).

The application of pharmacologic model substances is an innovative method for the study of cerebral metabolism and blood flow in schizophrenia and related psychotic disorders. It is well known that phencyclidine (PCP) and other *N*-methyl-D-aspartate (NMDA) antagonists may evoke symptoms resembling the positive and negative symptoms of schizophrenia in normal subjects. The PCP analogue ketamine, when given to normal subjects, was found to evoke mild psychotic symptoms that were accompanied by a focal increase of metabolic activity in the prefrontal cortex (Breier et al. 1997a).

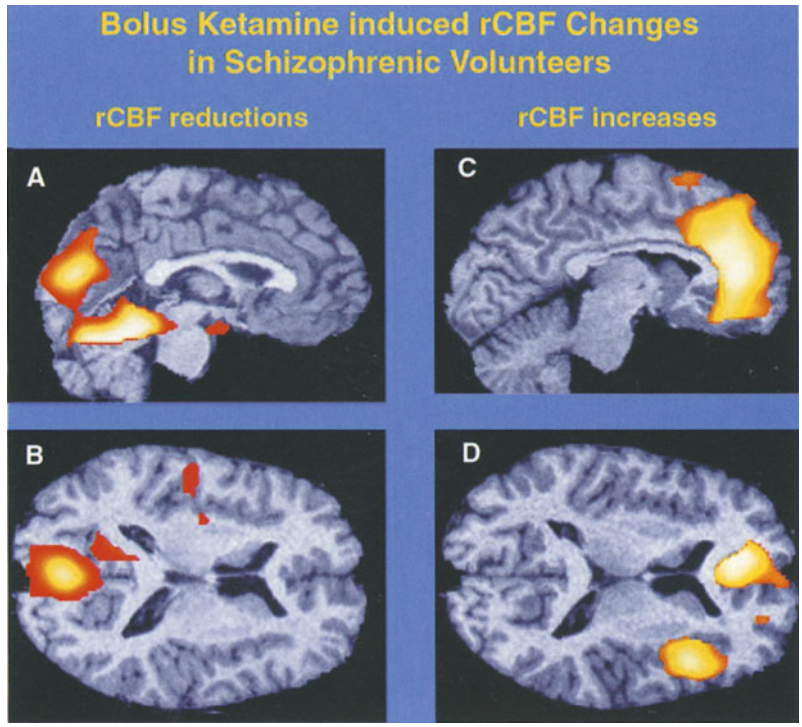


Fig. 1. Horizontal and sagittal images of the areas of significantly altered regional blood flow, as determined by $[^{15}O]H_2O$ positron emission tomography (PET), in five schizophrenic patients, before and 6 min after bolus administration of ketamine. **A,B** Enhanced regional cerebral blood flow (rCBF); **C,D** diminished rCBF. (After Lahti et al. 1995)

When given to schizophrenic patients, ketamine produced an activation of psychotic symptoms in tandem with a distinct pattern of alteration of rCBF, as determined by PET (Lahti et al. 1995). The results of $[^{15}O]H_2O$ PET scans of five schizophrenic patients before and 6 min after bolus administration of ketamine are shown in Fig. 1. Areas of significant change are color coded and superimposed on anatomical MRI images. As may be seen in these statistical parametric images, there is an enhancement of local rCBF in the anterior portion of the cingulate gyrus and in the inferomedial prefrontal cortex (A,B). In contrast, areas of diminished rCBF are shown (C,D); such areas are recognizable in the visual cortex and in the right hippocampus. These findings support the conclusion that the prefrontal cortex and limbic structures participate in the generation of NMDA receptor-mediated psychotic states.

The effect of the hallucinogen psilocybin, a 5-HT_2 and 5-HT_1 agonist, on regional glucose metabolism was studied in normal subjects using $[^{18}F]$FDG-PET. The findings are shown in Fig. 2 (color coded). The regional metabolic rate for glucose is shown under baseline conditions (*left*) and after a psilocybin challenge (*right*). Oral administration of psilocybin led to a global increase in rCMRglc that was most pronounced in the medial frontal (FRM), lateral frontal (FRL), and medial temporal cortex (TEMPORAL) and in the anterior portion of the cingulate gyrus (CGA). Subcortical structures, including the caudate nucleus, putamen, and thalamus, had a smaller increase in rCMRglc than the cortical areas mentioned.

These findings show that excessive activation of 5-HT receptors leads to a hyperfrontal metabolic pattern that is clearly distinct from the hypofrontality described in schizophrenic patients. The severity of the psilocybin-induced psychotic symptoms in normal subjects was correlated with the extent of the metabolic alteration (Vollenweider et al. 1997).

The studies cited regarding ketamine and psilocybin provide evidence that acute psychotic states are associated with a transient hyperfrontal metabolic pattern. In contrast, a hypofrontal pattern is described in schizophrenics.

The modulating effect of dopaminergic, serotonergic, and cholinergic pharmacological challenges on cognitive brain activation was investigated in a series of studies. Significant results were found in studies involving a combination of pharmacological and neuropsychological interventions. The effect of the muscarinic cholinergic antagonist scopolamine on CBF was studied by $[^{15}O]H_2O$-PET in normal subjects during performance of a memory task. The application of scopolamine was found to result in a widespread diminution of activation in the prefrontal cortex bilaterally as well as in the anterior cingulate region (Grasby et al. 1995).

Drug-naive schizophrenic patients were studied with $[^{15}O]H_2O$-PET before and after administration of the dopamine agonist apomorphine in conjunction with a

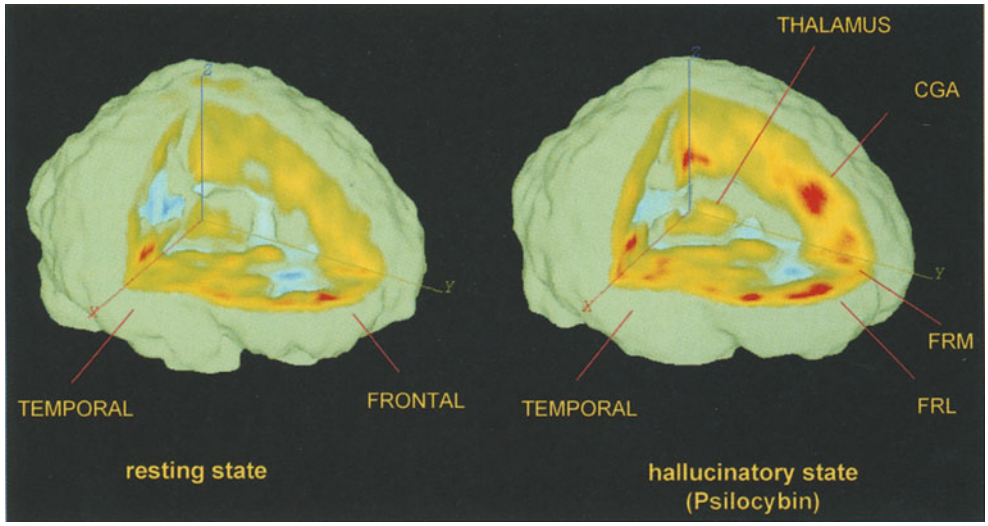


Fig. 2. Fluorodeoxyglucose ([^{18}F]FDG) positron emission tomography (PET) images under baseline conditions (*left*) and after psilocybin challenge (*right*). *CGA*, cingulate gyrus; *FRM*, medial frontal cortex; *FRL*, lateral frontal cortex. (After Vollenweider et al. 1997)

word-finding task. When no drug was given, schizophrenics had a relative reduction of activation of the anterior portion of the cingulate gyrus as compared to the normal controls. After the administration of apomorphine, however, the patient group showed an increased activation of the anterior portion of the cingulate gyrus that was, in fact, more pronounced than that of the normal controls (Dolan et al. 1995). These findings suggest that faulty dopaminergic regulation in the anterior portion of the cingulate gyrus is a feature of schizophrenia.

Affective Disorders

Baseline studies with [^{18}F]FDG- or [^{15}O]H_2O-PET in affective disorders revealed a reduction in either rCMRglc or rCBF in the inferior and superior frontal as well as the temporal and parietal cortex (Baxter et al. 1989; Hurwitz et al. 1990; Martinot et al. 1990; Bench et al. 1992; Drevets et al. 1992). There is increasing evidence for involvement of the anterior portion of the cingulate gyrus in depression. Decreased rCBF in this area was found in a direct comparison of depressed and normal subjects (Bench et al. 1993). Mayberg et al. (1997) found that the rCMRglc in the anterior portion of the cingulate gyrus of depressed patients differed in responders and nonresponders to antidepressant therapy: this area was hypometabolic in nonresponders, but hypermetabolic in responders. This finding is illustrated in color-coded fashion in Fig. 3.

A study of patients with bipolar and unipolar depression and a positive family history revealed a reduction in rCBF in the prefrontal cortex ventral to the genu of the corpus callosum. Lower CBF was, however, partly accounted for by a diminution of the volume of this area, which was revealed by an MRI study performed in parallel (Drevets et al. 1997).

In an [^{18}F]FDG study, administration of the serotonin releaser and reuptake inhibitor fenfluramine to

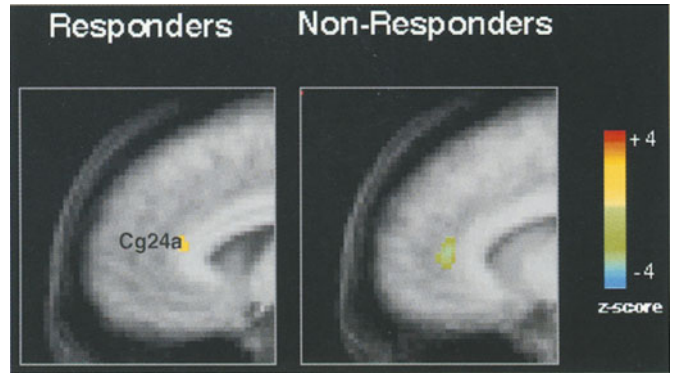


Fig. 3. Fluorodeoxyglucose ([^{18}F]FDG) positron emission tomography (PET). Statistical representation of differences in glucose metabolism in the rostral portion of the cingulate gyrus (Brodmann area 24a) in depressed patients who responded (*left*) or did not respond (*right*) to treatment with an antidepressant. (After Mayberg et al. 1997)

normal subjects caused a significant increase in rCMRglc, particularly in the left frontal and temporoparietal cortex, while rCMRglc was diminished in the right prefrontal cortex and elsewhere. In contrast, depressed patients showed no significant changes after fenfluramine administration, and their response pattern thus resembled that of normal subjects without a pharmacological challenge (Mann et al. 1996). This result supports the hypothesis of reduced serotonergic transmission in depression.

The induction of a sad mood brought about an increase in rCBF in the inferior orbitofrontal cortex of normal subjects (Pardo et al. 1993). The same result was found in another study regardless of whether a depressed or an elevated mood had been induced (Baker et al. 1997). In general, it may be concluded that higher-order functional neural networks responsible for the processing of emotional information are predominantly located in the anterior portion of the cingulate area and in the inferior frontal gyri (George et al. 1993). Activation during a word-finding task was diminished by either an elevated or a depressed mood throughout the left prefrontal, premotor, and cingulate cortex and in the thalamus, in association with an impairment of cognitive performance (Baker et al. 1997). This suggests that cognitive ability is linked to mood; indeed, cognitive distortion during depressive episodes is frequently observed in clinical practice. In sum, studies of normal subjects have yielded valuable information about the anatomical and functional substrates of emotional stimuli. The studies cited here provide a basis for further research in this area.

PET has also been used to study the effects of antidepressants on cerebral metabolism. Ten weeks of treatment with sertraline, a selective serotonin reuptake inhibitor (SSRI), led to increased activity in the middle frontal gyrus. In drug-naive depressed patients, this region exhibited reduced activity in comparison to temporal and some occipital areas. Other areas whose activity differed between depressed and control subjects (including the medial portion of the frontal lobe, the cingulate gyrus, and the thalamus) displayed a normalization of metabolism after sertraline treatment (Buchsbaum et al. 1997). These findings suggest that antidepressant therapy can reverse a depression-related reduction in cerebral metabolism.

Anxiety Disorders and Obsessive-Compulsive Disorder

In a PET study of patients with generalized anxiety disorder required to perform an active viewing task, the administration of benzodiazepines resulted in diffuse reduction of the absolute metabolic rate in cortical regions, the limbic system, and the basal ganglia (Wu et al. 1991). Nonetheless, this did not represent a normalization of the rCMRglc pattern present before benzodiazepine administration. The clinical effectiveness of benzodiazepines seems not to be associated with complete normalization of pathologically altered cerebral metabolism.

Baxter et al. (1987) carried out extensive studies of metabolic alterations in OCD patients before and after treatment. Two studies of untreated OCD patients revealed a significant elevation of glucose metabolism in the cerebral hemispheres, the heads of the caudate nuclei, and the orbital gyri in comparison to normal subjects. Responders to either pharmacotherapy or behavioral therapy were found to have a diminished rCMRglc in the head of the caudate nucleus relative to the ipsilateral hemisphere (Baxter et al. 1992). In another study, OCD patients had a significantly elevated rCMRglc compared to normal subjects in the cingulate gyrus, thalamus, and lentiform nucleus (putamen–pallidum complex). Improvement of OCD symptoms by SSRI treatment was associated with a significant reduction in rCMRglc in the cingulate gyrus (Perani et al. 1995). In summary, OCD is associated with a hyperactivity of specific neural circuits, which may be modulated or normalized by successful pharmacologic or behavioral therapy.

A number of symptom-provocation paradigms have been used to study anxiety disorders. While initial studies of patients with simple phobia revealed no association of induced anxiety with measured changes in rCBF (Mountz et al. 1989), later studies clearly showed that rCBF alters with anxiety: patients with simple phobia in whom phobic symptoms had been induced had significant elevations of rCBF in the anterior cingulate gyrus, the insula, the anterior temporal, postcentral, and orbitofrontal cortex, and the thalamus as compared with the nonanxious control state (Rauch et al. 1995). These findings indicate that paralimbic structures participate in the pathogenesis of simple phobia.

The amygdala seems to play an important role in the processing of emotional, anxiety-arousing information. The visual processing of anxious facial expressions by normal subjects is associated with significant activation of the amygdala (George et al. 1993).

The recall of traumatic events by patients with posttraumatic stress disorder has been reported to be associated with an elevation of rCBF (compared to the symptom-free control state) in right-sided limbic and paralimbic areas as well as in visual areas (Rauch et al. 1996).

Dementia

Alzheimer's disease and the associated neuropsychological deficits have been the subject of intensive study using PET. Imaging studies of brain metabolism in Alzheimer's disease consistently show a reduction in neocortical metabolism, particularly in the prefrontal, temporal, and parietal association cortex (Frackowiak

1989; Smith et al. 1992). Except in the early stages of the disease, both hemispheres are usually involved, though sometimes asymmetrically. The severity of dementia is primarily correlated with the extent of reduction of the temporoparietal rCMRglc (Herholz 1995).

In contrast, the typical pattern of altered glucose metabolism in vascular dementia consists of scattered focal cortical and subcortical areas of hypometabolism (Mielke et al. 1996). Functional imaging, like structural imaging, leaves a residual uncertainty in the differential diagnosis of Alzheimer's disease from vascular and other types of dementia (Tedeschi et al. 1995; Brooks 1996).

Three months of treatment with the cholinesterase inhibitor tacrine were shown, in a PET study, to result in an increase in rCMRglc (Nordberg et al. 1992).

The modulation of cholinergic transmitter systems in patients with Alzheimer's disease leads to altered performance on memory tasks, including those involving working memory. Physostigmine, a short-acting cholinesterase inhibitor, was found to improve performance of a working memory task while simultaneously reducing rCBF in the right middle frontal gyrus, an area thought to be involved in working memory. The authors of this study propose, as an explanation of this finding, that enhancement of cholinergic transmission may improve the efficiency of working memory and thereby reduce the effort required to perform a given task (Furey et al. 1997).

3.1.3 *Imaging of Specific Neurotransmitter Systems*

Studies of Drug-Binding Sites

In studies of the receptor binding profile of neuroleptic agents, PET is used to measure the displacement of a receptor-specific radioligand by the applied, nonlabeled drug. With the aid of this technique, it was found that a normally effective dose of a typical neuroleptic (such as 2–6 mg haloperidol) results in a 65%–89% occupancy of putaminal D_2 receptors (Farde et al. 1989). This finding implies that a high-dose treatment strategy would not be expected to increase receptor occupancy significantly.

PET studies of the atypical neuroleptic clozapine revealed a significantly lower D_2 receptor occupancy than with typical neuroleptics (Farde et al. 1992). However, even at relatively low doses of clozapine (125–250 mg/day), more than 80% of the 5-HT_2 receptors were found to be occupied. This finding supports the hypothesis that the atypical clinical effects of clozapine are largely produced by 5-HT_2 receptor antagonism (Nordström et al. 1993). A relatively high 5-HT_2 receptor occupancy has also been found for other atypical neuroleptics and seems to be a characteristic property of these agents. For example, a PET study using the radiolabeled 5-HT_2 ligand [^{18}F]setoperone in normal subjects showed that ziprasidone, an atypical neuroleptic, is a highly potent blocker of 5-HT_2 receptors (Fischman et al. 1996). Seven hours after normal subjects received a single dose of olanzapine, the D_2 receptor occupancy was 59%–63%, and, after 9.5 h, the 5-HT_2 receptor occupancy was 74%–92% (Nyberg et al. 1997).

The preclinical and clinical evaluation of the pharmacokinetic and pharmacodynamic properties of new drugs will certainly be an important future application of PET in psychiatry.

Schizophrenia

When D_2 receptor ligands became available for use in PET studies, scientific attention focused on the question of specific differences in receptor density (B_{max}) and affinity (K_D) among patients with different neuropsychiatric disorders. Initial studies in untreated schizophrenics yielded inconsistent findings. Wong and colleagues at Johns Hopkins repeatedly reported an increased density (B_{max}) of D_2 receptors in these patients compared with normal controls (Wong et al. 1986c; Tune et al. 1993).

These findings could not be replicated, however, by Farde and colleagues (1987) of the Karolinska Institute or by other groups, and the excitement they had initially aroused was considerably dulled. A number of possible explanations for these discrepant findings were subsequently discussed (Andreasen et al. 1988).

In their initial studies, the Karolinska group used [^{11}C]raclopride, a highly specific substituted benzamide, as the radioligand for the D_2 receptor. [^{11}C]Raclopride competes with endogenous dopamine for postsynaptic D_2 receptor sites (Wong et al. 1986a,b; Seeman et al. 1989), while [^{11}C]NMSP is less sensitive to the influence of endogenous dopamine (L.T. Young et al. 1991; Volkow et al. 1994).

A further difference between these two ligands, demonstrated by an in vitro study, is that [^{11}C]NMSP has a higher affinity for D_4 dopaminergic receptors than [^{11}C]raclopride. An autopsy study of the brains of schizophrenic patients revealed increased D_4 receptor density (Seeman et al. 1993). Ligand differences could thus at least partly account for the conflicting results of the PET studies.

Unfortunately, however, the reported differences in receptor density between schizophrenic patients and normal controls could not be replicated even in a study (Nordström et al. 1995) using the same radioligand that had been used in the original study by Wong et al. (1986c), i.e., [^{11}C]NMSP. Likewise, no quantitative alteration of striatal D_2 receptor density was found in studies using the ligands [^{76}Br]bromospiperone (Maziere et al. 1985) and [^{76}Br]bromolisuride (Martinot et al. 1991).

The effort to find an alteration of receptor density in schizophrenic patients tended to obscure other issues, as receptor density is merely one aspect of neurochemical regulatory processes. In view of the physiological ability of neurochemical processes to adapt to internal and external stimuli, B_{max} and K_D are probably variable both in normal individuals as in patients. Alterations in these parameters seem to reflect the general adaptive capability of the system rather than any specific pathophysiologic process.

In a recent PET study of normal subjects using [^{11}C]raclopride, a statistically significant interindividual variation was found in B_{max}, but not in K_D (Farde et al. 1995). This finding supports the hypothesis of interindividual variability of dopaminergic tone.

Other PET studies using [^{11}C]raclopride have demonstrated both short- and long-term stability of D_2 receptor binding parameters in individual subjects (Volkow et al. 1993; Schlösser et al. 1998a). This stability is, however, limited by effects of aging, which have been described for numerous pre- and postsynaptic binding and reuptake sites (Wang et al. 1996).

It has been consistently shown that [^{11}C]raclopride PET is sufficiently sensitive to detect changes in endogenous dopamine resulting from pharmacological intervention. Amphetamine, for example, has been found to lower the binding potential of [^{11}C]raclopride, presumably because of increased release of endogenous dopamine.

A clinical study using this activation paradigm revealed that amphetamine induces a significantly greater reduction of specific striatal [^{11}C]raclopride binding in schizophrenic patients than in normal control subjects (Breier et al. 1997b). Similar findings had been obtained earlier in a SPECT study with [^{123}I]iodobenzamide (IBZM) based on a similar amphetamine challenge paradigm (Laruelle et al. 1996).

These data provided the first direct, in vivo confirmation of the hypothesis of an amphetamine-induced elevation of synaptic dopamine concentration in schizophrenia. They further imply a dopaminergic hyperresponsivity in schizophrenics as compared to normal controls.

After more than a decade of PET studies using specific radiotracers for different neurotransmitter systems, the interactions of these systems are now drawing more attention. Earlier studies of this topic were performed on nonhuman primates (Dewey et al. 1992, 1993). A pharmacological study in healthy men revealed a significant reduction in specific striatal [^{11}C]raclopride binding after the acute administration of fenfluramine, a serotonin releaser and reuptake inhibitor. This finding confirms that serotonin stimulates the dopaminergic system and elevates the intrasynaptic dopamine concentration (Smith et al. 1997).

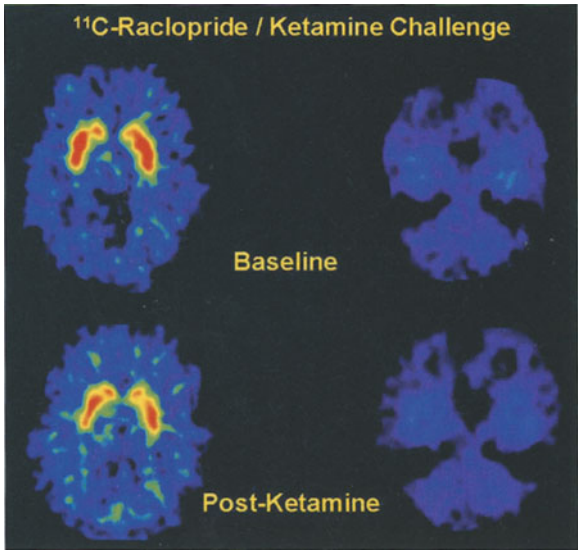


Fig. 4. ^{11}C-raclopride positron emission tomography (PET) images of a normal subject before (*top*) and after (*bottom*) the intravenous administration of ketamine. (After Smith et al. 1998)

The glutamatergic–dopaminergic axis was studied in an analogous fashion after acute administration of the NMDA antagonist ketamine. In Fig. 4, representative [^{11}C]raclopride PET images of a normal subject before and after the acute administration of ketamine are shown. The observed bilateral reduction of [^{11}C]raclopride binding in the basal ganglia is the result of increased release of endogenous dopamine (Smith et al. 1998).

This finding may be relevant to the pathophysiology of the psychotomimetic activity of ketamine, which may be mediated by transient alterations in the dopaminergic system. Future study designs will extend these paradigms to the patient population and thereby enable the investigation of neurotransmitter interactions in various diseases and forms of treatment.

Aside from the controversial discussion of D_2 receptor changes in schizophrenia, other receptor systems, too, have been the target of specific radiotracer binding studies. A study with the D_1 receptor ligand [^{11}C]Sch23390 revealed no difference in striatal binding between schizophrenic patients and normal controls. In the prefrontal cortex, however, ligand binding to D_1 receptors was reduced in schizophrenic patients as compared to the controls. The extent of this reduction was correlated with the severity of negative symptoms and with impairment of performance on the Wisconsin Card Sorting Test, which is used to assess deficits of frontal and other related brain functions. The authors of this study suggest that dysfunction of the D_1 receptor system in

the prefrontal cortex may contribute to the negative symptoms and cognitive deficits seen in schizophrenia (Okubo et al. 1997).

Presynaptic dopaminergic function is another focus of interest in schizophrenia research. Initial studies with [^{18}F]fluorodopa demonstrated a higher fluorodopa influx constant (K_i) in the caudate nucleus of schizophrenic patients than in normal controls (Hietala et al. 1995). In a more recent study, however, K_i did not differ significantly between schizophrenics and controls (Dao-Castellana et al. 1997). The latter study did reveal a higher variability of [^{18}F]fluorodopa uptake values among schizophrenics, which suggests a heterogeneity of presynaptic dopaminergic functional states in schizophrenia.

Affective Disorders

PET neurotransmitter studies have yielded less consistent results for the affective disorders than for schizophrenia, partly because suitable radiotracers for the relevant neurotransmitter systems are lacking. [^{11}C]NMSP, [^{18}F]setoperone, and [^{18}F]altanserine, among other compounds, have been used as 5-HT$_2$ ligands, but their low specificity limited their value for PET studies. In a group of depressed patients studied with [^{18}F]altanserine PET, a significantly lower uptake of the tracer was found, especially in the posterolateral orbitofrontal cortex and the anterior portion of the right insula. There was, however, no correlation between the extent of tracer uptake and the severity of depression (Biver et al. 1997).

As for presynaptic aspects of neurotransmitter systems in depression, a study of patients with unipolar depression revealed a diminished uptake of [^{11}C]5-HT and [^{11}C]L-dopa across the blood–brain barrier. Increased utilization of [^{11}C]5-HT, but not of [^{11}C]L-dopa, was detected in the inferior portion of the medial prefrontal cortex, particularly on the left side. The authors of this study interpret these findings as a reflection of increased serotonin synthesis in this area, possibly as a local compensatory mechanism in the context of a generalized serotonergic hypometabolism (Agren and Reibring 1994).

Dementia

Patients with Alzheimer's disease were found to have a decreased density of nicotinic acetylcholine receptors in a PET study using the radiotracer [^{11}C]nicotine (Nordberg et al. 1990). This finding confirms previous observations from autopsy studies. As for the dopaminergic system, a significant correlation was found between cognitive performance, as measured by the Mini-Mental Status Test, and uptake values (K_i) of [^{18}F]fluorodopa. The last result implies that progression of Alzheimer's disease is associated with increasing impairment of dopaminergic metabolism (Itoh et al. 1994).

The question of whether regional specific decreases in cerebral metabolism in Alzheimer's disease are due to local neuron loss or to a decrease in synaptic activity was addressed by a study of the accumulation of [^{11}C]methionine in protein structure in vivo. No significant decrease was found in the incorporation of [^{11}C]methionine into protein in the temporoparietal and frontal cortex. In the same group of patients, [^{18}F]FDG-PET revealed a 45% decrease in temporoparietal glucose metabolism. It thus seems that decreased synaptic activity plays a larger role than neuron loss in explaining the decrease in cerebral metabolism.

It was demonstrated in one study that treatment with the cholinesterase inhibitor tacrine led to increased uptake of [^{11}C]nicotine into brain tissue, which implies that cholinesterase inhibitor treatment can restore nicotinic cholinergic function. This finding was accompanied by an increase in the local rCMRglc as well as by an improvement in neuropsychological performance parameters (Nordberg et al. 1992, 1997).

In summary, the findings of specific neurotransmitter studies imply that different functional systems may be involved in the pathogenesis of Alzheimer's disease. Further, it seems that a specific pharmacotherapeutic intervention may partially restore both the impaired cholinergic transmission and the reduced cerebral metabolism of this disease.

3.2 Single Photon Emission Computed Tomography

3.2.1 Technical Principles

SPECT is based on the use of proton-rich radionuclides that have the ability to capture an electron, whereupon one of their protons is transformed into a neutron. The daughter nucleus generated by this process then decays further and emits a single photon (i.e., a particle of light or, in equivalent terms, a packet of gamma radiation). The gamma radiation emitted by the commonly used isotopes ^{99m}Tc and ^{123}I is sufficiently intense to traverse the body for external detection.

A collimator is positioned between the radiation source and the scintillation detectors so that only photons with parallel trajectories can reach the detectors. These detectors are either arrayed in several detector heads that rotate around the patient or else directly integrated into a ring detector. After the amount of radiation in many individual trajectories is measured, the distribution of the radiotracer can be calculated with suitable mathematical algorithms. Intra-arterially injected or inhaled ^{133}Xe was used as a radiotracer in many earlier studies, in which fixed

external radiation detectors were used. The spatial resolution of this technique was relatively poor, and low-lying areas of the brain could not be studied with it. The development of rotating gamma cameras was a major technical improvement. Currently available SPECT scanners provide a spatial resolution of 6–8 mm.

SPECT has the major advantage of not requiring a cyclotron at the site of the scanner, which spares the attendant costs of cyclotron installation and maintenance and of on-site radiochemistry. SPECT cameras are therefore in more widespread clinical use than PET systems. The tracers used in SPECT have relatively long half-lives and may thus be delivered from the site of synthesis to the SPECT scanner site, as is usually done with ^{123}I, if they are not synthesized at the scanner site itself, as is usually done with ^{99m}Tc.

Despite the progressive improvement of both the spatial resolution and the sensitivity of scanner systems, absolute quantification with SPECT remains problematic. A number of factors limit the sensitivity of the method. The photons relevant to SPECT possess a lower energy (80–160 keV) than those detected in PET (511 keV) and are thus more subject to attenuation, and the collimator eliminates a large number of photons. Because only a single photon is emitted, coincidence detection (as in PET) has no application to SPECT, which further limits its spatial resolution. The problems of correcting for scattering and attenuation, too, have been only partially solved to date.

One of the most commonly used radioligands for SPECT studies of rCBF and presynaptic metabolic activity is ^{99m}Tc-hexamethyl propyleneamine oxime (HMPAO). This tracer rapidly crosses the blood–brain barrier, is then transformed into a hydrophilic metabolite, and enters the cells. The tracer remains stable intracellularly for several hours, so that scanning may take place at any time within several hours after the injection.

A number of specific SPECT tracers are available for different neurotransmitter systems for both pre- and postsynaptic sites of transmitter activity. Some of the more important ligands are [^{123}I]IBZM for D_2 and D_3 receptors, [^{123}I]iomazenil for the benzodiazepine receptor, and [^{123}I]β-CIT for the serotonin, dopamine, and noradrenaline transporter (Bartenstein and Koepp 1995; Schlösser and Schlegel 1995; Schlösser et al. 1997). Two other ligands for the D_2 receptor are [^{123}I]IBF and [^{123}I]epidepride. The latter has markedly higher affinity for striatal D_2 receptors than [^{123}I]IBZM (Kornhuber et al. 1995) and is therefore not suitable for challenge studies in which the ligand competes for receptor sites with the endogenous neurotransmitter. Nonetheless, its higher affinity makes [^{123}I]epidepride suitable for the study of extrastriatal D_2 receptors (Kuikka et al. 1997).

Because absolute quantification was hardly possible until recently, semiquantitative methods were developed that were sufficient for many research questions. The so-called ratio approach and the analysis of radiotracer uptake and washout curves were the most commonly used methods for semiquantitative determination of the receptor binding parameters B_{max} and K_D.

In the ratio approach, the uptake of a radiotracer in a receptor-rich target region is compared with its uptake in a reference region that, ideally, contains no, or negligibly few, receptors of the type under study. In the case of [^{123}I]IBZM, the striatum is the target region and the cerebellum or specific cortical regions are used as the reference region. The striatal-to-cerebellar uptake ratio provides a means of estimating the occupancy of striatal D_2 receptors. The receptor occupancy by a drug is calculated from the change in this ratio before and after its administration.

By studying uptake and washout curves, the time to maximum uptake, and the washout rate, after the administration of a drug can be determined and the values of these parameters can be compared to those obtained without the drug. Continuous infusion of the radiotracer at a constant rate by means of a suitable perfusion device has been used successfully to maintain stable levels of striatal and occipital activity (Laruelle et al. 1994). The main advantage of the continuous infusion method is the reduction of variability in cortical uptake due to alterations in blood flow.

3.2.2 *Imaging of Blood Flow*

A number of research questions have been addressed in comparable ways with both PET and SPECT. To avoid redundancy, we shall largely confine the following discussion to results that were not already mentioned in the section above on PET.

Schizophrenia

Despite the technical limitations discussed above, the ^{133}Xe method allowed fundamental observations to be made concerning cortical activation during normal and pathologically altered states of cerebral function. Ingvar and Franzen (1974), using this method, made the important discovery that frontal cortical blood flow is relatively reduced in schizophrenic patients; the shorthand term "hypofrontality" expresses this finding. It was later found, again using the ^{133}Xe method, that schizophrenic patients who performed poorly on the Wisconsin Card Sorting Test had a lower degree of activation in the dorsolateral prefrontal cortex (DLPFC) than normal control subjects (Weinberger et al. 1986). A similar finding of deficient frontal activation in schizophrenics was made in a SPECT

study involving the Tower of London task (Andreasen et al. 1992). The tasks used in these two studies both rely on an intact working memory, and the observed results are in accord with the reported impairment of working memory in schizophrenia.

A ^{99m}Tc-HMPAO-SPECT study of untreated patients with acute schizophrenia revealed patterns of both hyper- and hypoperfusion, depending on the predominant form of psychopathology. Scores on a rating scale for formal thought disorders and grandiosity were positively correlated with bifrontal and bitemporal rCBF, while scores for delusional ideas, hallucinations, and suspiciousness were negatively correlated with rCBF in bifrontal and left temporal cortex as well as in the cingulate gyrus and the left thalamus. Stereotyped ideas were negatively correlated with left frontal, temporal, and parietal rCBF (Sabri et al. 1997).

Affective Disorders

^{99m}Tc-HMPAO-SPECT studies support the hypothesis of diminished metabolic activity in the frontal, temporal, and parietal cortex in depression. The severity of depressive symptoms was negatively correlated with regional uptake parameters (Schlegel et al. 1989a). Although a direct comparison is difficult because of differences in data acquisition and analysis, the results of the ^{99m}Tc-HMPAO-SPECT studies generally agree with those of the PET studies.

Anxiety Disorders

SPECT has been used successfully to study alterations of regional cerebral blood flow in various anxiety disorders. The results confirm and complement those obtained in PET studies.

It was shown, with use of the ^{133}Xe inhalation method, that patients with panic disorder have either a mild increase or a decrease in hemispheric blood flow after the induction of a panic attack by lactate infusion. In contrast, lactate infusion resulted in a significant increase in hemispheric blood flow both in normal control subjects and in patients with panic disorder who did not experience a panic attack after the infusion (Stewart et al. 1988).

In OCD patients, too, ^{99m}Tc-HMPAO-SPECT has revealed a hyperfrontal metabolic pattern. Treatment with the SSRI fluoxetine reversed this hypofrontality, while (despite contrasting PET findings) the metabolic rate in the basal ganglia was unchanged (Hoehn-Saric et al. 1991).

Dementia

^{99m}Tc-HMPAO-SPECT studies of Alzheimer's disease quite consistently reveal a temporoparietal perfusion deficit (Lang et al. 1990). Data on sensitivity and specificity make clear that the discriminating ability of SPECT improves with increasing severity of dementia. Alzheimer's disease could be diagnosed with 90% specificity on the basis of the SPECT perfusion pattern, while the sensitivity of SPECT varied from 42% to 79% depending on the severity of the disease (Claus et al. 1994). Modern pattern recognition algorithms, e.g., with the aid of artificial neural networks, seem to be superior to classical statistical approaches in the analysis of perfusion patterns of patients with Alzheimer's disease (Page et al. 1996).

Vascular dementia and the so-called pseudodementia of depression are important alternative diagnoses to Alzheimer-type dementia. The temporoparietal hypoperfusion pattern revealed by SPECT scan in most patients with Alzheimer's disease (discussed above) was also seen in other types of dementia and thus cannot be regarded as specific for Alzheimer's disease (Masterman et al. 1997). Patients with vascular dementia, however, often had a higher uptake in anterior parietal regions than patients with Alzheimer's disease.

Patients with relatively higher posterior temporal perfusion and frontal reduction in rCBF or rCMRglc may reportedly be more likely to suffer from depressive pseudodementia than from Alzheimer's disease (Curran et al. 1993). Once again, it must be emphasized that SPECT and PET do not currently provide a reliable diagnosis in individual cases, because there is considerable overlap between groups.

3.2.3 Imaging of Specific Neurotransmitter Systems

Schizophrenia

Like the D_2 receptor ligands used in PET studies, [^{123}I]IBZM competes with endogenous dopamine for binding to the postsynaptic D_2 receptor. An acute challenge with the dopamine releaser amphetamine caused a significantly greater increase in the [^{123}I]IBZM binding potential in schizophrenic patients than in normal controls, presumably because more endogenous dopamine was released in schizophrenic patients. There was simultaneous development of psychotic positive symptoms. These findings, depicted in Fig. 5, suggest that schizophrenic patients are more sensitive to the dopamine-releasing effect of amphetamine, which in turn may be due to a pathologically altered responsiveness of dopaminergic neurons in this disease (Laruelle et al. 1996).

Central benzodiazepine receptors were studied in schizophrenic patients with the aid of the benzodiazepine receptor ligand [^{123}I]iomazenil. No overall differences in benzodiazepine receptor binding were found between schizophrenic patients and normal control subjects. There were, however, significant correlations between the severity of schizophrenic symptoms and [^{123}I]iomazenil binding in the limbic cortex: positive symptoms were negatively correlated with benzodiazepine receptor binding in the left

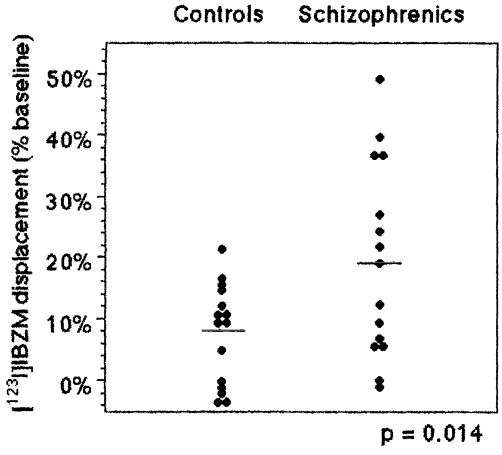


Fig. 5. Relative decrease in striatal [^{123}I]iodobenzamide (IBZM) binding after the acute administration of amphetamine to schizophrenic patients and control subjects, expressed as a percentage of the initial value. (After Laruelle et al. 1996)

medial temporal cortex, while negative symptoms were inversely related to benzodiazepine receptor binding in medial frontal areas. These results suggest that a reduction in inhibitory GABAergic tone contributes to the generation of schizophrenic symptoms (Busatto et al. 1997b).

Affective Disorders

Depressive patients were shown to have a diminished uptake of the 5-HT_2 antagonist [^{123}I]ketanserin in the parietal cortex, as well as a lateral asymmetry favoring the right side in the inferior frontal area, in comparison to normal controls. This finding suggests that alterations in the 5-HT_2 receptor system may play a role in the pathogenesis of depression (D'haenen et al. 1992).

Anxiety Disorders

Patients with panic disorder were found to have a lower [^{123}I]iomazenil uptake in frontal, occipital, and temporal cortex than patients with epilepsy (Schlegel et al. 1994). Another study of patients with panic disorder revealed an elevated right-to-left ratio of the uptake of the benzodiazepine receptor ligand in the middle and inferior frontal gyri (Kuikka et al. 1995).

The results described up to this point were obtained with the aid of semiquantitative techniques for the determination of [^{123}I]iomazenil uptake and thus might conceivably be due to alterations in rCBF rather than to specific benzodiazepine receptor effects. A study by Kaschka et al. (1995) provided evidence that this is so, to some extent. On the assumption that iomazenil uptake values 10 min after tracer injection largely represent perfusion-related effects, while the values after 90 min reflect specific ligand binding properties, the ratio of uptake values at these two times was used to correct semiquantitatively for perfusion-related tracer uptake. Even after this correction was made, patients with panic disorder were still found to have a hypoactivity in the left lateral temporal cortex.

Dementia

Patients clinically judged to have probable mild Alzheimer's disease were found to have reduced binding of the muscarinic acetylcholine receptor ligand [^{123}I]4-iododexetimide in temporal and temporoparietal cortical areas (Claus et al. 1997). This finding implies that the density of muscarinic acetylcholine receptors in these areas is decreased.

Chronic administration of scopolamine resulted in an increased uptake of the muscarinic acetylcholine receptor ligand [^{123}I]QNB in patients with Alzheimer's disease, but a decreased uptake of the same ligand in a control group of nondemented elderly subjects. Patients with Alzheimer's disease thus seem to be especially sensitive to chronic cholinergic blockade (Sunderland et al. 1995).

3.3 Functional Magnetic Resonance Imaging

3.3.1 Technical Principles

In recent years, fMRI has played an increasingly large role in the study of regional blood flow and neural activity. Focal changes in neural activity lead to focal changes in perfusion, blood oxygenation, and metabolism. The technique of fMRI exploits these changes to study brain activation during the performance of behavioral or neuropsychological tasks. The noninvasiveness of the technique, the absence of radiation exposure, and the ability to perform studies on widely available clinical scanners all contribute to the growing popularity of fMRI

Most fMRI studies at present use a method known as blood oxygenation level-dependent (BOLD) contrast. Perfusion studies have revealed that cerebral activation induces a local increase in oxygen delivery that actually exceeds the local metabolic demand (Fox and Raichle 1986). Increased oxygenation leads to a higher concentration of diamagnetic oxyhemoglobin and to a lower concentration of paramagnetic deoxyhemoglobin in the activated area. As a result, field inhomogeneity is decreased, and the local T_2*-weighted MR signal becomes stronger.

Rapid imaging sequences are favored because they enable high temporal resolution and reduction of motion artifact. The principle underlying echo-planar imaging (EPI) is that all of the desired information in the spatial frequency domain can be measured after a single RF pulse. The advantages of EPI include high

speed of acquisition (times as short as 50 ms) and high signal-to-noise ratio. The best spatial resolution that can be achieved at present with EPI using BOLD contrast is on the order of 1 × 1 mm in plane, with a slice thickness of 3 mm. The technical quality of such images is steadily improving as new scanners are being developed.

Multislice EPI fMRI of the whole brain with good spatial resolution can be performed in a few seconds. The signal-to-noise ratio can be improved further by the use of magnets stronger than 1.5 T. Scanners incorporating such strong magnets are not yet in widespread use. Further technical problems limiting the applicability of fMRI in clinical studies include field inhomogeneity and temperature gradients, which are increasingly troublesome as the field strength increases.

Fast fMRI techniques enable single images to be acquired practically as snapshots, in which motion artifact is not a significant problem. Nonetheless, there is still an artifact arising from movement in between image acquisitions, and technical measures are still necessary to limit the size of this artifact. Physical head immobilizers, such as head-holders, are used as well as computer algorithms that correct for movement (Woods et al. 1992).

Not only motion artifact, but also physiological artifacts – arising from respiration, heartbeats, and cerebral pulsations, among other sources – must be taken into account in the interpretation of fMRI data. There is as yet no fully satisfactory solution to the problem of signals arising from blood flow in large vessels such as cerebral veins. Microvascular changes related to neural activity occur at the site of activation, but altered blood flow in veins draining the site of activation might induce signal changes at other locations as well (Segebarth et al. 1994). Approaches to a solution of this problem include the use of more or less specific signal thresholds and of angiographic sequences for the subtraction of vascular signals.

The observed signal changes in fMRI activation studies at 1.5 T generally lie in the range of 1%–5%. A large number of images are obtained in rapid sequence so that activation paradigms can be carried out, which, as in PET, consist of alternating blocks of behavioral or neuropsychological tasks. In addition to this type of block design, it is now possible to study the hemodynamic correlates of single sensory or cognitive events. This technique is known as "event-related fMRI" (Rosen et al. 1998).

It should be noted that neither emission-tomographic techniques nor fMRI detect neural activity directly. Both approaches are used to measure correlates of neural activity that normally follow neural activation after a certain temporal delay. PET is characterized by relatively low temporal latency, with acquisition times on the order of 40 s to several minutes for a single image. Single fMRI images may be acquired in a considerably shorter time (currently about 50 ms).

MEG and EEG studies have shown that neural activity returns to its starting value 300–800 ms after the end of the activating stimulus. The hemodynamic response persists, however, for as long as 9 s (Bandettini et al. 1993). Thus the temporal resolution of fMRI appears to be limited more by the nature of the hemodynamic response curve itself than by technical aspects. This fact, in turn, implies the necessity of exact modeling of the underlying hemodynamic response function, as can now be done with the aid of suitable statistical techniques.

3.3.2 Findings in Normal Subjects

In recent years, functional activation imaging has been used to study motor, sensory, and higher association areas of the brain. Because of the absence of radiation exposure, fMRI is an ideally suitable technique for functional brain mapping. Reliable paradigms have been developed for the study of various functional modalities. In the following sections, we will present a few representative examples, in several modalities, among the many paradigms that have been used to date.

Language Processing

Studies using language activation paradigms have yielded new information regarding the neural networks underlying language processing. The cerebral areas that participate in language processing seem to extend considerably beyond the "classical" areas of Broca and Wernicke into frontal, temporal, and parietal cortex (Hinke et al. 1993; Imperato et al. 1994).

Initial fMRI studies seemed to support the hypothesis of sex differences in language processing, e.g., on the phonological level (Shaywitz et al. 1995). More recently, however, more sensitive studies failed to confirm the existence of such differences (Frost et al. 1997; Schlösser et al. 1998b). Figure 6 shows the results of an fMRI study using a word-finding task as its activation paradigm; areas in which the signal was increased during performance of the task are indicated. In this right-handed male subject, the left frontal and parietal cortex are seen to be activated.

Working Memory

Various tasks have been used as activation paradigms for the study, by fMRI, of brain areas involved in working memory.

The sequential letters task or, as it is also known, the *n*-back task is one of the most common activation paradigms for the study of various components of

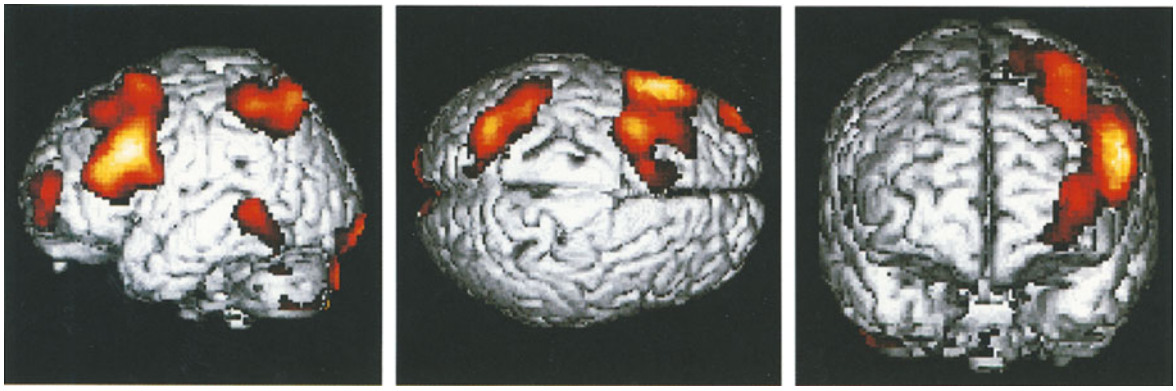


Fig. 6. Functional magnetic resonance imaging (fMRI) study using a word-finding task as the activation paradigm in a healthy right-handed male subject. Areas of increased signal during the activating condition are shown. (Data from Schlösser et al. 1998b)

working memory, having been used in many different studies, including imaging studies (Cohen et al. 1994). One version of this task involves a series of sequences of letters. In the 0-back condition, the subject is told to press a button each time a previously defined letter appears. In the 2-back condition, the subject must press the button whenever a letter appears that previously appeared two letters back, and so forth. The 2-back condition, when compared with the 0-back condition, was found to be associated with activation of the frontal and parietal association cortices (among other effects). Furthermore, when the demand on working memory was systematically altered by switching to the 1-back or 3-back condition, an increase in activation was found that depended on the degree of demand placed on working memory (Braver et al. 1997).

Further variations of applied neuropsychological paradigms allow the delineation of individual components of working memory and associated sensory processes. It was shown, with the aid of visual working memory tasks, that occipitotemporal areas respond only transiently to visual stimuli, in accordance with their primary role in the processing of sensory information. Prefrontal areas, in contrast, are continually activated over the entire period of information acquisition. This finding accords with their major attributed role in working memory (Courtney et al. 1997).

Long-Term Memory

fMRI has also been used to study long-term memory. In vivo studies have shown that the hippocampus and other medial temporal structures are involved in the processes of encoding and storage of information in long-term memory. During the encoding of novel images, statistically significant increases in the fMRI signal were seen bilaterally in the posterior portion of the hippocampus, in the parahippocampal gyrus, and in the lingual and fusiform gyri (Stern et al. 1996). Other studies have shown that the frontal cortex also participates in the encoding and recall of information stored in long-term memory (Busatto et al. 1997a).

Emotional Processing

fMRI is also suitable for the study of brain activation in connection with emotional processing. The left amygdala was found to be activated by the induction of either a sad or a happy mood, as had been found in an earlier PET study (Schneider et al. 1997).

3.3.3 Patient Studies

Schizophrenia

fMRI studies of schizophrenia focused at first on the primary sensory and motor cortices. Renshaw et al. (1994) used fMRI to study alterations in regional neural activity induced by visual stimulation. The mean change in signal intensity in the primary visual cortex was significantly higher in schizophrenic patients than in normal control subjects. The reason for this difference was unclear; an abnormality of the cerebral vasculature in schizophrenic patients was suggested as a possible explanation (Cohen et al. 1995). Schizophrenic patients were further found to have a lower than normal activation of both the sensorimotor cortex and supplementary motor areas during the performance of a finger-to-thumb opposition task (Schröder et al. 1995). Another study using the same task revealed no difference in the motor cortex of schizophrenic and control subjects (Buckley et al. 1997).

After these relatively simple studies of unimodal cortical areas, higher cognitive functions, too, were increasingly studied.

Studies of schizophrenic patients have repeatedly revealed multiple neuropsychological abnormalities,

particularly in the area of working memory. While performing the 2-back task described above, schizophrenic patients had a significantly lower degree of activation of the prefrontal cortex than normal control subjects (Callicott et al. 1998). In another study using the Wisconsin Card Sorting Test, schizophrenic patients were found to have a lower degree of activation in the right frontal lobe, and a higher degree of activation in the left temporal lobe, than normal control subjects (Volz et al. 1997a).

Word-finding tasks activate widely distributed neural networks and seem particularly suitable for the study of a number of patient groups. During such a task, schizophrenic patients had a significantly weaker left frontal activation, but a significantly stronger left temporal activation, than normal control subjects (Yurgelun-Todd et al. 1996). A more recent study, again using a word-finding task, revealed weaker left prefrontal activation and a reduced response in the left parietal cortex of schizophrenics as compared to normal control subjects (Curtis et al. 1998).

These findings all point to a diminution, in schizophrenics, of the frontal cortical activation associated with working memory and language processing and are thus in accordance with the findings of PET activation studies. Meanwhile, other functionally related areas also seem to be involved in the pathophysiology of schizophrenia.

Obsessive–Compulsive Disorder

The use of individually tailored provocative paradigms to induce obsessive–compulsive symptoms in patients with OCD was found, in an fMRI study, to produce an activation of various brain areas, including the medial orbitofrontal, lateral frontal, and anterior temporal cortices as well as the anterior portion of the cingulate gyrus. The caudate nucleus and the amygdala were also activated. Normal control subjects, in contrast, had no significant alteration of fMRI signal in response to these stimuli (Breiter and Rauch 1996; Breiter et al. 1996). These findings indicate a pathological overactivation of frontostriatal circuits in OCD and are thus consistent with the findings of earlier metabolic PET studies.

Pharmacological Modulation of the fMRI Signal

Preliminary fMRI studies have been performed concerning alterations of the cerebral activation pattern by the administration of drugs. In one such study, patients with narcolepsy were scanned during continuous presentation of periodic auditory and visual stimuli, before and after the administration of amphetamine. Amphetamine reduced the sensory-evoked activation of normal control subjects, while, in the narcoleptics, an increased activation was seen in the primary sensory cortex and the sensory association cortex (Howard et al. 1996). This finding demonstrates that drug challenge studies with fMRI are feasible. Similar study designs are likely to be used increasingly in the future.

fMRI is potentially a powerful tool for the study of cognitive activation in normal subjects and in patients with psychiatric diseases. This technique is likely to be used for many different applications in the future.

3.4 Magnetic Resonance Spectroscopy

3.4.1 Technical Principles

MRS enables the in vivo study of specific chemical substances and processes, including membrane components and substrates of energy metabolism, neurotransmitter metabolism, and the metabolism of pharmacologic agents.

The technical principles of MRS are essentially the same as those of MRI. As discussed above, each type of atomic nucleus resonates at a specific frequency. When the atomic nuclei reorient themselves along the axis of the static magnetic field, they emit a signal at their specific resonance frequency, which contributes to the generation of an MR signal. Regardless of the characteristics of the applied external gradient, however, not all of the atomic nuclei of a given type resonate at exactly the same frequency. This is because the local magnetic field experienced by each nucleus is slightly perturbed by its chemical microenvironment. This slight deviation of the resonance frequency is known as chemical shift. While MRI and fMRI make no special use of this additional frequency information, MRS resolves the full range of different resonance frequencies in the form of a spectrum. The power of MRS to resolve slightly different frequencies is a function of the strength of the applied magnetic field. A highly homogeneous field of at least 1.5 T is generally necessary to achieve adequate resolution in the frequency domain for the detection of differences in chemical composition.

The MRS techniques most commonly applied in psychiatric research are ^{1}H- and ^{31}P-spectroscopy. ^{1}H-Spectroscopy provides a chemical spectrum including peaks corresponding to *N*-acetyl aspartate (NAA), creatine (Cr), phosphocreatine (PCr), choline (Cho), *myo*-inositol (mIns), glutamine, glutamate, aspartate, γ-aminobutyric acid (GABA), and lactate. NAA is considered to be a marker of neuronal integrity, although the precise function of this substance is not yet completely understood (Tsai and Coyle 1995). Because the signal produced by these molecules is embedded in a signal arising from a very large number of water molecules, special techniques must be used to suppress the water signal.

^{31}P-Spectroscopy enables the detection of cerebral membrane phospholipids and high-energy phosphates. The phosphate spectrum contains peaks corresponding to adenosine triphosphate (ATP, β-ATP, and α-ATP), phosphocreatine (PCr), phosphodiesters (PDE), phosphomonoesters (PME), and inorganic phosphate (P_i). The PME and PDE peaks are considered to be indicators of membrane synthesis, and low values of the PME to PDE ratio are considered to be a manifestation of a low rate of phospholipid synthesis.

Other nuclei of importance to MRS for psychiatric research are ^{7}Li, ^{13}P, and ^{19}P.

3.4.2 Findings

Schizophrenia

A decreased level of PME and P_i and an increased level of ATP and PDE were found in the dorsolateral prefrontal cortex of drug-naive, first-episode schizophrenic patients (Pettegrew et al. 1991). A more recent study revealed a significantly diminished level of PME in the frontal regions of schizophrenic patients who had high scores on negative symptom rating scales (Shioiri et al. 1994). These findings suggest reduced synthesis and/or increased degradation of membrane phospholipids. Another study, however, showed a lower level of PDE in schizophrenic patients than in normal control subjects (Volz et al. 1997b). These contradictory findings may perhaps be explained by differences in MRS techniques and localization methods.

An ^{1}H-MRS study of schizophrenic patients revealed a decreased concentration of NAA in the frontal lobes (Buckley et al. 1994). Bertolino et al. (1996) found significantly decreased NAA to Cr and NAA to Cho ratios bilaterally in the hippocampus and dorsolateral prefrontal cortex of schizophrenic patients; there were no significant changes in the Cho to Cr, NAA to Cr, or NAA to Cho ratios in any of the other regions tested. Thus a lower NAA concentration was found exclusively in two areas that have been repeatedly implicated in the pathogenesis of schizophrenia. The regional distribution of metabolites obtained in this study is depicted in Fig. 7. In another study, the anterior portion of the cingulate gyrus also had a significantly lower NAA concentration in schizophrenic patients than in normal control subjects (Deicken et al. 1997). In addition, increased levels of glutamate were found in the medial prefrontal cortex and in the anterior portion of the cingulate gyrus. The authors of this last study interpret their findings as evidence of dysfunction of the glutamatergic system, because glutamine is the immediate precursor of glutamate in its synthetic pathway (Bartha et al. 1997).

In schizophrenic patients, memory functions were found to be closely related to creatine concentration in the temporal lobes, but not in the frontal lobes – a pattern of distribution different from that seen in normal control subjects. These findings suggest that disturbances of memory in schizophrenic patients may be associated with a specific abnormal pattern of temporal lobe metabolism (Buckley et al. 1994).

The basal ganglionic NAA to Cho ratio was found to be significantly reduced bilaterally in schizophrenic patients. This finding suggests that there is neuronal dysfunction in subcortical structures as well (Fujimoto et al. 1996).

Affective Disorders

^{31}P-MRS studies provided evidence of alterations of membrane metabolism and high-energy phosphate metabolism in patients with affective disorders. The PME level was elevated in patients in either the manic or the depressive phase of a bipolar I disorder and depressed in the same patients in the euthymic state. Bipolar II patients, in contrast, had reduced PCr values both during disease episodes and in the euthymic state (Kato et al. 1994b). The observed abnormalities had no correlation with the cerebral concentration of ^{7}Li, which was simultaneously determined by MRS (Kato et al. 1993b). ^{1}H-MRS revealed an elevated Cho to (Cr+PCr) ratio in the left basal ganglia of bipolar patients in the euthymic state, which may be interpreted as evidence of cell membrane damage in bipolar disorders (Kato et al. 1996b).

Anxiety Disorders

In a ^{31}P-MRS study, patients suffering from panic disorders were found to have a slightly decreased concentration of inorganic phosphate and a significant asymmetry (left side greater than right) of PCr concentration in the frontal cortex. Abnormalities of phosphorus metabolism in frontal cortex are thus possibly a feature of panic disorders (Shioiri et al. 1996b).

Dementia

MRS studies reveal that the early stages of Alzheimer's disease are associated with increased concentrations of cerebral PME. Later in the course of the disease, PDE and PCr concentrations increase (Pettegrew et al. 1988, 1994). The increase in PME may reflect alterations of membrane metabolism, while the increase in PDE and PCr indicates the presence of neuronal degeneration. Recent studies indicate an increase in the PME to PDE ratio by approximately 50%. Nonetheless, no differences of absolute PME or PDE concentration between schizophrenic patients and normal control subjects could be found (Gonzalez et al. 1996).

A lower NAA concentration in the temporal and frontal cortex of Alzheimer's disease patients, as compared to healthy age-matched control subjects, has been described (Parnetti et al. 1997). This decreased NAA concentration was significantly correlated with

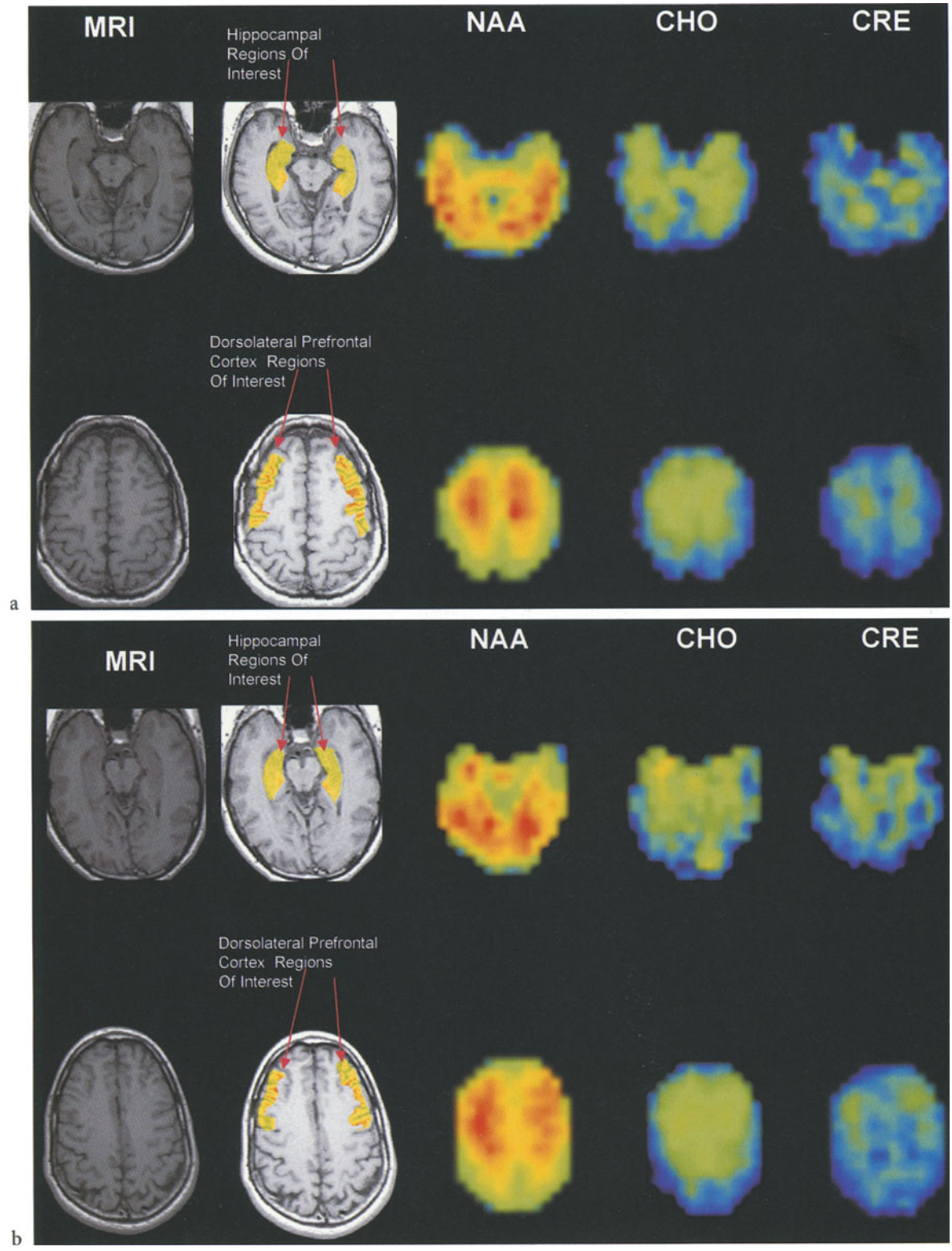


Fig. 7a–c. Studies using ^{1}H-magnetic resonance spectroscopy (MRS) in schizophrenic patients and normal control subjects. (After Bertolino et al. 1996). **a,b** Metabolite signal intensity of compounds containing *N*-acetyl aspartate (*NAA*), compounds containing choline (*CHO*), and creatine/phosphocreatine (*CRE*) in **a** a schizophrenic patient and **b** a normal control subject. Also shown are the areas used for regional analysis in the hippocampus and dorsolateral prefrontal cortex. **c** Representative ^{1}H-MRS spectra of a patient with schizophrenia (*two left columns*) and a normal control subject (*two right columns*) in 11 anatomical regions

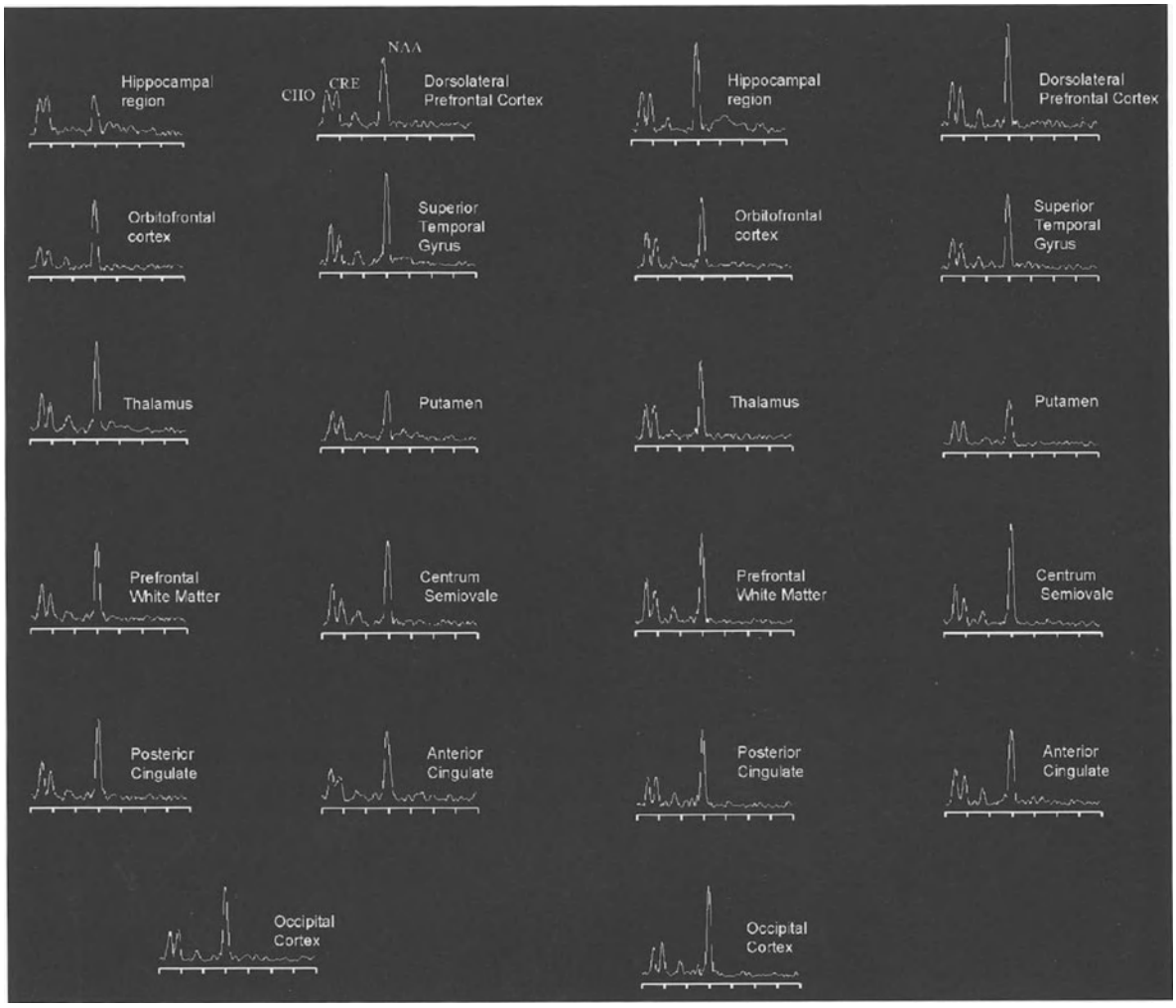


Fig. 7c

the extent of these patients' cognitive disturbance (Heun et al. 1997b; Parnetti et al. 1997). Elevated concentrations of *myo*-inositol, which were also revealed by an ^{1}H-MRS study of patients with Alzheimer's disease, may reflect changes in the polyphosphate second messenger system (Miller et al. 1993).

Treatment Studies

With the aid of ^{7}Li-MRS, it has been shown that the concentration of lithium in the brain is approximately half that in serum (Kato et al. 1993a). It was also shown in bipolar patients in the manic phase that the response to lithium treatment is more closely correlated to the brain concentration than to the serum concentration of lithium (Kato et al. 1994a). Furthermore, the occurrence of tremor was also found to be correlated to the brain lithium concentration (Kato et al. 1996a).

^{19}F has also been used in MRS studies for the determination of brain concentrations of various fluorine-containing drugs, such as fluvoxamine. Nonetheless, no clear relationship of these concentrations to the treatment response of psychiatrically ill patients has yet been established (Strauss et al. 1997).

MRS has, further, enabled the study of the effect of pharmacologic interventions on biochemical processes in the brain. Thus a positive correlation was found between the chlorpromazine equivalent dose of a neuroleptic treatment and the measured NAA concentration in the basal ganglia of schizophrenic patients. There was, at the same time, a negative correlation between the neuroleptic dose and the Cho to NAA ratio (Shioiri et al. 1996a).

The results presented here illustrate the special ability of MRS to probe various biochemical systems in vivo without prior injection of a tracer substance. The further development of MRS and its application to the study of psychiatric disease and, possibly, to the monitoring of treatment with psychotropic drugs

depend largely on the stability of the technique. At present, the intertest reliability of MRS still seems to be in need of improvement; in contrast, the intertest reliability of metabolic and neuroreceptor PET studies is now considered adequate (Marshall et al. 1996; Bertolino et al. 1998). This residual methodological uncertainty must be taken into account, particularly in view of the partially conflicting findings of MRS studies.

4 Prospects

Functional imaging is a new approach to the study of psychiatric diseases and the pathophysiologic processes that underlie them. The deeper understanding of these diseases thus acquired should enable the development of new biological concepts and innovative treatments. In the future, the classical nosology of psychiatric disease may perhaps be replaced, at least in part, by a functional classification.

Future research involving functional imaging with PET and fMRI will likely move increasingly in the direction of determining the role played by functional networks and their connectivity in the genesis of psychopathological abnormalities. The application of in vivo challenge paradigms may further lead to the establishment of dose–response relationships relating to the functional interaction of multiple neurotransmitter systems.

The various methods described in this chapter will, in future, be integrated into multimodal imaging systems in which anatomical, functional–metabolic, and specific neurochemical information will be displayed simultaneously. The monitoring of therapeutic interventions and the study of neuropsychological performance ability by means of functional imaging will be of major clinical utility.

It is also becoming evident that the fields of psychiatry and neurology are moving closer together with respect to the questions they pose and the methods they use to answer them, particularly the modern methods of functional imaging described in this chapter. Both clinical medicine and scientific research will benefit from this new, integrated neuropsychiatric approach.

5 References

Agren H, Reibring L (1994) PET studies of presynaptic monoamine metabolism in depressed patients and healthy volunteers. Pharmacopsychiatry 27: 2–6

Agren H, Reibring L, Hartvig P et al (1991) Low brain uptake of L-[11C]5-hydroxytryptophan in major depression: a positron emission tomography study on patients and healthy volunteers. Acta Psychiatr Scand 83: 449–455

Andreasen NC (1988) Evaluation of brain imaging techniques in mental illness. Annu Rev Med 39: 335–345

Andreasen N, Nasrallah HA, Dunn V et al (1986) Structural abnormalities in the frontal system in schizophrenia. A magnetic resonance imaging study. Arch Gen Psychiatry 43: 136–144

*Andreasen NC, Carson R, Diksic M et al (1988) Workshop on schizophrenia, PET, and dopamine D2 receptors in the human neostriatum. Schizophr Bull 14: 471–484

*Andreasen NC, Ehrhardt JC, Swayze VW, Alliger RJ, Yuh WT, Cohen G, Ziebell S (1990a) Magnetic resonance imaging of the brain in schizophrenia. The pathophysiologic significance of structural abnormalities. Arch Gen Psychiatry 47: 5–44

Andreasen NC, Swayze VW, Flaum M, Yates WR, Arndt S, McChesney C (1990b) Ventricular enlargement in schizophrenia evaluated with computed tomographic scanning. Effects of gender, age, and stage of illness. Arch Gen Psychiatry 47: 1008–1015

Andreasen NC, Rezai K, Alliger R et al (1992) Hypofrontality in neuroleptic-naive patients and in patients with chronic schizophrenia. Assessment with xenon 133 single-photon emission computed tomography and the Tower of London. Arch Gen Psychiatry 49: 943–958

Andreasen NC, Flaum M, Swayze V, O'Leary DS, Alliger R, Cohen G, Ehrhardt J, Yuh WT (1993) Intelligence and brain structure in normal individuals. Am J Psychiatry 150: 130–134

Andreasen NC, O'Leary DS, Flaum M, Nopoulos P, Watkins GL, Boles Ponto LL, Hichwa RD (1997) Hypofrontality in schizophrenia: distributed dysfunctional circuits in neuroleptic-naive patients. Lancet 349: 1730–1734

Aquilonius SM, Bergstrom K, Eckernas SA et al (1987) In vivo evaluation of striatal dopamine reuptake sites using 11C-nomifensine and positron emission tomography. Acta Neurol Scand 76: 283–287

Baker SC, Frith CD, Dolan RJ (1997) The interaction between mood and cognitive function studied with PET. Psychol Med 27: 565–578

Bandettini PA, Jesmanowicz A, Wong EC, Hyde JS (1993) Processing strategies for time-course data sets in functional MRI of the human brain. Magn Reson Med 30: 161–173

Bartenstein P, Koepp M (1995) Benzodiazepine receptor imaging with positron emission tomography and single photon emission tomography. Nervenarzt 66: 412–421

Bartha R, Williamson PC, Drost DJ et al (1997) Measurement of glutamate and glutamine in the medial prefrontal cortex of never-treated schizophrenic patients and healthy controls by proton magnetic resonance spectroscopy. Arch Gen Psychiatry 54: 959–965

Bartlett EJ, Barouche F, Brodie JD, Wolkin A, Angrist B, Rotrosen J, Wolf AP (1991a) Stability of resting deoxyglucose metabolic values in PET studies of schizophrenia. Psychiatry Res 40: 11–20

Bartlett EJ, Wolkin A, Brodie JD, Laska EM, Wolf AP, Sanfilipo M (1991b) Importance of pharmacological control in PET studies: effects of thiothixene and haloperidol on cerebral glucose utilization in chronic schizophrenia. Psychiatry Res 40: 115–124

*Bartlett EJ, Brodie JD, Simkowitz P et al (1998) Effect of a haloperidol challenge on regional brain metabolism in

neuroleptic-responsive and nonresponsive schizophrenic patients. Am J Psychiatry 155: 337–343

*Baxter LR Jr, Phelps ME, Mazziotta JC, Guze BH, Schwartz JM, Selin CE (1987) Local cerebral glucose metabolic rates in obsessive-compulsive disorder. A comparison with rates in unipolar depression and in normal controls. Arch Gen Psychiatry 44: 211–218

Baxter LR Jr, Schwartz JM, Phelps ME et al (1989) Reduction of prefrontal cortex glucose metabolism common to three types of depression. Arch Gen Psychiatry 46: 243–250

Baxter LR Jr, Schwartz JM, Bergman KS et al (1992) Caudate glucose metabolic rate changes with both drug and behavior therapy for obsessive-compulsive disorder. Arch Gen Psychiatry 49: 681–689

Behar D, Rapoport JL, Berg CJ (1984) Computerized tomography and neuropsychological test measured in adolescents with obsessive-compulsive disorder. Am J Psychiatry 141: 363–369

Bench CJ, Price GW, Lammertsma AA et al (1991) Measurement of human cerebral monoamine oxidase type B (MAO-B) activity with positron emission tomography (PET): a dose ranging study with the reversible inhibitor Ro 19–6327. Eur J Clin Pharmacol 40: 169–173

Bench CJ, Friston KJ, Brown RG, Scott LC, Frackowiak RS, Dolan RJ (1992) The anatomy of melancholia – focal abnormalities of cerebral blood flow in major depression. Psychol Med 22: 607–615

Bench CJ, Friston KJ, Brown RG, Frackowiak RS, Dolan RJ (1993) Regional cerebral blood flow in depression measured by positron emission tomography: the relationship with clinical dimensions. Psychol Med 23: 579–590

Bennett BA, Wichems CH, Hollingsworth CK, Davies HM, Thornley C, Sexton T, Childers SR (1995) Novel 2-substituted cocaine analogs: uptake and ligand binding studies at dopamine, serotonin and norepinephrine transport sites in the rat brain. J Pharmacol Exp Ther 272: 1176–1186

*Bertolino A, Nawroz S, Mattay VS et al (1996) Regionally specific pattern of neurochemical pathology in schizophrenia as assessed by multislice proton magnetic resonance spectroscopic imaging. Am J Psychiatry 153: 1554–1563

Bertolino A, Callicott JH, Nawroz S et al (1998) Reproducibility of proton magnetic resonance spectroscopic imaging in patients with schizophrenia. Neuropsychopharmacology 18: 1–9

Biver F, Wikler D, Lotstra F, Damhaut P, Goldman S, Mendlewicz J (1997) Serotonin 5-HT2 receptor imaging in major depression: focal changes in orbito-insular cortex. Br J Psychiatry 171: 444–448

Blin J, Sette G, Fiorelli M, Bletry O, Elghozi JL, Crouzel C, Baron JC (1990) A method for the in vivo investigation of the serotonergic 5-HT2 receptors in the human cerebral cortex using positron emission tomography and 18F-labeled setoperone. J Neurochem 54: 1744–1754

Braver TS, Cohen JD, Nystrom LE, Jonides J, Smith EE, Noll DC (1997) A parametric study of prefrontal cortex involvement in human working memory. Neuroimage 5: 49–62

Breier A, Malhotra AK, Pinals DA, Weisenfeld NI, Pickar D (1997a) Association of ketamine-induced psychosis with focal activation of the prefrontal cortex in healthy volunteers. Am J Psychiatry 154: 805–811

Breier A, Su TP, Saunders R, Carson RE et al (1997b) Schizophrenia is associated with elevated amphetamine-induced synaptic dopamine concentrations: evidence from a novel positron emission tomography method. Proc Natl Acad Sci USA 94: 2569–2574

*Breiter HC, Rauch SL (1996) Functional MRI and the study of OCD: from symptom provocation to cognitive-behavioral probes of cortico-striatal systems and the amygdala. Neuroimage 4: 127–138

Breiter HC, Rauch SL, Kwong KK et al (1996) Functional magnetic resonance imaging of symptom provocation in obsessive-compulsive disorder. Arch Gen Psychiatry 53: 595–606

Brooks DJ (1996) Functional imaging techniques in the diagnosis of non-Alzheimer dementias. J Neural Transm Suppl 47: 155–167

Buchsbaum MS, Wu JC, DeLisi LE, Holcomb HH, Hazlett E, Cooper-Langston K, Kessler R (1987) Positron emission tomography studies of basal ganglia and somatosensory cortex neuroleptic drug effects: differences between normal controls and schizophrenic patients. Biol Psychiatry 22: 479–494

*Buchsbaum MS, Haier RJ, Potkin SG et al (1992a) Frontostriatal disorder of cerebral metabolism in never-medicated schizophrenics. Arch Gen Psychiatry 49: 935–942

Buchsbaum MS, Potkin SG, Marshall JF et al (1992b) Effects of clozapine and thiothixene on glucose metabolic rate in schizophrenia. Neuropsychopharmacology 6: 155–163

Buchsbaum MS, Potkin SG, Siegel BV Jr et al (1992c) Striatal metabolic rate and clinical response to neuroleptics in schizophrenia. Arch Gen Psychiatry 49: 966–974

Buchsbaum MS, Wu J, Siegel BV, Hackett E, Trenary M, Abel L, Reynolds C (1997) Effect of sertraline on regional metabolic rate in patients with affective disorder. Biol Psychiatry 41: 15–22

Buckley PF, Moore C, Long H et al (1994) 1H-magnetic resonance spectroscopy of the left temporal and frontal lobes in schizophrenia: clinical, neurodevelopmental, and cognitive correlates. Biol Psychiatry 36: 792–800

Buckley PF, Friedman L, Wu D et al (1997) Functional magnetic resonance imaging in schizophrenia: initial methodology and evaluation of the motor cortex. Psychiatry Res 74: 13–23

Busatto G, Howard RJ, Ha Y et al (1997a) A functional magnetic resonance imaging study of episodic memory. Neuroreport 8: 2671–2675

Busatto GF, Pilowsky LS, Costa DC, Ell PJ, David AS, Lucey JV, Kerwin RW (1997b) Correlation between reduced in vivo benzodiazepine receptor binding and severity of psychotic symptoms in schizophrenia. Am J Psychiatry 154: 56–63

Buschong SC (1996) Magnetic resonance imaging: physical and biological principles, 2nd edn. Mosby, St Louis

Calabrese G, Colombo C, Bonfanti A, Scotti G, Scarone S (1993) Caudate nucleus abnormalities in obsessive-compulsive disorder measurements of MRI signal intensity. Psychiatry Res 50: 89–92

Callicott JH, Ramsey NF, Tallent K et al (1998) Functional magnetic resonance imaging brain mapping in psychiatry: methodological issues illustrated in a study of working memory in schizophrenia. Neuropsychopharmacology 18: 186–196

Chua SE, McKenna PJ (1995) Schizophrenia – a brain disease? Br J Psychiatry 166: 563–582

Claus JJ, van Harskamp F, Breteler MM et al (1994) The diagnostic value of SPECT with Tc 99m HMPAO in Alzheimer's disease: a population-based study. Neurology 44: 454–461

Claus JJ, Dubois EA, Booij J et al (1997) Demonstration of a reduction in muscarinic receptor binding in early Alzheimer's

disease using iodine-123 dexetimide single-photon emission tomography. Eur J Nucl Med 24: 602–608

*Cohen JD, Forman SD, Braver TS, Casey BJ, Servan-Schreiber D, Noll DC (1994) Activation of prefrontal cortex in a non-spatial working memory task with functional MRI. Hum Brain Mapp 1: 293–304

Cohen BM, Yurgelun-Todd D, English CD, Renshaw PF (1995) Abnormalities of regional distribution of cerebral vasculature in schizophrenia detected by dynamic susceptibility contrast MRI. Am J Psychiatry 152: 1801–1803

Cohen RM, Nordahl TE, Semple WE, Andreason P, Litman RE, Pickar D (1997) The brain metabolic patterns of clozapine- and fluphenazine-treated patients with schizophrenia during a continuous performance task. Arch Gen Psychiatry 54: 481–486

Courtney SM, Ungerleider LG, Keil K, Haxby JV (1997) Transient and sustained activity in a distributed neural system for human working memory. Nature 386: 608–611

Curran SM, Murray CM, Van Beck M et al (1993) A single photon emission computerised tomography study of regional brain function in elderly patients with major depression and with Alzheimer-type dementia. Br J Psychiatry 163: 155–165

*Curtis VA, Bullmore ET, Brammer MJ et al (1998) Attenuated frontal activation during a verbal fluency task in patients with schizophrenia. Am J Psychiatry 155: 1056–1063

Dao-Castellana MH, Paillere-Martinot ML, Hantraye P et al (1997) Presynaptic dopaminergic function in the striatum of schizophrenic patients. Schizophr Res 23: 167–174

Deicken RF, Zhou L, Schuff N, Weiner MW (1997) Proton magnetic resonance spectroscopy of the anterior cingulate region in schizophrenia. Schizophr Res 27: 65–71

DeLisi LE, Sakuma M, Kushner M, Finer DL, Hoff AL, Crow TJ (1997a) Anomalous cerebral asymmetry and language processing in schizophrenia. Schizophr Bull 23: 255–271

DeLisi LE, Sakuma M, Tew W, Kushner M, Hoff AL, Grimson R (1997b) Schizophrenia as a chronic active brain process: a study of progressive brain structural change subsequent to the onset of schizophrenia. Psychiatry Res 74: 129–140

Dewey SL, Smith GS, Logan J et al (1992) GABAergic inhibition of endogenous dopamine release measured in vivo with 11C-raclopride and positron emission tomography. J Neurosci 12: 3773–3780

Dewey SL, Smith GS, Logan J, Brodie JD (1993) Modulation of central cholinergic activity by GABA and serotonin: PET studies with 11C-benztropine in primates. Neuropsychopharmacology 8: 371–376

D'haenen H, Bossuyt A, Mertens J, Bossuyt-Piron C, Gijsemans M, Kaufman L (1992) SPECT imaging of serotonin2 receptors in depression. Psychiatry Res 45: 227–237

*Dolan RJ, Fletcher P, Frith CD, Friston KJ, Frackowiak RS, Grasby PM (1995) Dopaminergic modulation of impaired cognitive activation in the anterior cingulate cortex in schizophrenia. Nature 378: 180–182

Drevets WC, Videen TO, Price JL, Preskorn SH, Carmichael ST, Raichle ME (1992) A functional anatomical study of unipolar depression. J Neurosci 12: 3628–3641

Drevets WC, Price JL, Simpson JR Jr., Todd RD, Reich T, Vannier M, Raichle ME (1997) Subgenual prefrontal cortex abnormalities in mood disorders. Nature 386: 824–827

Dupont RM, Jernigan TL, Butters N, Delis D, Hesselink JR, Heindel W, Gillin JC (1990) Subcortical abnormalities detected in bipolar affective disorder using magnetic resonance imaging. Clinical and neuropsychological significance. Arch Gen Psychiatry 47: 55–59

Erkinjuntti T (1987) Differential diagnosis between Alzheimer's disease and vascular dementia: evaluation of common clinical methods. Acta Neurol Scand 76: 433–442

*Farde L, Ehrin E, Eriksson L et al (1985) Substituted benzamides as ligands for visualization of dopamine receptor binding in the human brain by positron emission tomography. Proc Natl Acad Sci USA 82: 3863–3867

*Farde L, Wiesel F, Hall H, Halldin C, Stone-Elander S, Sedvall G (1987) No D2 receptor increase in PET study of schizophrenia. Arch Gen Psychiatry 44: 672

Farde L, Wiesel FA, Nordström AL, Sedvall G (1989) D1- and D2-dopamine receptor occupancy during treatment with conventional and atypical neuroleptics. Psychopharmacology (Berl) 99[Suppl]: 28–31

Farde L, Nordström AL, Wiesel FA, Pauli S, Halldin C, Sedvall G (1992) PET analysis of central D1 and D2 dopamine receptor occupancy in patients treated with typical neuroleptics and clozapine: relation to extra-pyramidal side effects. Arch Gen Psychiatry 49: 538–544

Farde L, Halldin C, Muller L, Suhara T, Karlsson P, Hall H (1994) PET study of [11C]beta-CIT binding to monoamine transporters in the monkey and human brain. Synapse 16: 93–103

Farde L, Hall H, Pauli S, Halldin C (1995) Variability in D2-dopamine receptor density and affinity: a PET study with [11C]raclopride in man. Synapse 20: 200–208

Faulstich ME (1991) Brain imaging in dementia of the Alzheimer type. Int J Neurosci 57: 39–49

Fischman AJ, Bonab AA, Babich JW et al (1996) Positron emission tomographic analysis of central 5-hydroxytryptamine2 receptor occupancy in healthy volunteers treated with the novel antipsychotic agent, ziprasidone. J Pharmacol Exp Ther 279: 939–947

Fox PT, Raichle ME (1986) Focal physiological uncoupling of cerebral blood flow and oxidative metabolism during somatosensory stimulation in human subjects. Proc Natl Acad Sci USA 83: 1140–1144

Frackowiak RS (1989) PET: studies in dementia. Psychiatry Res 29: 353–355

**Frackowiak RS, Friston KJ, Frith CD, Dolan RJ, Mazziotta JC (1997) Human brain function. Academic, San Diego

Frey KA, Holthoff VA, Koeppe RA, Jewett DM, Kilbourn MR, Kuhl DE (1991) Parametric in vivo imaging of benzodiazepine receptor distribution in human brain. Ann Neurol 30: 663–672

Friston KJ (1992) The dorsolateral prefrontal cortex, schizophrenia and PET. J Neural Transm Suppl 37: 79–93

Friston KJ, Frith CD, Liddle PF, Frackowiak RS (1991) Comparing functional (PET) images: the assessment of significant change. J Cereb Blood Flow Metab 11: 690–699

Friston KJ, Holmes A, Poline JB, Price CJ, Frith CD (1996) Detecting activations in PET and fMRI: levels of inference and power. Neuroimage 4: 223–235

Frith CD, Friston KJ, Herold S et al (1995) Regional brain activity in chronic schizophrenic patients during the performance of a verbal fluency task. Br J Psychiatry 167: 343–349

Frost JA, Springer JA, Binder JR, Hammeke TA, Bellgowan PSF, Rao SM, Cox RW (1997) Sex does not determine functional lateralization of semantic processing: evidence from fMRI. Neuroimage 5: S564

Fujimoto T, Nakano T, Takano T, Takeuchi K, Yamada K, Fukuzako T, Akimoto H (1996) Proton magnetic resonance spectroscopy of basal ganglia in chronic schizophrenia. Biol Psychiatry 40: 14–18

Furey ML, Pietrini P, Haxby JV et al (1997) Cholinergic stimulation alters performance and task-specific regional cerebral blood flow during working memory. Proc Natl Acad Sci USA 94: 6512–6516

George MS, Ketter TA, Gill DS, Haxby JV, Ungerleider LG, Herscovitch P, Post RM (1993) Brain regions involved in recognizing facial emotion or identity: an oxygen-15 PET study. J Neuropsychiatry Clin Neurosci 5: 384–394

Gonzalez RG, Guimaraes AR, Moore GJ, Crawley A, Cupples LA, Growdon JH (1996) Quantitative in vivo 31P magnetic resonance spectroscopy of Alzheimer disease. Alzheimer Dis Assoc Disord 10: 46–52

Grasby PM, Frith CD, Paulesu E, Friston KJ, Frackowiak RS, Dolan RJ (1995) The effect of the muscarinic antagonist scopolamine on regional cerebral blood flow during the performance of a memory task. Exp Brain Res 104: 337–348

Gur RC, Gur RE (1995) Hypofrontality in schizophrenia: RIP. Lancet 345: 1383–1384

Gur RE, Mozley PD, Shtasel DL et al (1994) Clinical subtypes of schizophrenia: differences in brain and CSF volume. Am J Psychiatry 151: 343–350

Herholz K (1995) FDG PET and differential diagnosis of dementia. Alzheimer Dis Assoc Disord 9: 6–16

Heun R, Mazanek M, Atzor KR et al (1997a) Amygdala-hippocampal atrophy and memory performance in dementia of Alzheimer type. Dement Geriatr Cogn Disord 8: 329–336

Heun R, Schlegel S, Graf-Morgenstern M, Tintera J, Gawehn J, Stoeter P (1997b) Proton magnetic resonance spectroscopy in dementia of Alzheimer type. Int J Geriatr Psychiatry 12: 349–358

Hietala J, Syvalahti E, Vuorio K et al (1995) Presynaptic dopamine function in striatum of neuroleptic-naive schizophrenic patients. Lancet 1130–1131

Hinke RM, Hu X, Stillman AE, Kim SG, Merkle H, Salmi R, Ugurbil K (1993) Functional magnetic resonance imaging of Broca's area during internal speech. Neuroreport 4: 675–678

Hoehn-Saric R, Pearlson GD, Harris GJ, Machlin SR, Camargo EE (1991) Effects of fluoxetine on regional cerebral blood flow in obsessive-compulsive patients. Am J Psychiatry 148: 1243–1245

Hoffman WF, Ballard L, Turner EH, Casey DE (1991) Three-year follow-up of older schizophrenics: extrapyramidal syndromes, psychiatric symptoms, and ventricular brain ratio. Biol Psychiatry 30: 913–926

Hokama H, Shenton ME, Nestor PG et al (1995) Caudate, putamen, and globus pallidus volume in schizophrenia: a quantitative MRI study. Psychiatry Res 61: 209–229

Holcomb HH, Cascella NG, Thaker GK, Medoff DR, Dannals RF, Tamminga CA (1996) Functional sites of neuroleptic drug action in the human brain: PET/FDG studies with and without haloperidol. Am J Psychiatry 153: 41–49

Howard RJ, Ellis C, Bullmore ET et al (1996) Functional echoplanar brain imaging correlates of amphetamine administration to normal subjects and subjects with the narcoleptic syndrome. Magn Reson Imaging 14: 1013–1016

Huang S-C, Phelps ME, Hoffman EJ, Sideria K, Selin CJ, Kuhl DE (1980) Noninvasive determination of local cerebral metabolic rate of glucose in man. Am J Physiol 238: E69–E82

Hurwitz TA, Clark C, Murphy E, Klonoff H, Martin WR, Pate BD (1990) Regional cerebral glucose metabolism in major depressive disorder. Can J Psychiatry 35: 684–688

Imperato A, Dazzi L, Serra M, Gessa GL, Biggio G (1994) Differential effects of abecarnil on basal release of acetylcholine and dopamine in the rat brain. Eur J Pharmacol 261: 205–208

Ingvar DH, Franzen G (1974) Abnormalities of cerebral blood flow distribution in patients with chronic schizophrenia. Acta Psychiatr Scand 50: 425–462

Itoh M, Meguro K, Fujiwara T et al (1994) Assessment of dopamine metabolism in brain of patients with dementia by means of 18F-fluorodopa and PET. Ann Nucl Med 8: 245–251

Jones PB, Harvey I, Lewis SW, Toone BK, Van-Os J, Williams M, Murray RM (1994) Cerebral ventricle dimensions as risk factors for schizophrenia and affective psychosis: an epidemiological approach to analysis. Psychol Med 24: 995–1011

Kaschka W, Feistel H, Ebert D (1995) Reduced benzodiazepine receptor binding in panic disorders measured by iomazenil SPECT. J Psychiatr Res 29: 427–434

Kato T, Shioiri T, Inubushi T, Takahashi S (1993a) Brain lithium concentrations measured with lithium-7 magnetic resonance spectroscopy in patients with affective disorders: relationship to erythrocyte and serum concentrations. Biol Psychiatry 33: 147–152

Kato T, Takahashi S, Shioiri T, Inubushi T (1993b) Alterations in brain phosphorous metabolism in bipolar disorder detected by in vivo 31P and 7Li magnetic resonance spectroscopy. J Affect Disord 27: 53–59

Kato T, Inubushi T, Takahashi S (1994a) Relationship of lithium concentrations in the brain measured by lithium-7 magnetic resonance spectroscopy to treatment response in mania. J Clin Psychopharmacol 14: 330–335

Kato T, Shioiri T, Murashita J, Hamakawa H, Inubushi T, Takahashi S (1994b) Phosphorus-31 magnetic resonance spectroscopy and ventricular enlargement in bipolar disorder. Psychiatry Res 55: 41–50

Kato T, Fujii K, Shioiri T, Inubushi T, Takahashi S (1996a) Lithium side effects in relation to brain lithium concentration measured by lithium-7 magnetic resonance spectroscopy. Prog Neuropsychopharmacol Biol Psychiatry 20: 87–97

Kato T, Hamakawa H, Shioiri T, Murashita J, Takahashi Y, Takahashi S, Inubushi T (1996b) Choline-containing compounds detected by proton magnetic resonance spectroscopy in the basal ganglia in bipolar disorder. J Psychiatry Neurosci 21: 248–254

Kellner CH, Jolley RR, Holgate RC, Austin L, Lydiard RB, Laraia M, Ballenger JC (1991) Brain MRI in obsessive-compulsive disorder. Psychiatry Res 36: 45–49

Kesslak JP, Nalcioglu O, Cotman CW (1991) Quantification of magnetic resonance scans for hippocampal and parahippocampal atrophy in Alzheimer's disease. Neurology 41: 51–54

Kornhuber J, Brucke T, Angelberger P, Asenbaum S, Podreka I (1995) SPECT imaging of dopamine receptors with [123I]epidepride: characterization of uptake in the human brain. J Neural Transm 101: 95–103

Krishnan KR, McDonald WM, Escalona PR et al (1992) Magnetic resonance imaging of the caudate nuclei in depression. Preliminary observations. Arch Gen Psychiatry 49: 553–557

Kuikka JT, Pitkanen A, Lepola U et al (1995) Abnormal regional benzodiazepine receptor uptake in the prefrontal cortex in patients with panic disorder. Nucl Med Commun 16: 273–280

Kuikka JT, Akerman KK, Hiltunen J et al (1997) Striatal and extrastriatal imaging of dopamine D2 receptors in the living human brain with [123I]epidepride single-photon emission tomography. Eur J Nucl Med 24: 483–487

Kuwabara H, Cumming P, Reith J, Leger G, Diksic M, Evans AC, Gjedde A (1993) Human striatal L-dopa decarboxylase activity estimated in vivo using 6-[18F]fluoro-dopa and positron emission tomography: error analysis and application to normal subjects. J Cereb Blood Flow Metab 13: 43–56

*Lahti AC, Holcomb HH, Medoff DR, Tamminga CA (1995) Ketamine activates psychosis and alters limbic blood flow in schizophrenia. Neuroreport 6: 869–872

Lang C, Herholz K, Huk W, Feistel H (1990) Diagnostic differentiation of dementia diseases by modern imaging procedures. Fortschr Neurol Psychiatr 58: 380–398

Laruelle M, Abi-Dargham A, Al-Tikriti MS et al (1994) SPECT quantification of [123I]iomazenil binding to benzodiazepine receptors in nonhuman primates. II. Equilibrium analysis of constant infusion experiments and correlation with in vitro parameters. J Cereb Blood Flow Metab 14: 453–465

*Laruelle M, Abi-Dargham A, van Dyck CH et al (1996) Single photon emission computerized tomography imaging of amphetamine-induced dopamine release in drug-free schizophrenic subjects. Proc Natl Acad Sci USA 93: 9235–9240

Lawrie SM, Abukmeil SS, Chiswick A, Egan V, Santosh CG, Best JJ (1997) Qualitative cerebral morphology in schizophrenia: a magnetic resonance imaging study and systematic literature review. Schizophr Res 25: 155–166

Leenders KL, Palmer AJ, Quinn N et al (1986) Brain dopamine metabolism in patients with Parkinson's disease measured with positron emission tomography. J Neurol Neurosurg Psychiatry 49: 853–860

Liddle PF, Friston KJ, Frith CD, Frackowiak RS (1992) Cerebral blood flow and mental processes in schizophrenia. J R Soc Med 85: 224–227

Lundkvist C, Halldin C, Ginovart N et al (1996) [11C]MDL 100907, a radioligland for selective imaging of 5-HT(2A) receptors with positron emission tomography. Life Sci 58: 187–192

Mann JJ, Malone KM, Diehl DJ, Perel J, Cooper TB, Mintun MA (1996) Demonstration in vivo of reduced serotonin responsivity in the brain of untreated depressed patients. Am J Psychiatry 153: 174–182

Marshall I, Wardlaw J, Cannon J, Slattery J, Sellar RJ (1996) Reproducibility of metabolite peak areas in 1H MRS of brain. Magn Reson Imaging 14: 281–292

Martinot JL, Hardy P, Feline A et al (1990) Left prefrontal glucose hypometabolism in the depressed state: a confirmation. Am J Psychiatry 147: 1313–1317

Martinot JL, Paillere-Martinot ML, Loc'h C et al (1991) The estimated density of D2 striatal receptors in schizophrenia. A study with positron emission tomography and 76Br-bromolisuride. Am J Psychiatry 158: 346–350

Masterman DL, Mendez MF, Fairbanks LA, Cummings JL (1997) Sensitivity, specificity, and positive predictive value of technetium 99-HMPAO SPECT in discriminating Alzheimer's disease from other dementias. J Geriatr Psychiatry Neurol 10: 15–21

*Mayberg HS, Brannan SK, Mahurin RK et al (1997) Cingulate function in depression: a potential predictor of treatment response. Neuroreport 8: 1057–1061

Maziere B, Loc'h C, Baron JC, Sgouropoulos P, Duquesnoy N, D'Antona R, Cambon H (1985) In vivo quantitative imaging of dopamine receptors in human brain using positron emission tomography and [76Br]bromospiperone. Eur J Pharmacol 114: 267–272

Mielke R, Kessler J, Szelies B, Herholz K, Wienhard K, Heiss WD (1996) Vascular dementia: perfusional and metabolic disturbances and effects of therapy. J Neural Transm Suppl 47: 183–191

Miller BL, Moats RA, Shonk T, Ernst T, Woolley S, Ross BD (1993) Alzheimer disease: depiction of increased cerebral myo-inositol with proton MR spectroscopy. Radiology 187: 433–437

Mountz JM, Modell JG, Wilson MW, Curtis GC, Lee MA, Schmaltz S, Kuhl DE (1989) Positron emission tomographic evaluation of cerebral blood flow during state anxiety in simple phobia. Arch Gen Psychiatry 46: 501–504

Murphy DG, DeCarli CD, Daly E et al (1993) Volumetric magnetic resonance imaging in men with dementia of the Alzheimer type: correlations with disease severity. Biol Psychiatry 34: 612–621

Nordberg A, Hartvig P, Lilja A et al (1990) Decreased uptake and binding of 11C-nicotine in brain of Alzheimer patients as visualized by positron emission tomography. J Neural Transm Park Dis Dement Sect 2: 215–224

Nordberg A, Lilja A, Lundqvist H et al (1992) Tacrine restores cholinergic nicotinic receptors and glucose metabolism in Alzheimer patients as visualized by positron emission tomography. Neurobiol Aging 13: 747–758

Nordberg A, Lundqvist H, Hartvig P, Andersson J, Johansson M, Hellstrom-Lindahi E, Langstrom B (1997) Imaging of nicotinic and muscarinic receptors in Alzheimer's disease: effect of tacrine treatment. Dement Geriatr Cogn Disord 8: 78–84

Nordström AL, Farde L, Halldin C (1993) High 5-HT2 receptor occupancy in clozapine treated patients demonstrated by PET. Psychopharmacology 110: 365–367

Nordström AL, Farde L, Eriksson L, Halldin C (1995) No elevated D2 dopamine receptors in neuroleptic-naive schizophrenic patients revealed by positron emission tomography and [11C]N-methylspiperone. Psychiatry Res 61: 67–83

Nyberg S, Farde L, Halldin C (1997) A PET study of 5-HT2 and D2 dopamine receptor occupancy induced by olanzapine in healthy subjects. Neuropsychopharmacology 16: 1–7

Okubo Y, Suhara T, Suzuki K et al (1997) Decreased prefrontal dopamine D1 receptors in schizophrenia revealed by PET. Nature 385: 634–636

Page MP, Howard RJ, O'Brien JT, Buxton-Thomas MS, Pickering AD (1996) Use of neural networks in brain SPECT to diagnose Alzheimer's disease. J Nucl Med 37: 195–200

Pardo JV, Pardo PJ, Raichle ME (1993) Neural correlates of self-induced dysphoria. Am J Psychiatry 150: 713–719

Pearlson GD, Kim WS, Kubos KL et al (1989) Ventricle-brain ratio, computed tomographic density, and brain area in 50 schizophrenics. Arch Gen Psychiatry 46: 690–697

*Pearlson GD, Harris GJ, Powers RE et al (1992) Quantitative changes in mesial temporal volume, regional cerebral blood flow, and cognition in Alzheimer's disease. Arch Gen Psychiatry 49: 402–408

Parnetti L, Tarducci R, Presciutti O et al (1997) Proton magnetic resonance spectroscopy can differentiate Alzheimer's disease from normal aging. Mech Ageing Dev 97: 9–14

Perani D, Colombo C, Bressi S et al (1995) [18F]FDG PET study in obsessive-compulsive disorder. A clinical/metabolic correlation study after treatment. Br J Psychiatry 166: 244–250

Pettegrew JW, Moossy J, Withers G, McKeag D, Panchalingam K (1988) 31P nuclear magnetic resonance study of the brain in Alzheimer's disease. J Neuropathol Exp Neurol 47: 235–248

Pettegrew JW, Keshavan MS, Panchalingam K, Strychor S, Kaplan DB, Tretta MG, Allen M (1991) Alterations in brain high-energy phosphate and membrane phospholipid metabolism in first-episode, drug-naive schizophrenics. A pilot study of the dorsal prefrontal cortex by in vivo phosphorus 31 nuclear magnetic resonance spectroscopy. Arch Gen Psychiatry 48: 563–568

Pettegrew JW, Panchalingam K, Klunk WE, McClure RJ, Muenz LR (1994) Alterations of cerebral metabolism in probable Alzheimer's disease: a preliminary study. Neurobiol Aging 15: 117–132

Petty RG, Barta PE, Pearlson GD et al (1995) Reversal of asymmetry of the planum temporale in schizophrenia. Am J Psychiatry 152: 715–721

Pike VW, McCarron JA, Lammertsma AA et al (1996) Exquisite delineation of 5-HT1A receptors in human brain with PET and [carbonyl-11C]WAY-100635. Eur J Pharmacol 301: R5–R7

Rauch SL, Savage CR, Alpert NM et al (1995) A positron emission tomographic study of simple phobic symptom provocation. Arch Gen Psychiatry 52: 20–28

Rauch SL, van der Kolk BA, Fisler RE et al (1996) A symptom provocation study of posttraumatic stress disorder using positron emission tomography and script-driven imagery. Arch Gen Psychiatry 53: 380–387

Renshaw PF, Yurgelun-Todd DA, Cohen BM (1994) Greater hemodynamic response to photic stimulation in schizophrenic patients: an echo planar MRI study. Am J Psychiatry 151: 1493–1495

Robertson JS, Marr RB, Rosenblum M et al (1973) 32-Crystal positron transverse section detector. In: Freedman GS (ed) Tomographic imaging in nuclear medicine. Springer, Berlin Heidelberg New York

**Roland PE (1993) Brain activation. Wiley and Liss, New York

Rosen BR, Buckner RL, Dale AM (1998) Event-related functional MRI: past, present, and future. Proc Natl Acad Sci USA 95: 773–780

Sabri O, Erkwoh R, Schreckenberger M et al (1997) Regional cerebral blood flow and negative/positive symptoms in 24 drug-naive schizophrenics. J Nucl Med 38: 181–188

Sadzot B, Lemaire C, Maquet P et al (1995) Serotonin 5HT2 receptor imaging in the human brain using positron emission tomography and a new radioligand, [18F]altanserin: results in young normal controls. J Cereb Blood Flow Metab 15: 787–797

Schlaepfer TE, Harris GJ, Tien AY et al (1994) Decreased regional cortical gray matter volume in schizophrenia. Am J Psychiatry 151: 842–848

Schlegel S, Aldenhoff JB, Eissner D, Lindner P, Nickel O (1989a) Regional cerebral blood flow in depression: associations with psychopathology. J Affect Disord 17: 211–218

Schlegel S, Maier W, Philipp M, Aldenhoff JB, Heuser I, Kretzschmar K, Benkert O (1989b) Computed tomography in depression: association between ventricular size and psychopathology. Psychiatry Res 29: 221–230

Schlegel S, Steinert H, Bockisch A, Hahn K, Schloesser R, Benkert O (1994) Decreased benzodiazepine receptor binding in panic disorder measured by IOMAZENIL-SPECT. A preliminary report. Eur Arch Psychiatry Clin Neurosci 244: 49–51

Schlösser R, Schlegel S (1995) D2-receptor imaging with [123I]IBZM and single photon emission tomography in psychiatry: a survey of current status. J Neural Transm 99: 173–185

*Schlösser R, Simkowitz P, Bartlett EJ, Wolkin A, Smith GS, Dewey SL, Brodie JD (1996) The study of neurotransmitter interactions using positron emission tomography and functional coupling. Clin Neuropharmacol 19: 371–389

*Schlösser R, Schlegel S, Hiemke C, Nickel O, Bockisch A, Rao ML, Hahn K (1997) [123I]IBZM SPECT in patients treated with typical and atypical neuroleptics: relationship to drug plasma levels and extrapyramidal side effects. Psychiatry Res 75: 103–114

Schlösser R, Brodie JD, Dewey SL et al (1998a) Long-term stability of neurotransmitter activity investigated with 11C-raclopride PET. Synapse 28: 66–70

*Schlösser R, Hutchinson M, Joseffer S et al (1998b) Functional magnetic resonance imaging of human brain activity in a verbal fluency task. J Neurol Neurosurg Psychiatry 64: 492–498

Schneider F, Grodd W, Weiss U, Klose U, Mayer KR, Nagele T, Gur RC (1997) Functional MRI reveals left amygdala activation during emotion. Psychiatry Res 76: 75–82

Schröder J, Wenz F, Schad LR, Baudenstiel K, Knopp MV (1995) Sensorimotor cortex and supplementary motor area changes in schizophrenia: a study with functional magnetic resonance imaging. Br J Psychiatry 167: 197–201

Seeman P, Guan HC, Niznik HB (1989) Endogenous dopamine lowers the dopamine D2 receptor density as measured by [3H]raclopride: implications for positron emission tomography of the human brain. Synapse 3: 96–97

Seeman P, Guan HC, Van Tol HH (1993) Dopamine D4 receptors elevated in schizophrenia. Nature 365: 441–445

Segebarth C, Belle V, Delon C et al (1994) Functional MRI of the human brain: predominance of signals from extracerebral veins. Neuroreport 5: 813–816

Shaywitz BA, Shaywitz SE, Pugh KR et al (1995) Sex differences in the functional organization of the brain for language. Nature 373: 607–609

Seidman LJ, Yurgelun Todd D, Kremen WS, Woods BT, Goldstein JM, Faraone SV, Tsuang MT (1994) Relationship of prefrontal and temporal lobe MRI measures to neuropsychological performance in chronic schizophrenia. Biol Psychiatry 35: 235–246

Shioiri T, Kato T, Inubushi T, Murashita J, Takahashi S (1994) Correlations of phosphomonoesters measured by phosphorus-31 magnetic resonance spectroscopy in the frontal lobes and negative symptoms in schizophrenia. Psychiatry Res 55: 223–235

Shioiri T, Hamakawa H, Kato T, Murashita J, Fujii K, Inubushi T, Takahashi S (1996a) Proton magnetic resonance spectroscopy of the basal ganglia in patients with schizophrenia: a preliminary report. Schizophr Res 22: 19–26

Shioiri T, Kato T, Murashita J, Hamakawa H, Inubushi T, Takahashi S (1996b) High-energy phosphate metabolism in the frontal lobes of patients with panic disorder detected by phase-encoded 31P-MRS. Biol Psychiatry 40: 785–793

Silbersweig DA, Stern E, Frith C et al (1995) A functional neuroanatomy of hallucinations in schizophrenia. Nature 378: 176–179

Smith GS, de Leon MJ, George AE et al (1992) Topography of cross-sectional and longitudinal glucose metabolic deficits in

Alzheimer's disease. Pathophysiologic implications. Arch Neurol 49: 1142–1150

Smith GS, Dewey SL, Brodie JD et al (1997) Serotonergic modulation of dopamine measured with [11C]raclopride and PET in normal human subjects. Am J Psychiatry 154: 490–496

*Smith GS, Schlösser R, Brodie JD et al (1998) Glutamate modulation of dopamine measured in vivo with positron emission tomography (PET) and 11C-raclopride in normal human subjects. Neuropsychopharmacology 18: 18–25

Sokoloff L, Reivich M, Kennedy C et al (1977) The [14C]deoxyglucose method for the measurement of local cerebral glucose utilization: theory, procedure, and normal values in the conscious and anesthetized albino rat. J Neurochem 28: 897–916

Stern CE, Corkin S, Gonzalez RG et al (1996) The hippocampal formation participates in novel picture encoding: evidence from functional magnetic resonance imaging. Proc Natl Acad Sci USA 93: 8660–8665

Stewart RS, Devous MD Jr., Rush AJ, Lane L, Bonte FJ (1988) Cerebral blood flow changes during sodium-lactate-induced panic attacks. Am J Psychiatry 145: 442–449

Strauss WL, Layton ME, Hayes CE, Dager SR (1997) 19F magnetic resonance spectroscopy investigation in vivo of acute and steady-state brain fluvoxamine levels in obsessive-compulsive disorder. Am J Psychiatry 154: 516–522

Sunderland T, Esposito G, Molchan SE et al (1995) Differential cholinergic regulation in Alzheimer's patients compared to controls following chronic blockade with scopolamine: a SPECT study. Psychopharmacology (Berl) 121: 231–241

Tedeschi E, Hasselbalch SG, Waldemar G et al (1995) Heterogeneous cerebral glucose metabolism in normal pressure hydrocephalus. J Neurol Neurosurg Psychiatry 59: 608–615

Tsai G, Coyle JT (1995) *N*-acetylaspartate in neuropsychiatric disorders. Prog Neurobiol 46: 531–540

Tune LE, Wong DF, Pearlson G et al (1993) Dopamine D2 receptor density estimates in schizophrenia: a positron emission tomography study with 11C-N-methylspiperone. Psychiatry Res 49: 219–237

Van Horn JD, McManus IC (1992) Ventricular enlargement in schizophrenia. A meta-analysis of studies of the ventricle:brain ratio (VBR). Br J Psychiatry 160: 687–697

Videbech P (1997) MRI findings in patients with affective disorder: a meta analysis. Acta Psychiatr Scand 96: 157–168

Volkow ND, Brodie JD, Wolf AP, Angrist B, Russell J, Cancro R (1986) Brain metabolism in patients with schizophrenia before and after acute neuroleptic administration. J Neurol Neurosurg Psychiatry 49: 1199–1202

Volkow ND, Wolf AP, Van Gelder P, Brodie JD, Overall JE, Cancro R, Gomez-Mont F (1987) Phenomenological correlates of metabolic activity in 18 patients with chronic schizophrenia. Am J Psychiatry 144: 151–158

Volkow ND, Fowler JS, Wolf AP et al (1992) Distribution and kinetics of carbon-11-cocaine in the human body measured with PET. J Nucl Med 33: 521–525

Volkow ND, Fowler JS, Wang GJ et al (1993) Reproducibility of repeated measures of carbon-11-raclopride binding in the human brain. J Nucl Med 34: 609–613

Volkow ND, Wang GJ, Fowler JS et al (1994) Imaging endogenous dopamine competition with [11C]raclopride in the human brain. Synapse 16: 255–262

Volkow ND, Ding YS, Fowler JS et al (1995) A new PET ligand for the dopamine transporter: studies in the human brain. J Nucl Med 36: 2162–2168

*Vollenweider FX, Leenders KL, Scharfetter C, Maguire P, Stadelmann O, Angst J (1997) Positron emission tomography and fluorodeoxyglucose studies of metabolic hyperfrontality and psychopathology in the psilocybin model of psychosis. Neuropsychopharmacology 16: 357–372

Volz HP, Gaser C, Hager F et al (1997a) Brain activation during cognitive stimulation with the Wisconsin Card Sorting Test – a functional MRI study on healthy volunteers and schizophrenics. Psychiatry Res 75: 145–157

Volz HP, Rzanny R, May S et al (1997b) 31P magnetic resonance spectroscopy in the dorsolateral prefrontal cortex of schizophrenics with a volume selective technique – preliminary findings. Biol Psychiatry 41: 644–648

Wagner HN Jr, Burns HD, Dannals RF et al (1983) Imaging dopamine receptors in the human brain by positron tomography. Science 221: 1264–1266

Wang GJ, Volkow ND, Fowler JS et al (1996) Age associated decrements in dopamine D2 receptors in thalamus and in temporal insula of human subjects. Life Sci 59: PL31–PL35

Weinberger DR, Berman KF, Zec RF (1986) Physiologic dysfunction of dorsolateral prefrontal cortex in schizophrenia. I. Regional cerebral blood flow evidence. Arch Gen Psychiatry 43: 114–124

Williamson P (1987) Hypofrontality in schizophrenia: a review of the evidence. Can J Psychiatry 32: 399–404

Wong DF, Gjedde A, Wagner HN Jr (1986a) Quantification of neuroreceptors in the living human brain. I. Irreversible binding of ligands. J Cereb Blood Flow Metab 6: 137–146

Wong DF, Gjedde A, Wagner HN Jr, Dannals RF, Douglass KH, Links JM, Kuhar MJ (1986b) Quantification of neuroreceptors in the living human brain. II. Inhibition studies of receptor density and affinity. J Cereb Blood Flow Metab 6: 147–153

*Wong DF, Wagner HN Jr, Tune LE et al (1986c) Positron emission tomography reveals elevated D2 dopamine receptors in drug-naive schizophrenics. Science 234: 1558–1563

Woodruff PW, Pearlson GD, Geer MJ, Barta PE, Chilcoat HD (1993) A computerized magnetic resonance imaging study of corpus callosum morphology in schizophrenia. Psychol Med 23: 45–56

Woods R, Mazziotta J, Cherry S (1992) Automated algorithm for aligning tomographic images. II. Crossmodality MRI-PET images. J Comput Assist Tomogr 16: 620–633

Wu JC, Buchsbaum MS, Hershey TG, Hazlett E, Sicotte N, Johnson JC (1991) PET in generalized anxiety disorder. Biol Psychiatry 29: 1181–1199

Young AH, Blackwood DH, Roxborough H, McQueen JK, Martin MJ, Kean D (1991) A magnetic resonance imaging study of schizophrenia: brain structure and clinical symptoms. Br J Psychiatry 158: 158–164

Young LT, Wong DF, Goldman S et al (1991) Effects of endogenous dopamine on kinetics of [3H]N-methylspiperone and [3H]raclopride binding in the rat brain. Synapse 9: 188–194

Yurgelun-Todd DA, Waternaux CM, Cohen BM, Gruber SA, English CD, Renshaw PF (1996) Functional magnetic resonance imaging of schizophrenic patients and comparison subjects during word production. Am J Psychiatry 153: 200–205

CHAPTER 12

U. Baumann, M. Perrez

Psychology and Its Relevance to Psychiatry

Translator: E. Taub

1 Introduction

1.1 Psychology as a Science

There is no single definition of psychology, but the definitions given by various authors tend to make use of similar terms. Gray (1994, p. 3), for example, defines psychology as "the science of behavior and the mind." "Behavior" here refers to observable actions, while "mind" subsumes perception, memory, motivation, emotion, and so forth. Among German-speakers, psychology is often defined as the science of "Erleben" (experience) and "Verhalten" (behavior), which accords well with the American definitions. The science of psychology thus comprises a large number of subfields, of which the study of the mentally ill patient is but one.

There is a consensus among scientific psychologists that psychology is an *empirical* science (Baumann 1995; Perrez 1991), i.e. a discipline that draws its conclusions from experience, in which experimentation is particularly important and has been so from as early as the nineteenth century (for the history of psychology, see Evans et al. 1992; Ash and Geuter 1985). Psychology is considered to be partly a basic science and partly an applied science (Herrmann 1976).

The discoveries of those branches of psychology that are primarily basic science, such as general psychology, allow a better understanding of the mechanisms and principles that regulate human behavior and that underlie changes in behavior in both the normal and the pathologic situation. Such knowledge is necessary for the understanding of normal and pathologic processes (etiology and pathogenesis). The technical knowledge obtained from the applied-science subfields of psychology, such as clinical psychology, provides a rational basis for the application of methods of treatment by establishing the results that various methods are likely to achieve under various conditions (Perrez 1991, 1998b). Such knowledge is indispensable for therapy, and especially for prevention; it is fundamental to methods of both assessment and psychotherapy.

Only a small subset of the actual and potential contributions of psychology to psychiatry can be discussed in this chapter. Particular subfields will be dealt with separately in other parts of this text (for example, see Vol. 1, Part 1, Chaps. 13, 14). Individual findings will not be discussed in detail here, as this chapter aims to concentrate on more fundamental aspects.

1.2 The Science of Psychology and Its Relation to Other Fields

The scientific study of humans can be separated into basic categories, each of which concerns data collected on a different level (Seidenstücker and Baumann 1987): the biologic/somatic, the psychic/psychological, the social, and the ecological level. The use of expressions such as social psychology, psychophysiology, and psychosomatics shows that psychology attempts to link the psychic data level with other data levels. According to the German Scientific Council (Wissenschaftsrat 1983), psychology occupies a position between the social sciences and humanities, the natural and biological sciences, and medicine and is thus intrinsically multimodal and multidisciplinary. The analysis of the relation between the psychic data level and the biologic/somatic data level is relevant to the mind–body problem, for which philosophers have made various attempts at an ontological or epistemological solution.

In accordance with the complementarity principle enunciated by Fahrenberg (1981), we assume that the somatic and psychic levels of description in human science complement one another, even though each level possesses its own category systems, theoretical foundations, and basic methodologies (for the emergence concept, see Bunge 1984). We thus reject ontological reductionism and recognize that the discipline of psychology plays an important, independent role in the analysis of behavior, by addressing the entire range of topics lying between the social and biological levels.

1.3 Clinical Psychology and Psychiatry

Among all of the subspecialities of psychology, clinical psychology bears an especially close relation to psychiatry (for an overview, see Baumann and Perrez 1998a). Baumann and Perrez (1998b, p. 4) provide the following definition of clinical psychology: "Clinical psychology is the subfield of psychology that deals with psychic disorders and the psychic aspects of somatic disorders." The topics addressed by clinical psychology include fundamental issues (scientific/theoretical, ethical, and conceptual questions), epidemiology, classification, diagnosis, etiology, and analysis of perpetuating factors as well as interventions such as prevention, psychotherapy, rehabilitation, health care, and evaluation (Baumann and Perrez 1998a; for the history of clinical psychology, see Routh 1994; Walker 1991).

From the foregoing list it is clear that clinical psychology overlaps substantially with psychiatry. From the scientific point of view, both clinical psychology and psychiatry deal with psychic disorders; in other words, they differ little on a fundamental level, even though they have different emphases with regard to the area of study; psychiatry tends to stress the somatic level of description, and clinical psychology the psychic level. There are, nonetheless, significant and legally mandated differences between psychologists and psychiatrists in terms of their professional training and clinical practice.

Other border areas to psychiatry, besides clinical psychology, include health psychology, behavioral medicine, and medical psychology. Health psychology (see Schwarzer 1997) is understood by its practitioners as being oriented toward the prevention of somatic disorders and the preservation of health. It is also partly concerned with questions of treatment and rehabilitation. Behavioral medicine also overlaps with psychiatry (Blanchard 1992), inasmuch as it constitutes an interdisciplinary approach to the study and treatment of (primarily) somatic disorders. It involves psychologists, physicians, and scientists from other disciplines. The term "medical psychology" refers to psychology as a subject of instruction for physicians. Research in medical psychology mainly concerns physician–patient interaction and the situation of the patient (Schwenkmezger and Schmidt 1994).

2 Fundamental Methods

2.1 General Observations

Alongside the subject-oriented subspecialities discussed above, a further central subspeciality of basic psychology is that of methodology in psychological research. This subspeciality deals with questions such as the construction of theories, the planning of experiments, and the statistical assessment of results (Breakwell et al. 1995). The basic principles of psychological methodology require further specification in two respects when applied to particular situations in clinical psychology and psychiatry (Baumann 1991):

1. The particular concerns of clinical psychology and psychiatry require a content-based adaptation of methodological postulates to the specific problem at hand. For example, general methodology requires that all experiments have a control group; the nature of the control group in a particular experiment (e.g. placebo control group) must be decided on the basis of a field-specific methodology (see, e.g. Kazdin 1994; Sher and Trull 1996).
2. Particular research methods may be of greater or lesser value at different stages in the course of the research effort. For example, single case studies are useful in phase I medication trials, but of no value in phase III, in which comprehensive, multicenter studies are required.

The methodology of research in psychology has also given a strong impulse to psychiatry, making contributions not only to the planning of experimental research, but also to assessment (see Sect. 5) and to the evaluation of interventions and concepts of care (see Sect. 6.4). As an illustration of the discussion of method in psychology, which is also of importance to psychiatry, the following sections will deal with two general framework concepts: field versus laboratory research, and multimodality. There is not enough space to discuss further topics in detail.

2.2 Field Versus Laboratory Research

In psychological methodology, a distinction is often drawn between laboratory and field research (Breakwell et al. 1995). Laboratory research is prototypically viewed as an artificially constructed approach and equated with the performance of hypothesis-testing experiments, while field research is viewed as a correlative approach, with an emphasis on naturally occurring phenomena. Recent methodology (Patry 1982) views this dichotomy, and the associated criticisms of particular methods, as superfluous and untenable. According to Paltry (1982), research methods may be characterized as "natural" or "artificial" with respect to each of at least three criteria, namely, the setting of the investigation, the treatment given (independent variable), and the behavior investigated (dependent variable).

Artificially constructed methods have dominated psychological research up to the present, both in the laboratory and in the field. Because the artificiality of a method bears on the generalizability of the research findings, the naturalness or artificiality of the method used in any research study must be taken into account in the interpretation of its conclusions. In our opinion, too much research in clinical psychology and psychiatry falls into the category of artificial, with respect not only to setting, but also to the dependent variables used (often based on questionnaires or interviews). The conclusions of such studies are even less reliable because, frequently, only retrospective data are

obtained. A narrow focusing of research methods on artificial conditions appears problematic for a science that considers itself to be empirical; it is imperative that natural conditions be used as well (see Sect. 5.3). Studies performed under natural conditions are complementary, rather than alternative, to studies with artificial parameters.

2.3 Multimodality as a Fundamental Principle

The principle of multimodality is a central tenet of research in clinical psychology and in psychiatry (Seidenstücker and Baumann 1987). Multimodality means that a multi- rather than univariate approach is used, in which variation is possible in each of several individual categories. The following categories may be distinguished:

- *Data level* (basic categories of observation): Biologic/somatic, psychic/psychological, social, ecological (see above).
- *Data source* (informant): The proband him- or herself, other people (e.g. people bearing some relationship to the proband, trained judges, therapists, institutionally collected data), data from instruments or apparatus. When self-assessment and external assessment yield different results, the disagreement is not merely due to inaccuracies of method. Rather, self-assessment and external assessment are fundamentally different means of obtaining access to the individual, and they are informative to different extents in different areas. External assessment should not be considered more objective than self-assessment; the two variants are of equal theoretical value.
- *Data-gathering techniques* (see Sect. 5.2).
- *Constructs*: There is no generally accepted taxonomy of constructs in psychology; in each individual case, a reasonable choice of construct is made on the basis of the current status of research. Complex global constructs incorporating data from several different areas are specific to the clinical field (see Laireiter et al. 1994). These global constructs (social adaptation, social integration) primarily represent an attempt to characterize patients on the social level, but also frequently include indicators on the biological, psychic, and ecological levels (e.g. state of health, coping, and quality of residence, respectively) and are thus multimodal constructs. The construct of quality of life is similar to these global constructs, and to social adaptation in particular (Bullinger 1996; see also Vol. 1, Part 2, Chap. 7).

Multimodality is a general framework concept and does not provide specific details of the proper selection of data-gathering techniques in concrete, individual cases. This may lead to difficulties in interpretation (Seidenstücker and Baumann 1987); several simultaneously investigated data modalities will not necessarily yield matching results at a single point in time or in their overall course. There is no definitive solution to this problem, but difficulties in interpretation can be alleviated somewhat if the selection of data modalities is directed by particular hypotheses and theories. Furthermore, method studies such as "multitrait multimethod analysis" are necessary to elucidate complex relationships empirically.

3 Fundamental Concepts: Personality

3.1 Objectives of the Study of Personality

A further subspeciality of psychology that is relevant to psychiatric research is the study of personality. In this subspeciality, psychology provides theories for the general understanding of different types of people. These theories are important in psychiatry because an understanding of normal experience and behavior is a necessary prerequisite for the assessment of psychic disorders (Baumann 1993; Becker 1996; Butcher and Rouse 1996). The concepts of the psychology of personality are particularly important for the understanding not only of personality disorders, but also of the premorbid personality (Watson and Clark 1994; von Zerssen 1996).

The dimensions of person, personality characteristics, and situation (or time) are important sources of variation for the study of personality. As far as the dimensions of person and characteristics are concerned, two viewpoints can be distinguished:

- The analysis of the covariation of characteristics gives rise to traits or, in the clinical sector, to syndromes.
- The analysis of the covariation of persons, on the other hand, gives rise to typologies or, in the clinical sector, to diagnoses.

While the trait approach has supplanted the typological approach in the psychology of personality, we find that the typological approach dominates in psychiatry, because of the manner in which diagnoses are made; the trait, or syndrome, approach has a long tradition in psychiatry, but has receded into the background in the last few years because of the dominance of the ICD and DSM diagnostic systems.

3.2 Different Approaches to Personality

Concepts of personality may be classified into groups of related theories. Some of these concepts that are or were applied in psychiatry, either implicitly or explicitly, will be discussed here (Asendorpf 1996; Amelang and Bartussek 1997; Wiggins and Pincus 1992).

The psychodynamic approaches stress intrapsychic aspects, particularly motivation. The psychoanalytic concepts of personality (compare Freud), which are fundamental to psychoanalytic theories of etiology and treatment, are prototypical of psychodynamic approaches. Psychodynamic approaches often originate in the clinical setting and then are generalized to apply to the field of normal experience and behavior. From the viewpoint of empirical personality research, psychoanalytic personality concepts, in particular, have not received adequate empirical confirmation. The concept of motivation, however, has proved to be very fruitful, as has that of defense mechanisms. The latter has been developed further in coping research (Lazarus 1991; Perrez and Reicherts 1992) and in new concepts of defense strategies, such as vigilance and cognitive avoidance (Krohne 1996).

The trait approach has become extremely widespread in personality psychology since the 1930s (e.g. Guilford, Eysenck, and Cattell; see Amelang and Bartussek 1997). In this approach, behavior is said to be determined by persistent features of personality ("traits") that are largely independent of situational factors. Since the 1980s, integrative efforts have increasingly been made with regard to the temperamental factors, culminating in the concept of the "big five factors" (Wiggins and Pincus 1992). Extraversion, neuroticism, agreeableness, conscientiousness, and openness were postulated as overriding entities. Integrative approaches may also be found in intelligence research, where a few basic dimensions have been postulated (Carroll 1993). Integrative efforts are to be applauded, but they do not make up for the drawbacks of the trait approach, particularly its neglect of situational factors (see Butcher and Rouse 1996). Trait concepts also occur in the other kinds of theories discussed here, e.g. psychodynamic theories (where "orality" and other traits play a role) and cognitive approaches (compare "locus of control," control beliefs). Traits such as "hardness," social supportiveness, and hostility are also discussed among the factors that may protect against or predispose to psychic and somatic illnesses.

Behavioral approaches stress observable behavior, which is conditioned by the environment or by situational factors (see, e.g. Thorndike, Dollard, and Miller; also see Amelang and Bartussek 1997). The laws of the psychology of learning, as they apply to conditioning, operant learning, and model learning, are central to such approaches. These approaches have given a major impetus not only to interventional treatment (behavioral therapy), but also to the understanding of the origins of behavioral disturbances (Perrez and Zbinden 1996; Westmeyer 1995). Behavior-theory personality constructs serve as intermediate variables between environmental factors and learning processes and include, among others, conditionability (or introversion, according to Eysenck; see Amelang and Bartussek 1997) as a temporally stable personality trait, which, when it assumes extreme values, may result in either hyper- or hypoadaptability (dysthymic disorders or sociopathy). A further trait is the differential capacity for "delay of reinforcement," which is relevant to operant conditioning (Mischel 1983) and has been studied particularly with regard to the etiology of delinquent behavior. A further development of classical behavioral approaches may be seen in the cognitive concepts to be discussed below, which extensively incorporate concepts of learning theory as well as other concepts.

Internal processes that manifest themselves in planning, goal-setting, personal constructs, values (e.g. reinforcer preferences), self-regulating systems, and so on are important components of cognitive and information-processing approaches (see, e.g. Kelly, Rotter, Mischel, and Kanfer; also see Amelang and Bartussek 1997). These approaches are a major extension of the behavioral approach, by which behavioral therapy in the narrow sense is further developed into cognitive behavioral therapy. In information-processing approaches, cognitive theories are made more precise and given added depth.

According to humanistic and existential approaches, the most important avenue to the understanding of a person is his or her individuality, as stressed by Frankl, Maslow, and Rogers, among others; the central concepts here are those of the self and of self-regulating abilities (Buss 1995). This conceptual approach has found major resonance, particularly in humanistic psychotherapy. Its development has been, in part, in a direction away from empirical research; phenomenological, hermeneutic conceptual categories are preferred.

The advantages of several different approaches may be summed up under the heading "interactionism" (e.g. Magnusson 1990). Actual experience and behavior are the product of a dynamic interaction between internal processes and external factors. A given person does not enter a given situation as a tabula rasa, but rather brings to it his or her own stable framework conditions (traits). In the theoretical formulation of complex concepts of this type, the individual operationalizations have mostly been realized only for partial

aspects. Interactionalist approaches require a complex consideration of personality, situation, and environment.

All biopsychological models to date have been formulated primarily on the psychic data level. Despite numerous studies with interesting findings, there is still no convincing biopsychological concept of personality (see Zuckermann 1991; Fahrenberg 1996). Such concepts are indispensable not only for personality research, but also for the collaborative interaction of psychology and psychiatry. In addition, models of personality that delineate the range of normal states and the clinical spectrum with equal precision are necessary for research in the clinical sector. Most models of personality are formulated as "normal psychology" and can be applied to psychic disorders only as extreme variants, without differentiating adequately among the distinct disorders. Eysenck's concept (see Amelang and Bartussek 1997) is an exception; it postulates the dimensions of extraversion, neuroticism, and psychoticism as constructs allowing a consideration of both normality and various clinical disorders. The construct of psychoticism is, however, not as well elaborated as neuroticism, and the distinction between neuroticism and psychoticism has become problematic, as neurosis has been eliminated from both the DSM-IV and the ICD-10 diagnostic systems.

4 Etiology of Psychic Disorders and Factors that Perpetuate Them

4.1 Multiple-Phase Model

With respect to the generation of psychic disorders, clinical psychology distinguishes between their etiology, or initial cause, and the factors that perpetuate them. Several distinct phases in the development of psychic disorders have been described (Shepherd 1987; Baumann and Perrez 1998c), in each of which specific factors, including psychological factors, are at work:

1. *Pre- and perinatal phase*: Genetic, intrauterine, and perinatal influences are included here. The sum of all factors occurring during this phase is often subsumed under the term "predisposition."
2. *Socialization or developmental phase*: This phase includes those factors that come into play over the course of the individual's life. Great importance has often been attributed to early childhood influences, in line with psychoanalytic theory. The current understanding of development as a lifelong process implies, however, that later factors are also significant. The influence of socialization moderates the individual's predisposition and determines his or her vulnerability to psychic disorders in general or to specific psychic disorders.
3. *Phase immediately prior to the onset of the disorder*: It is assumed that the onset of a number of disorders may be brought about by critical life events, traumata, etc., among other factors (compare stress research).
4. *Phase after the onset of the disorder*: This fourth phase corresponds roughly to the medical concept of pathogenesis. In clinical psychology, the current intrapsychic and social conditions that perpetuate the disorder are considered to be the most important factors at work in this phase.

While an understanding of the first three phases is primarily important for the prevention of psychic disorders, an understanding of the fourth phase is particularly relevant to their treatment. An understanding of the perpetuating conditions includes disorder-specific mechanisms, which may be intrapsychic, social, or both.

As an example, we may consider depressive disorders, whose onset may be triggered by an experience of loss. We may assume that an individual may be more vulnerable to a depressive reaction because of predisposing factors and socialization factors. This can only be a working hypothesis, as these factors cannot be reconstructed exactly in the concrete case, and the etiology in this sense must remain unknown. A more secure understanding of the triggering factors can be gained, however, by analysis of the experience of loss and the corresponding functional mechanisms of the depressive reaction. Thus certain modalities of control belief (e.g. belief of external control) and certain forms of causal attribution (attribution of failure to self) may be important elements in the depressive functional mode. Even if these elements are not the etiology of depression, they are nevertheless conditioning factors whose modification is of therapeutic importance.

4.2 Models of Causation

The models of causation underlying clinical psychology are no different, in principle, from those of medicine. A multicausal, deterministic or probabilistic cause-and-effect structure is assumed, in which biological, psychological, social, and ecological factors (see Sect. 2.3) may influence psychosomatic processes. The historical/genetic type of explanation, which reconstructs a disorder as the result of a developmental

process, plays a central role in etiology. The stages of this development represent a series of prior dispositions on the one hand and newly occurring events on the other.

With regard to psychophysical aspects, it is possible in some groups of disorders that biological processes bring about psychic conditions, just as psychic conditions and processes may affect neurobiological processes. Causal chains often cannot be exactly defined, whether they span the somatic and psychic data levels or remain within the psychic data level itself. Psychic factors may be the cause, precondition, prodrome, symptom, accompaniment, or result of a disorder (consider, for example, the interpretation of depressive disorder by means of the concept of learned helplessness; Baumann and Perrez 1998a). Experiments with individuals with psychic disorders often fail to reveal the causal chain; longitudinal studies are necessary. Psychology, in collaboration with psychiatry, must develop precise experimental protocols that can optimally investigate the causal mechanisms at work.

4.3 Socialization Theories

Psychoanalysis was long considered the prototype of a socialization theory. The psychoanalytic theories of etiology were a major impulse to the field, but, because of their exclusive nature, failed to achieve empirical confirmation (see Ernst and von Luckner 1985). The psychoanalytic conceptions of bonding theory, which constitute an interactionistic developmental psychology of early childhood (mother–child interaction), have proved to be of greater significance (see also Vol. 1, Part 1, Chap. 22). The study of bonding initiated by Bowlby has yielded fruitful hypotheses about the consequences of inadequate bonding, and of separation, for the genesis of depressive and other disorders (Jones et al. 1997).

Another socialization concept of developmental psychology that is useful in the study of the genesis of psychic disorders is Havinghurst's concept of developmental tasks (Perrez 1998a). Here, development is understood as a sequence of more or less demanding readaptational or developmental tasks, which are experienced as phase-specific disturbances of equilibrium. These tasks are predetermined in a more or less obligatory fashion by the individual's development or by culture. They represent phases of high vulnerability; inadequate performance of a task raises the risk of failure to perform the next task in the sequence. Thus an inadequate performance of the first developmental task – bonding – can impair performance of the next developmental task – the infant's exploration of its immediate environment.

4.4 Learning Theories and Theories of Information Processing

Two traditions of psychological research have yielded valuable contributions to the improved understanding of the generation and perpetuation of psychic disorders: the study of learning and the concepts of information processing. The tradition of the study of learning began in Russia with the Pavlov school, which analyzed the disorganization of behavior by the conditioning of antagonistic inhibitory and excitatory processes. In the United States, Dollard, Meyer, Masserman, and Miller studied the neuroses through animal experiments, starting in the late 1930s. There exists today an extensive corpus of knowledge derived from the psychology of learning (see Perrez and Zbinden 1996): classical conditioning (e.g. the generation of anxiety disorders), operant conditioning (e.g. the generation of certain depressive disorders through the loss of reinforcement and control), and model learning (e.g. the acquisition of aggressive behavior) (for more on this subject, see Vol. 1, Part 1, Chap. 14).

Information-processing approaches (see also Sect. 3.2) are centered on the ascertainment of characteristics or present determinants of psychic disorders, with the emphasis to date on depressive and anxiety disorders (see Ehlers and Lüer 1996). Experimental findings on the relationship between mood and memory, for example, imply that depressive patients are more likely to report negative experiences, although they have not necessarily had such experiences more frequently than others. The concepts of information processing have also proved useful in the study of schizophrenic disorders.

The future development of psychological theories of etiology may be expected to link developmental aspects with behavioral aspects and those of cognitive psychology. Above all, multimodal theories of etiology are required, which will connect psychological findings with biological parameters (compare the approaches of psychoendocrinology and psychoimmunology).

5 Assessment

5.1 Objectives of Assessment in Clinical Psychology

Psychological assessment involves questions reaching over the individual disciplines of psychology (e.g. Jäger and Petermann 1995). In specific disciplines, however, specific problems occur. In this section, we shall discuss some of the aspects of assessment in

clinical psychology (see Stieglitz and Baumann 1994; Maruish 1994).

In medicine, assessment is often equated with the provision of a diagnosis, but in psychology this is but one of several objectives. According to Perrez (1985), assessment in clinical psychology has several distinct functions, with regard not only to individuals, but also to interpersonal systems (partners, families, groups, and organizations) (Reinecker-Hecht and Baumann 1998):

1. *Description*: The description of the present condition and of its alterations forms a basis for the other functions of assessment (e.g. classification); many different assessment procedures are available.
2. *Classification*: Descriptions of large numbers of individual patients may be used in the clinical field to construct systems of classification (e.g. ICD-10, DSM-IV), which often rely upon criteria derived from structured, standardized interviews.
3. *Explanation*: Assessment in clinical psychology may help to explain a disorder (see Sect. 4), particularly with regard to the factors that perpetuate it.
4. *Prognosis*: Assessment in clinical psychology helps to predict the future course of psychic disorders with or without intervention. The prognostication of the course of a disorder after intervention involves an estimate of the probability of success of treatment (compare predictor analysis).
5. *Evaluation*: In the clinical sector, both individual therapeutic approaches (see Baumann and Reinecker-Hecht 1998) and care models (compare quality assurance; Haug and Stieglitz 1995; see also Vol. 1, Part 2, Chap. 12) must be evaluated by means of psychological data-gathering techniques.

German-speaking psychologists use the term *Diagnostik* ("diagnostics") as the equivalent of the English "assessment," and *Verhaltensdiagnostik* as the equivalent of "behavioral assessment."

5.2 Data-Gathering Techniques in Psychology

A number of quite different classification schemes have been suggested for the systematization of data-gathering techniques in psychology, but none of them have achieved general recognition. These techniques are thus usually classified along pragmatic lines (see, e.g. Baumann and Stieglitz 1994). Psychological tests have proved to be an especially important group of techniques (e.g. Rost 1996; Lienert and Raatz 1994). Tests are characterized by the following: standardization (i.e. fixed rules) with respect to their performance, grading, and interpretation; the acquisition of random samples of behavior as evidence for the traits of the subject tested; the measurement (quantification) of traits; and the presence of evaluative criteria such as objectivity, reliability, validity, and norms (Jäger and Petermann 1995).

Although psychological tests are the most rigorous type of technique in clinical psychology, the diagnostic interview is very often used in practice (Wittchen et al. 1994). The diagnostic interview is a means not only of gathering information, but also of providing counseling and treatment.

Various data-gathering techniques and measurement instruments are used to acquire data on the psychic and social levels; these cannot be discussed here individually (see, e.g. AMDP and CIPS 1990; CIPS 1996; Stieglitz and Baumann 1994; Westhoff 1993). Most data-gathering techniques, particularly psychological tests, are intended for use in assessment of the individual. A further group of techniques deals with interindividual systems (such as social networks, partners, families) from the point of view of the individual ("egocentric approach," e.g. family atmosphere as viewed by the child). In fact, only a minority of techniques are intended to be systemic, i.e. to cover all elements of a system and their mutual relationships (e.g. cooperative problem-solving in the family; Cierpka 1996).

A new, particularly important group of techniques known as field assessment represent a complementary approach to classical assessment (see Sect. 2.2). In field assessment, both the data-gathering situation and the observed behavior are in a natural setting (see Perrez 1994; Fahrenberg 1994; Laireiter and Thiele 1995). Field assessment fulfills the objectives of assessment listed in Sect. 5.1. The various techniques of field assessment differ from one another in several parameters, of which the following are most important:

- *Data source*: Self-observation or external observation, automatic recording
- *Data level*: somatic, psychic, social, ecological level
- *Method of recording*: paper and pencil (compare the classical diary), audio/videotape, computer
- *Data sampling*: interval contingent (e.g. at predetermined times), event contingent (e.g. after stress, after contact), signal contingent (upon hearing a signal tone)
- *Temporal separation* between event and recording: immediate, delayed (retrospective)
- *Systems of reference* for the results of field assessment: empirical norms, ideal criteria, intrapersonal before-and-after comparisons.

The diary is the oldest technique of field assessment. It is relatively unstructured in its traditional form and can be of considerable clinical use in the generation of hypotheses as well as for exploratory purposes. More highly structured paper-and-pencil techniques include observation questionnaires or books, which contain

precise instructions about the modalities of observation (including the time of observation, which may be determined by a timer; de Vries 1992).

These paper-and-pencil techniques, however, are gradually being replaced by computerized techniques (pocket or palm-top computers; Fahrenberg and Myrtek 1997), which have a number of advantages. One advantage is that the computerized techniques reduce the latency interval between the actual experience or behavior and its diagnostic recording. The improved organization of the conditions of observation, the opportunity for branched questioning, and the greater evaluative complexity of the data are all further advantages. Electronic data storage makes the data directly transferable to other computers. Furthermore, the computer provides psychologically relevant information about the behavior of the person under assessment, e.g. with respect to the exactness with which he or she follows the instructions for observation. Not only can the computer give the auditory signal for the time of observation, but it can also record when the observation is actually made. Depending on the goals of assessment, the users can define behavioral protocols according to a temporal or event-based sampling plan, by which different types of data (e.g. motives of behavior, behavior, mood) can be obtained in different contexts (e.g. the evaluation of panic attacks, stressful experiences, contact structures, family stresses; Fahrenberg and Myrtek 1997).

In summary, computerized field assessment yields new possibilities for data collection that go far beyond what was previously feasible. These techniques are useful particularly for external assessment of both in- and outpatients, as well as for outpatient self-assessment; they have only a limited application to self-assessment in the hospital setting.

6 Intervention

6.1 Intervention in Clinical Psychology

Various forms of intervention may be used to produce changes in experience and behavior. In somatic medicine, medical, surgical, and physiotherapeutic forms of intervention occupy the foreground. Psychological intervention is a further form that finds application in educational, occupational, and clinical settings. The clinical psychology-based interventions of relevance to psychiatry (Bergin and Garfield 1994) are characterized primarily by the specificity of their methods (Perrez and Baumann 1998a), which are localized to the psychic level, i.e. that of experience and behavior. Typical psychological methods are the interview (e.g. interpretation, confrontation) or the application of behavioral elements (e.g. role-playing, confrontation with reality). What interventions in clinical psychology have in common is not the etiology of the disorders for which they are applied, or their objectives, but rather their methods. These interventions may thus be directed toward somatic characteristics as well.

Further determinants of interventions in clinical psychology which do not apply equally to all methods are their functions (prevention, treatment, rehabilitation), goal orientation, theoretical background (particularly psychological theories), verification of effectiveness, and professional realization (Perrez and Baumann 1998a). These methods of intervention are directed not only toward individuals with disordered functioning (e.g. memory disturbances) or disordered functional patterns (e.g. depressive symptoms), but also toward interpersonal systems such as couples and families (e.g. partner conflicts).

Depressive disorders may serve as an example for the way disordered functioning may be influenced by intervention. These disorders manifest themselves by loss of interest, loss of activities, feelings of worthlessness, guilt feelings, etc. Cognitive behavior-therapy approaches (see Hautzinger and de Jong-Meyer 1996) attempt to affect the psychic components of depressive disorders selectively and systematically through the use of psychological methods. The therapeutic arrangement consists, among other things, in the organization of learning processes that recreate the motivation to participate in gratifying activities and consolidate the quality and frequency of leisure and professional activities (e.g. through reinforcement schemes). Furthermore, cognitive components are selectively influenced that motivate the individual to increase his or her level of activity, promote gratifying emotions, enable appropriate causal attributions, etc. Self-confidence training is used to modify the individual's self-image. The various elements of therapy are thus systematically integrated and do not simply constitute an agglomeration of single aspects.

Clinical-psychology methods of intervention are traditionally largely restricted to the treatment of psychic disorders (e.g. Van Hasselt and Hersen 1994; Bellack and Hersen 1990). In the last 20 years, they have been increasingly applied to somatic disease as well (Sweet et al. 1991).

6.2 Psychotherapy

The traditional conception of psychotherapy covers the subset of interventional methods in clinical psychology

that are used for the treatment of disordered functional patterns (syndromes) and disordered interpersonal systems in the presence of psychic disturbances (Freedheim 1992). The various forms of psychotherapy display a similarity of structure (Perrez and Baumann 1998b):

1. Introductory phase, with thorough assessment and establishment of indications
2. Construction of the therapeutic relationship, which enables role-structuring and the formation of a therapeutic alliance
3. Realization of therapeutic learning
4. Evaluation of the therapeutic process

Behavioral therapy occupies a special position in psychotherapy because of its particularly close connection to scientific psychology (Baumann 1996). Classical behavioral therapy was primarily based on the theory of learning. In the early 1970s, new currents took on importance, such as the social-cognitive learning theory of Banduras (model-based learning) and the cognitive approach of Beck. There is no single definition of behavioral therapy. The characterization by Franks and Wilson (1975; compare Margraf and Lieb 1995) is often cited with approval; this characterization stresses the close relation of behavioral therapy to scientific psychology, especially regarding the application of principles of experimental and social psychology and the systematic evaluation of the effectiveness of treatment.

More than any other method of psychotherapy, behavioral therapy retains as a central element its orientation toward empirical psychology (Margraf and Lieb 1995). Current behavioral therapy still places an emphasis on learning psychology, but also makes use of other subdisciplines (e.g. cognitive, emotional, and social psychology).

Behavioral therapy has become an overarching term for a multitude of diverse methods that are also of major importance in psychiatry (e.g. Bellack and Hersen 1993; Margraf 1996). Behavioral therapy includes very diverse types of intervention and has proved its value remarkably well in multiple evaluative studies (see Sect. 6.5). Behavioral therapy is particularly helpful in psychiatry because of its focus on concrete behavior.

6.3 Combined Approaches in Psychotherapy

Although behavioral therapy is well established and has been thoroughly evaluated, it has nonetheless been criticized from a psychological viewpoint. Behavioral therapy has only partly fulfilled its basic aim of being founded on the entirety of empirical psychology, and in its structural development it has partly become an independent school of psychotherapy, with the associated drawbacks (Baumann 1996). A less school-dependent and more overarching concept of psychotherapy is required in order to provide the necessary depth. Combined approaches may be listed under the following headings (see Norcross 1995):

- *Integration*: This refers to a theoretically based combination of different approaches (e.g. behavioral therapy and psychoanalysis).
- *Eclecticism*: The effective elements of diverse approaches are applied, without regard to their theoretical compatibility.
- *Common, general, nonspecific factors*: These related terms have been used in a pejorative sense ("nonspecific factors as placebo effect") and in an explanatory sense ("nonspecific factors are responsible for the absence of differential findings"). Recently, the term "general factors" has been used to refer to important dimensions (central postulates) that should be realized in all forms of psychotherapy (Weinberger 1995; Grawe 1995, 1998).

6.4 Evaluation

The evaluation of individual therapeutic approaches and systems of care is especially important in the interventional sector (Baumann and Reinecker-Hecht 1998). Behavioral therapy has given a major impulse to the evaluation of therapeutic approaches, while the evaluation of systems of care and questions of quality assurance (Haug and Stieglitz 1995) have only recently received major attention in psychology.

Effectiveness is a central element of scientifically based forms of intervention (Perrez and Baumann 1998a). It should, however, be remembered (Baumann 1997) that the effectiveness construct has multiple components, so that we cannot speak of *the* effectiveness of a method of treatment. Effectiveness must be specified with respect to the following points, among others:

- Data level, data source, constructs (see Sect. 2.3)
- Level of abstraction of the therapeutic objectives (theoretical concepts, concepts of disposition and observation; Perrez and Baumann 1998a)
- Time of evaluation (during or at the end of therapy or at the time of catamnesis)

Given these multiple aspects of effectiveness, it is hardly surprising that there is sometimes controversy over the effectiveness of psychotherapeutic techniques.

Alongside the evaluation of effectiveness of psychotherapeutic methods by individual studies, the

meta-analysis of groups of studies has risen in importance in the last 15 years (Smith et al. 1980; Lipsey and Wilson 1993; Grawe et al. 1994). Meta-analyses attempt to evaluate the available literature systematically by means of so-called strengths of effect (for the advantages and disadvantages of meta-analyses, see e.g. Grawe et al. 1994). Meta-analyses in the narrow sense provide an important contribution to the evaluation of the effectiveness of psychotherapy, but an evaluation cannot rest on such meta-analyses alone (see Vol. 1, Part 2, Chap. 10). These primarily provide descriptions, but can also detect differences in the effectiveness of different forms of therapy. The evaluation of differences, however, like all statistical questions, depends on the particular conventions used (e.g. α-risk).

A normative approach to evaluation going beyond meta-analyses has been recommended by the American Psychological Association (Task Force APA 1993). A catalogue of criteria was created for well-founded interventional techniques, which can be drawn upon as a raster for the evaluation of individual forms of intervention. The Task Force acknowledged the arbitrariness of the criteria chosen, but justified their selection on the basis of expert knowledge and provided a list of psychotherapeutic techniques that satisfy the given criteria. Such lists must, of course, be tested and updated continually, so that "seals of approval" will not be awarded in an uncontrolled fashion.

As the discussions of the approach used by Grawe et al. (1994) reveal, the question of evaluation is central for intervention and remains unsettled in psychotherapy in many of its aspects. Further discussions of this question can profit greatly from a consideration of methodology in psychology.

7 Overview

The discussion in this chapter reveals that psychology can provide major contributions to psychiatry, both through an understanding of psychological methods and through psychological theories. Unless a reductionist position is assumed ("all psychic phenomena are ultimately biologically based and thus biologically explicable"), which we consider unacceptable, psychiatry will always be dependent upon psychology for such major contributions. To what extent this transfer of knowledge can actually be realized depends on the efforts undertaken by both sides. Psychology and psychiatry must make a commitment to multimodal thinking and must acquire competence in diverse modalities if communication between them is to take place. For psychology, this implies a deeper consideration of the modern biological concepts of psychiatry; psychiatry, on the other hand, must take adequate account of phenomena on the psychic level, despite its strong biological orientation. Dialogue with the individual patient can only occur on the psychic level. This perspective must, therefore, be optimized for the benefit of those who suffer from psychic disorders.

8 References

AMDP, CIPS (eds) (1990) Rating scales for psychiatry (European edition). Beltz Test, Weinheim

Amelang M, Bartussek D (1997) Differentielle Psychologie und Persönlichkeitsforschung, 4th edn. Kohlhammer, Stuttgart

Asendorpf JB (1996) Psychologie der Persönlichkeit. Springer, Berlin Heidelberg New York

Ash MG, Geuter U (eds) (1985) Geschichte der deutschen Psychologie im 20. Jahrhundert. Westdeutscher Verlag, Opladen

Baumann U (1991) Methodische Probleme in klinischer Psychologie und Psychiatrie. In: Schneider F, Bartels M, Foerster K, Gaertner HG (eds) Perspektiven der Psychiatrie. Forschung – Diagnostik – Therapie. Fischer, Stuttgart, pp 209–218

Baumann U (1993) Persönlichkeitsforschung in der Psychiatrie. In: Berger M, Möller HJ, Wittchen HU (eds) Psychiatrie als empirische Wissenschaft. Zuckerschwerdt, Munich, pp 40–50

Baumann U (1995) Bericht zur Lage der deutschsprachigen Psychologie 1994 – Fakten und Perspektiven. Psychol Rundschau 46: 3–17

Baumann U (1996) Wissenschaftliche Psychotherapie auf der Basis der Wissenschaftlichen Psychologie. Rep Psychol 21: 686–689

Baumann U (1997) Wie objektiv ist die Wirksamkeit der Psychotherapie. In: Mundt CH, Linden M, Barnett W (eds) Psychotherapie in der Psychiatrie. Springer, Berlin Heidelberg New York

Baumann U, Perrez M (eds) (1998a) Lehrbuch Klinische Psychologie – Psychotherapie, 2nd edn. Huber, Bern

Baumann U, Perrez M (1998b) Grundbegriffe – Einleitung. In: Baumann U, Perrez M (eds) Lehrbuch Klinische Psychologie – Psychotherapie, 2nd edn. Huber, Bern, pp 3–18

Baumann U, Perrez M (1998c) Ätiologie, Bedingungsanalyse: methodische Gesichtspunkte. In: Baumann U, Perrez M (eds) Lehrbuch Klinische Psychologie – Psychotherapie, 2nd edn. Huber, Bern, pp 135–148

Baumann U, Reinecker-Hecht C (1998) Methodik der klinisch-psychologischen Interventionsforschung. In: Baumann U, Perrez M (eds) Lehrbuch Klinische Psychologie – Psychotherapie, 2nd edn. Huber, Bern, pp 346–365

Baumann U, Stieglitz RD (1994) Psychodiagnostik psychischer Störungen: Allgemeine Grundlagen. In: Stieglitz RD, Baumann U (eds) Psychodiagnostik psychischer Störungen. Enke, Stuttgart, pp 3–18

Becker P (1996) Persönlichkeit. In: Ehlers A, Hahlweg K (eds) Grundlagen der Klinischen Psychologie. Hogrefe, Göttingen, pp 465–534 (Enzyklopädie der Psychologie, vol 1)

Bellack AS, Hersen M (eds) (1990) Comparative treatments for adult disorders. Wiley, New York

Bellack AS, Hersen M (eds) (1993) Handbook of behavior therapy in the psychiatric setting. Plenum, New York

Bergin AE, Garfield SL (eds) (1994) Handbook of psychotherapy and behavior change, 4th edn. Wiley, New York

Blanchard EB (1992) An update for the 1990s. J Consult Clin Psychol 60/4: 491–643

Breakwell GM, Hammond S, Fife-Schaw Ch (eds) (1995) Research methods in psychology. Sage, London

Bullinger M (1996) Lebensqualität – ein Ziel- und Bewertungskriterium medizinischen Handelns? In: Möller HJ, Engel RE, Hoff P (eds) Befunderhebung in der Psychiatrie. Lebensqualität, Negativsymptomatik und andere aktuelle Entwicklungen. Springer, Berlin Heidelberg New York, pp 13–29

Bunge M (1984) Das Leib-Seele-Problem. Mohr, Tübingen

Buss AH (1995) Personality. Allyn and Bacon, Boston

Butcher JN, Rouse SV (1996) Personality: individual differences and clinical assessment. Ann Rev Psychol 47: 87–111

Carroll JB (1993) Human cognitive abilities. Cambridge University Press, Cambridge

Cierpka M (ed) (1996) Handbuch der Familiendiagnostik. Springer, Berlin Heidelberg New York

CIPS (Collegium Internationale Psychiatriae Scalarum) (ed) (1996) Internationale Skalen für die Psychiatrie, 4th edn. Beltz, Weinheim

de Vries M (ed) (1992) The experience of psychopathology. Cambridge University Press, Cambridge

Ehlers A, Lüer G (1996) Pathologische Prozesse der Informationsverarbeitung. In: Ehlers A, Hahlweg K (eds) Grundlagen der Klinischen Psychologie. Hogrefe, Göttingen, pp 351–403 (Enzyklopädie der Psychologie, vol 1)

Ernst C, von Luckner N (1985) Stellt die Frühkindheit die Weichen? Eine Kritik an der Lehre von der schicksalshaften Bedeutung erster Erlebnisse. Enke, Stuttgart

Evans RB, Sexton VS, Caswallader TC (eds) (1992) The American Psychological Association – a historical perspective. American Psychological Association, Washington

Fahrenberg J (1981) Zum Verständnis des Komplementaritätsprinzips. Z Klin Psychol 29: 205–208

Fahrenberg J (1994) Ambulantes Assessment. Computerunterstützte Datenerfassung unter Alltagsbedingungen. Diagnostica 40: 195–216

Fahrenberg J (1996) Biopsychologische Unterschiede. In: Amelang M (ed) Differentielle Psychologie und Persönlichkeitsforschung. Hogrefe, Göttingen, pp 139–193 (Enzyklopädie der Psychologie, vol 2)

Fahrenberg J, Myrtek M (eds) (1997) Ambulatory assessment. Hogrefe and Huber, Seattle

Franks CM, Wilson GT (1975) Annual review of behavior therapy – theory and practice, vol 3. Brunner, New York

Freedheim DK (ed) (1992) History of psychotherapy. American Psychological Association, Washington

Grawe K (1995) Grundriß einer allgemeinen Psychotherapie. Psychotherapy 40: 130–145

Grawe K (1998) Psychologische Therapie. Huber, Bern

Grawe K, Donati R, Bernauer F (1994) Psychotherapie im Wandel – von der Konfession zur Profession. Hogrefe, Göttingen

Gray P (1994) Psychology, 2nd edn. Worth, New York

Haug HJ, Stieglitz RD (eds) (1995) Qualitätssicherung in der Psychiatrie. Enke, Stuttgart

Hautzinger M, de Jong-Meyer R (1996) Depression. Z Klin Psychol 25/2: 79

Herrmann T (1976) Die Psychologie und ihre Forschungsprogramme. Hogrefe, Göttingen

Jäger S, Petermann F (eds) (1995) Psychologiosche Diagnostik. Psychologie Verlagsunion, Weinheim

Jones EE, Main M, del Carmen R (eds) (1997) Attachment and psychopathology, part 1/2. J Consult Clin Psychol 64(1/2): 5–74, 237–294

Kazdin AE (1994) Methodology design and evaluation in psychotherapy research. In: Bergin AE, Garfield SL (eds) Handbook of psychotherapy and behavior change, 4th edn. Wiley, New York, pp 19–71

Krohne HW (1996) Angst und Angstbewältigung. Kohlhammer, Stuttgart

Laireiter AR, Thiele C (1995) Psychologische Soziodiagnostik. Tagebuchverfahren zur Erfassung sozialer Beziehungen, sozialer Interaktionen und sozialer Unterstützung. Z Diff Diagn Psychol 16: 125–151

Laireiter AR, Baumann U, Stieglitz RD (1994) Soziodiagnostik. In: Stieglitz RD, Baumann U (eds) Psychodiagnostik psychischer Störungen. Enke, Stuttgart, pp 21–33

Lazarus E (1991) Emotion und Adaptation. Oxford University Press, New York

Lienert GA, Raatz U (1994) Testaufbau und Testanalyse, 5th edn. Psychologie Verlagsunion, Munich

Lipsey MW, Wilson DB (1993) The efficacy of psychological, educational, and behavioral treatment. Am Psychol 7: 702–714

Magnusson D (1990) Personality development from an interactional perspective. In: Pervin LA (ed) Handbook of personality. Guilford, New York, pp 193–222

Margraf J (ed) (1996) Lehrbuch der Verhaltenstherapie, vol 1/2. Springer, Berlin Heidelberg New York

Margraf J, Lieb R (1995) Was ist Verhaltenstherapie? Versuch einer zukunftsoffenen Neucharakterisierung. Z Klin Psychol 24: 1–7

Maruish ME (1994) (ed) The use of psychological testing for treatment planning and outcome assessment. Erlbaum, Hillsdale

Mischel W (1983) Delay of gratification as process and person variable in development. In: Magnusson D, Allen VP (eds) Human development: an interactional perspective. Academic, New York, pp 149–166

Norcross JC (1995) Psychotherapie-Integration in den USA. Überblick über eine Metamorphose. Integr Ther 1: 45–62

Patry JL (1982) Laborforschung – Feldforschung. In: Patry JL (ed) Feldforschung. Huber, Bern, pp 17–42

Perrez M (1985) Diagnostik in der Psychotherapie – ein anachronistisches Ritual? Psychol Rundschau 36: 106–109

Perrez M (1991) The difference between everyday knowledge, ideology, and scientific knowledge. New Ideas Psychol 9: 227–231

Perrez M (1994) Felddiagnostik mit besonderer Berücksichtigung der computerunterstützten Diagnostik. In: Stieglitz RD, Baumann U (eds) Psychodiagnostik psychischer Störungen. Enke, Stuttgart, pp 149–161

Perrez M (1998a) Psychologische Faktoren: Einflüsse der Sozialisation. In: Baumann U, Perrez M (eds) Lehrbuch Klinische Psychologie – Psychotherapie, 2nd edn. Huber, Bern, pp 215-245

Perrez M (1998b) Wissenschaftstheoretische Grundbegriffe der klinisch-psychologischen Interventionsforschung. In: Baumann U, Perrez M (eds) Lehrbuch Klinische Psychologie – Psychotherapie, 2nd edn. Huber, Bern, pp 46–62

Perrez M, Baumann U (1998a) Systematik der klinisch-psychologischen Intervention. In: Baumann U, Perrez M (eds)

Lehrbuch Klinische Psychologie – Psychotherapie, 2nd edn. Huber, Bern, pp 309–319
Perrez M, Baumann U (1998b) Psychotherapie: Systematik. In: Baumann U, Perrez M (eds) Lehrbuch Klinische Psychologie – Psychotherapie, 2nd edn. Huber, Bern, pp 392–415
Perrez M, Reicherts M (1992) Stress coping and health. Hogrefe, Seattle
Perrez M, Zbinden M (1996) Lernen. In: Ehlers A, Hahlweg K (eds) Grundlagen der Klinischen Psychologie. Hogrefe, Göttingen, pp 301–349 (Enzyklopädie der Psychologie, vol 1)
Reinecker-Hecht C, Baumann U (1998) Klinisch-psychologische Diagnostik: allgemeine Gesichtspunkte. In: Baumann U, Perrez M (eds) Lehrbuch Klinische Psychologie – Psychotherapie, 2nd edn. Huber, Bern, pp 100–116
Rost J (1996) Testtheorie und Testkonstruktion. Huber, Bern
Routh DK (1994) Clinical psychology since 1917. Plenum, New York
Schwarzer R (ed) (1997) Gesundheitspsychologie, 2nd edn. Hogrefe, Göttingen
Schwenkmezger P, Schmidt L (1994) Gesundheitspsychologie: alter Wein in neuen Schläuchen? In: Schwenkmezger P, Schmidt L (eds) Lehrbuch der Gesundheitspsychologie. Enke, Stuttgart, pp 1–8
Seidenstücker G, Baumann U (1987) Multimodale Diagnostik als Standard in der Klinischen Psychologie. Diagnostica 33: 243–258
Shepherd M (1987) Formation of new research strategies on schizophrenia. In: Häfner H, Gattaz WF, Janzarik W (eds) Search for the causes of schizophrenia. Springer, Berlin Heidelberg New York, pp 29–38
Sher KJ, Trull TJ (1996) Methodological issues in psychopathology research. Ann Rev Psychol 47: 371–400
Smith ML, Glass GV, Miller TI (1980) The benefits of psychotherapy. John Hopkins University Press, Baltimore
Stieglitz RD, Baumann U (eds) (1994) Psychodiagnostik psychischer Störungen. Enke, Stuttgart
Sweet JJ, Rozensky RH, Tovian SM (eds) (1991) Handbook of clinical psychology in medical setting. Plenum, New York
Task Force APA (1993) Task force on promotion and dissemination of psychological procedures. American Psychological Association, Washington
Van Hasselt VB, Hersen M (eds) (1994) Advanced abnormal psychology. Plenum, New York
von Zerssen D (1996) Forschung zur prämorbiden Persönlichkeit in der Psychiatrie der deutschsprachigen Länder: die letzten drei Jahrzehnte. Fortschr Neurol Psychiatr 64: 168–183
Walker CE (ed) (1991) Clinical psychology: historical and research foundations. Plenum, New York
Watson D, Clark A (eds) (1994). Personality and psychopathology. J Abnorm Psychol 103/1: 6–158
Weinberger J (1995) Common factors aren't so common the common factors dilemma. Clin Psychol 2: 45–69
Westhoff G (eds) (1993) Handbuch psychosozialer Meßinstrumente. Hogrefe, Göttingen
Westmeyer H (1995) Lerntheoretische Persönlichkeitsforschung. In: Pawlik K (ed) Grundlagen und Methoden der Differentiellen Psychologie. Hogrefe, Göttingen, pp 205–239 (Enzyklopädie der Psychologie, vol 1)
Wiggins JS, Pincus AL (1992) Personality structure and assessment. Ann Rev Psychol 43: 473–504
Wissenschaftsrat (1983) Empfehlungen zur Forschung in der Psychologie. Wissenschaftsrat, Köln
Wittchen HU, Unland H, Knäuper B (1994) Interview. In: Stieglitz RD, Baumann U (eds) Psychodiagnostik psychischer Störungen. Enke, Stuttgart, pp 107–125
Zuckerman M (1991) Psychobiology of personality. Cambridge University Press, Cambridge

CHAPTER

13

G. Goldenberg

Neuropsychology of Memory and the Central Executive

Translator: S. Davison

1 Introduction

This chapter concentrates on current developments in two areas of neuropsychology that are of particular relevance in understanding psychiatric syndromes: disorders of memory and disorders of the central executive control of cognitive and emotional function. Both areas are strongly interlinked. The chapter begins with the description of a model of human memory that is integrated by the central executive. On the basis of this model, clinical aspects of relevant disorders are described and new approaches to their explanation presented for discussion.

2 A Neuropsychological Model of Human Memory

In the field of neuropsychology, the lay concept of "memory" is actually separated into several different memory functions (Tulving 1985; Baddeley 1990; Squire et al. 1993). In order to describe and distinguish these different functions, neuropsychologists construct models. These models are supposed to explain observed behaviour and bridge the gap between behaviour and neuronal function. However, they are, and always will be, theoretical constructs, whose contents and classification will remain the subject of heated debate and constant evolution (Goldenberg 1996).

Traditionally, a distinction is made between a primary type of memory, which serves to hold information in the short term, and a secondary type of memory, in which information is stored in the long term. While primary memory has a restricted capacity, secondary memory, in principle, has no restrictions on its capacity. The contents of primary memory are directly present in the conscious mind, whereas the contents of the secondary memory first have to be retrieved in order to enter consciousness.

Primary and secondary memory are also known as short-term and long-term memory, respectively. In the model presented below, primary memory will be referred to as working memory. Secondary memory will be subdivided into episodic and semantic memory.

2.1 Working Memory

Information held in primary memory can be processed cognitively. Primary memory can therefore be thought of as working memory. Theories about working memory postulate that it is comprised of specialised subsystems and a central executive.

The verbal subsystem of working memory holds verbal information, i.e. mainly words and digits, whereas the visuospatial subsystem stores visual and spatial information in the short term. The central executive is needed when information held within both subsystems needs to be processed simultaneously, such as when planning several moves ahead in a chess game. The central executive also co-ordinates the activities of the two subsystems when they are attending to several competing flows of information, such as when reading while also listening to the radio.

In the aforementioned example of the chess game, the planning ahead of possible combinations of moves is not an end in itself. It may serve to find a way out of a tricky situation or at least to lead to a decision about the next move. This example clearly demonstrates that problems are solved and decisions made about future behaviour in working memory.

It can therefore be seen that the central executive of the working memory is also a higher controlling authority of thoughts and behaviour. This aspect of the central executive will be discussed in more detail in the second part of this chapter.

2.2 Episodic and Semantic Memory

Within secondary memory, a distinction is made between episodic and semantic memory. Episodic memory contains memories of individual, time- and place-specific memories of discrete events. In contrast, semantic memory stores general knowledge whose validity is not restricted to individual episodes. We are often not aware of how and when such knowledge was acquired. The contents of episodic and semantic memory are closely linked to one another. During storage of an incident in episodic memory, individual incidents will be related to general knowledge and interpreted in this light. When the incident is retrieved, missing details will be plausibly filled in using general knowledge.

Most of the information that is taken up into episodic and semantic memory is initially labile and can still be forgotten. Really long-lasting storage requires a consolidation process that can last over several years. It can generally be assumed that information that has survived for about 5 years will be retained for the rest of one's life.

The retrieval of contents from episodic and semantic memory stores is co-ordinated and checked by the central executive. The central executive is particularly important for properly ordering and for checking the plausibility of the contents of episodic memory.

2.3
Explicit and Implicit Knowledge

Episodic and semantic memory share the fact that information retrieved from them enters conscious awareness. It is not only the retrieved information, but also the act of retrieval itself that is conscious. Episodic and semantic memory, as forms of explicit or declarative memory, have therefore been distinguished from implicit, non-declarative or procedural memory. These latter terms bring together a series of learning capacities and acquired skills that are used without being consciously sought and retrieved. Sometimes there is not even any conscious access to the contents. Pre-school children, for example, already have implicit knowledge of grammar rules, but this is only manifest in the fact that they build grammatically correct sentences. The children would be unable to state explicitly the rules that govern their sentence construction. In addition, motor skills such as skiing or playing the piano are only truly mastered when their rules have passed from explicit to procedural knowledge and can give rise to the behaviour in question without a detour via conscious retrieval.

3
Memory Disorders in Brain Lesions

3.1
Disorders of Working Memory

When one of the subsystems of the working memory fails, the capacity to retain the relevant information is reduced. The loss of function is manifest even in very easy tasks that do not require any intervention from the central executive. Such a task for the verbal subsystem is the digit span, where a patient has to repeat a sequence of digits immediately after they are presented to him or her. A parallel task for the visuospatial system is the block span, where a sequence of spatial locations have to be reproduced immediately after presentation. Significant reductions in performance on these simple memory tasks are usually linked to further disorders in processing the corresponding information: the digit span with aphasia, and the block span with visuospatial disorders.

If the central executive of working memory fails or is impaired, the simple span capacities are preserved, as they only require the short-term storage, and not the simultaneous processing of the information. However, such patients have difficulties with tests which require information to be simultaneously held and manipulated. An example of such a test is the Paced Auditory Serial Addition Test (Gronwall and Wrightson 1981), in which a patient hears numbers in quick succession and, on hearing each number, must answer with the sum of that number and the preceding one.

3.2
Anterograde and Retrograde Memory Disorders

Disorders of explicit long-term memory may be anterograde or retrograde. Anterograde disorders affect the ability to take new contents into long-term memory and to consolidate them there permanently. Retrograde memory disorders, on the other hand, consist of the loss of memories that were permanently stored and retrievable before the onset of the disorder. Retrograde memory disorders may be restricted in time, only affecting the memory contents that were still in the consolidation phase at the onset of the disorder, or they may be extensive and affect memories that had previously been stored permanently and had been retrievable for many years (Hodges 1995).

3.3
Amnesic Syndrome

Brain damage can cause an amnesic syndrome. The amnesic syndrome is characterised by a selective impairment of the acquisition and consolidation of new information into episodic and semantic memory. Primary memory and implicit learning capacity are preserved.

The amnesic syndrome is always associated with a time-restricted retrograde memory deficit that may stretch back from a few minutes to a few years. On the other hand, the retrieval of memory contents that were already permanently consolidated remains unharmed.

The anterograde memory impairment is seldom global. In the case of incomplete amnesias, the learning of new information is not impossible, but is slow and patchy. In some cases, memory may even appear normal if recall is tested soon after the acquisition of information. However, rapid forgetting over only a few days demonstrates the defective consolidation process. The impairment in the ability to retain information in memory may be specific to particular types of material, either mainly verbal or mainly non-verbal (Milner 1971).

In clinical practice and in everyday life, the lack of acquisition of new information in episodic memory is more noticeable and dramatic initially than the faulty storage in semantic memory.

The inability to store ongoing events in episodic memory can lead to loss of orientation in time and place. The fact that new learning in semantic memory

is also impaired may only become noticeable when patients try to return to training or study.

3.4 Confabulation

Neuropsychological syndromes are characterised by the fact that they are seldom "pure". This is because the causes of brain damage are rarely restricted to exactly those structures which are responsible for one particular neuropsychological function. The amnesic syndrome occurs particularly commonly in association with simultaneous impairment of the central executive. When this occurs, not only is the consolidation of new material faulty, but so are the mechanisms for systematic retrieval of existing information and for monitoring plausibility.

Confabulation may ensue. Patients do not search systematically for the correct answers to questions. Instead they give the first answer that comes to mind, without checking its plausibility. If the material is only being stored poorly from the outset, then the symptoms of the actual memory impairment will be exacerbated by the impairment of the retrieval mechanisms. However, the degree of confabulation depends less on the extent of memory impairment itself than on the extent of the accompanying impairment of executive functions (Fischer et al. 1995).

3.5 Extensive Retrograde Amnesia

Extensive retrograde memory disorders extend beyond time-limited retrograde amnesias and affect materials that were acquired and consolidated many years previously. They are an area that has been actively debated and researched for some years in the field of neuropsychology (Warrington and McCarthy 1988; Hodges 1995). They may affect episodic or semantic memory.

3.5.1 Retrograde Disorders of Autobiographical Memory

Episodic memory is usually referred to as autobiographical memory when discussed in connection with extensive retrograde memory impairment. Extensive retrograde impairment of autobiographical memory has been studied and analysed most extensively in alcoholic Korsakoff's syndrome.

In Korsakoff's syndrome, the memory gaps extend back over many years. However, they show a time gradient, with relative sparing of more remote memories. Serious anterograde memory impairment always accompanies the extensive retrograde impairment. It was therefore suggested that the apparent retrograde memory impairment might come about as a result of an anterograde memory impairment which worsens over years, hindering the learning of new information, such that the apparently "lost" past memories are never properly stored in the first place. However, convincing evidence to the contrary was provided by the observation of a patient who had written an autobiography shortly before the onset of Korsakoff's syndrome (Butters and Cermak 1986). In comparison with the written autobiography, it was evident that there was complete retrograde memory loss for the preceding 10 years. Beyond that, the impairment decreased with increasing remoteness of the memory, but nonetheless extended back over 50 years. The patient had therefore clearly lost access to memories that were previously stored and accessible.

It is a matter of controversy whether extensive impairment of autobiographical memory can occur in the absence of anterograde memory impairment. In the first instance, this combination would suggest a psychogenic amnesia or a feigned deficit. In the case of patients with "identity loss", who no longer know what they are called or who they are, this assumption is seldom called into question (Kopelman et al. 1994). Recently, however, several cases have been reported following head and brain injury where there was minimal or no anterogorade memory impairment but there was severe and extensive retrograde impairment of autobiographical memory, and investigators could find no psychiatric symptoms or motives for simulation (Kapur 1993). The fact that several patients with focal retrograde disorders of autobiographical memory occurring in isolation had lesions in the same parts of the brain has also been put forward as evidence for a possible organic aetiology for such disorders (see below).

Another unresolved question is whether extensive retrograde amnesia can selectively affect either autobiographical memory or semantic memory (Hodges and McCarthy 1993; De Renzi et al. 1987). Selective loss of autobiographical memory might be simulated because the reconstruction of one-off autobiographical events makes greater demands on the central executive than the retrieval of general semantic knowledge. Disorders of the central executive may therefore lead to apparently selective loss of autobiographical memory. This relationship appears particularly convincing when the retrograde impairment of autobiographical memory manifests itself less as a complete loss of memory and more as a confabulatory mixing and mistaking of memories (Hodges and McCarthy 1993). On the other hand, patients have also been described in whom retrieval of autobiographical events was better preserved than was that of general semantic

knowledge (De Renzi et al. 1987). This would suggest that autobiographical memories and general knowledge may in fact have different cerebral substrates.

3.5.2 Retrograde Disorders of Semantic Memory

Patients with extensive retrograde impairment of semantic memory lose access to knowledge that was previously self-evident to them. This loss of knowledge affects linguistic functioning. Spontaneous speech may be devoid of content and filled with set phrases. Disorders of word finding and naming mistakes become apparent when trying to name objects. As far as misnaming of objects is concerned, it is typical for generic categories to be preserved while differentiation within particular categories is lost. Thus a patient may call all four-legged animals horses. It hardly ever occurs, however, that a horse is mistaken for a type of fruit or vice versa. In addition, if the names of objects are provided and patients are asked to describe their characteristics, they cannot always answer. They have particular difficulty in describing those characteristics that distinguish a particular specimen in a given category from other specimens in the same category. On the other hand, knowledge of the qualities that are common to the whole category is retained. If they are asked about the characteristics of a peach, patients know that, like most fruits, peaches are round and juicy, but they do not know that their skin is velvety and they have large stones. Impairment of semantic memory is therefore more likely to manifest itself as an impoverishment and lack of differentiation of knowledge than as a complete loss of knowledge of particular things (Warrington 1975).

3.5.3 Category-Specific Disorders of Semantic Memory

Extensive retrograde impairment of semantic memory may be category specific (Warrington and McCarthy 1987; Gainotti et al. 1995). A series of patients were described who had lost knowledge of natural things such as animals, fruits and vegetables but had well-preserved knowledge about man-made things such as tools and items of furniture.

The opposite problem, i.e. preserved knowledge of natural things and missing knowledge about man-made things, has only been described in very few cases. All except one (Sacchett and Humphreys 1992) of the affected patients had severe aphasias which prevented all but a very limited examination of their semantic memory.

The differentiation between man-made and natural things describes only very approximately which types of knowledge are retained and which are lost in these patients. For example, patients who have lost knowledge about natural things often retain the ability to distinguish body parts, even though these are naturally occurring things. A plausible explanation for the pattern of the gaps in knowledge is that particular types of knowledge about the individual attributes of things are linked together (Warrington and McCarthy 1987). For example, one type of knowledge specifies the visual appearance of objects, whereas another type of knowledge specifies the motor actions and functions that are associated with them. The theory is therefore that examples from the category of naturally occurring objects are distinguished by differences in their appearance and rarely by differences in their function. An apple is distinguished from a pear by its curvature and not by its function as an edible fruit or by the motor action of eating it. In contrast, man-made things have differing functions and are associated with different motor actions. A pair of pliers is used differently from a hammer. Differences in appearance are largely determined by the different functions and uses: the head of the hammer has to be firmly attached to the shaft so that the force of the blow can be transmitted to the nail, whereas the pliers have to have a joint such that the jaws can close around the nail. Body parts likewise have specific and distinct functions which determine their shape: the nose has openings to draw in air, and the hand has fingers to grasp with. Category-specific deficits can therefore occur in which either knowledge about appearances or knowledge about functions is selectively lost. Categories of objects are affected to a greater or lesser extent by the deficit, depending on how important that type of knowledge is for distinguishing between different examples within that category. The loss of knowledge about appearances would tend particularly to affect natural things, whereas the loss of knowledge about function and usage would tend particularly to affect man-made objects and body parts.

The idea that semantic memory is composed of subsystems which each hold information about only one aspect or property of objects is supported by observations of patients who have selectively lost their knowledge about the appearance of objects while retaining knowledge about the non-visual properties of the same objects (Goldenberg 1993). In the same way, the failure of patients with apraxia to use tools properly has been interpreted as being caused by a selective loss of the knowledge about the correct use of objects and tools (De Renzi and Luchelli 1988; Goldenberg and Hagmann 1998).

Despite some methodological shortcomings (Gaffan and Heywood 1993), these observations present a coherent and convincing picture. It follows that semantic memory consists of a network in which similar types of information about particular

characteristics of things are linked together. When the knowledge is retrieved, the related components are simultaneously activated. If only individual parts of the network are damaged, then knowledge is lost selectively for individual categories of objects or for individual properties of things.

4
Anatomical Basis of Memory Disorders

Technical advances in the last 10 years have led to an enormous expansion and refinement of the depiction of the anatomical basis of psychological functions and their disorders. The impressive images that are publicised and the popularity which they enjoy should not, however, be allowed to obscure the fact that the mapping of psychological functions onto anatomical locations remains an uncertain enterprise.

The uncertainty arises from the fact that psychological functions are theoretical functions that can only be inferred by interpreting the data that are actually observed. Even when the anatomical findings themselves are accurate and reliable, relating them to particular psychological functions remains a matter of interpretation and controversy. Nevertheless, it is precisely these attempts to discover and verify such relationships that form the essence of and also the fascination of neuropsychology.

4.1
Cerebral Substrate of Working Memory

The measurement of cerebral blood flow in healthy subjects undertaking tasks which require the short-term storage of verbal or visuospatial material provides a clear and credible picture of the cerebral basis of working memory. The short-term storage of verbal information activates pre-motor and temporo-occipital speech areas of the left hemisphere in a similar way to speech. This is attributed to the fact that the active retention of verbal memory within working memory occurs through inner speech (Paulesu et al. 1993). In contrast, visuospatial tasks activate predominantly right brain regions, namely the frontal and parieto-occipital cortices (Jonides et al. 1993). However, if the information to be held relates to the appearance of objects rather than to their spatial relationships, the activity shifts from the right to the left parietal lobe with additional activation of the left temporo-occipital region (Smith et al. 1995). The anatomy of the central executive of working memory will be discussed in the last section of this chapter.

4.2
Amnesic Syndrome

The localisation of the structures which are responsible for the uptake and consolidation of information in declarative long-term memory have mainly been deduced from the examination of lesions in patients with the amnesic syndrome. It is clear from this work that these memory functions rely on the integrity of the limbic structures which surround the third ventricle on the inner surface of the cerebral hemispheres.

Within the limbic system, there are three strategic locations in which lesions typically give rise to an amnesic syndrome: the hipppocampal formation, the thalamus and the basal forebrain. The core symptoms of the amnesic syndrome should be the same for lesions in all three regions, but because of their differing anatomical relationships, each may have differing accompanying symptoms.

Lesions of the basal forebrain can also affect the frontal lobes and are then often associated with confabulation (Fischer et al. 1995). Lesions of the thalamus form the basis for Korsakoff's syndrome, which is likewise associated with confabulation and with extensive impairment of autobiographical memory. These connections are understandable if we accept that confabulation and extensive impairment of autobiographical memory occur as a result of insufficient central executive control of retrieval from long-term memory. The central executive depends largely on the function of frontal cortical areas and frontobasal lesions thus lead directly to disorders of the central executive. Thalamic lesions may conceivably interrupt the connection of the frontal lobes to the limbic system and with it the central executive's access to the memory store (Hodges and McCarthy 1993).

The hippocampal formation borders on the neocortical areas of the temporal lobe. Lesions of the temporal neocortex can cause extensive retrograde memory deficits (see Sect. 4.3). In addition, the hippocampal formation lies close to the amygdala, which plays an important role in the regulation of emotions. However, it has not yet been conclusively established which symptoms may be attributed to damage of the amygdala in humans (Aggleton 1993; Young et al. 1995).

Bilateral damage to the limbic system is necessary for global amnesia, while unilateral lesions lead to material-specific memory impairment. Verbal material is affected predominantly by left-sided lesions, and non-verbal material by right-sided lesions. However, the consequences of left-sided lesions are on the whole more marked than those of purely right-sided lesions. The few patients in which it has been suggested that unilateral lesions may have led to a transient global

amnesia have all without exception had left-sided brain lesions (Goldenberg 1995a).

4.3
Extensive Retrograde Impairment of Semantic Memory

The cases of category-specific deficits of semantic memory that were mentioned above have provided a fairly consistent picture of the anatomical basis of semantic memory (Gainotti et al. 1995). They suggest that the left hemisphere is more important than the right for nearly all types of knowledge. Unilateral left brain lesions can be sufficient to cause retrograde deficits of semantic memory. This asymmetry is seen irrespective of whether the knowledge is tested through speech or pictures (Gainotti et al. 1983; Goldenberg 1995b). Temporal lobe lesions form the basis of deficits specific to knowledge about living things, whereby the lateral parts of the inferior and middle temporal gyri may be important. On the other hand, all patients with a loss of knowledge about man-made things have had suprasylvian lesions. These lesions were very extensive and affected the frontal lobes, parietal lobes and the perisylvian speech area. In keeping with the extensive nature of the lesions nearly all patients had severe aphasias.

Isolated deficits in knowledge about the appearance of things have been described following lesions of the temporo-occipital gyri at the base of the left hemisphere (Goldenberg 1993), whereas apraxias with deficits in the knowledge about the correct use of objects have been linked to left suprasylvian areas (Basso et al. 1985). These areas are bordered on one side by the areas responsible for category-specific deficits and on the other side by both visual and motor cortices. Such topographical relationships lend plausibility to the hypothesis mentioned above that similar types of information are linked in semantic memory.

Studies of cerebral blood flow in healthy subjects concur in this respect with clinical findings. Activation during animal recognition is concentrated in the inferior temporal lobe, whereas during the recognition of tools it extends over the sylvian fissure into the parietal or frontal lobe (Damasio et al. 1996; Perani et al. 1995). Visual mental imagery activates left occipital and temporo-occipital areas (Farah 1989; Goldenberg 1993; D'Esposito et al. 1997).

4.4
Extensive Retrograde Deficits of Autobiographical Memory

Extensive autobiographical memory deficits following thalamic lesions have already been mentioned. These are associated with an anterograde amnesia and are best understood as a disorder of the systematic retrieval of available autobiographical memories. Extensive bilateral temporal damage, such as that seen after herpes simplex encephalitis, destroys not only semantic memory but also remote autobiographical memory.

It is less clear what forms the anatomical substrate of selective deficits of autobiographical memory. Such deficits might be caused by right lesions of the fronto-temporal cortex (Markowitsch et al. 1993; O'Connor et al. 1992). However, patients have been described in which no clear lesions can be detected.

4.5
Memory Disorders in the Degenerative Dementias

The combination of an amnesic syndrome with extensive retrograde memory impairment is particularly characteristic of dementia of the Alzheimer type. This fits well with the theory that the temporal lobe plays a key role in memory function, because the pathological changes in Alzheimer's disease appear to extend from the entorhinal cortex on the medial side of the temporal lobe to the hippocampus on one side and laterally into the neocortex on the other side (Braak and Braak 1991).

It appears that, in the early stages of the disease, the retrograde impairment of semantic memory affects natural things to a greater extent than man-made one (Silveri et al. 1991). This finding is compatible with the predominantly temporal location of the degenerative changes.

5
Disorders of the Central Executive

The disorders described in this section are traditionally referred to as the "frontal lobe syndrome". This term does not, however, fit well with the logic of neuropsychology. Neuropsychological method consists first of relating observed behaviours to psychological functions and then seeking the neurological substrate of such functions. The term "frontal lobe syndrome" suggests a direct relationship between behaviour and the neurological substrate. On the assumption that the disturbed functions are best described as those of the central executive, the term "dysexecutive syndrome" has gained acceptance in recent years (Baddeley 1996; Shallice 1988).

The central executive has already been discussed in an earlier section in relation to memory function. It is postulated that the same central executive that

co-ordinates working memory also co-ordinates and oversees other cognitive functions, namely emotional reactions and social behaviour. This theory is not without its problems. It can lead to the central executive being equipped with omnipotence and omniscience and thus becoming a "mind within the mind" or a homunculus. With some exaggeration, it could be suggested that the indivisible and inexplicable soul hides in the central executive from the attacks of neuropsychological analysis (see Baddeley 1996).

Faulty central executive function is assumed to cause a variety of neuropsychological symptoms. In the first instance, these can be divided into two categories: disorders of cognitive problem-solving and disorders of emotions and social behaviour.

5.1 Disorders of Problem-Solving

The central executive has to intervene when problems and obstacles interrupt routine activities, thereby calling for forward planning, the weighing up of alternative actions and rational decision-making. If the central executive fails, patients may continue their routine activities without recognising that any problem exists. If patients seek solutions to problems, they fail to take all the alternatives into consideration. They stick to previous ways of solving problems and make rash decisions without consideration of the consequences and also sometimes in violation of previously given rules.

Studies of the dysexecutive syndrome using psychological tests aim largely to demonstrate the cognitive problem-solving deficits. Several types of task are used for this purpose (von Cramon and Matthes-von Cramon 1993).

Sorting tests require a series of stimuli to be classified and sorted according to as many different criteria as possible. For example, coloured tiles of varying size, thickness and shape may be sorted by colour, size, thickness or shape. In order to find the similarities between the stimuli, the patient has to be able to abstract from the visible differences. Sorting tests therefore require both the ability to abstract and flexibility of thought. Tests aimed specifically at flexibility of thought are those in which the sorting criteria are learned over several attempts and then changed. If the central executive fails, then habitual reactions are not successfully inhibited.

So-called fluency tests examine the fluency of the production of ideas. For example, patients are asked to give as many words as they can think of beginning with a particular letter, to produce as many graphical figures as they can, or to name as many uses for a particular object such as a brick within a specified time. A further variation of the fluency test is the random generation test (Baddeley 1966; Spatt and Goldenberg 1993). In this test, patients are required to produce digits or letters in a random order. This apparently trivial exercise makes great demands on the central executive. The human mind cannot throw a dice. The human mind can only produce the digits or letters by using strategies (e.g. in ascending order or steps of two). If a strategy is employed for too long, then a predictable sequence of letters or digits results. In order to simulate chance, each strategy must be used for only a short time. The central executive must monitor for uninterrupted strategies and change them.

Perseveration appears as a central symptom of disorders of the central executive in both sorting and fluency tests; patients stick to previously selected answers and strategies. Perseveration may appear after an interval during which another answer was given.

A further line of investigation consists of administering planning exercises in which patients are confronted with complex problems and have to seek a solution. The problems may consist of mental puzzles in which counters or balls have to be moved from one position to another in accordance with particular rules. These games require patients to plan several moves ahead, and the final position can only be reached via a circuitous route. Another possibility is to present patients with tasks that involve planning activities similar to those encountered in daily life, such as the drawing up of a timetable in which a series of things to do have to be co-ordinated in the optimal way. The disturbance of executive control function manifests itself in these tasks as rash actions, inadequate consideration of all circumstances and the breaking of rules.

All cognitive "frontal tests" share the fact that patients receive clear instructions as to what they have to do; that patients have to concentrate fully on the task at hand; and that they are protected in the test situation from outside distractions.

The abilities required for such tests, namely abstraction, production of ideas, flexibility of thought and forward planning, are strongly reminiscent of a description of intelligence. In fact, a high correlation was found in normal subjects between results on frontal tests and culturally fair intelligence tests (Duncan et al. 1996).

5.2 Disorders of Behaviour

The consequences of disordered control of emotional and social behaviour do not lend themselves easily to systematic classification, especially as some produce apparently conflicting symptoms. Increased

"environmental dependency" may be manifest in differing levels of severity. Extreme and impressive manifestations are "utilisation behaviour" and "imitation behaviour".

The term "utilisation behaviour" describes the tendency to pick up objects within easy reach and use them, even when their use is inappropriate at that moment. For example, if a rubber stamp and a piece of paper are lying on a table, the paper will get stamped all over. Milder manifestations which lie at the boundary to normality are the tendency to play with objects that are lying around, such as match boxes or pencils.

In the case of imitation behaviour, patients mimic the actions of those facing them without being asked to, and sometimes in contravention of the rules of social conduct. Echolalia, the constant repetition of the last few words of the person one is talking to, belongs in this category.

Increased distractibility is a further symptom of so-called environmental dependency. It is difficult for patients to tune out irrelevant environmental stimuli and they attend to them, sometimes abandoning their original behavioural goal. Even patients who can resist this temptation complain that they are easily distracted by irrelevant environmental stimuli and are hindered in their activity.

Poor impulse control, manifest as unbridled outbursts of aggressive or sexual behaviour, is a symptom of failure of the central executive that has particularly serious social consequences. Pathological outbursts of anger may be precipitated by trivial things. Sometimes patients are able to recognise in retrospect that the precipitant was ridiculous and the anger out of proportion. At the time, however, they are incapable of controlling their behaviour. Inadequate control over sexual impulses may be manifest in socially unacceptable attempts to seek sex or, in extreme cases, as excessive masturbation.

Rash behaviour refers above all to the behaviour involved in solving cognitive problems. It is a matter of opinion whether this is classed as a cognitive or a behaviour disorder. When patients are confronted with problems they do not think through the consequences of several possible solutions. Instead they decide overhastily to take the first behavioural option they can think of. As a result of this rash behaviour, problem-solving is stopped before it has really begun and the need to solve a problem is not even recognised and assessed accordingly.

Rule-breaking as a type of error seen in frontal lobe tests has already been discussed. In such tests, patients must follow rules that are given to them at the time of the test but that otherwise have no general application. Patients may, however, break social rules that they have been familiar with for many years. Such rules exist, for example, for engaging in conversations. When patients disregard them, they do not let the other person get a word in; they do not take notice of objections; they talk round questions; and bring up topics that are not appropriate for the other person they are with or the situation they are in. Disregard for social rules relating to good manners and hierarchy is experienced by others as over-familiarity.

Flattening of affect and egocentric behaviour are subtle symptoms which may only be noticed by the closest relatives, but that nonetheless have particularly severe consequences because of their effect on interpersonal relationships. The patients appear coarse and show little sympathy for those closest to them. In conflicts they are egocentric and concentrate exclusively on their own interest or misfortune. It appears as though they are incapable of simultaneously taking into consideration their own opinions and feelings as well as those of other people.

The central executive is needed to initiate behaviour in the absence of an external impetus as well as to see behaviour through in the absence of external reinforcement. If the central executive fails, then lack of drive results. Patients do not start any activities for themselves. Instead, they wait for instructions and requests. Behaviours that have been started are broken off in the face of small difficulties. Patients make no long-term plans and do not actively contemplate the future.

Placidity (from the Latin *placidus*, meaning calm, peaceful) is related to lack of drive, but the two are not identical and do not necessarily occur in association with one another. Some patients with disorders of the central executive are excessively peaceable and contented. Remarkably, some patients experience this change as positive and rationalise it to the effect that the stroke of fate that was the brain damage has actually helped them to achieve peace with themselves and the world. It can be that relatives also see the change as advantageous if patients engaged in aggressive behaviour before their brain injury.

6 Dissociation of Different Aspects of Central Executive Deficit

It is an unresolved question whether the central executive represents a unified function or whether a whole series of controlling authorities and functions, related to one another to a greater or lesser extent, are hidden within this construct. To answer this question, it is essential to know whether different aspects of the disorders of executive control can occur independently of one another. Dissociation of the different aspects of the disorder of executive control would go against the idea of a unified central executive.

Clinical experience tends to support a division into several executive control functions that are independent of each other. The results of frontal lobe tests do not permit any conclusive predictions of behavioural disturbance. There are patients who manage normally on theoretical problem-solving tests but who founder in everyday life (Shallice and Burgess 1991; Saver and Damasio 1991). On the other hand, there are patients who fail on frontal tests but whose emotional and social behaviour in everyday life, at least according to clinical impression, are largely unremarkable. Patients with Parkinson's disease, for example, fall into this group (Goldenberg et al. 1990).

Below we present three theoretical approaches which attempt to explain the dissociation of cognitive problem-solving deficits and disorders of emotional and social behaviour. However, it must be borne in mind when assessing these theories that the assumption that the functions are not associated with one another is based on clinical impression and individual observations that are more or less anecdotal. It has not been systematically researched to date.

Two of the approaches presented are based directly on anatomical observations and equate the function of the central executive with that of the frontal lobes. In particular, they refer to the pre-frontal cortex, which occupies the anterior portion of the frontal lobe.

6.1 Integration and Inhibition

One anatomically based approach distinguishes two functions of the central executive which are each assigned to different sections of the pre-frontal cortex: one holds and integrates information in working memory, while the other prevents irrelevant information from entering working memory (Fuster 1989). This differentiation of the two functions is based mainly on experiments on apes who have parts of their frontal lobes removed.

Apes with dorsolateral frontal lesions fail to react to a stimulus after a delay when they are required to do so. Their difficulty appears be the holding of the association between the stimulus and the response in working memory. Primates with orbitofrontal lesions, on the other hand, have particular difficulties when the relationship between stimulus and response is constantly changed. They cannot prevent a previously learned association from influencing their reactions. The failure of the animals can be attributed to an inability to shield the working memory from irrelevant and distracting information.

Parallels can be drawn between the findings of these animal experiments and the clinical manifestations of impaired central executive functions in humans.

Frontal lobe tests call mainly for the storing and processing of information in working memory, whereas the control of everyday behaviour and emotions calls mainly for the suppression of irrelevant and digressive information and actions. Frontal tests are administered in a structured and distraction-free environment, whereas in everyday life, behaviour has to be controlled in unstructured situations and under the influence of environmental stimuli. Everyday life therefore calls more on the ability to shield from irrelevant stimuli and to inhibit digressive reactions than does the test situation. Conversely, the demands placed by frontal tests on the storing and comprehensive integration of information are probably greater than the requirements for controlling socially acceptable behaviour in everyday situations. The significance of this difference between "laboratory" and "life" in the manifestations of central executive deficits is demonstrated by patients who are able to perform without hesitation in problem-solving tests, but who behave unsystematically, perform rash and irrelevant actions and break rules when they have to plan and carry out real activities outside laboratory situations (Shallice and Burgess 1991). The observation mentioned earlier that patients with Parkinson's syndrome fail on frontal tests but are not conspicuous in everyday life for their chaotic behaviour or rule-breaking may also be an indication that everyday behaviour is better preserved than problem-solving in a laboratory.

6.2 Cognitive Planning and Behavioural Control

A further approach, based directly on anatomical observations, differentiates between the cognitive capacity of the central executive and its ability to intervene in the control of emotional reactions and behaviour. It is assumed that the latter abilities relate to the integrity of the orbitofrontal cortex and its connection to the limbic system. Saver and Damasio (1991) showed that a patient who had suffered a complete social decline as a result of "sociopathy" caused by brain damage, was not only able to complete cognitive problem-solving tests perfectly, but was also able to provide highly moral and socially acceptable solutions to theoretical cases of social conflict. The central executive was obviously able to plan and assess actions and their consequences, but could not carry this judgement over into real behaviour.

In a further study, Damasio et al. (1990) studied psycho-galvanic skin responses in this patient and five further patients with orbitofrontal lesions. The patients were shown pictures of catastrophes, maimings and nudity under two sets of conditions: once without

further instructions and once with the request that they describe their feelings on seeing the pictures. Whereas both normal controls and patients with non-frontal brain damage produced a psycho-galvanic response in both sets of conditions, the patients with frontal lobe damage showed a restricted response in the second set of conditions where they had to describe their feelings.

Damasio constructs a far-reaching theoretical speculation from these findings. He postulates that emotional reactions are the means whereby the prefrontal cortex is able to communicate biologically important decisions swiftly and effectively to the rest of the brain and translate them into appropriate behaviour (Damasio 1994). The anatomical basis for this behavioural control is the connection of the prefrontal cortex to the limbic system. This connection requires the integrity of the orbitofrontal cortex. The psycho-galvanic skin response is interpreted as the outward sign of emotional reactions that are controlled by the prefrontal cortex. Its loss in patients with orbitofrontal lesions means therefore that the pre-frontal cortex has lost access to emotions and thereby to effective control of behaviour. However, if we believe the experiment that forms the basis for this speculation, the challenge of describing one's own feelings is enough to suggest that the competence of the central executive is restored.

6.3
Control of Behaviour as a Dual Cognitive Task

The third approach presented here seeks an explanation at the level of psychological functioning without relating it directly to anatomical structures. It is postulated that the co-ordination of two simultaneous activities places a particularly large burden on the central executive. The control of emotional and social behaviour can be seen as just such a dual task. In order to control behaviour and emotions in everyday life, the information, demands and stimuli coming from the environment must be processed. Simultaneously, one's emotional reactions to the stimuli and one's behaviour must be observed, judged and controlled. This dual task overloads the central executive of working memory. The inability to supervise two activities simultaneously affects dual cognitive tasks as much as it does behavioural control. However, the impairment in dual cognitive task performance can only be demonstrated in specialised test situations, whereas the impairment of behavioural control has dramatic repercussions on patients' everyday lives.

The relationship between dual cognitive tasks and control of behaviour was examined in two studies by an English research group (Baddeley 1996; Alderman 1996). The ability to carry out two tasks simultaneously was examined using a combination of two very easy exercises, namely a tracking exercise and the repetition of sequences of digits. The ability to combine the two exercises was measured by assessing the drop in performance when the exercises were combined, compared with when the exercises carried out individually. In one study, this measure showed a far greater relationship with an assessment of behavioural disorder in brain-damaged patients than did a series of other frontal lobe tests. A further study looked at the effectiveness of behaviour modification techniques in patients with disturbed behaviour following brain injury (Alderman 1996). A few patients benefited from the therapy, whereas others were not able to bring their behaviour under control. The patients who were considered "non-responders" differed from the "responders" in their dramatic drop in performance on the dual-task experiment mentioned previously, but hardly differed at all in their performance in other frontal lobe tests. The authors inferred that these patients were incapable of simultaneously monitoring their own behaviour and the social sanctions it gave rise to, and therefore the therapeutic feedback was not able to exert any influence over their behaviour.

The assumption that behavioural control is a dual task and that the management of dual tasks is a particularly difficult and error-prone task for the central executive does not, however, explain the fact that there are patients who do badly in cognitive problem-solving tests but who nonetheless behave in socially appropriate ways. If the supervision of dual tasks places particularly large demands on the central executive, then all patients exhibiting any symptoms of impaired central executive function should show impairment in the supervision of dual tasks. By the same token, they should all exhibit disorders of behaviour. In order to explain the lack of association, it would be necessary to assume that the supervision of dual tasks can be impaired or preserved independently of the other tasks of the central executive. This assumption is supported by the fact that many studies have failed to demonstrate one clear impairment in dual cognitive tasks in patients with Parkinson's disease (Goldenberg 1990; Brown and Marsden 1991; Robertson et al. 1996).

There is, however, another possible explanation for the observation that some patients have problem-solving difficulties with otherwise good control of their behaviour. It is possible that the lack of behavioural impairment is only apparent. For example, patients with Parkinson's disease may in fact display poor impulse control and egocentricity, but this is considered by relatives as less important than the obvious motor symptoms; or the relatives excuse the poor impulse control and egocentricity as natural age-related changes. In any case, there are no systematic studies of putative behavioural disorders in these patients.

7
Anatomy of the Central Executive

Disorders of the central executive are frequently associated with lesions of the frontal lobes. Both animal experiments and clinical findings suggest that different kinds of lesions within the frontal lobes can damage different aspects of the central executive. In addition, disorders of the central executive can also be caused by lesions that do not lie in the frontal lobes.

7.1
Frontal Lobe Areas

Three areas of the frontal lobes are particularly relevant for disorders of the central executive: the dorsolateral pre-frontal cortex, the orbital pre-frontal cortex and the cingulate gyrus. Lesions in these three areas can be related to three different aspects of impairment of executive function.

Dorsolateral lesions lead to the severest deficits in cognitive problem-solving tasks. Selective behaviour disorders with preserved problem-solving are seen almost exclusively following orbitofrontal lesions. Lesions of the cingulate gyrus are particularly significant for lack of drive and placidity.

However, clinical experience shows that attributing symptoms to particular locations in the frontal lobes is more difficult than it is in other areas of the brain. Clinical cases are well known where extensive frontal lobe lesions were present in the absence of any noticeable effects on cognition or behaviour and, vice versa, where severe deficits did not correspond to any substrate on imaging.

Naturally, lesions may cross the boundaries between the three portions of the frontal lobe. The resulting combination of the different aspects of impairment of central executive function may lead either to symptoms potentiating one another or dampening one another. Thus disregard for rules, poor impulse control and increased distractibility in response to environmental stimuli may all worsen problem-solving ability and performance on frontal lobe tests. On the other hand, placidity and lack of drive may conceal the existing poor impulse control and egocentricity.

7.2
Laterality of Frontal Lobe Lesions

Localising symptoms in the frontal lobes differs from that in other parts of the brain in that differences between right- and left-sided frontal lobe lesions are less significant. In the case of dorsolateral lesions, it can be shown, using appropriate tests, that right-sided lesions may be more likely to interfere with the fluent production of graphic patterns or drawings and with the retention and processing of visuospatial information. On the other hand, left-sided lesions may be more likely to affect the fluency of speech and the storage and processing of verbal information. In addition, the effects of right- and left-sided frontal lesions on social behaviour and cognitive ability should differ. In lesions that are otherwise in exactly the same location in the frontal lobes, disturbed social behaviour and disregard for rules are more prominent in right-sided lesions, and difficulties in exercises which call for problem-solving and cognitive flexibility are more prominent in left-sided lesions.

7.3
Central Executive Deficits in Non-frontal Lobe Lesions

Impairment of central executive function is a clinical diagnosis which can and should be made independently of the location of the damage. In fact, such impairment is not necessarily associated with frontal lesions. Similar disorders can arise as a result of lesions that do not affect the frontal lobes at all. It is an unresolved question whether, in these cases, the central executive deficits arise because the connection to and from the frontal cortex has been interrupted or whether in fact the neuronal substrate of the central executive includes areas of brain outside the frontal lobes.

Impaired performance on frontal lobe tests is the most striking neuropsychological finding in the subcortical dementias which occur as a result of multiple white matter lesions, such as those seen in progressive supranuclear palsy, lacunar state and multiple sclerosis (Albert et al. 1974). "Frontal" cognitive impairment is also seen in patients with Parkinson's disease who show no other signs of dementia whatsoever (Goldenberg et al. 1990; Dubois et al. 1991). Even healthy normal individuals show decreased flexibility of thinking and inadequate planning ability on frontal lobe tests when they are tested after sleep deprivation (Horne 1988). Likewise, patients with schizophrenic disorders perform badly on frontal lobe tests. Moreover, they also show behavioural disorders such as affective flattening and apathy which resemble the sequelae of frontal lobe lesions (Weinberger 1988; Frith 1992). These symptoms are associated with memory and other cognitive deficits (Dunkley and Rogers 1994; McKenna et al. 1995).

It is a matter of debate whether a central executive disorder can be regarded as causing psychosis (Frith 1992; David 1992). It has been suggested that poor executive control of perception, thinking and affect

may lead to delusions, hallucinations and thought disorder. This idea is tempting, but not rigorous in its logic. It is reminiscent of the principle of political accountability; following a serious crime, the resignation of the Home Secretary is called for because the police were not able to prevent the crime. However, this is far from meaning that the Minister is suspected of having committed the crime him- or herself. The co-occurrence of psychosis with impairment of the central executive and of memory and other cognitive functions (Dunkley and Rogers 1994; McKenna et al. 1995) can be explained, without falling into any logical traps, as different manifestations of the same brain disease.

8 Conclusions

The neuropsychological study of schizophrenia is a growing field for the exchange of ideas and collaborative working between psychiatrists and neurologists. Psychiatric research overlaps with neurological research insofar as it is concerned with the study of the psychological manifestations of brain diseases. Neuropsychology brings methods and ways of thinking to the study of psychiatric disorders that were developed first and foremost from experience with patients with circumscribed, anatomically defined brain lesions. Alongside the emphasis on the cerebral localisation of psychological functions, the theoretical model-oriented analysis of psychological phenomena is a defining characteristic of the neuropsychological approach.

Neuropsychology starts from models of normal psychological functions and their disorders and tries to derive falsifiable hypotheses from these models and test them. The application of these methods to the study of dementias, psychotic disorders and affective disorders presents a challenge to neuropsychology and hopefully will serve to enrich psychiatry (Halligan and Marshall 1996).

9 References

Aggleton JP (1993) The contribution of the amygdala to normal and abnormal emotional states. Trends Neurosci 16: 328–333

Albert ML, Feldman RG, Willis AL (1974) The "subcortical dementia" of progressive supranuclear palsy. J Neurol Neurosurg Psychiatr 37: 121–130

Alderman N (1996) Central executive deficit and response to operant conditioning methods. Neuropsychol Rehab 6: 161–186

Baddeley A (1966) The capacity for generating information by randomization. Q J Exp Psychol 18: 119–129

Baddeley A (1990) Human memory – theory and implications. Erlbaum, Hove

*Baddeley A (1996) Exploring the central executive. Q J Exp Psychol 49A: 5–28

Basso A, Faglioni P, Luzzatti C (1985) Methods in neuroanatomical research and an experimental study of limb apraxia. In: Roy EA (ed) Neuropsychological studies of apraxia and related disorders. North Holland, Amsterdam, pp 179–202

Braak H, Braak E (1991) Neuropathological staging of Alzheimer-related changes. Acta Neuropathol 82: 239–259

Brown RG, Marsden CD (1991) Dual task performance and processing resources in normal subjects and patients with Parkinson's disease. Brain 114: 215–231

Butters N, Cermak LS (1986) A case study of the forgetting of autobiographical knowledge: implications for the study of retrograde amnesia. In: Rubin DC (ed) Autobiographical memory. Cambridge University Press, Cambridge, pp 253–272

Damasio AR (1994) Descartes' error – emotion, reason and the human brain. Putmans, New York

Damasio AR, Tranel D, Damasio H (1990) Individuals with sociopathic behavior caused by frontal damage fail to respond autonomically to social stimuli. Behav Brain Res 41: 81–94

Damasio H, Grabowski TJ, Tranel D, Hichwa RD, Damasio AR (1996) A neural basis for lexical retrieval. Nature 380: 499–505

David AS (1992) Frontal lobology – psychiatry's new pseudoscience. Br J Psychiatry 161: 244–248

D'Esposito M, Detre JA, Aguirre GK, Stallcup M, Alsop DC, Tippet LJ, Farah MJ (1997) A functional MRI study of mental image generation. Neuropsychologia 35: 725–730

De Renzi E, Luchelli F (1988) Ideational apraxia. Brain 111: 1173–1185

De Renzi E, Liotti M, Nichelli P (1987) Semantic amnesia with preservation of autobiographic memory. A case report. Cortex 23: 575–598

Dubois B, Boller F, Pillon B, Agid Y (1991) Cognitive deficits in Parkinson's disease. In: Boller F, Grafman J (eds) Handbook of neuropsychology, vol 5. Elsevier, Amsterdam, pp 195–240

Duncan J, Emslie H, Williams P (1996) Intelligence and the frontal lobe: the organization of goal-directed behavior. Cogn Psychol 30: 257–303

Dunkley G, Rogers D (1994) The cognitive impairment of severe psychiatric illness: a clinical study. In: David AS, Cutting JC (eds) The neuropsychology of schizophrenia. Erlbaum, Hove, pp 181–196

Farah M (1989) The neural basis of mental imagery. Trends Neurosci 12: 395–399

Fischer RS, Alexander MP, D'Esposito M, Otto R (1995) Neuropsychological and neuroanatomical correlates of confabulation. J Clin Exp Neuropsychol 17: 20–28

Frith CD (1992) The cognitive neuropsychology of schizophrenia. Erlbaum, Hove

Fuster JM (1989) The prefrontal cortex – anatomy, physiology, and neuropsychology of the frontal lobe. Raven, New York

Gaffan D, Heywood CA (1993) A spurious category-specific visual agnosia for living things in normal human and nonhuman primates. J Cogn Neurosci 5: 118–128

Gainotti G, Silveri MC, Villa G, Caltagirone C (1983) Drawing objects from memory in aphasia. Brain 106: 613–622

*Gainotti G, Silveri MC, Daniele A, Giustolisi L (1995) Neuroanatomical correlates of category-specific semantic disorders: a critical survey. Memory 3: 247–264

Goldenberg G (1990) Performance of concurrent non-motor tasks in Parkinson's disease. J Neurol 237: 191–196

Goldenberg G (1993) The neural basis of mental imagery. In: Kennard C (ed) Baillières clinical neurology: visual perceptual defects. Baillière Tindall, London, pp 265–286

Goldenberg G (1995a) Transient global amnesia. In: Baddeley A, Wilson BA, Watts F (eds) Handbook of memory disorders. Wiley, Chichester, pp 109–133

Goldenberg G (1995b) Aphasic patients' knowledge about the visual appearance of objects. Aphasiology 9: 50–56

**Goldenberg G (1996) Neuropsychologie – Grundlagen, Klinik, Rehabilitation. Fischer, Stuttgart

Goldenberg G, Hagmann S (1998) Tool use and mechanical problem solving in apraxia. Neuropsychologia 36: 581–589

Goldenberg G, Lang W, Podreka I, Deecke L (1990) Are cognitive deficits in Parkinson's disease caused by frontal lobe dysfunction? J Psychophysiol 4: 137–144

Gronwall D, Wrightson P (1981) Memory and information processing capacity after closed head injury. J Neurol Neurosurg Psychiatry 44: 889–895

*Halligan PW, Marshall JC (1996) Method in madness – case studies in cognitive neuropsychiatry. Psychology, Hove

Hodges JR (1995) Retrograde amnesia. In: Baddeley AD, Wilson BA, Watts FN (eds) Handbook of memory disorders. Wiley, Chichester, pp 81–108

Hodges JR, McCarthy RA (1993) Autobiographical amnesia resulting from bilateral paramedian thalamic infarction. Brain 116: 921–940

Horne JA (1988) Sleep loss and "divergent" thinking ability. Sleep 11: 528–536

Jonides J, Smith EE, Koeppe RA, Awh E, Minoshima S, Mintun MA (1993) Spatial working memory in humans as revealed by PET. Nature 363: 623–625

Kapur N (1993) Focal retrograde amnesia in neurological disease: a critical review. Cortex 29: 217–234

Kopelman MD, Christensen H, Puffett A, Stanhope N (1994) The great escape: a neuropsychological study of psychogenic amnesia. Neuropsychologia 32: 675–692

Markowitsch HJ, Calabrese P, Liess J, Haupts M, Durwen HF, Gehlen W (1993) Retrograde amnesia after traumatic injury of the fronto-temporal cortex. J Neurol Neurosurg Psychiatry 56: 988–992

McKenna P, Clare L, Baddeley AD (1995) Schizophrenia. In: Baddeley AD, Wilson BA, Watts FN (eds) Handbook of memory disorders. Wiley, Chichester, pp 271–292

Milner B (1971) Interhemispheric differences in the localization of psychological processes in man. Br Med Bull 27: 272–277

O'Connor MO, Butters N, Miliotis P, Eslinger P, Cermak LS (1992) The dissociation of anterograde and retrograde amnesia in a patient with Herpes encephalitis. J Clin Exp Neuropsychol 14: 159–178

Paulesu E, Frith DD, Frackowiak RSJ (1993) The neural correlates of the verbal component of working memory. Nature 362: 342–344

Perani D, Cappa SF, Bettinardi V, Bressi S, Gorno-Tempini M, Matarrese M, Fazio F (1995) Different neural systems for the recognition of animals and man-made tools. NeuroRep 6: 1637–1641

Robertson C, Hazlewood R, Rawson MD (1996) The effects of Parkinson's disease on the capacity to generate information randomly. Neuropsychologia 34: 1069–1078

Sacchett C, Humphreys GW (1992) Calling a squirrel a squirrel but a canoe a wigwam: a category-specific deficit for artefactual objects and body parts. Cogn Neuropsychol 9: 73–86

*Saver JL, Damasio AR (1991) Preserved access and processing of social knowledge in a patient with acquired sociopathy due to ventromedial frontal damage. Neuropsychologia 29: 1241–1250

Shallice T (1988) From neuropsychology to mental structure. Cambridge University Press, Cambridge

*Shallice T, Burgess PW (1991) Deficits in strategy application following frontal lobe damage in man. Brain 114: 727–741

Silveri MC, Daniele A, Giustolisi L, Gainotti G (1991) Dissociation between knowledge of living and nonliving things in dementia of the Alzheimer type. Neurology 41: 545–546

Smith EE, Onides J, Koeppe RA, Awh E, Schumacher EH, Minoshima S (1995) Spatial versus object working memory: PET investigations. J Cogn Neurosci 7: 337–356

Spatt J, Goldenberg G (1993) Components of random generation by normal subjects and patients with dysexecutive syndrome. Brain Cogn 23: 231–242

**Squire LR, Knowlton B, Musen G (1993) The structure and organization of memory. Ann Rev Psychol 44: 453–495

Tulving E (1985) How many memory systems are there? Am Psychol 40: 385–398

von Cramon DY, Matthes-von Cramon G (1993) Problemlösendes Denken. In: von Cramon DY, Mai N, Ziegler W (eds) Neuropsychologische Diagnostik. VCH, Weinheim, pp 123–152

Warrington EK (1975) The selective impairment of semantic memory. Q J Exp Psychol 27: 635–657

*Warrington EK, McCarthy RA (1987) Categories of knowledge – further fractionations and an attempted integration. Brain 110: 1273–1296

Warrington EK, McCarthy RA (1988) The fractionation of retrograde memory. Brain Cogn 7: 184–200

Weinberger DR (1988) Schizophrenia and the frontal lobes. Trends Neurosci 11: 367–370

Young AW, Aggleton JP, Hellawell DJ, Johnson M, Broks P, Hanley JR (1995) Face processing impairments after amygdalotomy. Brain 118: 15–24

CHAPTER 14

D. Hellhammer, U. Ehlert

Behavioural Psychology

Translator: S. Johnson

1 Definitions and Development of Theoretical Concepts

Learning and behaviour are fundamental concepts in psychology and are part of the theoretical framework on which aetiological models of various psychological disorders are based (see this volume, Part 1, Chap. 12). This has wide implications for psychological treatments in these disorders. Learning can be defined as change in behaviour based on experience (Correll 1987). Learning can occur both by trial and error and through insights gained in the course of structured processes. Behaviour refers to actual actions, which are driven by physiological processes, cognitive judgements and also motivational and emotional factors. The study of behaviour needs to take into account the full range of factors which modify it. Description and explanation of learning processes are based on the theoretical assumption that people develop theoretical concepts inductively on the basis of experiments.

At the end of the nineteenth century, Thorndike (1878–1958) set out the foundations for behavioural psychology in his treatises about the association between sensory impressions and impulses to act. His observations laid the foundation for stimulus-response psychology (see Bower and Hilgard 1983). Two models of learning are fundamental to this type of psychology: classical and operant conditioning. The Russian physiologist and Nobel prize-winner Pavlov (1849–1936) worked in the tradition known as "reflexology" and demonstrated through experiments with dogs the formation of associations between different stimuli (classical conditioning). Building on these results, the emphasis in Skinner's (1904–1990) experiments is on investigation of the connections between behavioural responses and their consequences (operant conditioning).

More recent extensions of learning theory have involved greater investigation of the mediating processes between learning and behaviour, such as motivational factors and level of performance in attention, memory and imitation. For example, these are among the concepts which form the basis of the theory of observational or social learning (Bandura 1977; Miller and Dollard 1941; Mowrer 1960). Social learning theory is based on the observation that changes in behaviour result from observing the behaviour of others. Seligman's learned helplessness model (Seligman 1975) demonstrates the influence of cognitive processes on learning processes. According to this model, experiencing uncontrollable aversive stimuli leads to passive behaviour, which is also displayed in subsequent controllable situations.

In the course of the past century, theoretical concepts explaining human behaviour have been developed on the basis of a multitude of experimental investigations. Beginning from simple stimulus-response patterns, various modulating factors have been added to the models so that they now incorporate complex patterns of behaviour such as systematic, methodical actions and problem-solving (see Ehlers 1996; Perrez and Zbinden 1996; Reinecker 1987). In this chapter, we will give a brief account of the fundamental paradigms in behavioural psychology and will describe how they have been used as a basis for models of the aetiology of psychiatric disorders and for methods of treatment.

2 Theoretical Concepts

2.1 Classical Conditioning

The concept on which classical conditioning is based is the association of pairs of stimuli. Such association formation was successfully demonstrated by Pavlov (1928) using an animal experimental model (Pavlov's dog). Classical conditioning can be encapsulated as a process through which, after repeated presentation of an unconditioned stimulus (UCS) together with a neutral stimulus, the subject – in the original instance an experimental animal – that the previously neutral stimulus is a marker of the UCS. As a result of this learning process, the previously neutral stimulus becomes a conditioned stimulus and the previously unconditioned response can be triggered by the conditioned stimulus (conditioned response).

This process may be clarified by comparison with the conditioning of immune responses (Buske-Kirschbaum et al. 1992). Under experimental conditions, a sherbet sweet (conditioned stimulus) was placed on the tongue of healthy subjects just before intravenous administration of adrenaline (unconditioned stimulus). After four combined presentations of the sherbet sweet and the adrenaline injection (unconditioned stimulus), at the fifth time point the increase in natural killer cell activity (unconditioned response) normally associated with adrenaline administration (unconditioned stimulus) could be achieved solely through the conditioned stimulus (increase in natural killer cell activity has become a conditioned response) (Fig. 1). This conditioning experiment demonstrates that an originally physiological process is clearly susceptible to influence from a psychological learning process. With this conditioning experiment, it could be demonstrated that an originally physiological process could clearly

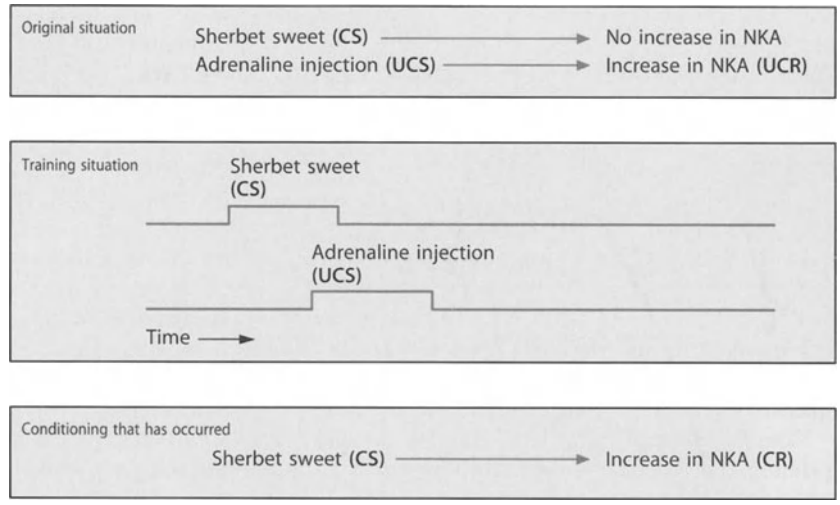


Fig. 1. Classical condition in relation to the example of natural killer cell activity

be influenced through the psychological process of learning.

The formation of conditioned responses is influenced by the motivational, cognitive and dispositional characteristics of the individual. In addition, temporal proximity between the conditioned and unconditioned stimuli is also a requirement for conditioning. Finally, the qualities of the stimuli themselves need to be taken into account. "Neutral" stimuli which already lead by themselves to clear unconditioned responses are very much less suitable for successful conditioning than stimuli not already associated with responses.

Once a successful association has been formed between a neutral and an unconditioned response, the conditioned response fades out (extinction) if the conditioned stimulus is repeatedly presented on its own. Interestingly, after this extinction phase, the conditioned response may nevertheless reappear at a further presentation of the conditioned stimulus (spontaneous return). This phenomenon should be noted, as an anxiety response which appeared to have been eliminated (e.g. through treatment of anxiety) may reappear to the patient's understandable consternation (see Schonecke 1996).

Further characteristic processes in classical conditioning are generalisation of stimuli, stimulus discrimination and higher-order conditioning. A conditioned response may be triggered not only by the conditioned stimulus, but also by another stimulus which resembles it (generalisation). In a patient with anxiety, a physiological response (e.g. sweating, increase in pulse rate) may be triggered only in lifts at the start of the illness. However, as the disorder advances, tunnels, escalators or aeroplanes (situations from which an immediate escape is not possible) may also trigger the conditioned response.

Discrimination is a process which complements generalisation; not every stimulus which resembles the conditioned stimulus triggers the conditioned response. Higher-order conditioning is the term applied to the observation that, where a stable association already exists between a conditioned and an unconditioned stimulus, the first conditioned stimulus may function as an unconditioned stimulus for a second conditioned stimulus. For example, the verbalisation of the first conditioned stimulus can become a second conditioned stimulus. For someone with a snake phobia, the word snake may trigger a conditioned response, even if he or she neither actually sees a snake (first conditioned stimulus) nor is bitten by one (unconditioned stimulus; see Reinecker 1987).

2.2 Operant Conditioning

The concept of operant or instrumental conditioning is based on Skinner's (1938) observations that while many of an organism's responses are triggered by stimuli, a variety of behaviours without observable stimuli are also displayed. The term "operant behaviour" was applied by Skinner to such behaviour, as the likelihood that it will be displayed is influenced by its consequences. Where the consequences lead to a behaviour subsequently occurring more frequently, positive reinforcement has taken place, whereas if it occurs less frequently, the behaviour has been punished. Strategies for increasing or reducing a behaviour can be derived from its positive and negative consequences. A behaviour can be increased through positive reinforcement (e.g. praise) or through discontinuation of a negative consequence (withdrawal of a punishment, also called negative reinforcement). Reduction of behaviour may be achieved through a negative consequence (punishment) or removal of a positive consequence (e.g. removing positive

Table 1. Promotion and reduction of a behaviour through operant conditioning

Promotion of a behaviour	Reduction of a behaviour
Positive reinforcement	Punishment through an aversive stimulus
Removal of an aversive stimulus (negative reinforcement)	Punishment through removal of a pleasant stimulus

conditions). The four forms of operant learning based on different reinforcers (negative or positive consequences of behaviour) are summarised in Table 1.

Consequences of behaviour may be divided into primary and secondary reinforcers. The first of these satisfy basic drives (e.g. food, sexual activity), while the second relate to conditioned (social or material) reinforcers. If a variety of consequences follow the behaviour, the likelihood of the behaviour being displayed again will be influenced most by the consequences with the closest temporal relationship to it and the greatest intensity. In experimental conditions, the reinforcement of a particular response may follow an interval schedule, occurring either after a fixed interval or after an interval of variable duration has elapsed. In ratio schedules, reinforcement occurs after a certain number of responses, with the number of responses to be made before reinforcement either fixed or variable (Skinner 1938, 1953). In real-life situations, reinforcement of behaviour usually follows variable schedules of reinforcement. A behaviour which is no longer reinforced is eventually extinguished. Responses prove to be particularly resistant to extinction if they have been conditioned through a combination of variable ratio and variable interval schedules (Ferster and Skinner 1957).

Complex patterns of behaviour are learned through shaping, i.e. through division of the behaviour into its constituent elements. Each element within the desired pattern is individually reinforced, and once all the elements have been successfully learned, reinforcement is provided only for successful combination of all of them (see Wippich 1984).

Miller's research group has shown through animal experiments that operant conditioning also applies to physiological forms of activity. For example, electric stimulation of the reward centre in the central nervous system (CNS) or punishment using electrical shocks can lead to increases or decreases in blood pressure (DiCara and Miller 1968a) or pulse rate (DiCara and Miller 1968b) in curarised rats. The target behaviour was achieved in these experiments by progressively increasing the degree of change in pulse or blood pressure required to trigger reinforcement. Even though these results could not be clearly replicated in later experiments (see Köhler 1995), the work of Miller and colleagues laid the foundations for research on biofeedback (on this subject, see Rau 1996; Rief et al. 1996).

The outstanding nature of Skinner's achievement in the description of human behaviour is evident from the techniques for functional analysis of complex behaviour which were derived from it. However, his model does not take account of the mediating processes between learning and manifest behaviour. In Skinner's view, behaviour is malleable and, if appropriate training is given, virtually every person is capable of achieving an acceptable performance in virtually every form of behaviour (see Bower and Hilgard 1983). Psychological concepts such as personality traits, insight, attribution, intrinsic motivation or motivation provided by external incentives (see below) play no part in the explanation of human behaviour following Skinner's principles. However, these concepts have been taken into account in the further development of the learning paradigms outlined above. To illustrate these further developments, we will now discuss observational learning as an example of a social learning theory and the concept of learned helplessness as an example of a cognitive theory.

2.3 Observational Learning

In many situations in life, learning can be shown to occur through imitation of the behaviour of others. In contrast to conditioned or operant learning processes, in the observational learning model (also referred to as imitative learning, vicarious learning, modelling, identification or social learning), cognitive and social aspects of the learning process are explicitly taken into account. From the theories of Bandura (1969, 1977), four components or elements may be identified in observational learning:

1. In order for a behaviour to be learnt by imitation, targeted attention from the observer is required (attention process).
2. The observed behaviour is coded in the memory as a stimulus event, symbolically represented and stored until the imitative behaviour takes place (retention process).
3. Assuming that a behaviour has been adequately observed and cognitively represented, in many cases particular motor abilities will be required for it to be imitated successfully (reproduction process).
4. Anticipation of positive reinforcement for the behaviour influences the likelihood that imitation will actually take place (motivational process).

The learning process is promoted where the model is attractive to the observer, but also similar enough to the

observer's own model so that imitation lies within his or her capabilities, i.e. the observer needs to be competent to carry out the behaviour (see Spada et al. 1990).

Social learning is directly linked to the principles of developmental psychology, in that observational learning is a basis for development of a capacity for self-regulation in human behaviour. According to Bandura (1976), there are three phases in development of this capacity. In the first phase, ways of behaving are adopted through observational learning. In the second phase, people in an individual's social environment react to these forms of behaviour as described in the operant conditioning model, with either positive or negative consequences. As a result, the individual forms expectations about future consequences of the behaviour. In the third phase, the behaviour is either continued or abandoned on the basis of these internalised expectations.

An individual's hypotheses about the consequences of actions can be influenced not only by his or her own experiences but also by the experiences of others communicated to the individual verbally. The learned helplessness model shows how assumptions about the consequences of actions (causal attributions) can lead to disturbances of behaviour.

2.4 Learned Helplessness

The concept of learned helplessness originates from the animal and human experiments of Seligman and colleagues (Hiroto and Seligman 1975; Overmier and Seligman 1967; Seligman and Hager 1975; Seligman and Maier 1967). In the human studies, the experimental means of producing helplessness involved two conditions. In the first condition, the subjects were subjected to an aversive stimulus which, regardless of their actions, they could not escape (uncontrollable and unavoidable negative situation). In the second condition, they were again subjected to a negative situation, but this situation could be terminated by a simple response (the negative situation is controllable and avoidable). Subjects who had previously experienced the first condition showed the poorest performance in learning how to end the aversive stimulus. The experience of a negative situation being uncontrollable (helplessness) leads in later controllable situations to passive behaviour. The subjects show no motivation for active coping with the situation because they have formed unfavourable assumptions about their own strategies for action when subjected to stress (Fig. 2).

According to Abramson et al. (1978), three dimensions can be distinguished in attributional styles: internal versus external, specific versus global and stable versus variable. Individuals' assessments of a helplessness-generating situation in relation to these dimensions indicate the extent of learned helplessness. Three questions may be used to indicate the nature of these assessments (see Davison and Neale 1996):

1. Does the failure have personal (internal) or environmental (external) causes?
2. Is this an instance of a specific or a general (global) inability to avoid failure?

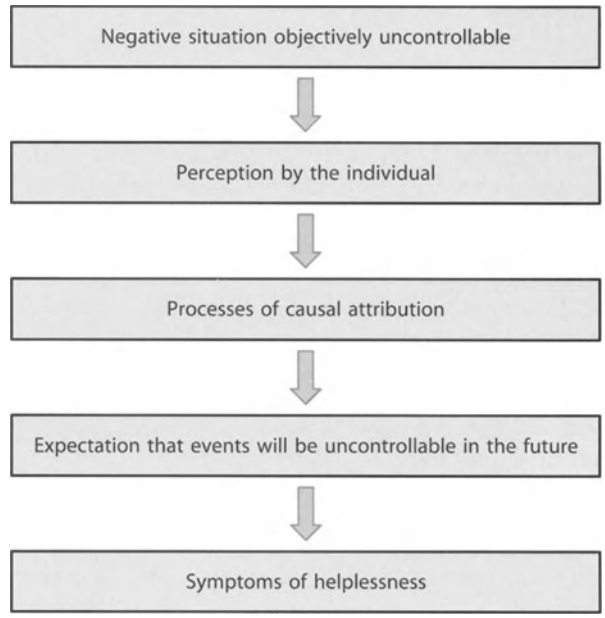


Fig. 2. Model assumptions regarding learned helplessness in humans

3. Is the failure to be seen as a transient (variable) or a lasting (stable) problem?

The least favourable combination is where the attribution made is internal, global and stable, resulting in a marked feeling of helplessness. Overmier (1988, p. 240 ff.) identifies frustration, fear, disturbances of concentration, disturbances of motor coordination, loss of appetite, raised levels of cortisol, reduced noradrenaline response, gastric ulcerations and an increased susceptibility to illness as characteristic psychological and physiological effects of learned helplessness.

Thus, according to the learned helplessness model, learning experiences can lead to the development of certain styles of attribution and emotion which are unfavourable for the individual. These are accompanied both by physiological dysfunctions and by a passive style of coping with future situations. In cases with poorer outcomes, learning experiences lead to deviations from physical and psychological homeostasis and thus to pathological states (see Overmier and Hellhammer 1988). If we accept this principle that illness may develop through learning, then presumably the inverse also applies, i.e. pathological conditions may be alleviated by changing learning experiences. In the following, we will discuss the contribution of behavioural psychology both to causal explanation and to treatment of a selection of disorder types.

3 Influence of Behavioural Psychology on Models of the Aetiology of Abnormal Behaviour and on Its Treatment

3.1 Aetiology

Concepts from behavioural psychology regarding the basis of psychological characteristics are applied with the aim of understanding the (abnormal) behaviour of a patient in terms of his or her individual responses to various situations. According to Hautzinger (1996, p. 36), individual behaviour is influenced by social learning history, situational conditions (stimuli) and the consequences of the behaviour. Thus three dimensions need to be taken into account as determinants both of healthy behaviour and of behaviours characteristic of psychological disorders:

1. Characteristics of the situation (circumstances which make it easier or more difficult to achieve a goal)
2. Characteristics of the individual (physiological characteristics, attitudes and experiences, behavioural characteristics, such as cognitive and motor skills), which lead to particular responses to the situational conditions
3. Short- and long-term consequences of the behaviour (positive or negative effects)

In a range of psychiatric and psychosomatic illnesses, detailed analysis of the patient's behaviour allows a variety of statements to be made about the disorder's origin and the treatment required.

Depending on the particular illness, behavioural principles allow explanation of part of the range of behavioural abnormalities observed. This may be understood by considering schizophrenic illnesses, drug dependence or somatoform disorders. In aspects of behaviour such as secondary gain from illness, reinforcement of the symptom behaviour also plays a decisive role. In the following, the value of learning psychology principles will be illustrated using depressive illness and anxiety disorders as examples. While the contribution of behavioural psychology in depressive illnesses needs to be placed in context of the overall framework of causal elements, concepts from learning theory are the main basis for effective treatment of anxiety.

Hautzinger and de Jong-Meyer (1990), reviewing the current status of research on depression, argue that the heterogeneity of depressive illnesses means that a satisfactory explanatory model needs to encompass multiple causes. In addition to biological factors, social, personality-related and behavioural aspects influence the occurrence and maintenance of the illness. Current research suggests that one of the explanations of depressive behaviour is in terms of an emotional state of hopelessness and a loss of or shortage of positive reinforcers.

Extending the theory of learned helplessness described above, Abramson et al. (1989) propose that the origin of some forms of depressive illness lies in hopelessness. This emotional state results from learning experiences and the cognitive judgements made subsequent to these. The assumptions are as follows: (a) desirable events do not occur, (b) unwanted events do occur and (c) this situation cannot be altered. According to the theory of learned helplessness, hopeless humans show a passive coping style in situations experienced subjectively as stressful and almost inevitably ascribe negative consequences to such situations. A general reduction in level of activity constitutes a form of passive coping.

Lewinsohn et al. (1984) established through observation of the behaviour of depressive individuals that these people obtain fewer positive reinforcers for their behaviour than non-depressive individuals in similar situations. Because of this lack of positive reinforcement, behaviours stop being manifested, resulting in a

reduction in level of activity. Because of their limited activity, these individuals obtain still fewer positive reinforcers, so that behaviours which might be favourable to mood are reduced still more. Lewinsohn's principles are much more a description of the maintaining factors for depressive behaviour than of its original causes.

Behavioural assumptions about the origins of anxiety disorders are the basis for the two-factor theory of the development of anxiety and avoidance put forward by Mowrer (1947). According to this model, there are likely to be some innate associations between particular pain–fear stimuli and pain–fear responses. If a neutral stimulus precedes such a stimulus–response pair once or more than once, then following the principle of classical conditioning, this neutral stimulus itself becomes a trigger for the pain–fear response. This response is followed by particular motor actions, which lead to a reduction in the pain–fear response. In future pain–fear responses, these actions are therefore more likely to occur, as, according to the operant conditioning model, reinforcement has been obtained in the form of avoidance of negative consequences. In order to understand the development of anxiety as a pathological condition, two assumptions behind this theory need to be made explicit (see Krohne 1982):

1. The conditioned pain–fear response can be differentiated from the unconditioned one, in that the actual pain component is missing. The conditioned response represents an anticipatory response based on expectation of a pain–fear response. On the basis of this distinction, Mowrer (1940) designates the conditioned response an anxiety response.
2. In contrast to the extinction principle for conditioned responses, anxiety responses show a high level of resistance to extinction. On the basis of animal experimental findings, Solomon and Wynne (1953) describe the development in several stages of "self-protection against anxiety". With repeated presentation of a conditioned stimulus, the time latency between presentation of the conditioned stimulus and the operationally conditioned avoidance behaviour becomes shorter and shorter. Because of the longer latency time between the conditioned stimulus and the conditioned response, this classically conditioned anxiety response can no longer occur. As this stimulus–response association does not occur, it can also not be extinguished, so that the intensity of the potential anxiety response remains unaltered. Thus individuals react so swiftly to a danger signal that they escape before they can really feel disturbed by it.

The specificity of the stimuli that trigger phobic behaviour can be explained through the concept of physiological preparedness for an anxiety response. This principle accounts for the fact that particular stimuli (e.g. spiders, rats, cats) are more likely to lead to anxiety behaviour than other stimuli (e.g. butterflies). In experimental investigations of stimulus specificity (hman et al. 1985; McNally 1987), it has been shown that in classical conditioning the conditioned response can be obtained to neutral stimuli with the same intensity as with a more uncomfortable stimulus (pictures of snakes), but that the extinction phase is clearly more prolonged with the uncomfortable than with the neutral stimuli. Finally, animal experiments have demonstrated that phobic behaviour can also be produced by learning from models (observational conditioning) (Mineka et al. 1984).

While the two-factor theory of anxiety is invoked to explain a variety of forms of anxiety disorder, the literature also contains critical discussions of the overall value of the theory in explaining the genesis of disorders (see Ehlers and Margraf 1990). For example, in relation to the development of post-traumatic stress disorder, there has been debate about whether a single episode of trauma ("one-trial learning") can really lead to such extensive psychological disturbances as intrusive thoughts, dissociative episodes of flashbacks or hypervigilance (see DSM-IV; Saß et al. 1996). In defence of the behavioural explanatory model, Quirk (1985) observes that the single appearance of a stimulus–response association certainly can lead to successful learning provided the connection is simple and the latent period short. This applies especially to a life-threatening situation, which is characterised by particular endocrinological features and a natural anxiety reaction. Although for ethical reasons, no experimental evidence has been obtained regarding these principles, Steil and Ehlers (1996) state that, whatever model is used to explain the occurrence of post-traumatic stress disorder, avoidance of stimuli associated with the trauma has a decisive role in maintaining this disorder.

Another reason for criticising the conditioning model of the aetiology of anxiety disorders is that for a significant proportion of patients, phobias or generalised anxiety disorders develop without any evidence of an initial triggering event. The reasons for this may lie either in real absence of a triggering situation or in limited ability to recall an unspectacular event associated with the first appearance of anxiety.

This can be illustrated by an example. An agoraphobic patient reported that the unpleasant physical sensations which characterise her anxiety disorder first appeared around 8 years before the time of assessment. At the time, she was attending a lecture in a large auditorium. No psychological stresses or other remarkable characteristics were associated with this situation. Only with exploration of her activities prior

to being in this situation did it emerge that around 2 h before going to the lecture, the patient had voluntarily given blood. On previous questioning, the patient had never mentioned this blood-taking, as she had not made a causal connection between this event and the dizziness, nausea and shortness of breath she experienced in the lecture theatre.

It should certainly not be concluded on the basis of a single case study that a patient with anxiety only needs to be questioned for long enough for a connection between physical and psychological discomfort and a corresponding triggering situation to be found. However, in evaluating models of anxiety disorders based on behavioural psychology, it is important not to lose sight of the importance of high-quality and extensive exploration.

Finally, there is evidence that, in addition to learning experiences, the development of an anxiety disorder is influenced by predisposing variables. These variables include lack of self-confidence (Bates 1990), cognitions, e.g. about the uncontrollability of a situation (Barlow 1988) or the "fear of fear" (Goldstein and Chambless 1978), and also as a genetically determined instability of the autonomic nervous system (Gabbay 1992). The original two-factor theory of the acquisition of anxiety and avoidance therefore needs to be extended in considering an individual's anxiety disorder to take into account personality, cognitive, physiological and genetic variables.

3.2 Treatment

Many individual methods of altering behaviour have been developed by a variety of therapists, based on the behavioural psychology theories described in Sect. 2 above. These procedures are collectively known as behaviour therapy (see Wolpe and Lazarus 1966). There are wide variations among procedures in behaviour therapy in their specificity and range of applications. Thus positive reinforcement of behaviours which promote a reduction in symptoms, for example, is used in a great variety of disorders (Blöschl 1996), whereas waking programmes are designed specifically for the treatment of enuresis (see Grosse 1993).

Behaviour therapy methods constitute an element within an overall treatment plan. While use of specific, partially standardised treatment procedures results in a clear structuring of the treatment session, the personal influence of the therapist is still of decisive importance for the success of therapy. Producing motivation for behavioural change is achieved less by applying the right techniques than by the therapist genuinely becoming involved with the client's overall welfare and being able to show the potentially negative consequences of maladaptive behaviours and to propose alternative behavioural strategies (Goldfried and Davison 1979, p. 43).

The techniques shown in Table 2 are derived directly from behavioural psychology theories.

Table 2. Behaviour therapy methods derived from learning psychology theories (based on Reinecker 1991, p. 132)

Type of technique	Method
Techniques of stimulus control	Systematic desensitisation Graded exposure Exposure and response prevention Introduction of a delay between stimulus and response Anxiety management training Flooding Paradoxical injunction
Techniques of control of consequences	Selective reinforcement of desirable responses Cue exposure Contingency contracting Token economies Aversive techniques Time out
Observational learning techniques	Observational learning in vivo Covert observational learning Provision of role models
Techniques for self-regulation	Self-monitoring Self-control Contingency control Setting up of contracts

The techniques of stimulus control enable the patient to develop a capacity for avoiding situations which are potentially problematic for him or her. The best-known technique of this kind is probably confrontation management. In techniques involving control of consequences of behaviour, selective reinforcement is used to minimise the problem behaviour and maximise the goal behaviour. The techniques of observational learning are used particularly in group sessions, e.g. in improving social skills. In relation to this, the therapist has substantial significance as a model. The aim of self-control techniques is to make the patient aware of particular ways of behaving through self-observation. Symptom management protocols are a frequently used application of self-observation. More extensive descriptions of specific techniques used in behaviour therapy can be found in the work of Linden and Hautzinger (1996) and Margraf (1996). To illustrate the management of specific clinical disorders, we will now describe a selection of techniques which are in use for the management of depressive illnesses and anxiety disorders.

3.2.1 Depression

In the behavioural treatment of depression, the main effective method based on the principles described in Sect. 3 is a systematic programme of increasing social and physical activity (a primary behavioural technique), together with modification of negative attributions and patterns of thinking (primary cognitive techniques, see Beck et al. 1992).

The first step in building up activity involves observation of activity; on this foundation, an agreement is made with the patient on a gradual increase in activities. All activities shown following this are positively reinforced and quality and quantity of activities are gradually built up. The experience of initiating actions themselves has a lasting positive impact on patients, and this generally has a positive influence on the overall course of treatment. Lack of interpersonal skills may be an obstacle to engaging in social activities. If this is so, role plays or training in communication skills may be used to modify problematic ways of behaving (see Ullrich and Ullrich de Muynck 1996).

The effectiveness of cognitive-behavioural treatment of depressive illnesses has been convincingly demonstrated in controlled trials (see Hautzinger 1993; Hollon et al. 1993).

3.2.2 Anxiety Disorders

In behavioural techniques for the treatment of anxiety, whatever subtype of anxiety disorder is present, the aim is repeated exposure of the patient to the imagined or actual anxiety-provoking stimuli. According to Birbaumer and Schmidt (1996, p. 660), such exposure to the anxiety-provoking stimulus leads to loosening of the associational link between the conditioned and the unconditioned anxiety stimuli, as the negative consequences fail to appear. At the level of cortical and subcortical processes, treatment leads to extinction of those memory connections which were associated with excessive arousal at the beginning of treatment.

Exposure treatments can be divided into prolonged exposure in the imagination, prolonged exposure in vivo and flooding. In the first two of these procedures, the patient is exposed to the anxiety-provoking situations in a graded way, either in the imagination or in reality. In flooding, the patient is immediately confronted with the situation which provokes the most severe anxiety. Accompanying exposure techniques, anxiety management techniques are used, including relaxation techniques, breathing exercises, distraction techniques, social skills training and protection against stress using positive self-instruction, self-observation and self-strengthening. The complaints of patients with a range of anxiety disorders have been alleviated by this combination of exposure and anxiety management (see Becker 1995; Ehlers and Margraf 1990; Foa et al. 1991; Steil and Ehlers 1996).

4 Outlook

In this chapter, we have discussed the central theoretical concepts in behavioural psychology. These theories form the basis for psychological explanations of behaviour. The precise analysis of human behaviour on behavioural psychological principles allows conclusions to be drawn about the causes of abnormal behaviour and is thus particularly important for aetiological deliberations regarding psychiatric disorders (such as behavioural disorders in childhood, schizophrenic illnesses, substance dependence, eating disorders, somatoform disorders and psychosomatic illnesses).

For many psychiatric illnesses, it is now widely accepted that a variety of factors have proven influences on illness development and maintenance. Taking depressive illnesses and anxiety disorders as an example, it has been shown that, in addition to biological factors, learning processes and ways of behaving that result from these influence the origin of illness. For both illnesses, behavioural treatment methods, combined with cognitive restructuring techniques, are the treatment of choice for patients who participate actively in treatment. It needs to be

emphasised that this description by way of an example of aetiological concepts and forms of treatment derived from learning psychology represents only a small excerpt from the large number of empirically grounded explanatory and treatment concepts for psychiatric and psychosomatic illnesses, and the significance of these approaches for the welfare of the patient should not be underestimated.

In future clinical research based on learning psychology, it is important that these concepts should not only be referred to when explaining psychiatric and psychosomatic illnesses, but also that they should be taken into account in the maintaining conditions for physical illnesses. In the area of care, behavioural therapy treatment procedures are to be assessed as especially effective (Grawe et al. 1994). Even so, on the grounds of health politics, the effectiveness of short-term treatments should be demonstrated more convincingly through cost-effectiveness studies. Longitudinal studies (outcomes at 1 and more years) should further be used to examine the long-term effectiveness of psychotherapeutic measures.

5 References

Abramson LY, Seligman MEP, Teasdale JD (1978) Learned helplessness in humans: critique and reformulation. J Abnorm Psychol 87: 49–74

Abramson LY, Metalsky GI, Alloy LB (1989) Hopelessness in depression: a theory-based subtype of depression. Psychol Rev 96: 358–372

Bandura A (1969) Principles of behaviour modification. Holt, New York

Bandura A (1976) Lernen am Modell. Ansätze zu einer sozialkognitiven Lerntheorie. Klett, Stuttgart

Bandura A (1977) Social learning theory. Prentice Hall, Englewood Cliffs

Barlow DH (1988) Anxiety and its disorders: the nature and treatment of anxiety and panic. Guildford, New York

Bates GW (1990) Social anxiety and self-presentation: conversational behaviors and articulated thoughts of heterosexually anxious males. Doctoral dissertation, University of Melbourne, Australia

*Beck AT, Rush AJ, Shaw BF, Emery G (1992) Kognitive Therapie der Depression, 3rd edn. Beltz, Weinheim

Becker ES (1995) Ätiologie und Therapie des Generalisierten Angstsyndroms. Verhaltenstherapie 5: 207–215

**Birbaumer N, Schmidt RF (1996) Biologische Psychologie, 3rd edn. Springer, Berlin Heidelberg New York

Blöschl L (1996) Grundlagen und therapeutisches Basisverhalten: Verstärkung. In: Linden M, Hautzinger M (eds) Verhaltenstherapie, 3rd edn. Springer, Berlin Heidelberg New York

Bower GH, Hildgard ER (1983) Theorien des Lernens I/II, 1st–5th edn. Klett-Cotta, Stuttgart

Buske-Kirschbaum A, Kirschbaum C, Stierle H, Lehnert H, Correll W (1987) Verstehen und Lernen. Grundlagen der Verhaltenspsychologie. mvg, Landsberg am Lech

Buske-Kirschbaum A, Kirschbaum C, Stierle H, Lehnert H, Hellhammer D (1992) Conditioned increase of natural killer cell activity (NKLA) in humans. Psychosom Med 54: 123–132

Correll W (1987) Verstehen und Lernen: Grundlagen der Verhaltenspsychologie. mvg, Landsberg am Lerch

**Davison GC, Neale JM (1996) Klinische Psychologie. Beltz, Weinheim

DiCara LV, Miller NE (1968a) Changes in heart-rate instrumentally learned by curarized rats as avoided responses. J Comp Physiol Psychol 65: 8–12

DiCara LV, Miller NE (1968b) Instrumental learning of systolic blood pressure responses by curarized rats: dissociation of cardiac and vascular changes. Psychosom Med 30: 489–494

Ehlers A (1996) Psychologische Grundlagen der Verhaltenstherapie. In: Margraf J (ed) Lehrbuch der Verhaltenstherapie. 1. Grundlagen, Diagnostik, Verfahren, Rahmenbedingungen. Springer, Berlin Heidelberg New York, pp 49–65

Ehlers A, Margraf J (1990) Agoraphobie und Panikanfälle. In: Reinecker H (ed) Lehrbuch der Klinischen Psychologie. Hogrefe, Göttingen, pp 73–106

Ferster CB, Skinner BF (1957) Schedules of reinforcement. Appleton-Century-Crofts, New York

*Foa EB, Rothbaum BO, Riggs DS, Murdock T (1991) Treatment of post-traumatic stress disorder in rape victims: a comparison between cognitive-behavioral procedures and counselling. J Consult Clin Psychol 47: 715–723

Gabbay FH (1992) Behavior genetic strategies in the study of emotion. Psychol Sci 3: 50–55

Goldfried MR, Davison GC (1979) Klinische Verhaltenstherapie. Springer, Berlin Heidelberg New York

Goldstein AJ, Chambless DL (1978) A reanalysis of agoraphobic behavior. Behav Ther 9: 47–59

Grawe K, Donati R, Bernauer F (1994) Psychotherapie im Wandel. Von der Konfession zur Profession. Hogrefe, Göttingen

Grosse S (1993) Enuresis. In: Steinhausen HC, von Aster M (eds) Handbuch Verhaltenstherapie und Verhaltensmedizin bei Kindern und Jugendlichen. Beltz, Weinheim

Hautzinger M (1993) Kognitive Verhaltenstherapie und Pharmakotherapie bei Depressionen: Überblick und Vergleich. Verhaltenstherapie 3: 26–34

Hautzinger M (1996) Depressionen. In: Linden M, Hautzinger M (eds) Verhaltenstherapie, 3rd edn. Springer, Berlin Heidelberg New York, pp 367–372

Hautzinger M, de Jong-Meyer R (1990) Depressionen. In: Reinecker H (ed) Lehrbuch der Klinischen Psychologie. Hogrefe, Göttingen, pp 126–165

Hiroto DS, Seligman MEP (1975) Generality of learned helplessness in man. J Pers Soc Psychol 31: 311–327

Hollon SD, Shelton RC, Davis DD (1993) Cognitive therapy for depression: conceptual issues and clinical efficacy. J Consult Clin Psychol 61: 270–275

Köhler T (1995) Psychosomatische Krankheiten. Eine Einführung in die Allgemeine und Spezielle Psychosomatische Medizin, 3rd edn. Kohlhammer, Stuttgart

Krohne HW (1982) Theorien zur Angst, 2nd edn. Kohlhammer, Stuttgart

Lewinsohn PM, Antonuccion DO, Steinmetz JL, Teri L (1984) The coping with depression course. Castalia, Eugene

*Linden M, Hautzinger M (1996) Verhaltenstherapie, 3rd edn. Springer, Berlin Heidelberg New York

*Margraf J (ed) (1996) Lehrbuch der Verhaltenstherapie. Springer, Berlin Heidelberg New York

McNally RJ (1987) Preparedness and phobias: a review. Psychol Bull 101: 283–303

Miller NE, Dollard J (1941) Social learning and imitation. Yale University Press, New Haven

Mineka S, Davidson M, Cook M, Keir R (1984) Observational conditioning of snake fear in rhesus monkeys. J Abnorm Psychol 93: 355–372

Mowrer OH (1940) Anxiety reduction and learning. J Exp Psychol 27: 497–516

Mowrer OH (1947) On the dual nature of learning – a reinterpretation of conditioning and problem-solving. Harvard Educ Rev 2/2: 102–148

Mowrer OH (1960) Learning theory and behavior. Wiley, New York

hman A, Dimberg V, Ost LG (1985) Animal and social phobias: biological constraints on learned fear responses. In: Reiss S, Bootzin R (eds) Theoretical issues in behavioral therapy. Academic, New York

Overmier JB (1988) Psychological determinants of when stressors stress. In: Hellhammer D, Florin I, Weiner H (eds) Neurobiological approaches to human disease. Huber, Toronto

Overmier JB, Hellhammer DH (1988) The learned helplessness model of human depression. In: Simon P, Soubrie P et al (eds) An inquiry into schizophrenia and depression. Animal models of psychiatric disorders, vol 2. Karger, Basel, pp 177–202

Overmier JB, Seligman MEP (1967) Effects of inescapable shock upon subsequent escape and avoidance learning. J Comp Physiol Psychol 63: 23–33

Pavlov IP (1928) Lectures on conditioned reflexes. International, New York

Perrez M, Zbinden M (1996) Lernen. In: Ehlers A, Hahlweg K (eds) Psychologische und biologische Grundlagen der klinischen Psychologie. Hogrefe, Göttingen (Enzyklopädie der Psychologie, subject area D, vol II/1)

Quirk DA (1985) Motor vehicle accidents and post-traumatic anxiety conditioning. Ontario Psychol 17: 11–18

Rau H (1996) Biofeedback. In: Margraf J (ed) Lehrbuch der Verhaltenstherapie, vol 2. Störungen, Glossar. Springer, Berlin Heidelberg New York, pp 415–422

Reinecker H (1987) Grundlagen der Verhaltenstherapie. Beltz, Weinheim

Reinecker H (1991) Verhaltenstherapeutisch orientierte Intervention. In: Perrez M, Baumann U (eds) Klinische Psychologie. 2. Intervention. Huber, Bern, pp 129–145

Rief W, Heuser J, Fichter M (1996) Biofeedback – ein therapeutischer Ansatz zwischen Begeisterung und Ablehnung. Verhaltenstherapie 6: 43–50

Saß H, Wittchen H-U, Zaudig M (1996) Diagnostisches und statistisches Manual psychischer Störungen DSM-IV. Hogrefe, Göttingen

Schonecke OW (1996) Lernpsychologische Grundlagen. In: von Uexküll T (ed) Psychosomatische Medizin, 5th edn. Urban and Schwarzenberg, Munich, pp 231–251

Seligman MEP (1971) Phobias and preparedness. Behav Ther 2: 307–320

Seligman MEP, Hager JL (1975) Helplessness. On depression, development and health. Freeman, San Francisco

Seligman MEP, Maier SF (1967) Failure to escape traumatic shock. J Exp Psychol 74: 1–9

Skinner BF (1938) The behavior of organisms. Appleton-Century-Crofts, New York

Skinner BF (1953) Science and human behavior. Macmillan, New York

Solomon RL, Wynne LC (1953) Traumatic avoidance learning: acquisition in normal dogs. Psychol Monogr 67/4: 1–19

Spada H, Ernst AM, Ketterer W (1990) Klassische und operante Konditionierung. In: Spada H (ed) Lehrbuch Allgemeine Psychologie. Huber, Bern

Steil R, Ehlers A (1996) Die Posttraumatische Belastungsstörung: Eine Übersicht. Verhaltensmod Verhaltensmed 3: 169–212

Ullrich R, Ullrich de Muynck R (1996) Aufbau sozialer Kompetenz: Sicherheitstraining, Assertiveness-Training. In: Linden M, Hautzinger M (eds) Verhaltenstherapie, 3rd edn. Springer, Berlin Heidelberg New York, pp 85–92

Wippich W (1984) Lehrbuch der angewandten Gedächtnispsychologie. Kohlhammer, Stuttgart

Wolpe J, Lazarus AA (1966) Behavior therapy techniques: a guide to the treatment of neuroses. Pergamon, London

CHAPTER 15

J. Siegrist

Sociology and Psychiatry

Translator: E. Taub

1 Introduction

The usual generic classification of the empirical sciences distinguishes the natural sciences (and, in medicine, the biological sciences) from the social, behavioral, cognitive, and cultural sciences. Economics, political science, sociology, and psychology are usually counted among the social and behavioral sciences. Sociology is that area of the social sciences concerned with the structural characteristics and developmental dynamics of human societies and socialization processes. Sociology shares certain characteristic features with all other scientific disciplines:

1. It possesses a specific set of concepts and methods with which a portion of the real world can be systematically described.
2. It contains a cumulative body of knowledge built up over decades by research work carried out in many countries on the basis of these concepts and methods.
3. It provides a number of more or less well-tested, generalizable theories by means of which societal phenomena can be explained.

Sociology arose historically out of the twofold revolution – both industrial and political (democratization) – that took place in the late eighteenth and early nineteenth centuries and was, perhaps, the most significant societal change of all times. The "crisis science" of sociology assumed the tasks of interpreting the new societal dynamics (e.g. classes and stratification, social mobility, urbanization, change in family structure) that had emerged after the collapse of the traditional social order and of making its own, informed contribution to the process of societal development.

Two aspects of the historical development of sociology are particularly significant in the context of this chapter. First, in the last 150 years, sociology has proved particularly susceptible to exploitation on behalf of normative sociopolitical programs, whether "progressive" or "conservative." Such coupling of scientific work with sociopolitical activity has split the discipline unproductively into camps and slowed its progress by blocking the development of cumulative paradigms. Political involvement has also made the process of professionalization more difficult and injured the reputation of sociology as a free and independent scientific field.

Second, since its early days as a scientific discipline, sociology has been divided into a number of subdisciplines, the so-called hyphenated sociologies. These have enriched the mother discipline – "general sociology" – in some ways, but have led mainly to a heterogeneity of theoretical approaches that now characterizes sociology as a whole. In the main, each subdiscipline is concerned with an individual aspect of society or an individual major function of social systems in developed societies, such as education, labor, economy, law, and health. The subdiscipline concerned with societal aspects of health and disease has developed historically into medical sociology (lately also called health sociology), and a major part of this subdiscipline deals with psychiatric topics (psychiatric sociology) (Cockerham 1997; Mechanic 1978; Siegrist 1995).

2 Psychiatric Sociology in Retrospect

The first epoch-making research publication in psychiatric sociology appeared just over 100 years ago: Emile Durkheim's study of societal influences on suicidal behavior in 1897 (Durkheim 1951). To explain the remarkable differences in suicide rates among different religious and occupational groups, Durkheim examined the hypothesis that the loss of binding social norms and institutions, and the lack of representation of societal values and patterns of thought in situations of crisis, very likely generates feelings of desperation that may end in suicidal behavior. On the basis of this idea, Durkheim developed a typology of suicidal behavior, in which the egoistic and anomic types of suicide are considered characteristic of modern societies. Durkheim's concept of anomie prepared the way for later psychiatric sociology. His major innovation lay in the decision to analyze suicidal behavior as a societal fact, rather than as a product of individual motivation. In methodological terms, he applied so-called ecological analysis to this problem, using aggregate rather than individual data.

Using similar methods, Faris and Dunham (1939), in the United States, demonstrated a relationship between markers of disadvantaged socioeconomic status and high rates of psychiatric disease. The classic studies in psychiatric sociology, however, were not carried out until the late 1950s and early 1960s. These included the analysis of the social component of treated psychiatric disease in New Haven by Hollingshead and Redlich (1958) (the lower the social status, the more severe the diagnosed illness), the Midtown Manhattan study (Srole et al. 1962), and the Stirling County Study (Leighton et al. 1963). The latter two studies provided a test of various sociological explanatory approaches, including a more refined concept of social anomie and sociocultural disintegration, and were the first to examine the role of critical life experiences and the significance of cumulative psychosocial stress in a life-cycle perspective. Two further studies should be mentioned in this

context: Kornhauser's analysis of the influence of occupational stress on mental health (Kornhauser 1965), and Brenner's time-series analyses of the relation between economic crises and rates of mental illness and psychiatric hospitalization (Brenner 1973).

A second line of development of this sociological subdiscipline, which has had considerably more practical impact, concerns the analysis of medical practice in psychiatry and the critical study of what happens during psychiatric hospitalization. In evaluating the achievements of earlier research today, we must be aware that institutional psychiatry was in a desolate condition before the introduction of effective medications and the coming of the reform movement in health policy, with its emphasis on deinstitutionalization. The unsparing sociological analysis of the psychiatric hospital as a "total institution" (Goffman 1961) and the convincing demonstration of "hospitalism" in chronic patients (Brown 1959) had an important signal effect on the formation of the social and community psychiatry movement of the 1960s and early 1970s. The same is true of the exemplary assessment studies of the course of illness of schizophrenic patients in a new model of care (Pasamanick et al. 1967).

The sociological analysis of the provision of psychiatric diagnoses was just as consequential in its ability to reveal deficiencies in medical practice, even though the conclusions drawn were excessively generalized. Researchers including Goffman (1963) and Scheff (1966) demonstrated the major effects of psychiatric diagnoses on the social identity of patients (stigmatization) and the negative course of social differentiation and exclusion. This work rested on theories of symbolic interactionism and on the theorem proposed by Thomas ("If a person defines a situation as real, then the consequences of that definition are also real"; Thomas 1965). The negative effects on patients were even more serious in that diagnoses were often provided without the requisite degree of care and nosological validity. Thus Scheff's analysis, for example, showed that psychiatrists (like other physicians), when faced with diagnostic uncertainty, tend to follow the rule that it is better to diagnose a healthy person as sick than to declare a sick person healthy – and this despite the very serious consequences of such a diagnosis. The far-reaching effects of this diagnostic principle were first made clear by Rosenhan's shocking report, published in the journal Science in 1973 (Rosenhan 1973). A group of pseudopatients, trained to dissimulate symptoms of insanity, were diagnosed as schizophrenic and committed to psychiatric institutions. Even though they resumed fully normal behavior immediately after hospitalization, they uniformly failed (with one exception) to convince the treating psychiatrists of the incorrectness of their original diagnosis.

This report and others lent credence to the psychiatrist Thomas Szasz' early, far-reaching and mordant critique of what he called "The Myth of Mental Illness" (Szasz 1960). He claimed that psychiatric diagnoses function primarily as metaphors for problems that are not rooted in biological psychopathology, but are rather essentially moral and social in nature.

Two aspects of this critique will be considered here. At the time Szasz' book was written, there were only very few etiologically well-founded psychiatric diagnoses, as indeed there are today, while the modern, standardized nosologic classification had not yet been widely put into practice. Because of this diagnostic near-vacuum, and the growing trend toward the "social disciplinization" and "medicalization" of deviant behavior, psychiatrists had considerable latitude in the assignment of diagnoses; there was thus considerable potential for abuse and for diagnostic errors fraught with serious consequences. A further and still current problem is that of "subdiagnostic" morbidity, e.g. in the area of depressive symptoms and anxiety disorders, and with it the problem of locating the border between mental health and illness. In fact, societal value judgments as well as economic interests exert a strong influence on physicians' determination of the threshold level that separates mere impairment of well-being from morbid functional disturbance.

In summary, psychiatric sociology went through a remarkable efflorescence in the relatively short period between 1955 and 1975, mainly in the United States and the United Kingdom, and had a major impact on health policy. It contributed to a society-wide reconsideration of the role of the physician and of diagnostic decision processes and provided an empirical foundation for some of the major ideas of new movements in social and community psychiatry. On the methodological and theoretical levels, it provided new evaluative approaches and explanatory concepts, particularly in the area of social epidemiology and follow-up studies. It may well be asked, therefore, why these fruitful developments had no comparable sequels in the two decades from the mid-1970s to the present.

3 Recent Developments

3.1 Social Determinants of Mental Illness

After the failure of empirically untenable attempts to ascribe the origin of mental illness mainly or exclusively to societal reaction processes (so-called labeling theory; for a critical account, see Robins 1975) or to disturbed interpersonal relationship patterns (Bateson

et al. 1956), a theoretically and methodologically sophisticated interdisciplinary approach gained prominence in psychiatric sociology, largely through the efforts of the British sociologist George W. Brown. He developed a comprehensive evaluative technique for the assessment of critical life events and chronic difficulties and thus gave empirical social research a new importance in psychiatry (Brown and Harris 1989). Furthermore, his group succeeded in identifying specific social determinants of affective disorders ("major depressions") in a research effort spanning 20 years (Brown and Harris 1978; Brown et al. 1990; Bifulco and Brown 1996).

The empirically supported model of affective disorders maintains that critical life events and chronic difficulties are most likely to lead to intense experiences of emotional distress, and possibly to depressive episodes, when the affected individual possesses certain vulnerability factors. The most important factors are the lack of a confidant who can provide emotional support and negative self-esteem. These two factors, when combined, lead to inappropriate coping behavior (primarily self-blaming and general helplessness). They not only reinforce each other (persons with negative self-esteem tend to blame themselves and often lose the support of those close to them), but also often share a common origin in the life history of the individual, as negative self-esteem is significantly more common among people who lost a parent in early life.

Brown's group obtained an impressive confirmation of this model in a longitudinal study of 353 women, all of them mothers living with at least one dependent child under unfavorable conditions. The occurrence of depression was best predicted by a model incorporating the two vulnerability factors mentioned, as well as negative coping style in the face of experienced social stresses. A total of 69% of women in the psychosocial high-risk group developed depression, as compared with only 9% in the group that had experienced little or no psychosocial stress (Bifulco and Brown 1996). These findings are all the more convincing, now that basic research on the physiology of stress has revealed a close relationship between emotional distress and dysregulation of the hypothalamic-pituitary-adrenal axis, in which an increased synthesis of corticotropin-releasing factor plays a major role (Axelrod and Reisine 1984).

The hypothesis that psychosocial factors influence the occurrence of depressive illnesses (particularly in women) has since been confirmed by other research groups (Bebbington et al. 1984; Parry 1986; for an overview, see Geyer 1998), but no such influence has been found for the second major group of mental illnesses, the schizophrenias. Rather, research has revealed the operation of social selection processes as a consequence of illness. This concept received solid support from an impressive longitudinal study of more than 4900 young adults in Israel (Dohrenwend et al. 1992).

The explanatory ability of sociological models in relation to the occurrence of mental illness goes further beyond the research results just cited. The past few years have witnessed intensive study of the role of morbid chronic stress in working life. Two theoretical approaches are most prominent: the so-called demand-control model, and the effort-reward imbalance model.

The demand-control model posits that occupations that place high quantitative demands (e.g. time pressure) on workers, and simultaneously offer little opportunity for control and decision-making, have long-term pathogenic effects (Karasek and Theorell 1990). In contrast, the model of effort-reward imbalance identifies chronic stress as a consequence of the lack of reciprocity between high cost spent and reward received; the model considers not only situational aspects, but also personal, intrapsychic processes, in particular a specific pattern of coping with demands termed "overcommitment" (Siegrist 1996). The model distinguishes three types of reward: wages/salary, esteem, and career opportunities/job security.

The two models were recently prospectively tested in the Whitehall II study, in which some 10,000 civil servants in London participated. The work-related stresses assessed upon each subject's entrance into the study were related to onset of psychiatric disorder over a mean period of 5.3 years, as determined by repeated administration of the General Health Questionnaire. Men who had initially suffered from effort-reward imbalance were found to be 2.6 times more likely to develop new mental illness, while the corresponding relative risk for women was 1.7. Both effects remained significant after controlling for age, socioeconomic status, and initial state of mental health (Stansfeld et al. 1999). Weaker independent effects were found for the variable "high demand" of the demand-control model, while a high level of control and good support in the workplace exerted a protective effect (also see Borrill 1996).

On the basis of these findings, we may presume that sociological research incorporating theory-driven, standardized methods of evaluation, similar to those successfully applied to chronic degenerative illnesses and cardiovascular illnesses (Bosma et al. 1998; Siegrist 1998; Theorell and Karasek 1996), can be profitably pursued in psychiatry.

3.2 Social Influences on Coping and the Course of Mental Illness

The present state of knowledge of the origins of major mental disorders, especially the schizophrenias

(Batesonet al. 1956), seems to leave little room for explanatory models derived from the social sciences, as the endogenous processes of neurotransmitter dysregulation are recognized to play a dominant etiologic role (for a summary, see Gaebel and Wölwer 1996). The situation is different, however, as regards the course of mental illnesses and the coping strategies used once they are manifest. Studies of the socioeconomic and psychosocial condition of the mentally ill after release from hospital have clearly shown that the dynamic interaction of the disease manifestations with the coping behavior of the patient and people close to him or her has a major influence on the further course of the illness and the patient's quality of life. Selected studies of the course of illness in schizophrenic patients will be used to illustrate this point briefly.

Deficiencies in social skills, such as a disturbance of nonverbal-empathic patterns of interaction and a reduced drive for the initiation of social contacts, leading to total social withdrawal in extreme cases, have been described as characteristic of the premorbid phase of schizophrenia (Goldstein et al. 1992). These deficiencies become especially important, however, in the further course of the illness, because they can reduce social competence and thereby increase the risk of a relapse. Negative styles of family interaction described by the "expressed emotions" concept (Leff and Vaughen 1984), i.e. either a disapproving style or one of extreme emotional engagement, leading to excessive solicitousness, increase the risk of relapse, at least partly because they permanently interfere with the patient's self-initiated social-communicative activities, whose proper "dosage" only he or she can determine (Kavanagh 1992).

A recent longitudinal study from Sweden is also of interest. Patients who succeed in actively seeking social support and who wish to overcome their social isolation have a markedly better prognosis than those who are satisfied with relatively undemanding interpersonal relationships (Hultman et al. 1997). Such patients have particularly severe anhedonia related to a disturbance of the dopaminergic mesolimbic-mesocortical system (see Blanchard et al. 1994).

The socioeconomic consequences of an unfavorable dynamic interaction of disease manifestations, coping strategies, and further course of illness are often severe and exacerbate the intrapsychic and interpersonal stresses already present. A comprehensive German follow-up study of more than 300 schizophrenics showed that 60% of them were single at a mean age of 35, more than half lived alone or with their parents, and one third were socially very isolated. Half of the patients became prematurely unemployed, and, after 2 years of observation, a progressive long-term economic impoverishment was observable in two thirds of the study cohort (Müller et al. 1998).

The findings just discussed are highly relevant to the design of sociotherapeutic measures, which make up an important part of an integrated treatment plan. Such measures include behavior-therapy approaches to employment, to the improvement of social skills, and to the learning of appropriate responses to illness (including actively seeking help and avoiding stressful situations), as well as systemic interventions such as family therapy and occupational rehabilitation.

The methodological and conceptual contributions of sociology and psychology to interdisciplinary studies of disease course and therapy will also be increasingly important in the future. One of the special challenges in this area is a valid operationalization of the self-assessment of mentally ill patients, several components of which are important in course-of-illness research as predictors or criterion variables (e.g. assessment of satisfaction, level of demand, understanding of illness, and social role functioning). Furthermore, the current, highly refined understanding of psychopathology will have to be complemented by a similarly differentiated assessment of social handicaps. The theoretical and methodological approaches of symbolic interactionism and phenomenological sociology may be of assistance here, as well as the more recent microsociological approaches.

3.3 Sociological Evaluation Research in Psychiatry

Sociological evaluation research in the health sector was developed in the United States in the 1960s (Suchman 1967) and provides an empirical evaluation of the structures, processes, and outcomes of medical care. It is concerned both with the analysis of in- and outpatient care, including epidemiologic determination of the number of individuals requiring treatment, and with the assessment of specific therapeutic interventions by means of clinical studies with extended outcome criteria. Sociological evaluation research in psychiatry to date has mainly been devoted to the analysis of in- and outpatient treatment methods and system-determined "patient careers" (referral processes). The resulting findings provide an estimation of the quality and efficiency of the current system of care and may thus yield useful ideas for improvement, including far-reaching innovations such as organizational models for the optimal integration of in- and outpatient care, decentralization measures involving "case management," and multiprofessional teamwork concepts in near-home care (Goldberg and Huxley 1980; Üstün and Sartorius 1995). Evaluation research in medical care is also increasingly expected to provide empirically founded arguments for rational health policy with

regard to the setting of priorities and the allocation of resources (Mechanic 1996).

Outcome evaluation research in psychiatry, in the narrow sense, has been undertaken on a large scale only very recently, understandably in health systems where economic constraints make it necessary to justify the expense of intervention. As in other medical specialties, the conviction that treatments should be maximally "evidence based" has not yet taken firm root in psychiatry. Efforts to promote this concept have run into a number of difficulties, including the limited current knowledge concerning effective treatments, the variability of courses of illness, multiple confounding factors in the manner of care and coping strategies, and, finally, the methodological problem of a valid and reliable determination of the patient's disease history and current status. These problems can be overcome only by giving adequate attention to the assessment of the treatment process itself, and not only of its outcome, because the quality of care and the organizational framework within which it is provided exert a major effect on the adequacy of coping and the course of illness.

The methodological advances and empirical findings of sociological evaluation research thus take their place in psychiatry alongside conventional clinical studies and health-economic analyses of treatment outcome.

4 Concluding Observations

The foregoing makes clear that sociology has not only directed its critical attention to the concept of mental illness and the role of the psychiatrist, thereby contributing to progress in the field, but has also, as an interdisciplinary science, made important empirical and methodological contributions to psychiatric research. These contributions may be roughly divided into three major areas, as described above: etiology; disease course and coping; and the results of treatment. We should not, however, overlook the fact that psychiatric sociology has not yet attained an appropriate level of professionalization, even in the countries where it is most widely practiced.

We will not analyze here to what extent the inadequate professionalization of psychiatric sociology is due to professional politics in sociology or to the paradigm of biological psychiatry that has acquired major economic importance over the last 20 years. There is reason for cautious optimism in the current increased effort to incorporate psychiatry into social-science aspects of public health research, whose aim is to study the health and illness of entire populations, and thereby determine the need for medical care and the development measures necessary for its provision (Schwartz et al. 1998).

There are promising developments, too, in basic research in biological psychiatry, where the vulnerability-stress model of mental illness is being further refined. Connections are being made between new discoveries of biopsychological and neurophysiological stress research on one hand and neurological and pharmacological basic research into the origins of mental illness on the other, including the interdisciplinary study of the progression of neurogenerative diseases afflicting the elderly. Thus Liu et al. (1996), for example, present novel strategies for research in Alzheimer's disease and Parkinson's disease, aimed at elucidating the possible synergistic effects of endogenous oxidative damage and stress-induced neurohumoral activation.

We may yet hope that avenues of successful interdisciplinary cooperation and knowledge transfer will be found, so that a theoretically and methodologically sound psychiatric sociology can assume its rightful place and continue to assist in the development of patient-centered, need-oriented, and effective psychiatric care.

5 References

Axelrod J, Reisine TD (1984) Stress hormones. Science 224: 452–459

Bateson G, Jackson DM, Haley J, Weakland J (1956) Towards a theory of schizophrenia. Behav Sci 1: 251–264

Bebbington P, Tennant C, Sturt E, Hurry J (1984) The domain of life events: a comparison of two techniques of description. Psychol Med 14: 219–222

Bifulco A, Brown GW (1996) Cognitive coping response to crises and onset of depression. Soc Psychiatry Psychiatr Epidemiol 31: 163–172

Blanchard JJ, Bellack AS, Mueser KT (1994) Affective and social-behavioral correlates of physical and social anhedonia in schizophrenia. J Abnorm Psychol 103: 719–728

Borrill, CC, Wall TD, West MA et al (1996) Mental health of workforce in NHS Trust, phase 1, final report. National Institute of Mental Health, Washington, DC

Bosma H, Peter R, Siegrist J, Marmot M (1998) Two alternative job stress models and the risk of coronary heart disease. Am J Public Health 88: 68–74

Brenner MH (1973) Mental illness and the economy. Harvard University Press, Cambridge

Brown GW (1959) Social factors influencing length of hospital stay of schizophrenic patients. Br Med J 2: 1300–1302

Brown GW, Harris TO (1978) Social origins of depression. Tavistock, London

Brown GW, Harris TO (eds) (1989) Life events and illness. Guilford, New York

Brown GW, Bifulco A, Andrews B (1990) Self-esteem and depression. III. Aetiological issues. Social Psychiatry Psychiatr Epidemiolo 25: 235–243

Cockerham WC (1997) Medical sociology, 7th edn. Prentice-Hall, Englewood Cliffs

Dohrenwend BP, Levav I, Shrout PE et al (1992) Socioeconomic status and psychiatric disorders: the causation-selection issue. Science 255: 946–952

Durkheim E (1951) The suicide. Free Press, New York

Faris RE, Dunham HW (1939) Mental disorders in urban areas. University of Chicago Press, Chicago

*Gaebel W, Wölwer W (1996) Affektstörungen schizophren Kranker. Kohlhammer, Stuttgart

Geyer S (1998) Belastende Lebensereignisse, Vulnerabilitätsfaktoren und die Entwicklung von Erkrankungen. Doctoral thesis, Düsseldorf University, Düsseldorf

*Goffman E (1961) Asylums. Anchor, New York

Goffman E (1963) Stigma. Prentice-Hall, Englewood Cliffs

Goldberg D, Huxley P (1980) Mental illness in the community. The pathway to psychiatric care. Tavistock, London

Goldstein MJ, Talovic SA, Nuechterlein KH et al (1992) Familieninteraktion versus individuelle Psychopathologie: sind beide Ausdruck derselben Prozesse in Familien schizophrener Erkrankungen? In: Brenner HD, Böker W (eds) Verlaufsprozesse schizophrener Erkrankungen. Huber, Bern, pp 209–219

Hollingshead AB, Redlich FC (1958) Social class and mental illness. Wiley, New York

Hultman CM, Wieselgren IM, Ohman A (1997) Relationships between social support, social coping and life events in the relapse of schizophrenic patients. Scand J Psychol 38: 3–13

Karasek RA, Theorell T (1990) Healthy work: stress, productivity, and the reconstruction of working life. Basic, New York

Kavanagh DJ (1992) Recent developments in expressed emotion and schizophrenia. Br J Psychiatry 160: 601–605

Kornhauser A (1965) The mental health of the industrial worker: a Detroit study. Wiley, New York

Leff JP, Vaughen C (1984) Expressed emotions in families. Guilford, New York

Leighton DC, Harding JS, Macklin DB et al (1963) The character of danger: psychiatric symptoms in selected communities. Basic, New York

Liu J, Shigenaga MK, Mori A, Ames BN (1996) Free radicals and neurodegenerative diseases: stress and oxidative damage. In: Packer L, Hiramatsu M, Yoshikawa T (eds) Free radicals in brain physiology and disorders. Academic, New York, pp 403–437

Mechanic D (1978) Medical sociology. Free Press, New York

Mechanic D (1996) Emerging issues in international mental health service research. Psychiatr Serv 47: 371–375

Müller P, Gaebel W, Bandelow B et al (1998) Zur sozialen Situation schizophrener Patienten. Nervenarzt 69: 204–209

Parry G (1986) Paid employment, life events, social support, and mental health in working-class mothers. J Health Soc Behav 27: 193–208

Pasamanick B, Scarpitti FR, Dinitz S (1967) Schizophrenics in the community: an experimental study in the prevention of hospitalization. Appleton, New York

Robins LN (1975) Alcoholism and labelling theory. In: Gove WR (ed) The labelling of deviance: evaluating a perspective. Halsted, New York, pp 21–38

**Rosenhan DL (1973) On being sane in unsane places. Science 179: 250–258

Scheff T (1966) Being mentally ill. Aldine, Chicago

*Schwartz FW, Badura B, Leidl R, Raspe H, Siegrist J (eds) (1998) Das Public Health Buch. Urban and Schwarzenberg, Munich

Siegrist J (1995) Medizinische Soziologie, 5th edn. Urban and Schwarzenberg, Munich

Siegrist J (1996) Soziale Krisen und Gesundheit. Hogrefe, Göttingen

Siegrist J (1998) Reciprocity in basic social exchange and health: can we reconcile person-based with population based psychosomatic research? J Psychosom Res 45: 99–105

Srole L, Langner TS, Michael ST et al (1962) Mental health in the metropolis: the Midtown Manhatten Study, vol 1. McGraw-Hill, New York

Stansfeld SA, Fuhrer R, Shipley MJ, Marmot MG (1999) Do work characteristics predict psychiatric disorder? Prospective results from Whitehall II Study. Occup Environ Med (in press)

Suchman EA (1967) Evaluation research. Russell Sage Foundation, New York

Szasz TS (1960) The myth of mental illness. Am Psychol 15: 113–118

Theorell T, Karasek RA (1996) Current issues relating to psychosocial job-strain and cardiovascular disease research. J Occup Health Psychol 1: 9–26

Thomas WI (1965) Person und Sozialverhalten. Luchterhand, Neuwied

Üstün TB, Sartorius N (1995) Mental illness in general health care. An international study. Wiley, Chichester

CHAPTER

16

M. Knapp

Economic Evaluation of Mental Health Care

Dr. Annette Donath made enormous contributions to the writing of this chapter and, but for various protocol considerations beyond her and my control, would have been a co-author. Dr. Thomas Becker and Klaudia Werth made very helpful comments on the penultimate draft of the manuscript. My thanks to them all. Some parts of this chapter are necessarily summary accounts of issues, methods or examples dealt with in more detail elsewhere (Knapp 1995).

1 Introduction

There has long been recognition of the enormous and persistent personal and social consequences of mental illness. The chronicity of disorders such as schizophrenia and dementia, and the long-term if lower-level debilitation of depression and many neuroses, have substantial impacts on cognition, health, functioning and quality of life. They can leave many sufferers and their families with constrained, deprived, often devastated lives. The care of people with mental health problems in hospital or in community settings, support for their families and the treatment of symptoms as well as clinical and social sequelae generate considerable costs for health care, social services and other agencies. Directly or indirectly, there are also cost burdens for sufferers and their families. The consequences of mental illness, in other words, can be far-reaching and expensive.

The high costs of treatment and support have generated interest in, and demand for, economic evaluations of alternative policies and practices. Although this interest (or *concern*) is relatively new in most countries, it is fairly pervasive in that many treatment modes are now the subject of economic research. It would be accurate to say that, in *all* countries of the developed world, there is widespread recognition of the need for economic insights to inform health care policies and intervention strategies, although controversy abounds as to the influence that cost considerations should have in individual treatment decisions. The growing recognition of the relevance of economics and the controversy this can engender are clearly evident from experience in countries such as Germany, the United States and the United Kingdom.

The humane demands of the German government's expert enquiry on psychiatry, Psychiatrie *Enquête*, reported in 1975, influenced the structural developments of mental health care for more than two decades (Deutscher Bundestag 1975). The *Enquête* found over-reliance on hospital-based care, few community-based services and a separation between psychiatry and general health care which often hampered the provision of good-quality treatment. The key principles recommended were to provide services for rehabilitation; to promote equal rights for somatic and mental illness; to introduce community-based, decentralised and comprehensive care for all patients; and to create active coordination and collaboration between all mental health care services. A later report, based on a comprehensive research project, *Modellprogramm des Bundes und der Länder*, made further recommendations for the organisation of community mental health care, including patient-oriented cooperation between services (Empfehlungen der Expertenkommission 1988). While it was possible to implement psychiatric reform at a time of relatively high prosperity among the various funds or sponsors, a second and ultimately dominant force appeared with the growing need for cost controls, resulting in two major health care reforms enacted by the German government in 1989 (the *Gesundheitsreformgesetz*, GRG; the Health Care Reform Act) and 1993 (the *Gesundheitstrukturgesetz*, GSG; the Health Care Structure Act). The dominant system of pursuing an expenditure-orientated income policy turned to the pursuit of an income-orientated expenditure policy. As we shall see, these developments are associated with changes to the economic structure of mental health care in Germany and clearly generate numerous demands for economic evaluations and perspectives.

Legislative reforms and other system changes in other countries have also brought about some marked changes to the balance of economic power within mental health care systems and have simultaneously generated a range of demands for economic analysis. In the United States, for example, concerns have been expressed about limitations on clinical practices imposed by certain forms of "managed care". As we shall see later, managed care has introduced a variety of arrangements – from benefit limits and cost sharing to case management – with, among other aims, the intention of reducing the costs of treating health problems. Arguably, the balance of power – clinical and economic – has shifted from the clinician (and the provider side of the market generally) to the insurance company or the state. Among the consequences has been an upsurge in the expressed needs for economic information and insights, but at exactly the time when (and in exactly the context where) many clinicians have become rather worried about constraints on their clinical freedom.

In the United Kingdom, the advent of the so-called internal market following major parliamentary legislation in 1990 brought the most fundamental changes to the health care system for more than 40 years. The previously bureaucratic or hierarchical decision-making structure was transformed into a public sector quasi-market, with most clinicians employed as providers, but primary care general practitioners acting as both providers of services and as purchasers within the "market." A new economic agenda was set by this market-inspired structure, albeit one that is still almost completely dominated by the public sector. The quite legitimate and very sensible growth in the demand for economic data – on the costs of treatment, the economic and other benefits, the incentives and

motivations which propel inter-agency relations, the structures and forces inherent within expenditure decisions and programmes – has thus been associated (at least by some people) with the politically loaded move from a welfare state model of health care to a form of health market. It was anticipated that "the money would follow the patients", tilting the balance of power towards service users and purchasers.

Many other countries have recently experienced health care reforms, although perhaps not as extensive in their scope (Klein 1995), and across the world there have, of course, been quite substantial developments in treatment approaches and modes. Each of these many changes poses questions about the role of economics: What, for example, has prompted this interest in the economics of mental illness? Where does the need for economic evaluation stem from? What modes of economic evaluation have been developed? And how have they been deployed? These are the core issues addressed in this chapter, starting with discussion of the needs and expressed demands for economic insights and then turning to the supply response: the main modes of economic evaluation and how they are conducted. Examples of completed mental health economics evaluations are then presented. The concluding section takes stock of the current and potential contributions of economics in psychiatry. A glossary of terms is provided in the Appendix.

2 Needs and Demands for Economics

2.1 Scarcity

What are the underlying needs for economic insights, and how do these manifest themselves in expressed demands for economic evaluation? Economics starts with recognition of the widespread prevalence of scarcity: there are too few resources relative to needs or demands. One should be careful not to overplay the scarcity argument or to use it as a convenient justification for unpopular forms of rationing (Light 1997), but only a fool would fail to recognise the endemic problem of scarcity. In many countries, there are shortages of psychiatrists, clinical psychologists, social workers and medication relative to the assessed needs and/or expressed wants of the population. In some parts of Britain, there are currently serious shortages of acute psychiatric in-patient beds (Johnson et al. 1997). Very rarely will individual and societal demands for health care be met fully by available supplies. In the face of scarcity, choices must be made between alternative uses of available services and resources.

Economics has given particular emphasis to understanding how such choices can be made. Indeed, as we now argue, the aim of economics in this context is to describe, evaluate, understand and guide the workings of health care systems so as to improve the efficiency with which health and other services are produced and distributed and to ensure the equitable funding of mental health services and their targeting on needs. These two criteria underpin the different modes of evaluation.

2.2 Efficiency and Equity

Efficiency and equity (justice) are examined in economic evaluations, either singly or in combination. Component criteria are effectiveness and economy (Fig. 1).

Economy is the saving of resources, and its pursuit requires detailed costs data, but the impact of lower spending upon patients or families (the outcomes of treatment) is not examined within this criterion. Effectiveness is the other side of the coin: it is conventionally defined in terms of improvements to patient and family health and quality of life, but it is also an incomplete criterion, for it pays no regard to costs. There are of course many well-developed and widely tested instruments to measure effectiveness, covering psychopathology (including measures related to specific aspects of a disorder), social and personal functioning, family and peer-group relations, and so on. Economic evaluations can (and generally *should*) base their effectiveness measures on such instruments, but a steadily increasing number are also including unidimensional ("utility") measures of users' and others' direct valuations of care processes or outcomes (see Sect. 3).

The criterion of efficiency combines the resource (cost) and effectiveness (outcome) dimensions. The pursuit of efficiency could mean reducing the cost of achieving a given level of effectiveness or improving the volume and quality of outcomes achieved with fixed budgets. Efficiency must not be used as a euphemism or excuse for "cutback", because it can sometimes be promoted by spending more, not less. Examples of efficiency-inspired evaluations will be provided later in the chapter. Efficiency is sometimes (and usually most helpfully) examined in combination with the fourth criterion of equity (justice). Allocating budgets or services so as to achieve a more equitable (fairer) distribution of access to treatments and/or a more equitable achievement of levels of health or quality of life are common aims of most health care systems. Targeting services on needs is an example of adopting an equity criterion, although even this has

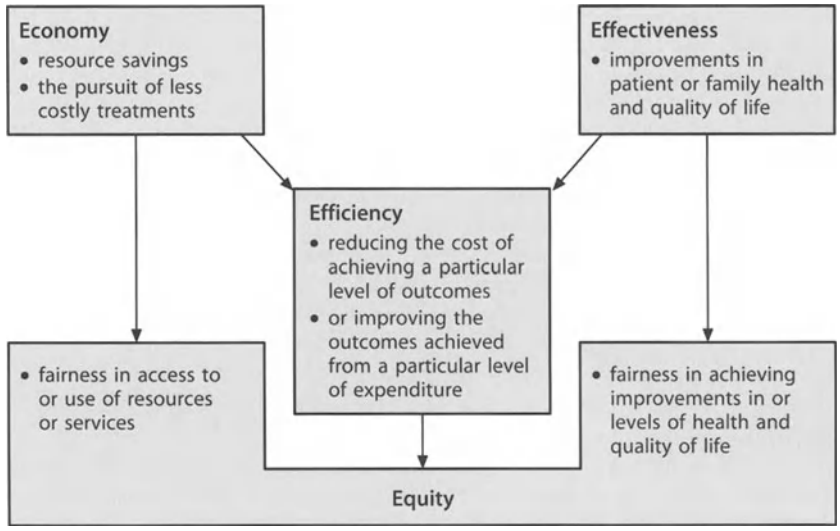


Fig. 1. Evaluative criteria

efficiency implications and interpretations, which illustrates the inter-connectedness of these important concepts and objectives.

2.3 Pressures of Growing Need

A number of factors are combining to increase the pressures of scarcity in mental health care systems, including those listed in Fig. 2. These underlying sources of scarcity, which can thus be called latent needs for economics, have fed into a number of expressed demands (manifest needs).

2.3.1 Growing Prevalence of Mental Illness

Many of the growing pressures on health and social care resources are especially important in mental health care, including the realisation that prevalence rates are higher than previously recognised and the concern that the proportion of identified need which is treated appropriately remains comparatively low. Unless there is an increase in resources, narrowing the difference between the true or underlying prevalence rate and the *treated* prevalence rate necessarily puts additional pressures on general practitioners, psychiatrists, pharmacy budget holders and others.

For example, affective disorders often go unrecognised and untreated. In the United Kingdom, the government and the main professional bodies have supported the Defeat Depression campaign – to raise awareness of the illness, among other things – even though success will undoubtedly further stretch available clinical time and budgets (at least in the short term). At the same time, health service budget holders and clinical decision makers are facing pressures from pharmaceutical companies and lobbyists to encourage wider utilisation of the selective serotonin reuptake inhibitors (SSRIs), whose market prices are much higher than those of the conventional tricyclic antidepressants (TCAs). Internationally, the apparently rising prevalence of depression in childhood and adolescence is causing further concern (Klerman and

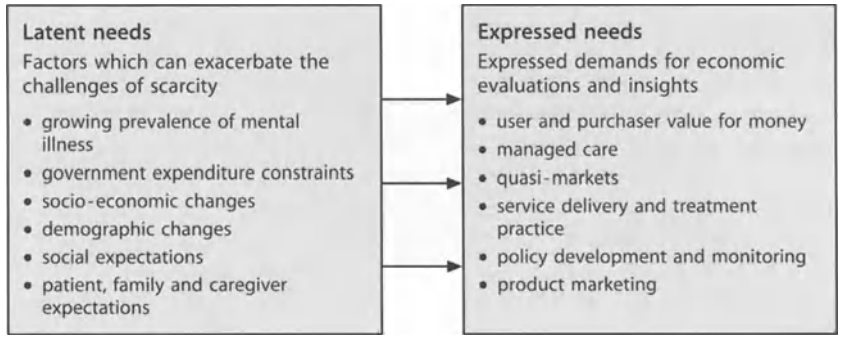


Fig. 2. Needs and demands for economics. (From Knapp 1997a)

Weissman 1989; Diekstra 1995). The biggest change in every country is the growing prevalence of Alzheimer's disease and other age-related dementias, which are already posing major resource allocation questions for national governments.

In Germany, the number of hospitalisations related to psychiatric diseases increased by 83% between 1980 and 1993, and the total number of hospital days increased by 40% over the same period. Even with the shortening of the duration of hospitalisation by about a quarter in the 1980s (Wissenschaftliches Institut der Ortskrankenkassen 1986), psychiatric patients still lead other specialties when it comes to mean duration of hospital stay (28 days in 1993; AOK-Bundesverband 1996), and mean length of stay has not shortened in recent years (Rössler and Salize 1994). This might be indicative of growth in treated prevalence (Rössler et al. 1996).

2.3.2 Government Expenditure Constraints

Growing underlying prevalence rates and the gap between underlying and treated prevalence stem in part from economic and socio-demographic trends. In countries in which a large proportion of health care is funded out of national or local taxation, the calls for economic evaluations to guide resource allocation are usually louder when the economy and/or public expenditure are under pressure. Pressures of this kind can be seen in Germany today. To understand how the current round of public expenditure cuts in Germany affects mental health care and perhaps to tease out some of the factors which have contributed to the scarcity of resources, we need to analyse the organisational and funding structures. The mixed economy of mental health care in Germany is a system of interdependent sectors:

- In-patient care and day-patient units (active psychiatric treatment) are financed by insurance companies (sickness funds).
- Custodial care is funded by patients (from their income and capital) and their relatives, complemented by funding from local authorities (*Sozialhilfeträger*).
- Out-patient medical treatment is provided by (freelance or private) psychiatrists and a few out-patient departments at psychiatric hospitals and is funded by the sickness funds.
- Most other out-patient care is not funded by health insurance but by social services agencies of local authorities (*Sozialhilfeträger*). This includes staffed hostels, sheltered homes, work rehabilitation or supported employment, drop-ins and socio-psychiatric services.
- Some work rehabilitation schemes are funded from the national insurance against unemployment (*gesetzliche Arbeitslosenversicherung*) and the pension scheme (*gesetzliche Rentenversicherung*).

The system is steered by regular negotiations between physicians' (Kassenärztliche Vereinigung, KV, i.e. the association of statutory health insurance physicians) and insurers' associations about the allocation of resources for ambulatory care, public planning for the hospitals and competition between pharmaceutical companies (both price and quality competition). One of the problems of this sectoral structure is a chronic lack of incentives both for cross-sector savings and for intra-agency savings (e.g. by shifting from in-patient to out-patient care). The GSG 1993 made doctors partly responsible for the quality of prescribed pharmaceuticals, which was apparently followed by a sharp decline in total expenditure on pharmaceuticals. As a consequence, the number of referrals to specialists and hospitals increased, as primary care physicians tried to shift the budgetary burden of complex and expensive cases. There is always the danger with these kinds of funding arrangements that they encourage supply-led decisions (fitting patients into available services) rather than needs-led decisions (shaping service responses to meet individual needs). There are also few incentives for coordinated action between different funding bodies.

Although government expenditure decisions have enormous implications for the level of available resources and the scarcity constraints faced by clinicians, it would be wrong to see scarcity as some kind of government plot or to believe that difficult resource allocation decisions could be obviated simply by spending more and more money.

2.3.3 Broader Socio-economic Changes

Increasing prevalence rates have been linked to cultural, social and economic changes. National suicide rates, for example, have been linked to rates for divorce, employment, homicide, female labour force participation, alcohol consumption and church membership, as well as demography (World Health Organization 1982; Diekstra 1995). Children in low-income families are more likely to have one or more psychiatric disorders, including conduct disorder, in later years (Offord et al. 1992). A high rate of unemployment (which is now the reality in all European countries) not only makes it harder for people with psychiatric problems to find jobs, but also increases the tax burden for the (smaller) number of people who are in employment. These are just some of the

direct economic impacts on scarcity. One of the most challenging changes for national and regional governments is population ageing. Between 1990 and 2000, the number of people aged 65–74 in Germany will have grown by around 40%. Pension payments will rise to around 18% of the gross domestic product by 2035, while pension contributions – on current trends – will remain at around 8% (The Economist 1996). The number of elderly people with dementia could increase by 50% (Beske and Kern 1995). This ageing of the population will inevitably mean some diversion of resources towards the support of elderly people, and away from younger age-groups. It could also mean increased tax burdens or higher insurance premiums for the (relatively smaller) working population.

2.3.4 Social Expectations

For a variety of reasons, the general public holds increasingly high expectations of its doctors and its health care services. There are other public concerns. The popular media in the United Kingdom, for example, have fanned concerns about seriously mentally ill people being discharged from in-patient settings to community care. These concerns have foundation, for there has been a catalogue of violent incidents in the United Kingdom perpetrated by people with psychoses, many of whom have not received adequate treatment (see, e.g. Ritchie et al. 1994), but rarely do the media report patient deaths associated with neuroleptic drugs or detention of people in secure settings when they no longer need to be there (Sayce 1995), nor do they publish many "good news" stories about successful treatments.

Patients and families (and other caregivers) are potentially important lobbyists for more and better services. The social stigma of mental illness is probably lower now than for a long time, and family pressures for higher standards of care or treatment are probably higher. The effect of new pharmacological developments on patient and family expectations can already be seen, for example, in relation to clozapine and fluoxetine: patients and families expect to have access to these new (and higher-priced) drugs partly because conventional treatments have failed (for them) and partly simply because they are new. In the circumstances, it is understandable if families question clinical or budgetary decisions which deny them access to particular drugs or courses of psychological therapy. In a study of prescription patterns for psychiatric disorders among general practitioners in Berlin, differences between the Eastern and the Western parts of the city were clear. The critical view of benzodiazepines held by the public was leading to a decrease in prescriptions of these drugs, but was correlated with a parallel increase in prescriptions of antidepressants, neuroleptics and psychotherapies (Heinrich et al. 1989; Linden and Gothe 1993).

A number of self-help groups have been formed by family members or caregivers, some of them in Germany represented by the federal organisation *Bundesverband der Angehörigen psychisch Kranker e.V.* By disseminating information on new treatment schemes, these groups can raise expectations and perhaps influence physicians' decisions. In 1992, a group of former psychiatric patients founded their own association, the *Bundesverband Psychiatrieerfahrener e.V.*, to help patients (who they call people "experienced in psychiatry") organise local groups and so defend their rights as patients. In cooperation with the *Deutsche Gesellschaft für soziale Psychiatrie*, these groups lobby to bring about further structural improvements for prevention and rehabilitation. In fact, with the GRG and the GSG, quality assurance has become a mandatory part of the health care system, and the patient perspective is an important part of outcome evaluation within quality assurance programmes. In social psychiatry research, of course, subjective patient and family criteria for the evaluation of care are of increasing interest and relevance (Priebe 1994; Lehman 1996).

2.4 Pressure of Demand for Economic Evaluations

The trends and changes summarised above generate what we might call the "latent" demands for economic data and evaluations; they explain the build-up of pressures on available resources. They are pervasive and cumulative. After a time, these latent demands are liable to become manifest: they change from underlying trends into real and expressed needs for economic insights and interpretations (Fig. 2). What triggers this change is not altogether clear, but it is associated with the recognition that economics can offer conceptual frameworks, relevant data and interpretations which can – in combination with other perspectives and data – inform, analyse and (ultimately) ease choices in the face of resource scarcity.

2.4.1 User and Purchaser Value for Money

The most fundamental concern of patients and their families is that the treatments they receive should be effective at alleviating the symptoms and/or compensating for the disabilities associated with mental illness. This, too, is the primary concern of the psychiatrist, psychotherapist, social worker or other

health care professional. However, in neither case is effectiveness likely to be the *only* concern. If patients or families are directly bearing some of the treatment cost, through co-payments in insurance-based systems or user charges in public health systems, what they pay could certainly influence what they decide to use (Brand et al. 1977; Frank and McGuire 1996). For example, the raising of co-payments and the reduction of coverage between 1978 and 1983 for U.S. federal employees and their dependents insured by Blue Cross and Blue Shield lowered both out-patient and in-patient mental health service use by children and adolescents (Padgett et al. 1993). Even when patients and families bear none of the direct costs, there could be indirect costs associated with particular treatments which could influence take-up and compliance (McPhillips and Sensky 1997). The influences of resource considerations on clinicians' decisions are discussed below.

2.4.2 Managed Care

Scarcity implies the need for choice, in particular, choice as to how available services are to be rationed between competing needs. Can these allocations be left to market forces? The moral legitimacy and efficiency of markets for health care services have been widely and heatedly debated. However, it is rare for the treatment of serious mental illness to be left exclusively to open market forces, where patients or their families must purchase treatments from profit-seeking providers, and there are numerous market intermediaries – private and social insurance schemes, brokerage, charities, and so on – and of course many governments operate "safety net" schemes to cover the pool of individuals (typically the chronically ill and low paid) without any other access to effective treatments. Need and ability to pay are often inversely correlated, and most mentally ill people have insufficient insurance coverage (Rössler and Salize 1994).

Mental health care markets are characterised by numerous difficulties. Prevailing incentive structures have been argued to lead to over-provision of certain services (especially in-patient care) and inefficient resource allocation. Other difficulties can arise because patients and families are poorly informed about their health status, health needs and the means to eradicate them. Professionals (psychologists, psychiatrists and others) act as their agents, but can patients and families judge the quality or monitor the effects of a professional's efforts? Professional self-interest or corporate profit margins may lead to negligence, in the form of failure to give proper attention to a patient, or to skimp, where professionals deliberately undertreat so as to save time or expense (Maynard et al. 1986).

"Managed care" has been the response in some insurance-based health care systems. For example, the main reason for the growth of managed care in the United States was that the health care costs falling to employers and other payers were increasing rapidly (Dorwart 1990; Frank et al. 1991). Limits were put on benefits, cost sharing was introduced and case management was promoted (Unützer and Tischler 1996). Managed care clearly raises the demand for reliable cost and cost-effectiveness evidence and, in so doing, raises practical and ethical issues, particularly around the influence which cost does and should have on clinical decision-making (Geraty et al. 1992; Sabin 1994).

This same link is evident in Germany. The German social insurance system has had to cope with a strong increase in demands for expenditure while facing a permanent decrease of revenues. In a system where health insurance coverage is financed by specific contributions calculated on the basis of gross monthly income, the overall deficit can be partly explained by reduced income from unemployed people. The two health care reforms – the GRG in 1989 regulating ambulatory care and the pharmaceutical sector, and the GSG in 1993 regulating hospital care and competition between the insurance companies – set in place current mechanisms to guide expenditure. The series of cost control interventions began in 1983 with the imposition of a "negative" pharmaceutical list, covering drugs for which the patient paid the full price (Hoffmeyer 1994). The list was further extended in 1989. Reference prices were introduced in 1989 and a drug budget for office-based physicians in 1993. To create better transparency, the standard daily rate for hospitalisations – until recently based on a mixed calculation including costs of care, hotel costs and medical treatment – was defined as a variable rate based on the hotel costs, specific rates for the different wards, *Fallpauschalen* (lump sum per patient) and *Sonderentgelten* (additional payments for specific medical services or assessments). Funding changes which are currently pending include capitation for office-based physicians, dependent on the subject area and the number of patients per practice, and more independence for the insurance companies in defining their services.

Discussion of managed care in Germany has focused on whether high-standard care with lower costs can be achieved by any of the following measures: better communication and cooperation between players in the health care system, particularly between primary and secondary care; elimination of duplication of effort and activity; incentives to encourage players to work together; quality assurance and standardised medical

practice; and provision of total health care for individual patients. In the recently introduced competition between insurance companies, insurers will not succeed in offering different contribution rates because of the *Risikostrukturausgleich* (risk-based fiscal equalisation, i.e. a system to balance out the expenditure and revenues among the individual regional health insurance companies; the amount to be transferred is calculated according to the individual "risk" of the insured, i.e. income, sex, age, number of family members who do not have to pay contributions), and they are unlikely to compete in terms of the range of provision because of consolidated financing mechanisms, but they probably *will* compete in terms of the structure of care they can offer. Encouraged by changes in the law, the major insurance companies have started evaluating new structural models of health care. Two pilot projects, based in Berlin and Baden-Württemberg, pursue similar strategies to managed care: both models – "networked practices" (combined budgets) and the *Hausarztmodell* (increased primary care) – have been developed to strengthen the out-patient sector by improving information flows between individual practices and funds. However, there are complications in Germany because of the different methods of funding and provision; the high level of government involvement; fragmented, uncoordinated health care delivery; and lack of, or prohibited access to, data on patient needs and outcomes.

2.4.3 Quasi-Markets

Expressed demands for economic evidence can also be seen in those countries which are moving away from a "socialised", welfare state or publicly dominated health care system to a mixed economy of market-like arrangements. A recent example is provided by the U.K. National Health Service (NHS), which moved from a "command and control" hierarchy to a set of internal or quasi-markets following the 1990 NHS and Community Care Act. It was intended that patterns of provision and use would be determined locally and – in principle – with greater opportunities for patient influence (for an excellent review, see Klein 1995). In this U.K. context, "managed care" developments have a somewhat different meaning and raise rather different issues both for the professional and the care system. Economic data are needed to inform, structure and monitor the contracts drawn up between purchasers and providers of services. In fact, cost awareness has been seen as an ethical duty for doctors in the British NHS for some time; a much-quoted British Medical Association (1988) statement, for example, included the principle that "it is the doctor's ethical duty to use the most economic and efficacious treatment available."

2.4.4 Service Delivery and Practice

The growth of forms of managed care in the United States, structural changes in Germany and the separation of public sector purchasers and providers in the United Kingdom offer examples of how economic considerations have entered the clinical domain at the practice level. One reaction to such constraints is to complain about and campaign against the collection and dissemination of economic evidence and to argue that a doctor's ethical duty is to achieve the greatest possible health gain for patients.

A more constructive and realistic view is to recognise that the scarce skills of general practitioners, psychiatrists and other mental health professionals should be used to best effect, and achieving that means taking account of the resource consequences of clinical decisions. In turn, this means gathering evidence on how available resources are deployed and with what effects. In other words, there is justifiably a practice-level need for economics. Part of that justification is in fact the argument that it would be unethical to take decisions without cognisance of the cost consequences, as to do so is potentially to waste scarce resources which might be better deployed so as to improve the health and quality of life of patients by greater amounts. Such pragmatic action does not mean that psychiatrists and others should not be lobbying for higher levels of government or other expenditure, but there is still a need to recognise the value of practice-level utilisation so as to make more effective and cost-effective use of those resources currently committed to the field.

In the United Kingdom today, for example, and also in countries such as Sweden, Canada, Australia and the United States, new treatments, new care settings and new ways of organising service systems must demonstrate not only their effectiveness in improving patient health and quality of life, but also their cost-effectiveness (see Sect. 2.4.6). In these countries, many of the major new practice innovations of recent years have been subjected to economic as well as clinical and social evaluation. Clinical decision-makers – and also case managers and keyworkers in social welfare settings – understand that they will be in a stronger position to make more effective placements and/or to offer better advice to their patients and families if they possess reliable data on needs, outcomes, costs and the links between them. The rapid growth of support for evidence-based medicine has had its economic dimension, as evidenced by the first mental health output from the Cochrane Collaboration on family interventions (Mari and Streiner 1996).

2.4.5 Policy Development and Monitoring

Cost-effectiveness demands at a macro-level in Germany are associated with the development and monitoring of both government policy and fund givers (*Leistungsträger*) policy. Governments and *Leistungsträger* require information about the resources needed nationally to care for people with mental health problems and the targeting of services on needs.

Recent changes, sometimes sweeping reforms, to health care systems in a number of European countries have been driven by concerns about rising costs, unknown outcomes, wide variations in assessment methods and treatment modes, inequitable allocations and perverse incentives. Health economics research which meets broader policy demands can therefore hope to inform and assist policy makers by uncovering the behaviours, constraints and motivating forces of different agencies or individuals. For example, there are evaluations under way in the United Kingdom which are looking at the new primary care arrangements (primary care groups), the move from hospital-based to community-based care systems and the prevalence of psychiatric problems among the growing number of homeless people. The illustrative studies in Sect. 5 offer further examples of policy-relevant economics research.

2.4.6 Product Marketing

A final source of expressed demand for cost-effectiveness and other economic evaluations is the pharmaceutical industry. Insurance companies, governments and individual pharmacy budget-holders are concerned about the rising prices of medications, especially with the arrival of the SSRIs and atypical antipsychotics. They want to know whether these high-priced medicines can be justified by resource savings from reduced hospitalisation and lower long-term costs and/or by improved patient outcomes (see Sect. 5).

At the same time, pharmaceutical companies naturally want to promote their products in both economic and clinical terms. Regulatory and funding bodies in some countries want similar evidence. In Australia, evidence of economic analysis is required before a new product can be accepted for government subsidy (Commonwealth of Australia 1990), and there are national guidelines and provincial regulations concerning the economic evaluation of new pharmaceutical products in Canada (Drummond 1992). Increasing attention to cost-effectiveness arguments (but without formal guidelines) can be seen in Italy and the United Kingdom. The same kinds of questions need to be asked of psychotherapy (Tillett 1996), but psychological therapies generally lack "product champions" to fight their corner, which might disadvantage them vis-à-vis pharmacological interventions if governments, purchasing authorities or clinicians become more cost-conscious.

3 Economic Evaluations

3.1 The Supply Response

The response from economists to these latent needs and expressed demands for economic evidence has been relatively limited. There are still comparatively few completed economic evaluations outside the United States, although many such evaluations are now under way in Europe. There are many reasons for this low supply response: a lack of mental health expertise and interest among economists, a dearth of suitable and relevant data, a lack of incentives for service providers and purchasers to seek economic insights, and few moves by clinical research teams to recruit an economist or add an economic dimension. There has also been the problem that policy-makers or decision-makers at various tiers in key agencies do not believe or recognise economic problems as being researchable or, by contrast, that they are so anxious to acquire cost data that they grasp unselectively at the first evidence that comes along, no matter how poor the quality (Knapp 1997b). As policy-makers have become more discerning, the problem has been that the supply of health economics expertise has not been able to keep pace with the demand.

3.2 Modes of Evaluation

The three most useful modes of economic evaluation are cost-effectiveness (CEA), cost-benefit (CBA) and cost-utility analyses (CUA) (Table 1; for a full account of principles, see Drummond et al. 1987; for an account of these evaluative modes in mental health applications, see Kavanagh and Stewart 1995). These modes have some common elements, but differ in two main respects: they measure outcomes differently, and consequently they address slightly different policy or practice questions. Other evaluative types are cost-offset and cost-minimisation analyses, but these do not measure outcomes in terms of the impacts on patients or families, and thus are (generally) less helpful.

CEA is concerned with ensuring that resources allocated to the mental health sector are used to

Table 1. Modes of economic evaluation

Mode of evaluation	Cost measurement	Outcome measurement
Cost-minimisation, cost-offset	Comprehensive	Not measured
Cost-effectiveness	Comprehensive	Clinical scales
Cost-benefit	Comprehensive	Monetary values
Cost-utility	Comprehensive	Utility (e.g. QALYs)

QALY, quality-adjusted life-years.

maximum effect. It is usually employed to help decision-makers choose between alternative interventions available to or aimed at specific population groups: If two treatments are of equal cost, which option has the greater effectiveness from a given budget? Or if two options have been found to be equally beneficial in terms of health outcomes (say), which is less costly? A CEA seeks to measure costs comprehensively (see Sect. 4) and measures outcomes along the dimensions usually distinguished by clinical researchers – such as symptomatology, behaviour, social role functioning, social networks, quality of life – and using the same clinical scales.

CBA seeks to show whether a particular course of action, such as a mental health treatment programme, is socially worthwhile. All costs and benefits are valued in the same units – usually monetary units – and can thus be directly compared: if benefits exceed costs, the evaluation would recommend the policy or project, and vice versa. The simple comparison of "costs incurred" with "costs saved" is not a CBA but what economists now call a "cost-offset analysis". Conducting proper cost-benefit analyses is particularly difficult because some very valid outcome indicators are not easily expressed in monetary terms. Recent methodological developments have sought direct measures of the values placed on outcomes (or care processes) by patients, families or others, perhaps using either willingness-to-pay methods (McGuire et al. 1988) or conjoint analysis (Ryan 1996).

The newest mode of economic evaluation, CUA, is similar to CEA with the important exception that it measures and then values the impact of an intervention as well as the cost of achieving that improvement. The value of health improvement from a treatment is measured in conflated units of "utility", in contrast to CBA, which uses monetary values. CUAs avoid the potential ambiguities with multi-dimensional outcomes in CEAs and can be applied to choices across a range of policies or practices for different target groups. Most progress with CUAs appears to have been made in health economics, where the most common value measure is the quality-adjusted life-year (QALY), which seeks to measure both the number of life-years gained following a health care intervention and the quality of those life-years (see below). Interesting alternatives are the disability-adjusted life-year (DALY) measure, pioneered by the World Bank (1993), and healthy year equivalents (HYEs; Mehrez and Gafni 1989).

These different modes of evaluation have a common aim in their approach to cost measurement. If a societal perspective is adopted, which is usually the case, this common aim is to range widely across all direct and indirect costs, no matter to what agencies or individuals they fall. The main evaluative modes obviously differ with respect to their measurement of outcomes. CEAs can rely on clinical outcome scales – which can make it easier to conduct an economic evaluation, especially if it is alongside a clinical trial – but cannot provide such powerful or widely applicable answers as CBAs or CUAs (Knapp 1995).

4 How to Conduct an Economic Evaluation

Generally, and for ease of execution, an economic evaluation should have the same broad design (criteria for eligibility, intervention modes, quasi-experimental approach, sample sizes) as the design chosen for the main research evaluation. It should employ the same data collection points and should span a similar period of time. There are unlikely to be strong reasons for adopting a different design in relation to each of these methodological considerations, except for sample size. In order to achieve enough power to test some economic hypotheses, a larger sample might be required, but currently there is little evidence upon which to base this argument (Gray et al. 1997).

It is helpful to distinguish six main stages to an economic evaluation:

1. Defining the alternative treatments or services to be evaluated
2. Identifying (listing) the costs and outcomes to be covered by the evaluation

3. Quantifying and valuing the costs and outcomes for each alternative
4. Comparing costs and outcomes between alternatives
5. Conducting sensitivity analyses and perhaps qualifying or revising findings
6. Examining the distributional implications and drawing conclusions

These stages are described in more detail in Knapp (1984) and in Knapp and Beecham (1996).

The first task in any evaluation – economic or clinical – must be to clarify the question to be addressed: What is the purpose of the study? All evaluations are comparative, whether they compare two or more alternative policies or modes of intervention, two or more groups of people or localities, or a single group of people before and after an intervention. It is obviously imperative from the outset that the alternatives being compared are agreed. At this point, therefore, the decision has to be taken as to the mode of evaluation to be employed.

The second evaluative stage is to draw up a comprehensive list of all relevant costs (regardless of the source of funding) and of all relevant outcomes or dimensions of effectiveness (and not just those relating directly to the mental health of the patient). These comprehensive lists are needed even if it is expected that some dimensions need not be measured (because a narrower perspective is to be adopted and/or because it is simply unnecessary to collect more data than can be used) or might subsequently prove impossible to measure in practice. In this way, the full breadth of the possible resource and effectiveness impacts of the alternative treatments can be drawn to the attention of decision-makers, agencies in service delivery networks, researchers and others, even if there are no data for some of those impacts. For example, very few completed evaluations have been able to obtain adequate measures of the costs borne by families, yet a substantial number of mentally ill people are financially dependent on their families (Bundesministerium für Jugend, Familie, Frauen und Gesundheit 1986).

The third stage in an economic evaluation is the measurement or quantification of each of the costs and outcomes. On the outcome or effectiveness side of the evaluation, the research task will clearly depend on the evaluative technique to be employed. Clinical scales for assessing outcomes are discussed in other chapters of this volume and would form the basis of a CEA. If a CBA is to be undertaken, these effects or outcomes would also need to be valued in monetary terms. We have already noted how difficult it would be to attach monetary values to most mental health outcomes, although some studies have used employment earnings as a (single and generally inadequate) benefit indicator. In a CUA, a single, uni-dimensional indicator of "utility" would need to be constructed. The most commonly used such measure is the QALY, but the instruments currently used for QALY assessment and comparison across specialties and diagnostic groups are largely irrelevant for mental health research (Chisholm et al. 1997). Recent advances in research on people with schizophrenia might eventually help (Revicki et al. 1996).

Cost measurement is probably easier, but unfortunately still not straightforward. Cost measures are most usefully based on service utilisation data obtained for each sample member. Service utilisation data can be collected using an instrument such as the Client Service Receipt Inventory (CSRI) (Beecham and Knapp 1992). Costs are then attached to each service or element of support in turn, using the best available estimates of long-run marginal opportunity cost. In this procedure, "marginal" refers to the addition to total cost attributable to the inclusion of one more user, and "opportunity cost" refers to the opportunities forgone by not using a resource in its best alternative use, which is the conventional economics definition. The short-run average revenue cost plus appropriately measured capital and overhead elements is usually likely to be close enough to the long-run marginal cost for most services, and it is conventionally taken as the cost estimate for empirical work in economic evaluations. (In the United Kingdom, there is an excellent annual compendium of many such unit cost measures, ranging across the most important services; Netten and Dennett 1997.)

The fourth stage is to compare the costs and outcomes. Comparisons between the alternatives being evaluated are made in relation to both costs and outcomes after whatever summation is possible. (In cost-benefit studies, the costs of a treatment or policy can be compared directly with its benefits, and cost-benefit differences or ratios can then be compared between alternative treatments or policies to see which offers the greatest net benefits to the health service or to society.) Difficulties can arise when costs and outcomes are distributed over time in different patterns: for example, costs may be incurred early, but outcomes may not accrue until rather later. The usual procedure is to discount future costs and outcomes back to a present value before making comparisons, using a suitable procedure to weight future costs and outcomes less highly than current costs and outcomes.

Most commonly in a CEA, average costs are calculated for each of the samples and compared with the outcomes. The option with the lowest cost per given level of outcome is then deemed to be the most desirable on the grounds of efficiency in the use of resources. Some CEAs rely on a single (dominant) outcome – perhaps the mortality rate or a measure of

the improvement in positive symptoms – but psychiatrists have generally found it more relevant (as a guide to policy or practice) to use multi-dimensional outcome measures. This rules out the use of a simple ratio measure to indicate comparative efficiency: reductionism could well misdirect decision-making. Complications will then arise if there is improvement along some outcome dimensions but deterioration along others. Complications will also arise if the cost and outcome comparisons point to different preferred solutions: treatment A may be more effective than treatment B, but also more expensive. In these circumstances, the role of the economic evaluator is to present all relevant findings to those policy-makers or practitioners whose task it is to make decisions. In turn, they will need to weigh up the clinical, social or political importance of those findings.

The fifth stage is to examine how sensitive the findings are to the assumptions that inevitably have to be made in the course of any evaluation. For example, costs may have been estimated with some error, or outcomes may have been weighted on the basis of one particular, but not generalisable perspective. Assumptions and perspectives should therefore be altered to see whether the conclusions change. This fifth stage thus provides the opportunity to check the assumptions made in all previous calculations to allow for possible error or bias.

The last stage is to draw conclusions from the evaluation. If, as is usually the case, the primary purpose of an economic evaluation is to examine efficiency, this final stage should also explore to what extent the distribution of the burden of costs and the enjoyment of outcomes which would follow if each of the alternative treatments or services were implemented is equitable or fair. One treatment may be the more cost-effective (efficient) of two alternatives, but it may leave patients or their families bearing a much higher cost burden. If this treatment were to be more widely adopted, it might therefore be necessary to find a way to compensate patients and families for their extra costs, not only for reasons of fairness but because these costs may create a disincentive for them to continue with treatment.

5
Examples of Economic Evaluations

The main types of economic evaluation will now be illustrated by describing a small number of completed studies. In principle, economic evaluations could address a wide range of questions about mental health policy and practice, but to keep the discussion manageable the section will focus on just two topics: (1) antidepressant drug therapies and (2) community care as an alternative to hospital treatment. Together, these allow us to look at each of the different evaluative modes and different research designs.

5.1
Antidepressant Drug Therapy

The treatment of depression has changed quite markedly since the launch of the first SSRI (fluoxetine) in Belgium in 1986. Since then, other SSRIs have been developed and marketed, usually at a price which is considerably higher than the typical prices of the TCAs. Two of the most pressing questions in the treatment of depression currently therefore concern relative effectiveness and cost. Are the higher prices of the SSRIs outweighed by either improved symptoms or better quality of life for patients or their caregivers? And are the overall treatment and support costs reduced either now or in the future? Neither question is easy to answer, yet without answers it would be premature either to support or to block the wider use of the new drugs on economic grounds (Hotopf et al. 1996). A new treatment mode should not necessarily have to demonstrate that it is less costly than alternative treatments in order to be cost-effective, but it must demonstrate that any increase in costs (compared to its closest alternative) is at least matched by increased effectiveness.

The economic questions raised by the SSRIs are certainly not more important than the clinical questions, but they are definitely pertinent. One reason is because depression is such a costly illness; for example, a few years ago, the direct and indirect costs were estimated at $44 billion annually in the United States (Rice et al. 1990). Another reason is that a high proportion of total expenditure on drugs in most countries is already accounted for by antidepressants (over 3% in England in 1991; National Health Service Executive 1996) even before the more widespread use of the higher-priced SSRIs. Although hospitalisation is costly in both the United Kingdom and the United States, these high "total burden" costs arise for a host of reasons, and similarly large figures would be found for other countries.

Four completed studies of treatment with fluoxetine illustrate a range of economic evaluative types and research designs (Table 2). Sclar et al. (1994, 1995) conducted two observational studies in an American health maintenance organisation (HMO), using multivariate statistical methods to interrogate a large retrospective data set. In the first study, they compared three TCAs (amitriptyline, nortriptyline and desipramine) with fluoxetine ($n = 555$ patients with depression); in the second, they compared three SSRIs (paroxetine, sertraline and fluoxetine; $n = 744$). They

Table 2. Economic evaluations of fluoxetine

Study	Research design	Evaluation mode	Treatments evaluated
Sclar et al. 1994, 1995	Retrospective, observational studies	Cost-offset analyses	Fluoxetine versus amitriptyline, nortriptyline and desipramine
Le Pen et al. 1994	Decision (tree) model	Cost-offset analysis	Fluoxetine versus variety of TCAs
Revicki et al. 1995	Decision (tree) model	Cost-utility analysis	Fluxetine versus nefazodone and imipramine
Simon et al. 1996	Prospective, randomised trial	Cost-effectiveness analysis	Fluoxetine versus desipramine and imipramine

TCA, tricyclic antidepressant.

were only able to compare costs, not outcomes (and the costs were not measured comprehensively), so that their evaluation fell into the category of a cost-offset analysis. There were other methodological weaknesses, but both studies at least endeavoured to avoid some of the biases which can sometimes arise with prospective trials. They found that the costs of all health care services used (including drug costs) in the year after initiating treatment were lower for fluoxetine than for the other drugs, partly because patients for whom fluoxetine was prescribed appeared to need dosage titration to a lesser extent.

Le Pen et al. (1994) and Revicki et al. (1995) both employed clinical decision models ("decision trees") built partly on trial data and partly on Delphi panel estimates of parameters for which no "real world" data could be obtained. Le Pen and colleagues pooled data from a number of short-term, randomised trials comparing fluoxetine with a variety of TCAs. Although the title of their paper suggests a cost-benefit study, in fact they merely compared costs incurred with costs saved, so that this is again a cost-offset analysis. They found that fluoxetine was potentially cost saving in France (through improved tolerance), so that higher-priced SSRIs might nevertheless bring about short-term financial savings. The authors are cautious about their methods and their findings – and clinical decision models are notoriously prone to error and bias – so that their deliberately conservative assumptions are sensible. Revicki et al. (1995) constructed their decision tree for Ontario, Canada, again using data pooled from completed randomised trials, estimates in the published literature and a Delphi panel. In this case, however, they sought to make cost-utility and not just cost comparisons (between fluoxetine, nefazodone and imipramine) over a 20-year period. They concluded that nefazodone is a cost-effective alternative to imipramine, but they found no difference between nefazodone and fluoxetine. Of course, decision trees of the kind reported by Le Pen and Revicki and their co-workers are not peculiar to economic evaluations – indeed many health economists are rather suspicious of them (Sheldon 1996) – but they often provide a starting point for discussion in the absence of better (prospective) data.

For fluoxetine, better data are now emerging from a large prospective, randomised trial underway in the United States. Simon et al. (1996) reported the results for the first 6 months of a 24-month trial to compare the costs and effects of desipramine, imipramine (both TCAs) and fluoxetine for a sample of 536 HMO patients. In common with the earlier studies, Simon and colleagues failed to measure costs comprehensively, but they employed a number of robust clinical and social outcome measures. They used these in a CEA. They found that fluoxetine patients had fewer medication changes and fewer side-effects, higher drug costs and lower in-patient hospitalisation costs. However, there were no differences between groups on the main clinical measures of depression, symptoms or quality of life, or on total costs. They therefore concluded that the higher drug costs associated with this particular SSRI did not lead to higher total costs in the short term, but equally it did not improve outcomes. However, patients suffered fewer side-effects, which probably explained the smaller number of medication changes.

Drug trials are well made for economic evaluations, and there are now many CEAs, CUAs and cost-offset analyses being conducted alongside prospective, randomised trials which should considerably help physicians and other decision-makers to make better use of the finite resources at their disposal. In other contexts – as we now illustrate in relation to hospital and community care – the design, implementation and interpretation of economic evaluations is not always so straightforward.

5.2 Hospital and Community Care

One of the major concerns expressed in the *Psychiatrie Enquête* was the heavy reliance on treatment in the

large, remote state mental hospitals (Deutscher Bundestag 1975). Among the report's recommendations was the development of community-based care tailored to individual needs. Between 1970 and 1988, the number of psychiatric beds in West Germany fell by 29%, with a more rapid decrease in the specialist mental hospitals and increases in the numbers of beds in psychiatric departments of general hospitals and in university hospitals (Rössler and Salize 1994). The decline and relocation of in-patient places is likely to continue, especially in the former GDR, where the system of treatment was dominated by this form of institutional care (Rössler et al. 1996). In fact, community-based provision is the preferred locus of care in most developed countries.

It has been difficult to gather evidence for Germany to shed light on the patient outcome and cost effects of shifting the locus of care from the hospital to the community, partly because of data protection issues (but see the ongoing study by Priebe et al. 1996). The study by Häfner and an der Heiden (1989) is one of the few to offer insights. (Sadly, data protection considerations forced the closure of the case register on which their study was based, and consequently the termination of this research study.) For 148 people with schizophrenia supported in community settings in Mannheim, they found that costs were generally considerably lower than continued hospital care. For only eight patients were hospital costs lower than community costs. Hess et al. (1989) had found similar cost differences in Bern. Salize and Rössler (1995) studied a smaller sample of Mannheim patients, each of whom had a diagnosis of schizophrenia and who had been discharged to the community from hospital, and reached the same conclusion: community care costs were markedly lower than hospital costs. Häfner and an der Heiden (1989) found that people living in sheltered accommodation derived more benefit from out-patient treatment than those living with their families or those living alone. Salize and Rössler took this question further, looking at cost predictors and cost-outcome linkages. Statistical interrogations of this kind can be very informative.

It has been easier to conduct economic research on the hospital/community balance in the United Kingdom and the United States. In the United Kingdom, for example, the number of people resident as in-patients under the psychiatric specialty today is less than one third of that 30 years ago, prompting many small and a few large studies (Fig. 3). Viewed historically, the hospital rundown policy in the United Kingdom has been part medical preference, part public opinion and part "hard political economy" (Korman and Glennerster 1990). Thus, although the rundown of the old psychiatric asylums was consistent with most psychiatrists' recognition that many people with long-term mental health problems could be successfully rehabilitated in community settings, economics also played its part in the reprovision or rehabilitation process. This alone justifies economic evaluation to check whether the *assumed* advantages of community care are in fact correct.

The most comprehensive U.K. study has been conducted by the Team for the Assessment of Psychiatric Services (TAPS) in North London, with hospital assessments, 1-year and 5-year outcomes after relocation to the community and a large economic element (Leff 1997). The clinical and social outcomes for former long-stay in-patients after 1 year in the community were at least as good as their "matched" counterparts who had remained in hospital. Moreover, their mental state and social disabilities were stable: the leavers had more diverse social networks (but fewer contacts with relatives), lived under less restrictive conditions and preferred life in the community (Leff et al. 1996). Of the 737 discharged long-stay patients (without dementia), 24 had died within the first year in the community, and virtually no one else was "lost" to services or to the research.

For most of these people leaving hospital for community residence, the costs of care were substantially lower (Beecham et al. 1997), although for the full

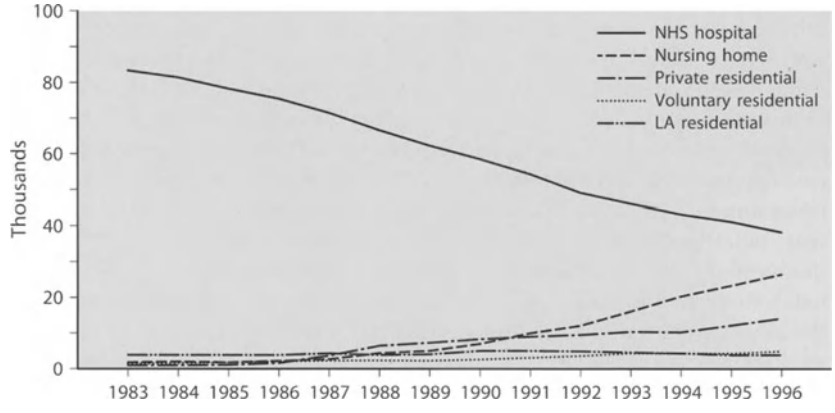


Fig. 3. Changing balance of mental health accommodation in England. *LA*, local authority

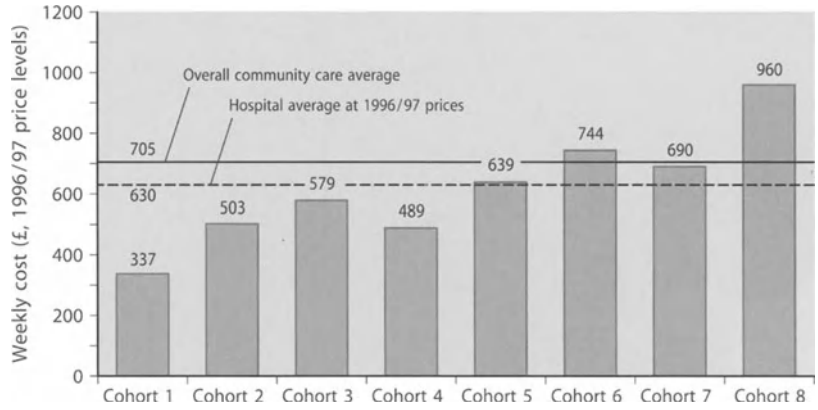


Fig. 4. Hospital and community care costs

hospital long-stay population there were no overall cost savings (Fig. 4). Five-year follow-up data suggest that costs increase slightly in the longer term. There was a clear link between costs and outcomes. For the early leavers, higher community care costs were associated with greater improvements in health and welfare (Beecham et al. 1991). In particular, improvements in negative symptoms, delusions and hallucinations, social networks (broadening) and the general need for care (from an index of physical health) were all associated with higher costs. Thus the hospital closure programme in Britain – which continues to gain momentum amid controversy about the effectiveness of treatment for the acutely ill – could offer both social and economic benefits for most people who would formerly have remained in hospital as long-stay in-patients.

Reproviding long-term care in the community for chronically ill people (many of whom no longer display active symptoms of mental illness) may not be straightforward, but it can be successful in terms of patient quality of life and it should be affordable from the savings from hospital rundown. However, successful and cost-effective community-based alternatives to hospitalisation for acutely ill patients have been more elusive. One model which has attracted a lot of attention is the assertive community treatment (ACT; sometimes called PACT) approach pioneered in Wisconsin by Stein and Test (1980) and recently reviewed by Scott and Dixon (1995). The economic evaluation of the original ACT model was a CBA conducted over a 12-month period (Weisbrod et al. 1980). Costs ranged beyond health care to include criminal justice, maintenance and family costs; benefits were measured (somewhat narrowly) as the monetary value of increased participation in the workforce. The study found a higher net benefit (i.e., benefit minus costs) for the community-based programme.

A replication of the ACT model in London (with some modifications) – the Maudsley Hospital's Daily Living Programme (DLP) – was the subject of a two-phase, 4-year evaluation. In the first phase (randomised allocation of 189 people with acute schizophrenia or severe affective disorder to either the DLP or standard in-patient care), the evaluation found better outcomes, higher patient and family satisfaction and lower costs over a 20-month period (Marks et al. 1994; Knapp et al. 1994). The second phase extended the research period to 4 years and included randomised withdrawal of DLP from 33 patients, continued DLP for another 33 and continued control care (in-patient/out-patient care) for 70 patients. At the end of this second phase, all of the earlier clinical gains had been lost, and there was no longer a cost advantage to the DLP (Audini et al. 1994; Knapp et al. 1998). The cost profile is illustrated in Fig. 5. The economic evaluation was a CEA: it included the fully comprehensive costing of all health, social care and other services; it examined the effects of DLP and standard care on patient employment and carer support; and it assessed outcomes in terms of symptoms, behaviour, social adjustment, personal functioning and patient and family satisfaction. No single outcome measure was assumed to dominate, and no attempts were made to obtain uni-dimensional "utility" measures or to express the outcomes in monetary terms.

6 Conclusions

There is a large unmet need for economic evaluation in the mental health field. Cost-effectiveness and similar insights are increasingly being recognised as relevant for drug licensing, budget-constrained case management, the increasing emphasis on evidence-based medicine and the more general and virtually perennial search for greater efficiency and equity in the delivery of treatment interventions and services. As discussed

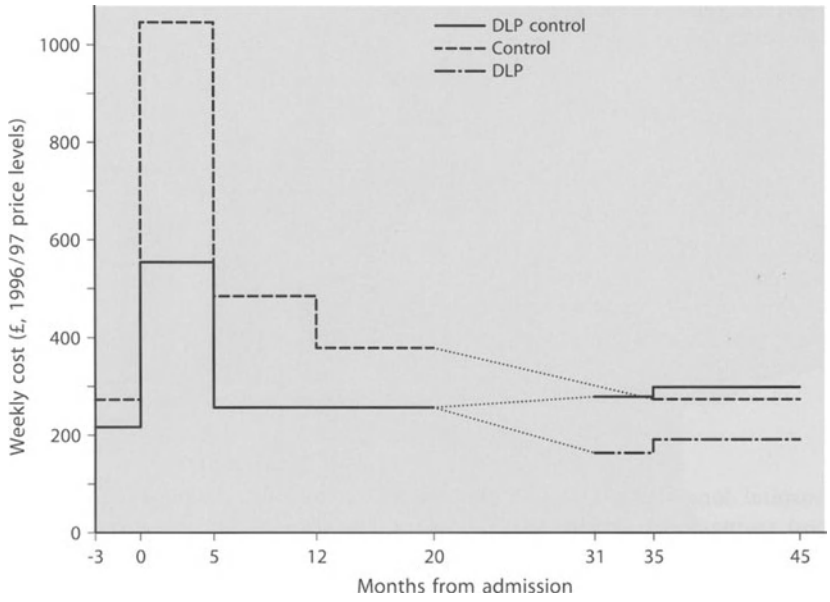


Fig. 5. Daily Living Programme (*DLP*): costs over 4 years

in this chapter, a number of factors are currently combining to raise the demands for health economics perspectives in psychiatry. These demands are likely to be greater when a country is in some economic difficulty, when government expenditure decisions produce a less generous settlement for a publicly funded health service, when employers, individuals or others are trying to negotiate lower health insurance premiums or when sickness funds are looking to cut providers' reimbursement rates. However, it must be emphasised that the relevance of economics does *not* stem solely from these fiscal difficulties and political pressures. Economics has relevance whenever there is scarcity of resources relative to needs.

Such scarcity characterises every country in the world, even if manifest demands for health economics are more commonly expressed in the developed world. In fact, there are clear signs of a recognition of the relevance of economic insights in the developing world.

Presently, probably the hardest task in conducting an economic evaluation of mental health care is finding a suitable economist. The demand for health economists worldwide – and particularly in Europe – outstrips the available supply. This might tempt clinical research teams to conduct their own economic evaluations, but they should take note of various potential complexities outlined here. Whoever seeks to undertake the evaluation should follow the well-rehearsed and widely accepted stages of a study of the kind set out in this chapter. They should also endeavour to learn from the mistakes and the successes of previous evaluations, as internationally there is now a gradual accumulation of experience upon which to draw. Many such experiences could be mentioned as relevant, but I will confine my closing remarks to just three.

Firstly, the best mental health economics research – and potentially the most influential – will be fully integrated into the work of clinical research teams. Although economic evaluations necessarily draw on a different disciplinary base from clinical research, there are many commonalities and complimentarities between the two which make it not merely sensible but almost imperative that integrated research is pursued.

Secondly, comprehensiveness of measurement is usually necessary. Clinical researchers would not willingly drop an outcome dimension from an evaluation without very good cause, and there should be an equivalent reluctance to omit any cost components from an economic study. Mental ill-health impacts on many aspects of an individual's and a family's lifestyle and functioning, and the treatment of mental ill-health often draws on the resources of many health care and non-health-care agencies and stakeholders. Evaluations should aim to assess all such impacts and resource consequences. Too many of the completed economic evaluations to be found in the literature today fail to measure costs comprehensively, which inevitably casts some doubt on their usefulness.

Thirdly, there are probably many causal and other links between the resource and outcome sides of mental health interventions – between costs and effectiveness. Some of these links will be direct and straightforward, others indirect and complex. Relatively few economic evaluations to date have explored those links, but those that have bear testimony to the insights that can be revealed and the benefits that may emerge for clinical practice and strategic policy.

7 Appendix

Average cost
: The cost of treatment per patient, equal to the total cost divided by the number of patients treated (e.g., average cost per in-patient day)

Cost-benefit analysis (CBA)
: Economic evaluation which compares the costs and the consequences of two or more treatment, placements or policies, where those consequences are measured in monetary units

Cost-effectiveness analysis (CEA)
: Economic evaluation which compares the costs and the level of effectiveness (see below) of two or more treatments, placements or policies

Cost-minimisation analysis
: Economic evaluation which takes as its criterion the minimisation of the cost of treating patients with particular clinical conditions

Cost-offset analysis
: Economic evaluation which compares costs incurred with costs saved, usually across two or more treatments, placements or policies

Cost-utility analysis (CUA)
: Economic evaluation which compares the costs and the level of utility (health-related quality of life) of two or more treatments, placements or policies

Economy
: Saving of resources; the pursuit of lower costs

Effectiveness
: Improvement in the health or quality of life of people with mental health problems or their caregivers

Efficiency
: Improvement in the level of effectiveness (see above) achieved from given expenditure or resources, or reduction in the cost of achieving a given level of effectiveness

Equity
: Justice of fairness in the distribution of resources, access to treatment, or achievement of levels of health or quality of life

Marginal cost
: The incremental cost of treating one more patient or the decremental cost of treating one less patient

Opportunity cost
: The cost of a service or resource expressed in terms of the value of the benefits forgone by not employing it in its best alternative use

Outcomes
: Changes in the health or quality of life of patients or caregivers (see "Effectiveness")

Quality-adjusted life-year (QALY)
: A measure of extended life expectancy adjusted for the patient's quality of life; used in cost-utility analyses (see above)

Value for money
: Another term for efficiency (see above)

8 References

AOK-Bundesverband (1996) Krankheitsartenstatistik 1993. AOK-Bundesverband, Bonn

Audini, B, Marks IM, Connolly J, Lawrence RE, Watts V (1994) Home-based versus out-patient/in-patient care for people with serious mental illness. Br J Psychiatry 165: 204–210

Beck B (1996) The economics of ageing. The Economist (27th January 1996): 3–16

*Beecham JK, Knapp MRJ (1992) Costing psychiatric interventions. In: Thornicroft G, Brewin C, Wing J (eds) Measuring mental health needs. Gaskell, London

Beecham JK, Knapp MRJ, Fenyo AJ (1991) Costs, needs and outcomes. Schizophr Bull 17: 427–439

*Beecham JK, Hallam A, Knapp MRJ, Baines B, Fenyo AJ, Asbury M (1997) Costing care in the hospital and in the community. In: Leff J (ed) Community care: illusion or reality? Wiley, Chichester

Beske F, Kern OA (1995) Kosten-Nutzen-Analyse der Behandlung dementieller Erkrankungen. Psycho 21: 724–730

Bundesministerium für Jugend, Familie, Frauen und Gesundheit (1986) Modellprogramm Psychiatrie – regionales Psychiatriebudget (181). Kohlhammer, Stuttgart

Bundesministerium für Jugend, Familie, Frauen und Gesundheit (eds) (1988) Empfehlungen der Expertenkommission der Bundesregierung zur Reform der Versorgung im psychiatrischen Bereich auf der Grundlage des Modellprogramms Psychiatrie der Bundesregierung. Bundesministerium für Jugend, Familie, Frauen und Gesundheit, Bonn

Brand FN, Smith RT, Brand PA (1977) Effect of economic barriers to medical care on patients' noncompliance. Public Health Rep 92: 72–78

British Medical Association (1988) Philosophy and practice of medical ethics. British Medical Association, London

Chisholm D, Healey A, Knapp, MRJ (1997) QALYs and mental health care. Soc Psychiatry Psychiatr Epidemiol 32: 68–75

Commonwealth of Australia (1990) Guidelines for the pharmaceutical industry on preparation of submissions to the Pharmaceutical Benefits Advisory Committee: including submissions involving economic analyses. Department of Health, Housing and Community Services, Woden

Deutscher Bundestag (1975) Bericht über die Lage der Psychiatrie in der Bundesrepublik Deutschland. Deutscher Bundestag, Bonn

Diekstra RFW (1995) Depression and suicidal behaviours in adolescence: sociocultural and time trends. In: Rutter M (ed) Psychosocial disturbances in young people. Cambridge University Press, Cambridge

*Dorwart RA (1990) Managed mental heath care: myths and realities in the 1990s. Hosp Community Psychiatry 41: 1087–1091

Drummond MF (1992) Cost-effectiveness guidelines for reimbursement of pharmaceuticals: is economic evaluation ready for its enhanced status? Health Econ 1: 85–92

**Drummond M, Stoddart G, Torrance G (1987) Methods for the economic evaluation of health care programmes. Oxford Medical, Oxford

Frank RG, McGuire TG (1996) Introduction to the economics of mental health payment system. In: Levin BL, Petrila J (eds) Mental health services: a public health perspective. Oxford University Press, New York

Frank RG, Salkever D, Sharfstein S (1991) A new look at rising mental health insurance costs. Health Affairs 10: 116–123

Geraty RD, Hendren RL, Flaa CJ (1992) Ethical perspectives on managed care as it relates to child and adolescent psychiatry. J Am Acad Child Adolesc Psychiatry 31: 398–402

Gray A, Marshall M, Lockwood A, Morris J (1997) Problems in conducting economic evaluations alongside clinical trials: lessons from a study of case management for people with mental disorders. Br J Psychiatry 170: 47–52

Häfner H, an der Heiden W (1989) Effectiveness and cost of community care for schizophrenia patients. Hosp Community Psychiatry 40: 59–63

Heinrich K, Linden M, Müller-Oerlinghausen B (1989) Werden zu viele Psychopharmaka verordnet? Methoden und Ergebnisse der Pharmakoepidemiologie und Phase-IV-Forschung. Thieme, Stuttgart

Hess D, Ciompi L, Dauwalder H (1989) Nutzen- und Kostenevaluation eines sozialpsychiatrischen Dienstes. Nervenarzt 57: 204–213

Hoffmeyer U (1994) The health care system in Germany. In: Hoffmeyer U, McCarthy T (eds) Financing health care I. Kluver, Dordrecht

Hotopf M, Lewis G, Normand C (1996) Are SSRIs a cost-effective alternative to tricyclics? Br J Psychiatry 168: 404–409

Johnson S, Ramsay R, Thornicroft G, Brooks L, Lelliott P, Peck E, Smith H, Chisholm D, Audini B, Knapp MRJ, Goldberg D (1997) London's mental health. King's Fund, London

Kavanagh S, Stewart A (1995) Economic evaluations of mental health care. In: Knapp MRJ (ed) The economic evaluation of mental health care. Arena, Aldershot

Klein R (1995) Big bang health care reform – does it work? The case of Britain's National Health Service reforms. Milbank Q 73: 299–337

Klerman GL, Weissman MM (1989) Increasing rates of depression. JAMA 261: 2229–2235

Knapp MRJ (1984) The economics of social care. Macmillan, London

Knapp MRJ (1995) The economic evaluation of mental health care. Arena, Aldershot

Knapp MRJ (1997a) Economic evaluation and interventions for children and adolescents with mental health problems. J Child Psychol Psychiatry 38(1): 3–26

Knapp MRJ (1997b) Economics and mental health: a concise European history of demand and supply. In: Tansella M (ed) Making mental health services rational. Pensiero Scientifico, Rome

Knapp MRJ, Beecham JK (1996) Programme-level and system-level health economics considerations. In: Knudson HC, Thornicroft G (eds) Mental health service evaluation. Cambridge University Press, Cambridge

Knapp MRJ, Beecham J, Koutsogeorgopoulou V, Hallam A, Fenyo A, Marks IM, Connolly J, Audini B, Muijen M (1994) Service use and costs of home-based versus hospital-based care for people with serious mental illness. Br J Psychiatry 165: 195–203

Knapp MRJ, Marks IM, Wolstenholme J, Beecham JK, Astin J, Audini B, Connolly J, Watts V (1998) Home-based versus hospital-based care for serious mental illness: a controlled cost-effectiveness study over four years. Br J Psychiatry 172: 506–512

Korman N, Glennerster H (1990) Hospital closure. Open University, Milton Keynes

Leff J (1997) Community care: illusion or reality? Wiley, Chichester

Leff J, Trieman N, Gooch C (1996) Prospective follow-up study of long-stay patients discharged from two psychiatric hospitals. Am J Psychiatry 153: 1318–1324

Lehman AF (1996) Measures of quality of life among persons with severe and persistent mental disorders. In: Thornicroft G, Tansella M (eds) Mental health outcome measures. Springer, Berlin Heidelberg New York, pp 75–92

Le Pen C, Levy E, Ravily V, Beuzen JN, Meurgey F (1994) The cost of treatment dropout in depression. A cost-benefit analysis of fluoxetine vs tricyclics. J Affective Disord 31: 1–18

Light D (1997) The real ethics of rationing. Br Med J 315: 112–115

Linden M, Gothe H (1993) Benzodiazepine substitution in medical practice: analysis of pharmacoepidemological data based on expert interviews. Pharmacopsychiatry 26: 107–113

Mari JJ, Streiner D (1996) The effects of family intervention on those with schizophrenia. In: Adams C, Anderson J, Mari JJ (eds) Schizophrenia module. Cochrane Database of Systematic Reviews. Cochrane Library, London

Marks IM, Connolly J, Muijen M, McNamee G, Audini B, Lawrence RE (1994) Home-based versus hospital-based care for people with serious mental illness. Br J Psychiatry 165: 179–194

Maynard A, Marinker M, Gray DP (1986) The doctor, the patient and their contract. Br Med J 292: 1438–1440

*McGuire A, Henderson J, Mooney G (1988) The economics of health care. Routledge, London

McPhillips MA, Sensky T (1997) Coercion, adherence or collaboration? Influences on compliance with medication. In: Wykes T (ed) Outcome and innovation in psychological treatment of schizophrenia. Wiley, Chichester

Mehrez A, Gafni A (1989) Quality-adjusted life years, utility theory and healthy-year equivalents. Med Decis Making 9: 142–149

National Health Service Executive (1996) Burdens of disease. National Health Service Executive, London

Netten A, Dennett JH (1997) Unit costs of health and social care 1997. Personal Social Services Research Unit, University of Kent, Canterbury

Offord DR, Boyle MH, Racine YA, Fleming JE, Cadman DT, Munroe Blum H, Byrne C, Links PS, Lipman EL, MacMillan HL, Rae Grant NI, Sanford MN, Szatmari P, Thomas H, Woodward CA (1992) Outcome, prognosis and risk in a longitudinal follow-up study. J Am Acad Child Adoles Psychiatry 31: 916–923

Padgett DK, Patrick C, Burns BJ, Schlesinger HJ, Cohen J (1993) The effect of insurance benefit changes on use of child and adolescent outpatient mental health services. Med Care 31: 96–110

Priebe S (1994) Die Bedeutung der Lebensqualität für psychiatrische Versorgung und Forschung. Psychiatr Prax 21: 87

Priebe S, Hoffman K, Isermann M, Kaiser W (1996) Klinische Merkmale langzeithospitalisierter Patienten. Psychiatr Prax 23: 15–20

Revicki DA, Brown RE, Palmer W et al (1995) Modelling the cost effectiveness of antidepressant treatment in primary care. Pharmacoeconomics 8: 524–540

Revicki DA, Shakespeare A, Kind P (1996) Preferences for schizophrenia-related health states: a comparison of patients, caregivers and psychiatrists. Int Clin Psychopharmacol 11: 101–108

*Rice DP, Kelman S, Miller LS (1990) The economic costs of alcohol and drug abuse and mental illness 1985. National Institute of Mental Health, Rockville, MD

Ritchie J, Kick D, Lingham R (1994) The report of the inquiry into the care and treatment of Christopher Clunis. HMSO, London

Rössler W, Salize JH (1994) Longitudinal statistics of mental health care in Germany. Soc Psychiatry Psychiatr Epidemiol 29: 112–118

Rössler W, Salize HJ, Bauer M (1996) Psychiatrische Abteilungen an Allgemeinkrankenhäusern – Stand der Entwicklung in Deutschland. Psychiatr Prax 23: 4–9

Ryan M (1996) Using conjoint analysis to establish consumer preferences. Office of Health Economics, London

Sabin JE (1994) A credo for ethical managed care in mental health practice. Hosp Community Psychiatry 45: 859–860

*Salize HJ, Rössler W (1995) The cost of comprehensive care of people with schizophrenia living in the community: a cost evaluation from a German catchment area. Br J Psychiatry 169: 42–48

Sayce L (1995) Response to violence: a framework for fair treatment. In: Crichton J (ed) Psychiatric patient violence: risk and response. Duckworth, London

Sclar DA, Robison LM, Skaer TL, Legg RF, Nemec NL, Galin RS, Hughes TE, Buesching DP (1994) Antidepressant pharmacotherapy: economic outcomes in a health maintenance organisation. Clin Ther 16: 715–730

Sclar DA, Robison LM, Skaer TL, Galin RS, Legg RF, Nemec NL, Highes TE, Buesching DP, Morgan M (1995) Antidepressant pharmacotherapy: economic evaluation of fluoxetine, paroxetine and sertraline in a health maintenance organisation. J Int Med Res 23: 395–412

Scott JE, Dixon LB (1995) Assertive community treatment and case management for schizophrenia. Schizophr Bull 21: 657–668

Sheldon T (1996) Problems of using modelling in the economic evaluation of health care. Health Econ 5: 1–12

*Simon GE, Von Korff M, Heiligenstein JH, Revicki DA, Grothaus L, Katon W, Wagner EH (1996) Initial antidepressant choice in primary care: effectiveness and cost of fluoxetine vs tricyclic antidepressants. JAMA 275: 1897–1902

Stein LI, Test M (1980) Alternative to mental hospital treatment. Arch Gen Psychiatry 37: 392–397

Beck B (1996) The economics of ageing. The Economist (27th January 1996): 3–16

Tillett R (1996) Psychotherapy assessment and treatment selection. Br J Psychiatry 168: 10–15

Unützer J, Tischler GL (1996) The many faces of managed care. In: Moscarelli M, Rupp A, Sartorius N (eds) Handbook of mental health economics and policy: schizophrenia. Wiley, Chichester

*Weisbrod BA, Test MA, Stein LI (1980) Alternative to mental hospital treatment. II. Economic benefit-cost analysis. Arch Gen Psychiatry 37: 400–405

Wissenschaftliches Institut der Ortskrankenkassen (1986) Krankenhäuser 1984: Ein statistischer Spiegel. Wissenschaftliches Institut der Ortskrankenkassen, Bonn

World Bank (1993) World development report: investing in health. Oxford University Press, Oxford

World Health Organization (1982) Changing patterns in suicide behaviour. WHO/Euro Report no 74. World Health Organization, Geneva

CHAPTER

17

T. Becker, N. Sartorius

Environmental Aspects of Psychiatry

Translator: T. Becker
I would like to thank Prof. Manfred Bauer (Offenbach, Germany), Dr. Georg Becker (Würzburg, Germany), Dr. Douglas Bennett† (Oxford, UK), Prof. Klaus Lange (Regensburg, Germany), Prof. George W. Brown, Prof. P. Graham and Dr. J. Thompson (London, UK), Prof. Y. Nakane (Nagasaki, Japan), Dr. J. Novikov (Hamburg, Germany) and Dr. Dirk K. Wolter-Henseler (Münster, Germany).

1 Introduction

In 1866, Ernst Haeckel coined the concept of "ecology" (Greek: *oikos* = house, household, and *logos* = study; i.e. study of household [of nature]): the study of "all relationships of the organism with all other organisms with which it is in contact". As a scientific concept in biology, ecology had been clearly defined by 1870 as a branch of the biological sciences focusing on the relationships between animals and plants on the one hand and their animate and inanimate environment on the other hand. Human ecology is a discipline of environmental sciences that deals with the interaction of humans and their environment.

One hundred years after being coined by Haeckel, the concept of "ecology", along with that of "environment", has experienced widespread popularity and a change in meaning. Over the past two decades, the adjective and prefix (ecological, eco-) have come to mean "in line with environmental requirements/precautions" (Streit 1994) and often have political overtones.

The ecological approach, in sociology, has developed into a field of work in its own right. Urban sociology, on the basis of the work of the Chicago School, has evolved into a multi-faceted area of research (Burgess and Park 1925; Burgess and Brogue 1967). It first focused on the question whether urban surroundings can be divided into zones or natural areas in a valid way. These areas were supposed to be a result of the "natural" growth of cities, which was believed to be the reflection of interactions of unplanned forces. Studies of this line of thought have looked into issues relating to ethnic and religious groups, neighbourhoods and rules governing communal life within them, social/cultural prejudices, delinquency, the role of police forces and family life. Links with psychiatry result from the assumption that psychiatric morbidity and suicide rates are influenced by local living conditions in urban surroundings; there is much less emphasis on the study of the influences that mental illness or abnormality may have on urban life.

Faris and Dunham (1930) studied the distribution of psychiatric disorders in Chicago. They distinguished five urban zones: a central "business district" virtually without stable resident population, but with a high proportion of hotel/hostel lodging residents and homeless people; in a neighbouring, more peripheral zone, there were immigrants and unskilled workers' families; a third zone was characterised by a more stable resident population with a transitional zone leading on to areas of better housing conditions; the fourth and fifth zones contained residential areas of the upper middle class. The dependent variable, in the study of these authors, was the annual number of psychiatric admissions to Elgin State Hospital, the mental hospital admitting all mentally ill from the catchment area in the decade 1922–1931. Risk of admission for schizophrenia, alcohol-related psychoses and neurosyphilis was higher in the "downtown" central urban zones, whereas those admitted due to manic-depressive illness were more evenly distributed in the five zones of residence. Similar studies in other large cities subsequently confirmed the skewed distribution of mental health service utilisation by patients with schizophrenia; there have, however, also been a few negative studies (Schwab and Schwab 1978).

Methodological problems implicit in studies such as those by Faris and Dunham (1939) are discussed by Riley (1963). The approach chosen by the Chicago School aims at identifying the impact of "certain undetermined types of social processes" on urban psychiatric morbidity. The underlying assumption is that urban zones or geographically defined populations (aggregates) are characterised by homogeneous qualities which can be distinguished from qualities of the components (individuals, families) that make up their constituent parts. The zones are characterised by qualities such as mobility of the resident population or cost of housing/level of rents. The independent variable in the model (the impact of the urban zone) cannot be measured easily as it is itself the result of socio-demographic analysis and boundary definition.

One of the key elements of the ecological approach is that environment and health (or disease) are viewed as having a reciprocal relationship. Modern urban development can be associated with an increase in the number of homeless people, e.g. because of an increase in the cost of care. The presence of homeless people in the urban ecosystem can contribute to changes in urban living conditions which may be associated with population instability, changes in social relationships and delinquency, for example. These in turn may affect the state of health of all segments of the population, including the group of homeless people. The ecological approach suggests that we should look at individual and group characteristics simultaneously in clarifying such interactions.

2 Ecology and Epidemiology

Ecology and epidemiology are neighbouring disciplines which complement each other. Both epidemiological and ecological studies often have their starting point in aetiopathogenetic or prognostic hypotheses. "Epidemiology is concerned with the distribution of

diseases in a population and the factors that influence that distribution" (Mann 1993). The ecological approach concentrates on the analysis of reciprocal relationships between physical, biological and social factors on the one hand and mental health or ill health on the other hand.

The concept of the community based on "locality and community sentiment" (Frankenberg 1966) is a core component of the field of ecology. The traditional themes of ecological research include differences in social roles and role relationships in urban versus rural communities (Frankenberg 1966). Festinger et al. (1959), in a sociometric study, found an influence of the spatial distribution of student flats/apartments on their social contacts. Similar types of study have been performed in psychiatric research. Population density and residential unit density, residential living conditions, urban–rural differences, population mobility and noise were investigated with respect to their relationship with psychiatric morbidity and service utilisation. An ecological long-term study in Mannheim analysed the urban distribution of treated mental disorders in 20 urban zones in the years 1965 and 1974–1980. In spite of substantial social changes (increase in proportion of ethnic minorities, conspicuous urban structural change, setting up of community mental health services), the study reported a high degree of stability in the geographical distribution of treated mental disorders both in the long term (1974–1980 vs. 1965) and in individual comparisons of the years 1974–1980 (Weyerer and Häfner 1989).

The examination of mental disorders in urban areas illustrates the complementarity of epidemiological and ecological approaches to a problem. Epidemiological studies find differences in the prevalence of mental disorders among urban regions, and epidemiologists then explore whether these differences are the result of migration which may have occurred before illness onset or in prodromal or early disease states, which in turn may reflect pre-morbid deficits or early signs of the disease process (Schwab and Schwab 1978). People in the prodromal or early phases of a psychotic illness, from this perspective, would change their residence and choose inner city areas with cheap housing, space for social withdrawal and an unstable resident population. Ecologists use these results to explore further whether the increase in numbers of people with mental illness in a particular part of the town could result in more tolerance and/or indifference towards unusual or bizarre behaviour in inner city areas. Goldberg and Morrison (1963) studied 509 male patients admitted to English mental hospitals with respect to social class. The proportion of lower social class membership was higher than in the general population. When the comparison was based on the social class of patients' fathers (at the time of birth of the index patients), there was no significant difference. Detailed analysis of 94 patients with schizophrenia revealed a downward social trend in education/training and job careers, and both findings tended to strengthen the social drift hypothesis. This does not, however, diminish the importance of the ecological approach, because context variables are important in the manifestation, course and prognosis of mental disorders. Causal effects and modulation of outcome are in no way mutually exclusive (Schwab and Schwab 1978).

In examining the relationships between the context and individuals, state of health is important in order to study the rate of change, the duration of change and its impact (Fig. 1). The effects of environmental agents on the mental health of populations or population groups differ in accordance with the time course, degree and length of impact. A comparison of natural disasters and chronic low-dose lead exposure highlights the wide range of possibilities.

In the following, the relationships between mental health and various environmental situations and influences will be presented. For reasons of clarity, the interaction between mental health and environmental factors will be described separately, although the latter

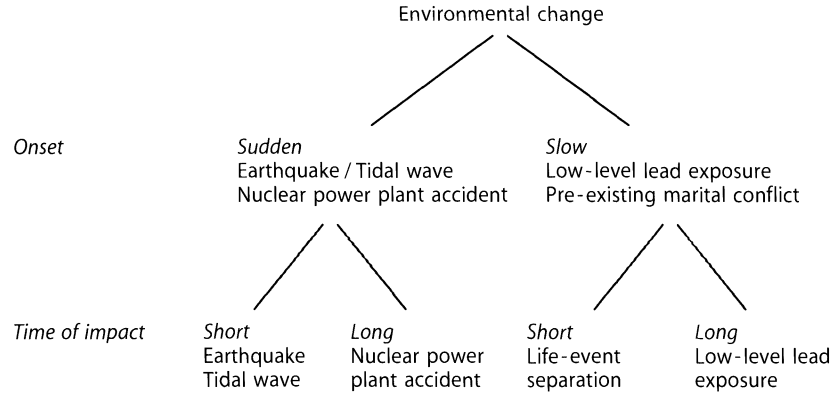


Fig. 1. Time course of environmental changes and their impact

often exert their effects on an individual or group simultaneously. An example may serve to illustrate this point: A life event may come up for a particular patient while he or she is in hospital; both the life event and hospital stay have an impact on the patient's mental state and social situation; this may occur in an overcrowded ward environment with significant noise problems; and simultaneously environmental hazards and socio-cultural stressors may be at work in the wider environment (e.g. because of environmental hazards in the immediate surroundings of the hospital). Environmental exposure and socio-cultural environment have simultaneous effects on the patient and interact with each other and with the above-mentioned factors.

3 Ecological Developmental Psychology

The study of human development in childhood and adolescence highlights a range of questions concerning context (family, school, neighbourhood). The importance of context results from the continuous interaction of the individual and the environment. Ecological developmental psychology or socialisation research (Bronfenbrenner 1979) pursued this approach to human development. The concept addresses interactions between individual and social context variables (person–context–time model; Moen 1995). Rutter et al. (1995), in line with this concept, describe a programme geared at minimising developmental risk in children. The elements of this programme comprise the following:

1. Identification of protective psychosocial factors (e.g. social support, confidants)
2. Reduction in psychosocial risk factors
3. Avoiding cumulation of risk factors
4. Improved problem-solving strategies

Prospective long-term studies by Rutter and his group revealed that emotional and behavioural abnormalities documented at the age of 10 years were related to life events which occurred in adulthood and to enduring psychosocial problems in the third decade of probands' lives. Behavioural abnormalities had a stronger predictive effect than emotional disorders. The likelihood of choosing a partner with antisocial behaviour or alcohol/drug abuse was higher in disturbed 10-year-olds. There were also relationships between adolescent peer group characteristics on the one hand and later practical life skills and choice of partner. Behavioural disturbances during adolescence were associated with later familial tension; school achievement was related with success in finding a job (Rutter et al. 1995). Wilson (1995) studied everyday adolescent life in North American ghettos and underlined how much deviant behaviour was embedded in the ecological context.

Rutter and Smith (1995) surveyed the literature on the prevalence of psychosocial disorders of adolescence in an attempt to elucidate secular trends. They found evidence for an increase in delinquency, substance misuse, depression and suicide among adolescents in the second half of the twentieth century. The evidence regarding eating disorders was not consistent. In summarising the factors which may have contributed to this secular trend, the authors emphasise the following points:

- It is unlikely that factors such as low social status, unemployment or health problems are responsible for the above-mentioned increases, as social parameters have on the whole improved during the time span considered.
- It appears plausible that problems in families, and parental conflicts in particular, have contributed to the increase in morbidity.
- The authors found factors whose importance was not quite clear. This applies to increased expectations in the young and the prolongation of adolescence (earlier onset of puberty, later end of school).
- The authors felt that media impact on the frequency of psychosocial disturbances in adolescence could not be ruled out, although it was likely that their contribution to increased prevalence was of minor importance.
- A number of mental disorders show secular increases in prevalence, and factors that could be responsible for these changes seem to differ from disorder to disorder.

One of the secular changes in our century is an increase in the variability of all social, economic, educational and family structures. Fergusson (1996) suggested that it is possible that an increase in variability of adolescent behaviour is an expression of both increased deviance and of the increase in successful forms of adaptation. This illustrates how a shift in focus from psychosocial disturbances to the broad spectrum of dysfunctional and functional juvenile behaviour might facilitate understanding of the data presented by Rutter and Smith (1995).

4 Mental Hospital and Illness Outcome

A series of reports from the 1950s and 1960s criticised living conditions and institutional standards in mental hospitals (Goffman 1961). Although it was not possible to find direct causal links between institutional

conditions and clinical state or symptom level (Wing and Brown 1970), negative effects of institutional care were plausible. An early study by Ullman (1967) found associations between hospital size and number of mental hospital staff on the one hand and mean length of stay on the other. Barton (1959) formulated the hypothesis that some aspects of institutional care, such as isolation from the wider environment, loss of social contacts and personal belongings, lack of stimuli in ward routine and loss of perspective, might lead to a syndrome he termed "institutional neurosis."

Wing and Brown (1970) examined therapeutic conditions in a longitudinal study of three English mental hospitals (Mapperley, Severalls, Netherne). They collected data on four occasions (1960, 1962, 1964 and 1968), which enabled them to analyse relationships between social impoverishment and a "clinical poverty syndrome" as well as poverty of milieu and patients' attitudes towards hospital discharge. They found: (a) differences between hospitals over time, (b) clinical improvement between 1960 and 1962, (c) patients receiving more social stimulation in 1962 experienced clinical improvement, (d) clinical improvement from 1960 to 1962 was a good predictor of hospital discharge between 1962 and 1968 and (e) most clinical and social improvement was observed between 1960 and 1962. Conditions in hospitals after this period stagnated or deteriorated. The authors concluded

> that a substantial proportion, though by no means all, of the morbidity shown by long-stay schizophrenic patients in mental hospitals is a product of their environment. The social pressures which act to produce this extra morbidity can to some extent be counteracted, but the process of reform may itself have a natural history and an end (Wing and Brown 1970, p. 177)

The findings of the study provide substantial evidence of a social hypothesis of so-called negative symptoms, i.e. of loss of drive, loss of interest, affective flattening and poverty of speech. They suggest caution where progress achieved in the recent past is extrapolated into the future; the maintenance of improved therapeutic conditions and clinical state requires continued and deliberate clinical effort.

5 Long-Term Studies of Psychiatric Morbidity

Longitudinal studies offer the opportunity to assess the effects of complex long-term environmental change on psychiatric morbidity and thereby enable the investigators to study interactions between environmental change and manifestation of illness. Sartorius et al. (1989) assessed the question of associations of economic and social/cultural change on the one hand and the incidence and prevalence of psychiatric disorders on the other by summarising longitudinal studies of mental disorders. Table 1 summarises the methodology and key findings, and Table 2 summarises environmental change on the one hand and changes in morbidity on the other hand. "Common" change indicates developments which are in keeping with broader social change in the same culture, while "significant" changes went beyond the average measure of change over periods of one to three decades (Strömgren et al. 1989). The tables show that associations between changes in the environment and changes in morbidity were rather weak. Taiwan, Berlevåg and Shetland experienced major environmental change. In the period between the studies, there had been immigration, social change, rapid economic development and foreign policy tension in Taiwan; in Berlevåg, the stresses of the Second World War had been felt drastically; and in Shetland, the off-shore oil industry had changed social conditions. However, changes in morbidity were observed only in the numbers of neurotic disorders (in older women in Shetland); however, they were not seen in endogenous psychoses. Environmental change in New York, Bavaria, Lundby and Stirling County was less dramatic. Neurotic disorders in New York decreased in women, and depressive disorders increased in Lundby; substance misuse and personality disorders increased in Bavaria. Regions with stable environmental conditions nevertheless experienced changes in morbidity of mental disorders.

Strömgren et al. (1989) conclude that, on the basis of the long-term studies reported, there is little evidence for the assumption that social changes such as industrialisation or disintegration of family structures lead to an increase in mental disorders. However, they arrive at the following conclusions:

- There is some data indicating a decrease in the prevalence of schizophrenia (Taiwan, Bornholm [for women]).
- Depressive disorders are increasing in prevalence; this is likely to apply to the group of so-called neurotic depressions rather than to severe depression with melancholic features (e.g. Lundby).
- An increase in neurotic disorders in response to socio-economic change appears possible. The authors mention the hypothesis of a U-shaped curve with a low incidence of neurotic disorders in instances of low environmental pressure and during particularly pressing external pressures (e.g. wartime).

Table 1. Long-term studies of psychiatric morbidity (Baxter 1990; Feldman 1991; Langolf et al. 1978)

Place	Target criterion	Sample	Years of sampling	Population	Main results
Taiwan	Prevalence	Census studies	1946/1948	19,931	No difference in psychosis prevalence rates except for schizophrenia (decrease), increase of non-psychotic disorders
			1961/1963	39,024	
New York (Midtown Manhattan)	Prevalence	Cohort study (probability sample)	1954	1,660	No difference in prevalence rates from 1954 to 1974, decrease in overall morbidity in women
			1974	Smaller sample at follow-up	
Stirling County	Prevalence	Census study (household sampling)	1952	1,003	Prevalence of depression and anxiety: no difference over time
			1970	1,094	
Lundby	Prevalence, incidence, morbid risk	Cohort study (cross-section of population)	1947	1: 2,550	Increase in morbid risk for depression 1947–1957 to 1957–1972, decrease in dementias
			1957	Cohort of 1947 plus 1,013 additional proband	
			1972		
Berlevåg	Prevalence	Census studies (all local residents)	1944	1,325	No difference in psychosis prevalence, 50% increase neurotic disorders
			1974	1,647	
Iceland	Morbid risk	Birth cohort	1957	5,395	Overall stability
			1971		
			1977		
Samsø	Prevalence	Census studies (island population)	1961	6,823	Reduction in frequency of organic disorders (most likely decreased incidence)
			1972	5,008	
			1977	4,907	
Denmark (learning disability)	Prevalence	Census studies (total population Denmark)	1888	2,172,380	Total prevalence 1965 and 1979 higher than in 1888
			1965	4,767,597	
			1979	5,111,534	
Upper Bavaria	Prevalence, incidence	Cohort study (cross-section population)	1975	1,668	No prevalence difference overall, increase in personality disorders and alcohol abuse
			1980		
Shetland	Prevalence	Cohort study	1975	Target population, 376; control group, 413	Increase in psychiatric symptoms in older women, otherwise no differences
			1978		
Bornholm	Prevalence	Census studies	1935	45,684	Decrease in prevalence of schizophrenia in 15- to 34-year-olds
			1983	47,313	

Table 2. Environmental change and changes in morbidity in long-term studies on the frequency of psychiatric disorder

Place	Environmental change	Morbidity				
		Marked decrease	Marked increase	Slight decrease	Slight increase	No change
Substantial change						
Taiwan	Political change, migration (from mainland), industrialisation, urbanisation, change in life-style and values		All psychiatric disorders neuroses	Schizophrenia		All psychoses
Berlevåg (Norway)	1944: poor fishing village 1974: More jobs in industry, increased housing		Neuroses			All psychoses
Shetland (Scotland)	–				Substance misuse, (neuroses)	All psychiatric disorders
"Usual" change						
New York (USA)	Significant changes in gender roles during first half of century			All psychiatric disorders (women)		No significant change in total sample
Upper Bavaria	–				Personality disorders, substance misuse	All psychiatric disorders
Lundby (Sweden)	Changes as in country as a whole, T2 vs. T1: more service jobs, more unemployment	Organic disorders (elderly population)	Depression, Substance misuse			
Stirling County (Canada)	T2 vs. T1: small-scale urbanisation, increased tertiary sector and living standard, families smaller					Depression, neuroses
Relative stability						
Bornholm (Denmark)	–	Schizophrenia				
Samsø (Denmark)	–	Organic disorders				
Iceland	–					All psychiatric disorders
Denmark	–				Learning disability	

- The Danish national census of learning disability (mental impairment) shows an increase in the overall rate from 1888 to 1966 which then levels off. The authors consider that this increase is not surprising, since (a) there has been a reduction in excess mortality and (b) diagnosis has been improved (i.e. the disorder is more readily recognised).
- There is some data suggesting a decrease in organic psychiatric disorders in the elderly, which is most likely the result of increased life expectancy due to improved health and social care. This trend is probably balanced, however, by a concomitant increase in functional psychoses of old age.

6 Life Events

The impact of life events on the manifestation of mental disorders has been studied by many authors. In schizophrenia, for example, several studies which followed the pioneer study by Brown and Birley (1968) confirmed that there was an excess of life events during the interval of 3 weeks to 6 months before manifestation of symptoms (Bebbington and Kuipers 1993). Despite methodological problems (e.g. those linked to the retrospective assessment of life events and precise definition of onset of an illness episode), the evidence obtained so far suggests a triggering effect of life events for the manifestation of episodes of psychotic disorders rather than a simple causal relationship between life events and psychotic illness.

Associations between stressful life events and depressive disorders were reported in different countries in several continents (Brown 1996; Creed 1993). Life events that are of particular importance include those life events which touch upon areas of life of particular subjective importance, life events associated with role conflicts, life events associated with pre-existing stresses such as sudden deterioration in the state of health of a relative (e.g. diagnosis of malignancy in a spouse) against the background of pre-existing long-term illness, separation following long-term marital problems and life-threatening experience (e.g. car accidents). Vulnerability factors have been described. Brown and Harris (1978) reported that only 10% of women with a severe life event and an intact confiding relationship developed depression, whereas depression occurred in 41% of women who experienced such a life event without a confiding relationship. Further vulnerability factors include loss of mother, unemployment, domestic care for three or more children and negative self-concept. Life events can also be relevant for the remission of depressive symptoms. Thus finding a solution for long-term difficulties or experiencing a "fresh start" (e.g. new confiding relationship) might facilitate remission. Moreover, having a positive self-concept can help.

Life events are probably not distributed randomly. Their incidence is influenced by individual and environmental factors. Such factors include social class and neuroticism scores, and personality variables may be of importance (Fergusson and Horwood 1987). There are data suggesting that different regions differ with respect to the type and frequency of life events and vulnerability factors. On a remote Scottish island, integration in the local community (church attendance, "crofting" as the particular type of organisation of rural economy) was important, but it was not important in London; in contrast, both in London and in the island community, lack of a confiding relationship and bringing up three or more children contributed to overall vulnerability. Some studies found that differences in prevalence of depression in different countries (South London, Outer Hebrides, Basque country, Zimbabwe) tallied with corresponding differences in frequency of life events (Brown 1996). On the other hand, an international study of life events before schizophrenic relapse found no significant differences in the frequency of life events between the countries concerned (Day et al. 1987).

7 Social Class

The incidence and prevalence of schizophrenia seems to be higher in the members of lower social classes. Hypotheses of social causation and social selection or drift have been formulated. The social causation hypothesis underscores a higher number of life event stressors, increased exposure with environmental, job and infectious stressors, disadvantages in prenatal care and lower social support under conditions of stress. The social selection or drift hypothesis rests upon data suggesting disadvantages in upward social mobility from one generation to the next and a downward social trend following onset of symptoms in people with schizophrenia. Since the publication of pivotal studies by Faris and Dunham (1939) and Hollingshead and Redlich (1958), causal inferences have become more cautious, and recent work has tended to strengthen the social drift hypothesis (Karno and Norquist 1995; Goldberg and Morrison 1963). Ample evidence suggests the association of socio-demographic factors with psychiatric service utilisation, and psychiatric hospital admissions in particular (Thornicroft 1991). Cox (1993) points out that the manifestation of traumatic life events in children is associated with lower social status.

The prevalence of child psychiatric disorders is higher in families with low socio-economic status than in upper-class families. However, the association is less marked than in adults, and the above-mentioned relationship is also weaker than the association between mental disorder and familial conflicts (Cox 1993). On the other hand, high socio-economic status may have a protective effect with regard to intellectual development (e.g. regarding the effects of perinatal stress on intelligence; Werner 1985).

8 Urban–Rural Differences

Affective and neurotic symptoms can become manifest following change of residence or under the impact of stressors in the immediate living environment (Freeman 1994). A large number of studies in the 1960s and 1970s assessed the mental health effects of living in England's "new towns" (conurbations with rapidly constructed, low-grade high-rise buildings, lack of social infrastructure and sense of community). Non-specific negative psychosocial effects of living in high-rise buildings have been documented. These effects, however, are not independent of other factors, in particular the specific social context. Simple causal relationships have not been established (Freeman 1994).

The rates of child psychiatric disorders are higher in urban areas, particularly in inner city areas, than in rural environments (Cox 1993). Differences are probably related to differences in mental health of parents, parent relationships, the living situation and schools. Pathogenetic links with child psychiatric disorders are not easily understood. The incidence of maternal depression is a possible moderating variable. Covariation of antisocial behaviour of male partners with living environment has been described, which suggests that simple causal models should be avoided (Coleman 1985, Wolkind and Rutter 1985).

Some earlier studies found psychiatric morbidity higher in urban regions. On the other hand, morbidity rates were higher in rural, isolated settlements or so-called "small boom towns" than in more diversified and stable communities (Webb 1984). In the 1950s, a review paper concluded that urban living by itself was not associated with mental ill health more frequently than a rural environment (Leacock 1957). Dohrenwend and Dohrenwend (1974) later came to the conclusion that mental disorders were more prevalent in urban areas. The Epidemiologic Catchment Area Program study of the NIMH (ECA study) did not show a difference in the prevalence of schizophrenia between urban and rural regions once factors such as ethnic group, gender and age had been controlled for (Shapiro et al. 1984). Recent studies have reported associations of urban place of birth and upbringing with a higher risk of developing schizophrenia later in life (Lewis et al. 1992). McCreadie et al. (1997) reported urban–rural differences between South London and a rural area in Scotland, which was however restricted to non-white ethnic groups.

Differences in the prevalence of non-psychotic disorders between urban and rural areas (in Taiwan) suggest interactions of residence with other psychosocial stressors (Cheng 1989). A study in Chicago assessed relationships between residential density (persons per dwelling unit), population density (persons per square mile) and psychiatric admission rates. The association of low residential density with high rates of hospital admission only applied to areas with low population density; in the elderly, there were interaction effects between population density and residential density (Magaziner 1988). Migration from rural into urban areas could result in a relative increase of mental retardation in rural areas and of schizophrenic disorders in urban areas; similar results were reported from China (Cooper and Sartorius 1996). In summary, a multitude of findings argue against simple causal relationships between living environment and mental disorder, but suggest some association between urban residence and morbid risk.

9 Industrial and Developing Countries

Prior to the NIMH ECA study, a total of about 50 studies covering the period 1931–1983 suggested the point prevalence of schizophrenia world-wide to be between 0.6 and 7.1 per 1000 population. The ECA study collected interview data from almost 18,000 individuals in five regions of the USA, and the study resulted in 1-year and life-time prevalence rates of 1.0 and 1.4%, respectively (Karno and Norquist 1995, Shapiro et al. 1984). In the 1960s, the World Health Organization (WHO) initiated and implemented the International Pilot Study of Schizophrenia (IPSS). In this study, more than 1200 patients were interviewed in nine countries (China, Colombia, CSSR, Denmark, India, Nigeria, USSR, UK and USA). It showed that schizophrenia can be found in similar frequency in different countries and that, in terms of symptomatology, similarities outweigh differences across countries. A follow-up study assessing outcome showed a large proportion of patients with good remission in all countries and some cross-national differences, with better clinical course in developing than in developed countries (Sartorius et al. 1987).

The follow-up study (Determinants of Outcome of Severe Mental Disorders, DOS) looked at estimates of incidence of schizophrenia in different parts of the world; 12 centres in ten countries participated (Denmark, India, Colombia, Ireland, USA, Nigeria, USSR, Japan, UK, CSSR; Jablensky et al. 1992). Incidence rates could be calculated in seven centres, and clinical data on treated patients with schizophrenia were reported from all 12 centres. A total of 1379 patients were included, and there was a follow-up study of patients over 5 years. In almost 80% of all patients, follow-up interviews were obtained at the end of 2 years. Broad diagnostic criteria of schizophrenic disorders were fulfilled in 80%, and more than half the patients had a syndrome which fulfilled the criteria for strictly defined "core" schizophrenia. Annual incidence rates for the broader diagnostic concept ranged from 1.5 to 4.2 per 10,000 inhabitants. In the broad diagnostic group, the highest incidence rate was found in India. There were significant differences in incidence across countries, but the rates for the more strictly defined core group varied comparatively little and were all between 0.7 and 1.4 per 10,000. There were no statistically significant differences in incidence rates of schizophrenia. The age of onset was higher in women than in men in both developing and developed countries. A total of 50% of patients had a single episode during the 2-year follow-up period, and 31% had two or more episodes followed by remission; 16% of patients had a chronic course without remission. The type of illness onset ("acute" better than "subacute" or "insidious") and country of origin were found to be the most important predictors of 2-year outcome; favourable outcome was significantly more frequent in developing countries (48%–57%) than in developed countries (6%–26%). Several studies have been undertaken to explain these differences. A study analysing expressed emotion (EE) data from Chandigarh, Aarhus and London found EE in India markedly less conspicuous than in either of the European centres (Wig et al. 1987). Other possible reasons for the differences in outcome were also studied (e.g. the frequency of life events, the perception of patients by relatives), but no definitive explanation of the differences was found. In addition to the type of onset and centre membership, outcome also depended on primary clinical diagnosis, marital state, gender, adolescent adaptation, social network (friends) and misuse of illegal drugs. In the final discussion of their data, the authors emphasise the cross-national and cross-continental similarities of incidence of schizophrenia and the importance of the cultural context for the course and outcome in schizophrenia.

There is no doubt that aspects of culture in general, and the perception of mental health and ill health in particular, are important for outcome in schizophrenia (Leff 1988; Pfeiffer 1994). There are culture-bound differences in psychiatric syndromes, some culture-specific differences in the symptomatology of endogenous psychoses and differences in treatment and outcome. This theme is among the ecological aspects in psychiatry and is addressed in other chapters of this book.

10 Unemployment

A number of papers address psychological and psychiatric sequelae of unemployment. Where mental health service utilisation is concerned, Brenner (1973) reported a much discussed study (Bhugra 1993). He found that unemployment was the most important single factor to explain variations in psychiatric hospital admission in New York over a period of 127 years. Hildebrandt (1994), on the basis of what was formerly Prussia, confirmed this relationship for the period 1876–1906 for alcohol-related psychoses but not for endogenous psychoses. Further authors reported associations between total psychiatric service utilisation and unemployment (Italian case register data; Thornicroft et al. 1993). Kammerling and O'Connor (1993) identified unemployment as a strong predictor of psychiatric admission rates in the UK.

Those who are currently unemployed report more physical and psychological complaints than those previously unemployed who have found a job (Studnicka et al. 1991). In Danish, Spanish and Finnish population samples, associations were found between unemployment and prevalence of total mental disorders (Fink et al. 1995; Vázquez-Barquero et al. 1992 [men only]; Viinamäki et al. 1995a). A Finnish study reported that re-employment and improvement of mental health were associated with each other (Lahelma 1992). There are data on increased prevalence of substance misuse and depressive disorders in the unemployed (Dooley et al. 1996). In schizophrenia, unemployment was a predictor of point prevalence, this relationship being more marked for wide diagnostic criteria than for the group of more narrowly defined schizophrenic disorders (Harvey et al. 1996). A significant association with non-affective endogenous psychoses was found for unemployment in the National Comorbidity Survey in the USA (Kendler et al. 1996). Causal relationships, despite the wealth of evidence available, cannot be inferred, and selection effects need to be considered (Goldney 1996). Viinamäki et al. (1994) attempted to distinguish those people unemployed and mentally ill from those unemployed but not mentally ill. They identified self-respect, social support and the degree of general

somatic complaints as discriminators. Personality was important in coping with unemployment (Viinamäki et al. 1995b).

At the aggregate level, positive associations were found between unemployment and suicide rates in the UK (Appleby 1992; Dooley et al. 1996). Increase in male suicide rates during the 1980s was associated with an increase in unemployment (Pritchard 1995); however, this relationship did not apply when 1970s data for men and women were analysed (Schmidtke 1997). Unemployment is a risk factor for suicide in younger people (Graham and Burvill 1992), and the same applies to the repetition of parasuicidal acts (Morton 1993) and to completed suicide following a prior suicide attempt (Hawton et al. 1993). Schmidtke (1997) reviewed European studies of suicide and parasuicide and concluded that simple causal models cannot deal with the complex findings.

Unemployment was also examined with respect to outcome. There were associations with psychotic relapse (e.g. Scottish Schizophrenia Research Group 1992). In a study by Rabinowitz et al. (1995), unemployment was a predictor of "revolving door" patterns of clinical contact. Tehrani et al. (1996) found loss of contact with psychiatric services associated with unemployment, and Os et al. (1995) reported on associations between unemployment and chronicity. Thornicroft et al. (1992) found the accumulation of so-called new long-stay patients related to unemployment rates. In a study by Wessely et al. (1994), unemployment and the diagnosis of schizophrenia were risk factors for delinquency of the mentally ill.

11 Migration

Mental disorders in migrants and refugees are discussed in other chapters of this book (e.g. Vol. 2, Part 1, Chap. 18). Migrations often involve large segments of a population crossing national boundaries. They imply drastic environmental change for all those involved. Looking at the level of social group alone is not sufficient; it is necessary to understand the interaction of individual- and group-level effects. Some studies reported a higher incidence of schizophrenia among migrant populations. On the other hand, findings in Australia suggest complex interactions between the situation in the country of origin, the level of life stress prior to emigration, stress associated with migration, the age distribution of migrants, growth of the migrant community and integration in the new social environment (Krupinski 1984). Methodological problems include the instability of family ties, the impact of additional factors such as social class, age or gender and the lack of a reference population in the country of origin. In addition, individual circumstances of the migration process are of great importance as they affect subjective appraisal (degree to which migration was voluntary, degree of preparation, time course). Some countries require evidence that applicants for an immigration visa do not suffer from a mental disorder, which clearly biases samples of those who arrive in the recipient country.

In London, the ethnic minority group of black Caribbean origin has a higher risk of developing non-affective functional psychoses (see Davies et al. 1995). However, methodological problems suggest that simple causal links between migration process and prevalence rates are unlikely (Harrison et al. 1988). In addition, the NIMH ECA study reported lower prevalence rates for schizophrenia among Mexican immigrants in Los Angeles. Selection effects need to be taken into consideration either way (i.e. in the case of reduced or increased prevalence rates). In a classic study (in Norwegian immigrants in Minnesota), Ödegaard (1932) differentiated between the effects of immigration stress on the one hand and selective migration on the other and weighed their impact in different disorders. Re-emigration to Norway was also associated with increased psychiatric morbidity. A review of a wide range of studies by Schwab and Schwab (1978) resulted in no clear picture, and this field of research is highly complex, as it requires the simultaneous examination of individual- and group-level phenomena. A critical methodological review by Leff (1988) highlights the range of methodological problems. The situation concerning migrants is clearly different from that concerning refugees and displaced persons (individuals forcibly relocated in the same country). These groups are at higher risk for mental disorders than migrants.

12 Disasters and Mental Health

The issues raised by psychiatric sequelae of disasters are covered by Bromet (Vol. 2, Part 1, Chap. 17). Environmental stressors cover a wide range: war and persecution, severe industrial accidents, natural disasters (e.g. volcanic eruption), other disasters (bursting of dam, see Buffalo Creek disaster) and possibly terrorist attacks. According to Thompson (1990), the sequels of the Three Mile Island accident in Harrisburg underscore the psychological dimension of the effects of such accidents. The prevailing feeling in the population concerned was described as a sense of insecurity about which information regarding the events at the nuclear power plant could be relied

upon. Eighteen months after the accident, there was a clear, albeit not dramatic increase in psychological complaints. There was an increase in the number of medical consultations and an increase in frequency of anxiety disorders. Goldhaber et al. (1983) describe the Three Mile Island accident as primarily exerting psychological stress.

In the case of the nuclear accident in Chernobyl in 1986, there were severe immediate, medium- and long-term health hazards. Severe health risks applied to a population of 600,000 people (Pivak 1992); there was acute radiation sickness in more than 200 individuals, and there were 28 deaths in the acute phase. Pivak (1992) reports on a study of 1572 patients who were surveyed during the acute phase, in the course of the first 6 months and in further follow-up examinations. Acute stress reactions (post-traumatic stress disorder, PTSD; anxiety; retardation/agitation) were found in about 75% of patients both in the acute phase and after 6 months, and this proportion decreased to 29% in the course of follow-up. While stress reactions decreased, neurotic disorders (asthenia, hypochondriasis, apathy and histrionic syndromes) increased from 13% to 29%. Psychoses were relatively rare (3.8% during the acute phase, 0.7% at 6 months: described as "reactive" and "hysterical" psychoses). In a study organised by the International Atomic Energy Agency (IAEA) which examined a large sample of about 1300 inhabitants in 13 villages, no immediate clinical radiation effects were found. However, the general level of anxiety and stress appeared high. The importance of detailed information for the population on exposure and consecutive risks was emphasised (Ginzburg 1993). Ginzburg and Reis (1991), in a large population sample of people exposed to the accident, reported reactive depression, PTSD and familial disorganisation as the most prevalent mental health consequences. Alexandrowski et al. (1992) reported on a study of 300 people exposed to the accident. They describe an increase in chronic medical disorders (e.g. hypertension), the phenomenon of "worrying tension" in many probands and five groups of maladaptation states: one group had neurasthenia (67 probands), and a second group psychosomatic disorders (77 probands); a third group was found to have irritability and histrionic features. There was a fourth group without clinical symptoms (the authors referred to this group as presenting with "preclinical" forms of disadaptation). Mistrust of official policy and guidelines were frequently reported. A total of 15 probands were considered to be in good mental health. Alexandrowski et al. (1992) conclude that their observations most likely reflect abnormal personality development under chronic stress exposure. Viinamäki et al. (1995c) reported on a study which included a control group without exposure to radiation. Seven years after the Chernobyl accident, they used the General Health Questionnaire (GHQ) in the highly contaminated region of Bryansk in a group of 325 individuals who had been resident in the area throughout, and they reported long-term impairment in mental well-being, this difference being significant in women compared with the control group without history of radiation exposure.

Ecological aspects are conspicuous, e.g. in studies of the Buffalo Creek mud slide disaster which killed 125 people in 1972 and made 4000 of the local residents homeless in a very brief time period (Titchener and Kapp 1976). Traumatic neurotic reactions (PTSD: grief, guilt, shame, rage or hopelessness) were found in 80% of those surviving and persisted over a 2-year period ("Buffalo Creek syndrome"). There are descriptions of how the disaster disrupted the close network of social support in the mining village, i.e. the local community which had been sustained by and had sustained the local residents. Six hundred and twenty-five of the Buffalo Creek survivors sued the mining company and claimed compensation due to the psychological impairment experienced. In an out-of-court settlement, they were awarded $13.5 million, of which $6 million was assigned on the basis of a scoring system based on proximity to the disaster (Stern 1976).

The nosological justification for the concept of PTSD has been confirmed by a variety of reports and studies from the USA and the UK referring to studies on individuals involved in the Vietnam war, natural disasters and ship and aircraft disasters. Psychiatric consequences of particularly stressful individual experience (e.g. torture and rape) can also be included in this category (Thompson 1990; Titchener and Kapp 1976; Williams et al. 1993). PTSD was incorporated in DSM-III, DSM-III-R, DSM-IV and ICD-10 (Williams et al. 1993). The syndrome comprises recurrent inner "re-enactment", dreams and behaviour related to the event (re-playing the experience in children). Those experiencing such syndromes tend to report avoiding associated stimuli, loss of interest, experiences of alienation and "numbness", irritability, insomnia and over-alertness. There are risk and vulnerability factors – not all those exposed develop PTSD. Some factors are related to the traumatic event itself (type of stimulus, closeness to victims, type and intensity of threat), while other vulnerability markers can be found in those exposed to trauma (interpretations, attributional style, additional life events, coping strategy and psychiatric history). There is some evidence that social support can reduce the likelihood of PTSD (Viinamäki et al. 1995c). The literature on PTSD covers a range of ecological mechanisms in looking at the effects of a serious stressor on social communities at a defined point in time under similar conditions.

The mental health consequences of the holocaust cannot be described in detail here. This is the focus of

another chapter in this handbook (Vol. 2, Part 1, Chap. 19). A comprehensive review on this topic is given by Peters (1989). In their book entitled *Psychiatrie der Verfolgten* ("Psychiatry of the Persecuted"), von Baeyer et al. (1964) presented their findings from a large series of medico-legal reports on psychiatric sequelae following Nazi persecution and concentration camp stays in particular. There are also a variety of chapters and monographs on the topic (Eitinger 1980; Niederland 1977; Stoffels 1991). von Baeyer et al. (1964) found a

Relatively homogeneous core syndrome with chronic anxiety, depressive mood and asthenia which after a latent period with symptoms of fatigue and exhaustion becomes manifest and is explained by the various authors according to their orientation in terms of neuropathology, psychosomatics or psychodynamic theory.

Peters (1989) used the concept of the survivors' syndrome coined by Niederland (1977) and summarised the literature on the multiple mental health consequences of the survivor syndrome. Symptoms comprise anxiety, fighting memory, tension, ruminations, a feeling of survival guilt, irritability and instability of mood, affective numbness, lack of initiative, apathy, restlessness, concentration and other cognitive deficits, non-responsiveness of mood, quest for meaning, lack of self-value and self-assertiveness, social insecurity and psychosomatic disorders. In addition, Peters (1989) described the peculiarities of Nazi persecution with respect to their role in the survivors' syndrome, mentioning the following: loss of home and culture of origin, non-respect, discrimination, social isolation, uprooting, mistreatment, lack of food, cold and other extreme physical conditions. Additional factors were the lack of judicial security, loss of individuality, permanent and recurrent acute fear of imminent death, the small number of survivors and a situation of individual, family and collective lack of identifiable burial sites.

13 Intoxication, Neurotoxins and Exogenous Psychoses

Huber (1988) wrote a comprehensive handbook chapter on the organic psychiatric disorders. He reported on psychiatric syndromes following intoxication and the effects of neurotoxins and of nutritional and vitamin deficiency states. Environmental medicine deals with acute and chronic intoxications of the nervous system through environmental agents. Acute intoxications as a rule entail central nervous system symptoms. The non-specific character of organic psychiatric syndromes in acute intoxication (disorientation, confusion, delirium, somnolence, stupor, coma) and the multitude of toxic substances make it difficult to establish firm causal connections between symptoms and toxins. In the case of some substance groups, the transition from medically tolerable exposure to chronic intoxication is not clearly defined, and multiple intoxications present particular problems. The clinical syndromes in acute and chronic intoxication are varied, and all organ systems can be affected (acute hepatic or renal failure, lung oedema, arrhythmia, cardiovascular shock, bone marrow depression, skin alterations). Delirium and impaired consciousness which may deteriorate to coma are the most important acute neuropsychiatric syndromes, while neurological findings include seizures, extrapyramidal features, ataxia, cerebellar symptoms or neuropathies. Chronic psychiatric consequences of intoxication include so-called pseudo-neurasthenic syndromes, organic personality change, organic affective disorders and dementias.

Neurotoxins lead to functional and/or structural impairment of the central nervous system. The range of noxious agents includes metals, organic solvents and pesticides. Table 3 summarises the neuropsychiatric sequelae of exposure to heavy metals, organic solvents and pesticides. The neurotoxic effects of solvents provide an example of the psychiatric effects of neurotoxins, and there has been considerable public and media coverage of what has been referred to as Danish painters' disease (Mikkelsen 1980; Triebig 1990). Long-term exposure to organic solvents (e.g. paint, varnish, glue) can induce organic brain syndromes which range from fatigue, irritability, subtle cognitive impairment and headache to marked dementia syndromes with neurological impairment, sulcal atrophy and ventricular enlargement on computed tomography (CT) (Triebig 1990). The pathogenesis of these states is still unclear because there are no clear correlations between dose and duration of solvent exposure and the severity of encephalopathy.

In the 1950s, significant lead exposure from fish led to severe neuropsychiatric syndromes in Minamata Bay, Japan (see Table 3). There was marked organic psychopathological impairment in some patients, and the degree of psychiatric impairment varied with the degree of exposure, but also with the compensation status of patients, i.e. it varied according to whether patients had claimed compensation and whether their claim had been successful (Sugisawa 1994).

Low-level exposure is another topic relevant in this context, and a series of studies have examined the issue of low-level lead exposure in children. Needleman et al. (1979) reported an inverse relationship between intelligence (as assessed by IQ) and the degree

Table 3. Selection of characteristic neurotoxins and psychiatric effects

Noxious agent	Industrial/commercial use	Psychiatric symptoms/syndromes	Neurological and other medical sequelae
Heavy metals			
Arsenic	Production of glass, semi-conductors, melting of copper, lead; insecticides, herbicides	Anxiety, amnestic deficits	Headache, peripheral neuropathy
Lead	Foodstuffs, water, air, dust, soil; metal industry, paint, petrol, pigments	Affective instability, irritability, hallucinations, lethargy, mnestic and concentration deficits, disorientation, word-finding and visual–spatial impairment	Polyneuropathy, myoclonus, intelligence deficit (early childhood exposure); brain oedema (acute/severe exposure)
Manganese	Manganese ore processing, production of paint, varnish, disinfectants	Irritability, obsessive–compulsive symptoms	Parkinsonian features, hyperreflexia
Mercury	Scientific instruments, alloys, paint, contaminated fish, pesticides	Depression, insomnia, fatigue, irritability, mnestic and concentration deficits, impairment in logical discourse	Peripheral neuropathy, ataxia, dysarthria, hearing loss, loss of vision, spasticity, extrapyramidal syndromes, coma, brain oedema, cerebellar and cortical atrophy, lethal complications
Tin	Widespread use	Affective instability, depression, loss of libido, insomnia, loss of orientation and mnestic functions, impairment of verbal memory and visual–motor coordination	
Organic solvents			
Carbon disulphide	Insecticides, plastic industry	Behavioural abnormalities, insomnia, irritability, affective instability, mnestic and attentional deficits	Peripheral neuropathies (acute intoxication), extrapyramidal syndromes, cerebral ischaemia
n-Hexanes, methyl-butyl-ketones	Shoe production, clothing industry, plastic, cosmetics		Peripheral neuropathies
Perchlorethylene (PCE)	Launderette, clothing industry	Irritability, sleepiness, confusion, impaired mnestic function and concentration	Dizziness, dysarthria, impairment of coordination
Trichlorethylene (TCE)	Industrial solvents	Fatigue, confusion, mental slowing, attentional impairment	Cranial nerve symptoms (trigeminal nerve)
Toluene	Printing, leather goods production, exhaust fumes	Irritability, fatigue, inadequate affect, affective instability, suicidal ideation, euphoria, cognitive impairment, confusion	Skin abnormalities
Pesticides			
Organophosphates	Insecticides (cholinesterase inhibitor)	Fatigue, restlessness, agitation, anxiety, emotional instability, irritability, impaired mnestic function and concentration	Cholinergic syndrome, coma
Carbaryl	Insecticides (reversible cholinesterase inhibitor)	Emotional instability, irritability, loss of drive, impaired mnestic function	Cholinergic syndrome, coma

of low-level lead exposure. Since that early study, there has been a range of studies on the topic, and Pocock et al. (1994) published a meta-analysis of the data from five prospective studies, including a total sample of more than 1100 children. Plasma lead levels in maternal or umbilical cord blood showed a lack of association with IQ in children over 5 years of age, while blood lead levels at the age of 2 resulted in a weak but significant inverse relationship. Cross-sectional analyses (14 studies comprising 3499 children) showed a significant negative relationship with considerable variation in findings. The seven cross-sectional studies based on lead content in teeth were more consistent in terms of findings and showed smaller effect size. The authors concluded that doubling blood lead levels (from 0.48 to 0.97 μmol/l) was associated with an average reduction in total IQ of 1–2 points. The methodological problems of these studies include representativeness of samples, comprehensive correction for confounding variables, selection bias and reverse causality. Pocock et al. (1994) are cautious regarding the priority which should be given to the recognition and reduction of slight increases in blood lead levels. An Australian study published since found a conspicuous association between environmental lead exposure and the IQ of 11- to 13-year-olds (Tong et al. 1996).

14
Noise and Mental Health Sequels

Exposure to noise is possibly the one environmental stressor which an urban population cannot escape. Among the consequences of long-term exposure to noise, deterioration of hearing and deafness need to be considered. Hearing impairment and deafness are associated with psychiatric morbidity in general and with manifestation of delusions in particular (Tarnopolsky and Clark 1984). Urban noise is associated with subjective reactions, most frequently the subjective impression of intrusion and anger. Irritability, insomnia, headache and tension can be observed following intensive noise exposure, and those who are exposed to noise at the workplace report similar complaints. Research findings suggest that anger is an important mediator between noise as a stressor and symptoms. Subjective appraisal can affect the intensity of stress reactions which are non-specific at a symptom level. Noise exposure, anger and socio-demographic variables interact with each other; there are, however, no simple causal relationships between noise and mental illness. In women with the maximum level of sensitivity to noise, Stansfeld et al. (1985) found significantly more psychiatric symptoms, higher neuroticism scores and a more marked reactivity to other (non-noise) stressors when they were compared with the least noise-sensitive women. In summary, subjective sensitivity to noise is associated both with sensitivity to stress and with neurotic depression, but not with other psychiatric disorders (Stansfeld 1992).

15
Protective Factors

Social networks and social support are of particular importance in protecting against the impact of psychosocial and other stressors. Brugha (1995) distinguishes instrumental, mostly material support and emotional aspects of social support. Regular positive feedback as one component of social support can exert direct, positive effects on mental well-being, it can improve coping resources and thereby act as a stress buffer and eventually improve self-appraisal (Brugha 1995; Lloyd 1995). Social support is important not only in the mentally ill; there are clear associations of a network of supportive social relationships and general mental well-being (Coyne and Delongis 1986). Life event research points to the lack of an intimate, confiding partner as an important vulnerability factor for manifestation of depressive disorders, which in turn underlines the protective function of social support (Brown 1996). An up-to-date review of relationships between social networks and social support on the one hand and a range of health indicators on the other suggests that social integration contributes to a reduction in general mortality rate and to improved mental well-being (Seeman 1996). Social support has also been shown to have a protective effect in alleviating the mental health effects of unemployment (Viinamäki et al. 1993).

16
Conclusion

Environmental aspects are undoubtedly relevant to mental health and psychiatric disease. Many of the findings reported, however, reflect correlations in groups of patients; in other words, causal links cannot easily be established. Correlations may be misleading where they are aggregates of different mechanisms which account for different types of relationship. Such complex associations may not be unravelled by an analysis restricted to the aggregate level (Riley 1963).

Group effects (increased morbid risk in a population) need to be distinguished from associations at the individual level (illness- or relapse-associated factors).

An ecological fallacy is possible where effects at one of these levels of analysis are taken to support conclusions at the other level (Selvin 1958). In the past, group effects at the ecological level have been considered less binding in a hierarchical model than findings of biomedical research at the individual level, and it has been assumed traditionally that such group-level data can only be used to generate hypotheses. However, a variety of findings suggest that the group (or "ecosystem") level permits the establishment of associations which are relevant to the understanding of mental disorders (Schwartz 1994). In the context of suicidal behaviour, this implies that a high-risk region (with high suicide rates) is not merely a place where individuals with a combination of individual risk factors are found, but that such an ecosystem has some characteristics which are specific to the place and social system rather than to each individual residing within it. Schmidtke (1997), in a review paper, discusses variation in suicide rates in Europe, which is by a factor of 22 and 20 for men and women, respectively. A simple biological or socio-demographic approach does not account for these differences; complex socio-cultual variables such as attitudes towards suicidal behaviour, type of media coverage ("labelling"), the attitudes towards the elderly and cultural variation in individual coping strategies are clearly at play. Schmidtke (1997) considers cross-cultural differences in the attitudes towards suicidal behaviour a particularly important factor accounting for variation in suicide rates. Neeleman et al. (1997) also report on associations between suicide tolerance and suicide rates. In male patients in 19 Western countries, they found a modulating (ecological) effect of the strength of religious beliefs (at national level) on the strength of (negative) associations between individual suicide tolerance and religious belief. Not everyone in a high-risk environment develops a particular psychiatric disorder. There are some people in a community who are at higher individual risk of developing symptoms, and it is likely that there is an ecological contribution to the distribution of these individual risk factors too.

Ecological research, within psychiatry, provides important clues to the generation and examination of hypotheses. It complements work done in the field of cross-cultural psychiatry (Leff 1988; Pfeiffer 1994) and epidemiology. Methodological problems of selection and reverse causality need to be taken into consideration (e.g. urban–rural differences, importance of social class and migration). Life event research has made substantial contributions to research into the pathogenesis of depression. This field of work has permitted researchers to look into pathogenetic mechanisms in some detail, and the interaction of life events with individual vulnerability factors and ongoing difficulties is important. Comparative research in different socio-economic and cultural settings offers scope for further development (Brown 1996). The pathogenetic life event mechanisms established so far do not apply to delusional depression and depression with melancholic features (Brown et al. 1994). This does not mean that social and environmental factors are irrelevant, but it suggests that the approach is strongest at the less severe end of the spectrum of depressive disorders.

Ecological work has also come to fruition in the field of developmental research in children and adolescents. This type of research has looked at depression, substance misuse, delinquency and successful adaptation to environmental stressors. Ecological research, however, should not be restricted to mental disorders in the young; it should include the whole range of successful and failed adaptation to environmental challenges (Fergusson 1996). In schizophrenia, the importance of environmental variables (such as social deprivation, housing, population structure or unemployment) has been established. Utilisation of psychiatric services and outcome of schizophrenic disorders are modulated by social factors, and the clarification of outcome-moderating variables is one important aspect of psychiatric research.

Ecological studies of service use can be used in service planning. However, research into differences in incidence across communities does not clarify illness onset in any individual patient. The latter requires the analysis of aetiological factors, pathogenetic mechanisms and susceptibility factors. Interactions between the two levels are best considered in a vulnerability–stress model.

Susser and Susser (1996), in a comprehensive conceptual paper, describe a new paradigm for epidemiological research which they describe using the image of successively smaller Chinese boxes put into each other – from the broader ecosystem to the individual level. For instance, in considering mental health consequences of unemployment, different levels of analysis need to be taken into consideration. At the most aggregate social level, the macro-economic environment is important. Mental health consequences of unemployment are obviously more likely to be relevant in a period of mass unemployment than they are while full employment prevails. At the next level down, regional or local community characteristics are likely to affect both the extent of the problem and the availability of social support as a resource in coping with the stressor. Further down, coping in families or smaller social networks are relevant as protective factors. Furthest down in this hierarchy, the degree of vulnerability (self-appraisal, personality), at the individual level, is relevant to the manifestation of a mental disorder.

The complexity of such multi-level analyses is underlined by the fact that coping resources, relative risk and vulnerability at the various levels are not distributed at random, but related to the specific location of a person in his or her respective ecosystem (and also to the position of the local community in the wider ecosystem). The strength of the ecological perspective lies in the simultaneous examination of multiple influencing factors (patient – social group – environment) and their interactions. This degree of complexity both reflects the strength of the ecological approach and the methodological pitfalls. Susser and Susser (1996) coined the concept of "eco-epidemiology" and consider multi-level analyses particularly helpful. The ecological analysis of environmental aspects of mental health and psychiatric disorders presents a promising challenge to the field of psychiatry.

17 References

Alexandrowski JA, Rumjanzewa GM, Jurow WW, Martjuschow AA (1992) Dynamik der psychischen Desadaptationszustände unter chronischem Streß bei Bewohnern der Gebiete, die beim GAU im Kernkraftwerk Tschernobyl in Mitleidenschaft gezogen wurden. Psychiatr Prax 19: 31–34

Appleby L (1992) Suicide in psychiatric patients: risk and prevention. Br J Psychiatry 161: 749–758

Barton R (1959) Institutional neurosis. Wright, Bristol

Baxter PJ (1990) Review of major chemical incidents and their medical management. In: Murray V (ed) Major chemical disasters – medical aspects of management. Royal Society of Medicine Services, London, p 7

Bebbington P, Kuipers L (1993) Social causation of schizophrenia. In: Bhugra D, Leff J (eds) Principles of social psychiatry. Blackwell, Oxford, p 82

Bhugra D (1993) Unemployment, poverty and homelessness. In: Bhugra D, Leff J (eds) Principles of social psychiatry. Blackwell, Oxford, p 355

Brenner MH (1973) Mental illness and the economy. Harvard University Press, Cambridge, MA

Bronfenbrenner U (1979) The ecology of human development: experiments by nature and design. Harvard University Press, Cambridge, MA

**Brown GW (1996) Genetics of depression: a social science perspective. Int Rev Psychiatry 8: 387–401

Brown GW, Birley JLT (1968) Crises and life changes and the onset of schizophrenia. J Health Soc Behav 9: 203–214

Brown GW, Harris TO (1978) Social origins of depression: a study of psychiatric disorder in women. Tavistock, London

Brown GW, Andrews B, Harris TO, Adler Z, Bridge L (1986) Social support, self-esteem and degression. Psychol Med 16: 813–831

Brown GW, Harris TO, Hepworth C (1994) Life events and endogenous depression. A puzzle reexamined. Arch Gen Psychiatry 51: 525–534

Brugha TS (1995) Social support and psychiatric disorder: overview of evidence. In: Brugha TS (ed) Social support and psychiatric disorder. Cambridge University Press, Cambridge, p 1

*Burgess EW, Brogue DJ (eds) (1967) Urban sociology. University of Chicago Press, Chicago

Burgess EW, Park RE (1925) The city. University of Chicago Press, Chicago

Cheng TA (1989) Urbanisation and minor psychiatric morbidity. A community study in Taiwan. Soc Psychiatry Psychiatr Epidemiol 24: 309–316

Coleman A (1985) Utopia on trial: vision and reality in planned housing. Shipman, London

Cooper JE, Sartorius N (eds) (1996) Mental disorders in China. Gaskell, London

Cox A (1993) Social factors in child psychiatric disorders. In: Bhugra D, Leff J (eds) Principles of social psychiatry. Blackwell, Oxford, p 202

Coyne JC, Delongis A (1986) Going beyond social support: the role of social relationships in adaptation. J Consult Clin Psychol 54: 457–460

Creed F (1993) Life events. In: Bhugra D, Leff J (eds) Principles of social psychiatry. Blackwell, Oxford, p 144

Davies S, Thornicroft G, Leese M, Higgingbotham A, Phelan M (1995) Ethnic differences in risk of compulsory psychiatric admission among representative cases of psychosis in London. Br Med J 312: 533–537

Day R, Nielsen JA, Korten A, Ernberg G, Dube KC, Gebhart J, Jablensky A, Leon C, Marsella A, Olatawura M et al (1987) Stressful life events preceding the acute onset of schizophrenia: a cross-national study from the World Health Organization. Cult Med Psychiatry 11: 123–205

Dohrenwend BP, Dohrenwend BS (1974) Psychiatric disorders in urban settings. In: Arieti S, Caplan G (eds) American handbook of psychiatry, vol 29, 2nd edn. Basic, New York

Dörner K, Plog U (1996) Irren ist menschlich. Lehrbuch der Psychiatrie/Psychotherapie, 3. Aufl. Psychiatrie-Verlag, Bonn

Dooley D, Fielding J, Levi L (1996) Health and unemployment. Annu Rev Public Health 17: 449–465

Eitinger L (1980) The concentration camp syndrome and its late sequelae. In: Dimsdale JE (ed) Survivors, vicitims, and perpetrators. Essays on the Nazi holocaust. Hemisphere, Washington, p 127

Faris REL, Dunham HW (1939) Mental disorders in urban areas. University of Chicago Press, Chicago

Feldman RG (1991) Effects of toxins and physical agents on the nervous system. In: Bradley WG, Daroff RB, Fenichel GM, Marsden CD (eds) Neurology in clinical practice, vol 2. Butterworth-Heinemann, Boston, p 1186

Fergusson DM (1996) Critical notice on: Psychosocial disorders in young people: time trends and their causes. Edited by M Rutter & DJ Smith. Wiley, Chichester, 1995. J Child Psychol Psychiatry 37: 485–487

Fergusson DM, Horwood LJ (1987) Vulnerability to life events exposure. Psychol Med 17: 739–749

Festinger L, Schachter S, Back K (1959) Social pressures in informal groups. A study of human factors in housing. Tavistock, London

Fink P, Jensen J, Borgquist L, Brevik JI et al (1995) Psychiatric morbidity in primary public health care: a Nordic multicentre investigation. I. Method and prevalence of psychiatric morbidity. Acta Psychiatr Scand 92: 409–418

Frankenberg R (1966) Communities in Britain. Social life in town and country. Penguin, Harmondsworth

Freeman HL (1994) Schizophrenia and city residence. Br J Psychiatry 164[Suppl 23]: 39–50

Ginzburg HM (1993) The psychological consequences of the Chernobyl accident – findings from the International Atomic Energy Agency Study. Public Health Reports 108: 184–92

Ginzburg HM, Reis E (1991) Consequences of the nuclear power plant accident at Chernobyl. Public Health Reports 106: 32–40

*Goffman E (1961) Asylums: essays on the social situation of mental patients and other inmates. Doubleday, New York

Goldberg EM, Morrison SL (1963) Schizophrenia and social class. Br J Psychiatry 109: 785–802

Goldhaber MK, Tokuhata GK, Digon E, Caldwell GG, Stein GF, Lutz G, Gur D (1983) The Three Mile Island Population Registry. Public Health Reports 98: 603–609

Goldney RD (1996) Unemployment and health. Aust NZ J Psychiatry 30: 309–311

Graham C, Burvill PW (1992) A study of coroner's records of suicide in young people, 1986–88 in Western Australia. Aust NZ J Psychiatry 26: 30–39

Harrison G, Owens D, Holton A, Neilson D, Boot D (1988) A prospective study of severe mental disorder in Afro-Caribbean patients. Psychol Med 18: 643–657

Harvey CA, Pantelis C, Taylor J, McCabe PJ, Lefevre K, Campbell PG, Hirsch SR (1996) The Camden schizophrenia surveys. II. High prevalence of schizophrenia in an inner London borough and its relationship to socio-demographic factors. Br J Psychiatry 168: 418–426

Hawton K, Fagg J, Platt S, Hawkins M (1993) Factors associated with suicide after parasuicide in young people. Br Med J 306: 1641–1644

Hildebrandt H (1994) Mental disorders and economic crisis: a study on the development of admission into the psychiatric hospitals of Prussia between 1876 and 1906. Soc Psychiatry Psychiatr Epidemiol 29: 190–196

*Hollingshead AB, Redlich FC (1958) Social class and mental illness. Wiley, New York

Huber G (1988) Körperlich begründbare psychische Störungen. Endokrinopathien, Generationsvorgänge, Vitaminmangel und Hirntumoren. In: Lauter H, Kisker KP, Stömgren E (eds) Psychiatrie der Gegenwart, vol 6: Organische Psychosen. Springer, Berlin Heidelberg New York, p 197

Jablensky A, Sartorius N, Ernberg G, Anker M, Korten A, Cooper JE, Day R, Bertelsen A (1992) Schizophrenia: manifestations, incidence and course in different cultures. A World Health Organization Ten-Country Study. Psychol Med Monogr Suppl 20: 1–97

Kammerling RM, O'Connor S (1993) Unemployment rate as predictor of rate of psychiatric admission. Br Med J 307: 1536–1539

Karno M, Norquist GS (1995) Schizophrenia: epidemiology. In: Kaplan HI, Sadock BJ (eds) Comprehensive textbook of psychiatry VI, vol 1, 6th edn. Williams and Wilkins, Baltimore, p 902

Kendler KS, Gallagher TJ, Abelson JM, Kessler RC (1996) Lifetime prevalence, demographic risk factors, and diagnostic validity of nonaffective psychosis as assessed in a US community sample. The National Comorbidity Survey. Arch Gen Psychiatry 53: 1022–1031

Krupinski J (1984) Changing patterns of migration to Australia and their influence on the health of migrants. Soc Sci Med 18: 927–937

Lahelma E (1992) Paid employment, unemployment and mental well-being. Psychiatria Fennica 23: 131–144

Langolf GS, Chafflin DB, Henderson R, Whittle HP (1978) Evaluation of workers exposed to elemental mercury using quantitative tests of tremor and neuromuscular functions. Am Indust Hyg Assoc J 39: 976–984

Leacock E (1957) Three social variables and the occurrence of mental disorder. In: Leighton AH, Clausen JA, Wilson RN (eds) Explorations in social psychiatry. Basic, New York

**Leff J (1988) Psychiatry around the globe. A transcultural view. Gaskell, London

Lewis G, David A, Andreasson S, Allebeck P (1992) Schizophrenia and city life. Lancet 340: 137–140

Lloyd C (1995) Understanding social support within the context of theory and research on the relationship of life stress and mental health. In: Brugha TS (ed) Social support and psychiatric disorder. Cambridge University Press, Cambridge, p 41

Magaziner J (1988) Living density and psychopathology: a re-examination of the negative model. Psychol Med 18: 419–431

Mann A (1993) Epidemiology. In: Bhugra D, Leff J (eds) Principles of social psychiatry. Blackwell, Oxford, p 25

McCreadie RG, Leese M, Tilak-Singh D, Loftus L, Macewan T, Thornicroft G (1997) Nithsdale, Nunhead and Norwood: similarities and differences in prevalence of schizophrenia and utilisation of services in rural and urban areas. Br J Psychiatry 170: 31–36

Mikkelsen S (1980) A cohort study of disability pension and health among printers with special regard to disabling presenile dementia as an occupational disease. Scand J Soc Med 34: 16

Moen P (1995) Introduction. In: Moen P, Elder GH, Lüscher K (eds) Examining lives in context. Perspectives on the ecology of human development. American Psychological Association, Washington, DC, p 1

Morton MJ (1993) Prediction of repetition of parasuicide: with special reference to unemployment. Int J Soc Psychiatry 39: 87–99

Murphy GE, Wetzel RD, Robins E, McEvoy L (1992) Multiple risk factors predict suicide in alcoholism. Arch Gen Psychiatry 49: 459–463

Needleman HL, Gunnoe C, Leviton A, Reed R, Peresie H, Maher C et al (1979) Deficit in psychologic and classroom performance of children with elevated dentine lead levels. N Engl J Med 300: 689–695

Neeleman J, Halpern D, Leon D, Lewis G (1997) Tolerance of suicide, religion and suicide rates: an ecological and individual study in 19 Western countries. Psychol Med 27: 1165–1171

Niederland WG (1977) Das Überlebenden-Syndrom in klinischer und gutachterlicher Sicht. Freiheit und Recht, Bonn

Ödegaard Ö (1932) Emigration and insanity: a study of mental disease among the Norwegianborn population of Minnesota. Levin and Munksgaards, Copenhagen

Peters UH (1989) Die psychischen Folgen der Verfolgung. Das Überlebenden-Syndrom. Fortschr Neurol Psychiat 57: 169–191

Pfeiffer (1994) Transkulturelle Psychiatrie, 2nd edn. Thieme, Stuttgart

Pivak LI (1992) Psychiatric aspects of the accident at Chernobyl Nuclear Power Station. Eur J Psychiatry 6: 207–212

*Pocock S, Smith M, Baghurst P (1994) Environmental lead and children's intelligence: a systematic review of epidemiological evidence. Br Med J 309: 1189–1197

Pritchard C (1995) Unemployment, age, gender and regional suicide in England and Wales 1974–90: a harbinger of increased suicide for the 1990s? Br J Social Work 25: 767–790

Rabinowitz J, Mark M, Popper M, Slyuzberg M, Munitz H (1995) Predicting revolving-door patients in a 9-year national sample. Soc Psychiatry Psychiatr Epidemiol 30: 65–72

Riley MW (1963) Sociological research. I. A case approach. Harcourt, Brace and World, New York

*Rutter M, Smith DJ (eds) (1995) Psychosocial disorders in young people. Time trends and their causes. Wiley, Chichester

Rutter M, Champion L, Quinton D, Maughan B, Pickles A (1995) Understanding individual differences in environmental-risk exposure. In: Moen P, Elder GH, Lüscher K (eds) Examining lives in context. Perspectives on the ecology of human development. American Psychological Association, Washington, DC, p 61

Sartorius N, Jablensky A, Ernberg G, Leff J, Korten A, Gulbinat W (1987) Course of schizophrenia in different countries: some results of a WHO international comparative 5-year follow-up study. In: Häfner H, Gattaz WF, Janzarik W (eds) Search for the causes of schizophrenia. Springer, Berlin Heidelberg New York, p 107

**Sartorius N, Nielsen JA, Strömgren E (1989) (eds) Changes in frequency of mental disorder over time: results of repeated surveys of mental disorders in the general population. Acta Psychiatr Scand 79[Suppl 348]

*Schmidtke A (1997) Perspective: suicide in Europe. Suicide Life Threat Behav 27: 127–136

Schwab JJ, Schwab ME (1978) Sociocultural roots of mental illness. An epidemiologic survey. Plenum, New York

Schwartz S (1994) The fallacy of the ecological fallacy: the potential misuse of a concept and the consequences. Am J Public Health 84: 819–824

Scottish Schizophrenia Research Group (1992) The Scottish First Episode Schizophrenia Study. VIII. Five-year follow-up: clinical and psychosocial findings. Br J Psychiatry 161: 496–500

Seeman TE (1996) Social ties and health: the benefits of social integration. Ann Epidemiology 6: 442–451

Selvin H (1958) Durkheim's suicide and problems of empirical research. Am J Sociol 63: 607–619

Shapiro S, Skinner EA, Kessler LG, Von Korff M, German PS, Tischler GL, Leaf PJ, Benham L, Cottler L, Regier DA (1984) Utilization of health and mental health services. Arch Gen Psychiatry 41: 971–978

Stansfeld SA (1992) Noise, noise sensitivity and psychiatric disorder: epidemiological and psychophysiological studies. Psychol Med Monogr Suppl 22: 1–44

Stansfeld SA, Clark CR, Jenkins LM, Tarnopolsky A (1985) Sensitivity to noise in a community sample. I. Measurement of psychiatric disorder and personality. Psychol Med 15: 243–254

Stern GM (1976) From chaos to responsibility. Am J Psychiatry 133: 300–301

Stoffels H (1991) (ed) Schicksale der Verfolgten. Psychische und somatische Auswirkungen von Terrorherrschaft. Springer, Berlin Heidelberg New York

Streit B (1994) Ökologie kurzgefaßt. BI, Mannheim

Strömgren E, Nielsen JA, Sartorius N (1989) Discussion. In: Sartorius N, Nielsen JA, Strömgren E (1989) (eds) Changes in frequency of mental disorder over time: results of repeated surveys of mental disorders. Acta Psychiatr Scand 79 [Suppl 348]: 167–178

Studnicka M, Studnicka-Benke A, Wögerbauer G, Rastetter D, Wenda R, Gathmann P, Ringel E (1991) Psychological health, self-reported physical health and health service use: risk differential observed after one year of unemployment. Soc Psychiatry Psychiatr Epidemiol 26: 86–91

Sugisawa A (1994) Health conditions among fishermen living in the Minamata disease prevalent area. Nippon Koshu Eisei Zasshi 41: 428–440

**Susser M, Susser E (1996) Choosing a future for epidemiology. II. From black box to Chinese boxes and eco-epidemiology. Am J Public Health 86: 674–677

Tarnopolsky A, Clark C (1984) Environmental noise and mental health. In: Freeman HL (ed) Mental health and the environment. Churchill Livingstone, London, p 250

Tehrani E, Krussel J, Borg L, Munk-Jorgensen P (1996) Dropping out of psychiatric treatment: a prospective study of a first-admission cohort. Acta Psychiatr Scand 94: 266–271

Thompson J (1990) Psychological impact. In: Murray V (ed) Major chemical disasters – medical aspects of management. Royal Society of Medicine Services, London, p 197

*Thornicroft G (1991) Social deprivation and rates of treated mental disorder: developing statistical methods to predict psychiatric service utilisation. Br J Psychiatry 158: 475–484

Thornicroft G, Margolius O, Jones D (1992) The TAPS project. VI. New long-stay psychiatric patients and social deprivation. Br J Psychiatry 161: 621–624

Thornicroft G, Bisoffi G, De Salvia D, Tansella M (1993) Urban-rural differences in the associations between social deprivation and psychiatric service utilization in schizophrenia and all diagnoses: a case-register study in Northern Italy. Psychol Med 23: 487–496

Titchener JL, Kapp FT (1976) Family and character change at Buffalo Creek. Am J Psychiatry 133: 295–299

Tong S, Baghurst P, McMichael A, Sawyer M, Mudge J (1996) Lifetime exposure to environmental lead and children's intelligence at 11–13 years: the Port Pirie cohort study. Br Med J 312: 1569–1575

Triebig G (1990) Toxische Enzephalopathie als Berufskrankheit. Dtsch Med Wochenschr 115: 1287–1290

Ullman LP (1967) Institution and outcome. Pergamon, London

van Os J, Takei N, Castle DJ, Wessely S, Der G, Murray RM (1995) Premorbid abnormalities in mania, schizomania, acute schizophrenia and chronic schizophrenia. Soc Psychiatry Psychiatr Epidemiol 30: 274–278

Vázquez-Barquero JL, Diez-Manrique JF, Munoz J, Menendez-Arango J, Gaite L, Herrera S, Der GJ (1992) Sex differences in mental illness: a community study of the influence of physical health and sociodemographic factors. Soc Psychiatry Psychiatr Epidemiol 27: 62–68

Viinamäki H, Koskela K, Niskanen L, Arnkill R (1993) Social support in relation to mental well-being among the unemployed: a factory closure study in Finland. Nordic J Psychiatry 47: 195–201

Viinamäki H, Koskela K, Niskanen L, Tähkä V (1994) Mental adaptation to unemployment. Eur J Psychiatry 8: 243–252

Viinamäki H, Kontula O, Niskanen L, Koskela K (1995a) The association between economic and social factors and mental health in Finland. Acta Psychiatr Scand 92: 208–213

Viinamäki H, Niskanen L, Koskela K (1995b) How do mental factors predict ability to cope with long-term unemployment? Nordic J Psychiatry 49: 183–189

Viinamäki H, Kumpusalo E, Myllykangas M, Salomaa S, Kumpusalo L, Kolmakov S, Ilchenko I, Zhukowsky G,

Nissinen A (1995c) The Chernobyl accident and mental wellbeing – a population study. Acta Psychiatr Scand 91: 396–401

**von Baeyer W, Häfner H, Kisker KP (1964) Psychiatrie der Verfolgten. Springer, Berlin Göttingen Heidelberg

Vroom FQ, Greer M (1972) Mercury vapor intoxication. Brain 95: 305–318

Warner R, Appleby L, Whitton A, Faragher B (1966) Demographic and obstetric risk factors for postnatal psychiatric morbidity. Br J Psychiatry 168: 607–611

Webb SD (1984) Rural-urban differences in mental health. In: Freeman H (ed) Mental health and the environment. Churchill Livingstone, London, p 226

Werner EE (1985) Stress and protective factors in children's lives. In: Nicol AR (ed) Longitudinal studies in child psychology in psychiatry: practical lessons from research experience. Wiley, Chichester, p 335

Wessely SC, Castle D, Douglas AJ, Taylor PJ (1994) The criminal careers of incident cases of schizophrenia. Psychol Med 24: 483–502

Weyerer S, Häfner H (1989) The stability of the ecological distribution of the incidence of treated mental disorders in the city of Mannheim. Soc Psychiatry Psychiatr Epidemiol 24: 57–62

Wig NN, Menon DK, Bedi H, Leff J, Kuipers L, Ghosh A, Day R, Korten A, Ernberg G, Sartorius N, Jablensky A (1987) Expressed emotion and schizophrenia in North India. II. Distribution of expressed emotion components among relatives of schizophrenic patients in Aarhus and Chandigarh. Br J Psychiatry 151: 160–165

Williams R, Joseph S, Yule W (1993) Disaster and mental health. In: Bhugra D, Leff J (eds) Principles of social psychiatry. Blackwell, Oxford, p 450

Wilson JW (1995) Jobless ghettos and the social outcome of youngsters. In: Moen P, Elder GH, Lüscher K (eds) Examining lives in context. Perspectives on the ecology of human development. American Psychological Association, Washington, DC, p 527

*Wing JK, Brown GW (1970) Institutionalism and schizophrenia. A comparative study of three mental hospitals 1960–1968. Cambridge University Press, London

Wolkind S, Rutter M (1985) Socio-cultural factors. In: Rutter M, Hersov L (eds) Child and adolescent psychiatry: modern approaches. Blackwell, Oxford, p 82

CHAPTER

18

R. Gardner, W.T. McKinney

Ethology and the Use of Animal Models

1 Introduction

1.1 Definitions

Ethology describes animal behavior in the natural habitat and assumes operation of evolutionarily conserved basic plans encoded in the genome. Such basic plans determine behavioral patterns, including flexible variants involving learning. Human ethology has emerged as a subdiscipline, including observations of psychiatric patients. Animal models are experimental preparations developed in one species to study phenomena of another. For psychiatry, this refers to states in animals that resemble human mental disorders to foster laboratory studies of mechanism and treatment. These scientific areas supplement each other as part of psychiatry's basic science. The research involved has produced an enormous and varied literature. This chapter touches on principles and a few exemplary lines of research.

1.2 History

Present principles of psychopathology and pathophysiology stemmed in part from ethological approaches to animal behavior (McKinney et al. 1994). Early debates queried whether sets of animal behaviors represented specific human clinical syndromes (McKinney and William 1988). Field narratives used clinical terms when describing animal behavior. For instance, the concept of "experimental neurosis" in animals deployed terminology familiar to clinicians, yet the behaviors were foreign and the action patterns described likely stemmed from different experiential and brain mechanisms than did those operating in distressed people. On the positive side, early describers began to rate behavior quantitatively, now a central research practice. Presently investigators seek reaction patterns and their mechanisms without requiring that they be specific counterparts to particular psychiatric conditions.

1.3 Animal Models

No comprehensive animal model exists in that no nonhuman animal condition mirrors exactly etiology, symptoms, mechanisms, and treatment responsivity of any psychiatric syndrome. However, animal model research has instructed investigators and clinicians about across-species behavioral similarities, provided tests for etiological and pathogenetic theories, and pointed to new effective drugs (Cooper and Hendrie 1994).

Behavioral models in psychopharmacology are of three kinds:

- Screening tests
- Behavioral assays
- Simulations of clinical phenomena

For screening effective drugs, false-positive and false-negative results indicate lack of perfect modeling, but animals used in this way have produced the clinical relevance of developing a broader range of effective safer medications.

Pharmaceutical companies often focus on practical utility. Thus, if an animal demonstrates a possible counterpart to the human syndrome in question regardless of known mechanism, their experiments will typically study whether the medication response resembles that of humans. If it does, the method is considered to have empiric validity.

Empiric validity can be limiting, however. For instance, candidate antipsychotic drugs for many years had to have neuroleptic or motor properties as a shortcut indicator of efficacy. With more research on mechanism, however, dopamine receptor-blocking drugs clearly affected a variety of dopamine receptors. Drug action on some of these cell constituents affected behavior but caused no or fewer neuroleptic problems. This finding newly permitted development of a generation of antipsychotic drugs without motor effects, which were in fact deleterious. Indeed, at times, new drugs such as clozapine in fact reversed seemingly permanent conditions such as tardive dyskinesia.

2 General Principles

2.1 Central Concepts

Animal action and reaction patterns stem via darwinian natural selection from the survival tasks of food, defense, reproduction, and adaptation to habitat. This includes the social environment of other animals of the same species (conspecifics). Modeling depends upon brain–body factors that the animal and humans possess in common. Ethologically studied behavior patterns may be species specific, but core components shared by related animals are typically embroidered through natural selection to produce modified methods of survival and reproduction.

For example, human brains contrast in size to those of other animals, weighing three times more than

brains of surviving large primates or those of human ancestors 3 million years ago. Cerebral cortical area is four times greater, so the contrast with other primates especially stems from a massively enlarged neocortex, especially the frontal lobes. The increases likely stemmed from advantages of social functions. Mammalian cortex and overall brain size correlate with species-typical group size. Such cortical attributes as language obviously serve social functions.

Yet the human brain and behavior also show comparability to other species through widely shared, conserved features. This comparability makes continued use of animals important for the study of the pathogenesis, mechanisms, and treatment of mental disease. Conserved body features contributing to such illness are in the early stages of detection and exploration. Lack of a sufficient database as yet limits other modeling efforts such as computer simulations, mathematical models, and experimentally induced states in humans, and research on animal models remains indispensable.

The concepts of homology and convergence apply to across-species study. Homology means that a common ancestor once possessed a trait now shared by two species. At points in the past, humans shared common ancestors with monkeys, mice, chickens, fish, insects, and single-celled organisms, each such forebear more remote in biological history. In contrast, convergent traits are similar features that stem from environmental shaping through natural selection, although basic plan starting points vary. Wings of insects, bats, and birds illustrate this. Ancestors of each animal group had not flown, so airborne ability evolved separately and the three kinds of wings illustrate convergent evolution on the level of aerial locomotion. As vertebrate upper extremities, however, wings of bats and birds are homologous to each other, but not to the wings of insects. If the starting point of the basic plan generalizes to contractile tissue, however, locomotor body extensions of all three achieve homology.

In the nineteenth century, Darwin examined processes influencing variation in species traits conceptualizing what he called "descent with modification." He outlined natural selection to replace divine intervention, helping the then half-century old discipline of biology enter the scientific realm. The "Darwin machine" is now considered a general process operating not only for evolution of species, but also for processes as varied as neuron migration during development to the mental generation of particular thoughts. Its six components are as follows: (1) An entity or pattern must exist (2) for which there is a mechanism for copying; (3) variations on the pattern co-occur, (4) competition for a work space exists, (5) a multifaceted environment biases the competition for which pattern version will win (selection), and (6) the process reiterates: there are closed, repeating loops for the variation-and-selection steps.

2.2 Basic Plans

With molecular biology's explosive development, including genome projects of species at different phylogenetic levels, layered basic plans on molecular levels have taken on new meaning and importance. Much deciphering is necessary, but the genome seems to contain at least a partial record of the organism's ancestry. Decoding this will help determine evolutionary history (homology), which in turn fosters knowledge of proximal neuronal determinants of behavior. However, detection and description may be difficult. At the behavioral level, redundant, multiply determined brain–behavior adaptations complicate inference, and at the DNA level, genetic transformations such as chromosomal crossovers reduce certainty about genomic hypotheses.

Basic plans obviously operate in varied species, e.g., the body plan of anatomic structures such as anterior–posterior axis and limb number. Behaviorally, mammals show high investment in offspring. Body constituents may operate complexly at the molecular level, however. Serotonin, norepinephrine, and dopamine exemplify this. These monoamines are important for humans and are implicated in schizophrenia, affective disorders, and parkinsonism and involved in eating, sleep, and aggression. However, their body usefulness stemmed from the early evolution of metazoans, and they possess many functions in the diverse organisms in which they are found, acting, for example, as neurotransmitters, neuromodulators, neurohormones, and true hormones. They also operate in many ways within the mammalian body, sequestered in various body compartments and systems for extraordinary complexity and flexibility. Each impacts many receptors with differing downstream functions.

At a still more basic level, receptors are members of conserved superfamilies. For instance, some cell membrane receptors for serotonin, γ-aminobutyric acid (GABA), and glycine share common protein elements, although glycine and GABA, at least, have the opposite functions of stimulation and inhibition. Although likely derived from the same ancestral molecule with components in common, they mediate varied cellular communications.

Moving to higher organizational levels and behaviors familiar from anxiety disorders, flight–fight mechanisms involve the vertebrate locus coeruleus of the brain stem. This structure evolved early and has not disappeared for lack of usefulness, although its presumed activation in schizophrenia and paranoid

states in addition to panic and anxiety disorders may reflect maladaptation. This pontine structure supplies the brain with most of its norepinephrine and includes the subcortical temporal lobe nucleus, the amygdala, which is also activated during fearful states. This more recently evolved structure with increasing prominence in mammalian species overlays and modifies the functions of the locus coeruleus to which it directly and indirectly connects. From animal studies, we know the amygdala in association with the orbital frontal cortex to be also involved in maternal and familial bonding, so that disruptions of such states naturally affect the activities of this neuronal structure. Loss of social affiliation after lesions of this system caused speculation about their importance for the deficit syndrome of schizophrenia.

Jackson in the nineteenth century first highlighted such layered brain features, followed by MacLean in the twentieth century (MacLean 1990). MacLean resurrected Broca's term, "limbic system," for his triune brain formulation, correlating these brain components, including the amygdala, with paleomammalian evolution. These structures had little importance in reptiles. However, in the mammalian brain stem, he noted counterparts of a reptilian evolutionary stage including the locus coeruleus (actually such neuroanatomic features exist still more primitively in prereptilian fish). In the other direction, after limbic system expansion, neocortex development correlated with neomammalian evolution, including the orbital frontal cortex. Critics of the triune brain have noted that not only neocortex but also the brain stem and limbic structures changed with later developments, apparently modifying their functions in concert with the cerebrum's expanding capacities.

3
Animal–Human Continuity and Psychiatry

3.1
History and Pioneers

Thorndike and Pavlov, early twentieth-century research pioneers on animal modeling of human psychopathology, tested hypotheses about internal states with behavioral observations. Pavlov's experimental study of psychopathology included an interactive model with genetic and developmental factors interacting with current stressors and neurobiology, in fact initiating the modern biopsychosocial paradigm. He based his model on data from animals in the laboratory as did Cannon, whose research focused on principles of learning and autonomic responsivity generalizable to humans. Selye discovered a general stress response to noxious stimuli and founded the study of stress psychophysiology as a general adaptation of many systems in many species. This concept has been fundamental for psychosomatic medicine in its search for emotional factors in medical illness.

Darwin's *Expression of the Emotions in Animals and Man* (1873) had focused on continuities in animal behavior rather than the discontinuities among species emphasized in his earlier *Origins of the Species* (1859). A century after this pioneering work, still quoted in comparative behavior analysis, European ethologists extending his work won the Nobel Prize for Medicine and Physiology. Two winners, Lorenz (1965) and Tinbergen (1972), along with their peers of the time, paid special attention to stereotypical behavior patterns in natural settings, made across-species contrast comparisons, and described developmental sequences. Scott (1989) studied genetic influences on behavior by observing varied dog breeds. Harlow, Spitz, and Bowlby focused on attachment behaviors and found that primate and human infants react powerfully to loss of the mother (Bowlby 1969). This has had heuristic importance for clinicians and has fostered much animal model research in a separation paradigm. Hofer and coworkers extended this work to rats, discovering the importance of the mother in regulating infant autonomic processes such as heart rate and gastric secretion (Hofer and Myers 1996; this paper is an introduction to an entire journal number on the topic of research in psychosomatic medicine).

After World War II, Chance used animals to examine behavior patterns stimulated by wartime concerns: English aviators had used amphetamines without prior study of their effects. Chance found that mortality in mice given the drug varied greatly depending on housing. Joint housing caused more deaths. He concluded that an animal's social circumstances must be considered when investigating drug effects (Chance 1988).

3.2
Social Rank Hierarchy

Stimulated by Chance, Price's theoretical statement (Price 1967) related depressive behaviors to low social rank in nonhuman animals, such as that observed in low-ranking chickens observed by the definer of "peck order," Schjelderup-Ebbe (Price et al. 1994). Price noted that yielding behaviors that resemble defeat behaviors might blunt conspecific aggression. Fruitful work on depression and stress effects stemmed from the now extensively investigated resident–intruder paradigm. Tubular cage arrangements of multiple mice housed together produced fatal hypertensive heart disease in subordinates who could not escape

the punishment of the dominant (Henry and Stephens 1977). This was the first convincing animal model for "psychosomatic medicine." In a gentler version, a new male rat will alert to the olfactory presence of the potential territory holder when newly put into a cage previously used by a stranger male rat. This represents a model stressor for some studies, such as those on sleep stages and mechanisms. Sapolsky (1994) studied stress effects in primates in the wild and showed subordinates to be more affected than dominants, as measured by adrenocortical hormone levels. In nonprimate species also studied in the wild, however, dominants also exhibited elevations of stress hormones.

Returning to the physiology of social rank hierarchy, Raleigh et al. (1991) showed that vervet monkey groups in captivity exhibit marked variations in blood serotonin levels. Males with alpha status in mixed sex troops produced levels twice as high as other monkeys in the cage. This effect required the presence of both other males and females in the cage. Removed from the cage and isolated, the alpha male's increased level faded over time, but increased again if reinstated (if in time so that another would-be alpha male with increasing serotonin had not gained ascendancy). Humans with energetic type A behavior and fraternity leaders also have elevated blood serotonin. When subordinate vervet males were administered a selective serotonin reuptake inhibitor (SSRI), an antidepressant drug that increases synaptic serotonin, they increased their rank to that of alpha status. In addition to treating human depression and anxiety, SSRI medications increase ratings of self-directedness. Social rank effects were detected at the single-neuron level in invertebrates. Thus Yeh et al. (1996) found that "serotonin reversibly enhanced the response to sensory stimuli of the lateral giant tailflip command neuron in socially dominant crayfish, reversibly suppressed it in subordinate animals and persistently enhanced it in socially isolated crayfish."

Research on social rank hierarchy represents an integrative top-down, bottom-up approach that relates top/up behavioral and social dimensions of experience to bottom/down cellular and molecular levels of body mechanisms.

4
Behavioral Observations of Humans

4.1
Human Ethology

Human ethology uses direct observations of people. In *Human Ethology* (1989), Eibl-Eibesfeldt aimed at human universals presumably transmitted via genome inheritance, basing his inferences on cross-cultural observations. Other work showed that direct observations of sleep behaviors in conjunction with instrumented recordings characterize two distinctive forms of sleep, rapid eye movement (REM) sleep and non-REM forms (Gardner and Grossman 1975). Salter (1995) reviewed studies on the ethology of command and found it markedly affected by institutional context.

Time-sampled direct observations of behavior in depressed inpatients showed differentiated stages of recovery from major depression: motoric features recovered first and cognitive features later (Schelde 1998). These findings correlated with the paradoxical clinical finding that depressed patients die from suicide disproportionately frequently during the early stages of recovery.

Troisi (1994) pointed out that behavioral analysis of human conditions would enhance the field. For instance, he found that response to an antidepressant drug negatively correlated with observed coded nonverbal behaviors. Responders and nonresponders did not differ in sex, age, education, clinical diagnosis, or severity of disorder, but did differ in their coded behaviors: nonresponders were more assertive and affiliative. Moreover, his observations cast doubt upon depressive subtypes defined by the American Psychiatric Association's *Diagnostic and Statistical Manual* (DSM). He discussed the third edition of this manual (American Psychiatric Association 1980), but subsequent editions have not significantly augmented behavioral observations; rather, self-report and subjective histories–though gathered in a systematic manner with formalized interview schedules – have had greater sway with researchers.

Despite this contemporary and valid criticism of the recent DSM-IIIR (American Psychiatric Association 1987) and DSM-IV (American Psychiatric Association 1994), major developments in modern psychiatry have stemmed from the importance of systematically assessing behavior. Robins and colleagues strongly fostered operational criteria for diagnosis in the mid-twentieth century when psychiatry in the United States was dominated by theory free of verifiable observation. Behavior, thought, and feeling patterns assessed from systematic criteria in the DSM echo an observational approach to human behavior. That stereotypical patterns can be detected in the specialty helps dissect core behavior patterns and provide hypotheses about their origins and most appropriate treatment. Thus panic resembles a motivational state to escape fear-producing stimuli, mania behaviorally resembles dysfunctional leadership behavior, depression may reflect defeated or yielding animal behaviors, and persecutory delusions are fixed self-definitions of outgroup status (Gardner 1997).

4.2
Genome-Neural-Behavioral Analysis

Such behavioral descriptions represent hypothetical formulations to be tested further with cellular/molecular correlates examined as well. Advances in modern psychiatry in the form of highly effective drugs with known, or researched, brain actions provide avenues to explore these research domains. That animals of various sorts are comparable to humans has allowed considerable impetus to both drug development and increased knowledge of mechanism. Use of animal models moved clinical psychiatry from simple-minded deterministic views of psychopathology towards integrative ones. Moreover, the work fostered hypothesis testing with prospective studies.

4.3
Genome Deletion Syndromes

Eventual genome-neural-behavioral analysis may be assisted by the systematic observation of the behavior patterns of individuals with known genomic deletions or aberrations on the one hand and characteristic behaviors on the other. For example, Prader-Willi syndrome patients do not have paternal genes at chromosome 15q11-12; in addition, they overeat in a fashion typical of animals and people with hypothalamic lesions and demonstrate obsessive–compulsive symptoms, temper, and stubbornness. Among other genes, they are missing one that translates to a intracellular protein usually located primarily in the hypothalamus. Hyperactive Angelman syndrome patients missing maternal genes in 15q11-12 exhibit profound retardation with frequent laughter but absent language. Their missing genes include one that codes a $GABA_A$ receptor protein. Carrier fragile X syndrome patients with varying production of diminished fragile X mental retardation protein show social avoidance and depression as the first evidences of diminished protein. Missing genes on chromosome 7 result in verbal fluency, social skills, and creativity with mild mental retardation (Williams syndrome). Frontal structures apparently form adequately, but posterior cerebral cortical structures do not.

5
Illustrative Animal Models in Psychiatry

5.1
Bond Disruption Model

Disruption of important bonds such as mother and infant often cause states that appear to be similar to depression. Of course, clinical depression may follow the loss of a loved one in humans. In animals, these are considered experimental systems in which a clear social inducer results in a reliably seen response with species-dependent variation. Clearly, separation and loss do result in protest and despair phases in infant primates, including humans, though with considerable individual variability. Much work in primates and other mammals has characterized the response, which typically includes increased adrenocortical responses, increased arrhythmias, and decreased body temperature and immune system function. Antidepressant medication produces reversal of the pattern. At low doses, alcohol alleviates peer separation response but at high doses exacerbates it.

Formerly, an idea guiding clinical action held that depressive syndromes could be distinguished based on loss factors (reactive versus endogenous). Evidence for this distinction does not exist, however, so the nosology no longer persists. Animal models at times present confusing data. Closely related species may nevertheless demonstrate marked differences in social organization that may in turn differentially affect the reaction of animals to the same stimulus. Thus pigtail and bonnet macaque infants differed when separated from their mothers; the pigtails showed signs of despair and depression, but the bonnets did not. Bonnets ordinarily intermingle with their conspecifics, but by contrast the pigtail mothers typically focus only on their own offspring.

The characteristic behavior pattern that emerges when a monkey mother is lost can be "treated" by putting the infant with a peer, but later deprivations show the animal to have heightened vulnerability. Separation studies in animals have documented systematically the powerful effects of disrupting attachment systems and have fostered theories of attachment now part of the conceptual backdrop for general and child psychiatry. Specifically, this category of study has provided empirical support for long-term effects of early losses and the neurobiological impact of such experiences.

A rat version of the maternal deprivation model entails rat pups being removed from their mothers for 3 h per day during days 2–14 of the young rats' lives (Nemeroff 1998). At day 21, they are taken from the mother to live in groups. When exposed to an airpuff startle, the deprived animals have an increased hypothalamic–pituitary–adrenal response that persists for life. This is mediated in the animal model by corticotropin-releasing factor (CRF). Messenger RNA for this gene is increased, meaning that the gene has been turned on. Catheters in their epidural space show that CRH increases are parallel to those in depressed humans. Antidepressant medications reverse the condition, although if then removed, the rats become

hyperreactive again. With antidepressants, the increased gene response normalizes. This is one of the few animal models of depression in which the animal drinks additional alcohol and cocaine. The effect surely does not involve maternal deprivation only, however, because the mother rejects the pup upon its return to the litter, thus implicating agonistic features in the pathogenesis.

5.2 Uncontrollability Model

In the uncontrollability model, the experimenter first shocks the animal in a manner that allows predictability and control (e.g., the animal is forewarned of a shock and an escape route provided). The rules then change so that the shock recurs but with no control options. The animal then assumes a position of stolid, unmoving defeat and does not participate further in learning tasks. Norepinephrine lowers in the locus coeruleus, GABA lowers in the hippocampus, and the endogenous opiate system changes. Recovery from the state takes considerable time, but hastens if the animal is treated with antidepressants. Injection of imipramine into the rodent forebrain also normalizes the animal.

Often called learned helplessness (LH), this phenomenon was interpreted as "learning" because learning theorists discovered and publicized it. Development of an inbred rat strain with a low threshold for LH allowed other analysis of the mechanism. Findings implicated the hypothalamus–pituitary–adrenal axis. Study of a rat's social rank prior to imposition of the LH stressor showed that subordinates responded less than dominants did. This is congruent with a social rank interpretation of LH in that dominants would exhibit more change in status after experimenter-imposed defeats.

The clinical implications of animal studies of uncontrollability (learned helplessness models) include documentation of significant effects when an individual is unable to control what happens. This line of research has linked neurobiological substrates of uncontrollability with cognitive theories of depression and might provide information on how such factors operate in this treatment.

Depression may represent a psychobiological final common pathway, but numerous variables influence its stimulation as well as the specific form it takes. Influencing variables include genetic, developmental, interpersonal, personality, social, and physiological factors. Given this complexity, animal models are needed as experimental systems for controlled, prospective studies and provide approaches with clinical implications.

5.3 Olfactory Bulbectomy Model of Depression

The olfactory bulbectomy model of depression in rodents exemplifies empiric validity. The rat olfactory bulbs are a complex structure making up 4% of the total brain weight that directly connects to limbic system structures. After removal, the animal typically demonstrates many behavioral abnormalities. Delay in learning a stepdown passive avoidance task measures impairment quantitatively. This model does not possess face validity as the condition is not one that mimics the human condition. However, it exemplifies the opportunity to screen therapeutic agents for efficacy. For example, given chronically, 18 known antidepressant drugs improved the state, and nine non-antidepressants did not. One non-antidepressant gave a false-positive and a known antidepressant a false-negative result.

5.4 Approach-Avoidance Ratios and Threat Model Anxiety

Rodents confronted with the possibility of exploring open versus closed arms of a maze with two pairs of closed and open arms will theoretically experience approach-avoidance conflict. When more fearful, they will preferably explore the closed arms. Measures of drug effect include the ratio of open to closed arm explorations. Classical anxiolytics such as benzodiazepines increase this, whereas anxiogenic drugs decrease it. Considerable research experience has shown this to be pharmacologically valid with a range of drugs enhancing or diminishing the activities of the GABA receptor/chloride channel ionophore. Rodgers and Cole (1994) argue that an ethological approach to the data collection was important. Did the animal demonstrate typical antipredator defensive maneuvers? Thus, in addition to typical postures, fearing animals show diminished activity generally, not just avoidance of open arms.

Other animal modeling work aiming at anxiety-involved study of animals reacting to external threats, conflicts, trauma, etc. Operant techniques elicit a behavior. In a well-established response, the behavior was then punished, which suppressed it. Potential antianxiety drugs can be screened for response restoration to presuppression levels. This approach has empirical validity in terms of predicting clinical drug responsiveness and has more recently been used to begin investigations on the neurobiology of anxiety. No longer responding in punishment paradigms has been considered a passive avoidance response; in view of the findings, perhaps some human anxiety arises from passive avoidance of social situations.

Finally, work on the more fundamental neurobiology of the flight-flight response has used animals infected with a double-virus transneuronal labeling technique. Jansen et al. (1995) demonstrated that the hypothalamus and brain stem contain a set of central autonomic command neurons that dually innervate both the sympathetic neurons regulating cardiovascular function and adrenomedullary function, both of which are quickly needed in emergencies.

6 Genetic Alterations in Animal Models

Bottom-up research involves alterations of cells and molecules with behavioral measurements as the dependent variable, e.g., using mutants with missing genes. Thus environmental events trigger immediate-early genes that induce quickly. For instance, the *fos* family of genes are expressed in the brain in regions that produce nurturing behavior. Mutants without the *Fos*B variant fail to show usual nurturing behavior, although the reproductive tract, hormonal status, and brain show no obvious differences. However, staining for *Fos*B in the brains of normal animals shows that the key regions are the olfactory bulb and pyriform cortex, but not the cortex or, more generally, the amygdala or hippocampus (Brown et al. 1996).

New technology involves splicing into the genes of animals, usually mice, a DNA sequence without the gene under investigation. At the same time, an attached constituent renders it immune to an injected agent so that all other versions of the gene are eliminated ("knocked out") from the animal. What the gene product usually accomplishes in the resulting experimental animal can then be assessed using a kind of ablation procedure. Male mice lacking neuronal nitric oxide synthase, for instance, exhibit major increases in aggressive and inappropriate sexual behavior (Nelson et al. 1996). Moreover, indifference to cocaine and amphetamine in knockout mice lacking the dopamine transporter protein show that this molecule is usually the target of these drugs of abuse. In addition, replication of human gene defects in mice provides direct models of human disease, although when this does not happen, new information about the basic biology is revealed by the differences in the two species.

7 Conclusion

Psychiatry's basic science should have parallels with the basic sciences of other medical specialties that conceptualize their illnesses as deviations from normal biochemical and physiological processes. Emerging databases make this possible for psychiatry as well. We predict that the material reviewed in this chapter will make important contributions to that effort.

8 References

American Psychiatric Association (1980) Diagnostic and statistical manual of mental disorders, 3rd edn (DSM-III). American Psychiatric Association, Washington, DC

American Psychiatric Association (1987) Diagnostic and statistical manual of mental disorders, 3rd edn revised (DSM-IIIR). American Psychiatric Association, Washington, DC

American Psychiatric Association (1994) Diagnostic and statistical manual of mental disorders, 4th edn (DSM-IV). American Psychiatric Association, Washington, DC

Bowlby J (1969) Attachment and loss. I. Attachment. Basic, New York

Brown JR, Ye H, Bronson RT, Dikkes P, Greenberg ME (1996) A defect in nurturing in mice lacking the immediate early gene fosB. Cell 86: 297–309

Chance MRA (ed) (1988) Social fabrics of the mind. Erlbaum, Hillsdale

Cooper SJ, Hendrie CA (eds) (1994) Ethology and psychopharmacology. Wiley, London

Darwin C (1859) On the origin of species by means of natural selection or the preservation of favoured races in the struggle for life. Murray, London

Darwin C (1873) The expression of the emotions in man and animals. Murray, London

Eibl-Eibesfeldt I (1989) Human ethology. de Gruyter, New York

Gardner R (1997) Sociophysiology as the basic science of psychiatry. Theoret Med 18: 335–356

Gardner R, Grossman WI (1975) Normal motor patterns in sleep in man. In: Weitzman ED (ed) Advances in sleep research, vol II. Spectrum, New York, pp 67–107

Henry JP, Stephens PM (1977) Stress, health, and the social environment: a sociobiological approach to medicine. Springer, Berlin Heidelberg New York

Hofer MA, Myers MM (1996) Animal models in psychosomatic research. Psychosom Med 58: 521–522

Jansen ASP, Nguyen XV, Karpitsky V, Mettenleiter TC, Loewy AD (1995) Central command neurons of the sympathetic nervous system: basis of the fight-or-flight response. Science 270: 644–646

Lorenz K (1965) Evolution and modification of behavior. University of Chicago Press, Chicago

MacLean PD (1990) The triune brain in evolution: role in paleocerebral functions. Plenum, New York

McKinney WT, William T (1988) Models of mental disorders: a new comparative psychiatry. Plenum, New York

McKinney WT, Gardner R, Barlow GW, McGuire MT (1994) Conceptual basis of animal models in psychiatry: a conference summary. Ethol Sociobiol 15: 369–383

Nelson RJ, Demas GE, Huang PL, Fishman MC, Dawson VL, Dawson TM, Snyder SH (1996) Behavioural abnormalities in male mice lacking neuronal nitric oxide synthase. Nature 378: 383–386

Nemeroff C (1998) The neurobiology of depression. Sci Am 6: 42–57

Price J (1967) Hypothesis: the dominance hierarchy and the evolution of mental illness. Lancet 2: 243–246

Price J, Sloman L, Gardner R Jr, Gilbert P, Rohde P (1994) The social competition hypothesis of depression. Br J Psychiatry 164: 309–315

Raleigh MJ, McGuire MT, Brammer GL, Pollack DB, Yuwiler A (1991) Serotonergic mechanisms promote dominance acquisition in adult male vervet monkeys. Brain Res 559: 181–190

Rodgers RJ, Cole JC (1994) The elevated plus-maze: pharmacology, methodology, ethology. In: Cooper SJ, Hendrie CA (eds) Ethology and psychopharmacology. Wiley, London, pp 9–40

Salter FK (1995) Emotions in command: a naturalistic study of institutional dominance. Oxford University Press, Oxford

Sapolsky RM (1994) Why zebras don't get ulcers: a guide to stress, stress-related diseases and coping. Freeman, New York

Schelde T (1998) Major depression: behavioral markers of depression and recovery. J Nerv Ment Dis 186: 141–149

Scott JP (1989) The evolution of social systems. Gordon and Breach, New York

Tinbergen N (1972) The animal in its world: explorations of an ethologist, 1932–1972, vol 2: Laboratory experiments and general papers. Harvard University Press, Cambridge, MA

Troisi A (1994) The relevance of ethology for animal models of psychiatric disorders: a clinical perspective. In: Cooper SJ, Hendrie CA (eds) Ethology and psychopharmacology. Wiley, London, pp 329–340

Yeh SR, Fricke RA, Edwards DH (1996) The effect on serotoninergic modulation of the escape circuit of crayfish. Science 271: 366–369

CHAPTER

19

D. Ploog

Evolutionary Biology of Emotion

Men ought to know that from the brain and from the brain only arise our pleasures, joys, laughter, and jests as well as our sorrows, pains, griefs and tears. . . . It is the same thing which makes us mad or delirious, inspires us with dread and fear, whether by night or by day, brings sleeplessness, inopportune mistakes, aimless anxieties, absent mindedness and acts that are contrary to habit . . .

HIPPOCRATES (c. 460–375 B.C.), *The Sacred Disease*

Translator: E. Taub

1 Introduction

The idea that evolutionary considerations may aid in the understanding of mental processes in both healthy and diseased states has a long history, but has now receded into the background because of recent major advances in molecular and cellular neuro- and psychopharmacology. Kraepelin (1920) was thinking along these lines when he wrote that one could form, at that time, no more than an "exceedingly crude and incomplete" picture of the origins of mental illness. He reflected on the phylogenetic origins of personality and suggested an investigation of "the extent to which certain impulses, dormant since the prehistory of personal and phylogenetic development, may come back to life in diseased states" (Krapelin 1920, p. 29).

Kraepelin also referred directly to Darwin in his 19th lecture on hysterical mental disorders (Kraepelin 1916), in which he interpreted the forms of expression of the emotions as "remnants of ancient defense mechanisms." Ernst Kretschmer (1953) likewise made use of evolutionary theory when he spoke of "psychomotor templates," by which he meant standardized motor sequences constituting phylogenetically prefigured "reflexive and instinctive patterns." We refer to these as preprogrammed motor sequences (Ploog 1957, 1958), which correspond to the "fixed action patterns" discussed by Konrad Lorenz (1937, 1953, 1992), insofar as they are neither learned nor automatized.

2 Function of Emotion

The concept of emotion includes that of affect, i.e. the motor component of emotion, which typically consists of brief movements. A single cerebral system underlies both emotion and affect. These two concepts are not systematically distinguished in the literature. In what follows, we shall take an evolutionary approach alongside the conventional medical one, and we shall largely limit the discussion to emotion, although the evolutionary approach to psychopathology may also be applied to cognition. As is well known, Darwin himself laid the foundations for all future work on the evolutionary biology of human emotion in his book *The Expression of the Emotions in Man and Animals* (1872). He considered the emotions to be important biological adaptive responses in many species.

Until recently, the field of psychopathology has largely ignored or underestimated the importance of evolutionary processes for the biological foundations of human emotion in the context of our social behavior; the studies carried out by Bowlby (1969) and Trivers (1985) were exceptions. Even today, this subject is far from exhausted. It is discussed, sometimes controversially, by psychologists, anthropologists, human ethologists, and sociobiologists (Frijda 1986; Reeve 1992; Oatley and Jenkins 1996). The ethological and evolutionary approach to psychiatry was formulated as early as the 1960s (Ploog 1964; Plutchik 1962, 1994) and was recently revived by McGuire and Troisi (1998).

Issues remaining to be clarified include the functional relation between emotion and motivation, and that between emotion and cognition, as well as the correspondence between subjectively experienced emotion and its expression. It also needs to be clarified which emotions are primary and which secondary, i.e. derived from the primary emotions (Schneider and Scherer 1988).

Since the time of William James (1890–1950), all of these issues have been bound up with the argument over the causation of the emotions. The James–Lange theory holds that a suitable stimulus can release a physiological reaction, which is then internally perceived as emotion; the emotional experience follows the physical change. The physiologist Walter B. Cannon (1927) was the first of a long series of investigators to contradict this notion. He demonstrated, among other things, that an emotion is experienced *before* the corresponding physiological reaction is produced through the action of the autonomic nervous system. A further complication in the argument came from the hypothesis presented by Schachter and Singer (1962), according to which an emotion arises from the interaction of a stimulus-dependent arousal reaction and the recognition of the releasing stimulus. The subject's evaluation of the stimulus (e.g. as dangerous) determines the type of emotion experienced (such as fear), while the degree of arousal determines its intensity. Cognitive concepts of the causation of emotion remain in opposition to biological concepts even today (e.g. Buck 1988; Reeve 1992).

While the "nativists," invoking Darwin, assume a genetically programmed link between the basal emotions and their expression (Izard 1977; Tomkins 1962; Zivin 1989), the "cognitivists" maintain that emotional expression arises from emotional inner states only after passage through an amorphous matrix that can be influenced by the individual's cognitive development and social learning (Emde et al. 1976; Sroufe 1979). Although the nativist and cognitivist theories stand in opposition to one another even today, they nonetheless agree in considering the emotions (as inner states) to be psychophysiological entities within the organism, and the basal emotional expressions to

be their species-specific outward manifestations. Plutchik (1985) considers this theoretical debate to be pointless. He maintains that emotions result from a dynamic interaction of biological and cognitive factors, quite aside from the fact that cognitive processes, too, are biologically based and have their own evolutionary history.

Wolff (1987, 1993) is critical of both of the approaches outlined above and, proceeding from his own experiments with infants, arrives at a different conclusion. He found that a given experience of the environment may evoke any one of a number of qualitatively distinct motor responses, just as a single motor expression may be produced by qualitatively distinct environmental experiences. The internal state of the child at the moment of the stimulus determines which coordinated motor response is realized (Prechtl 1974). The relationship between a socially relevant event and its emotional expression is thus better and more specifically described by a nonlinear relation between a releasing stimulus pattern and several possible motor expressions, or between a single motor expression and several possible releasing stimulus patterns, than by any preprogrammed isomorphic correspondence between releasing stimulus pattern and motor expression. Wolff concludes that it is impossible to draw a clear boundary between emotions, i.e. "internal states," on the one hand and "outwardly-directed" motor expressions on the other – an important conclusion indeed for an understanding of the entire complex system.

The so-called facial feedback hypothesis, which describes the functional connection between subjectively experienced emotions and the corresponding facial expressions, is particularly relevant to psychopathology in this context (Steiner 1974). According to Tomkins (1963), certain emotions are experienced when certain facial expressions are fed back into consciousness. More precisely, feelings can be induced by mimetic movements and by the changes in facial musculature that compose them. Facial expressions are activated by subcortical centers in which emotion-specific "programs" are localized, a different one for each emotion. These programs are genetically fixed and phylogenetically old. The emotion-specific programs can activate specific expressive movements and vice versa.

Much experimental energy has been consumed in attempts to confirm the facial feedback hypothesis. The first demonstration of functional relationships of the type posited by Tomkins was provided by Ekman et al. (1983), who correlated certain autonomic response patterns with certain facial expressions as well as with certain experienced emotions. Reeve (1992) drafted a circuit diagram representing the neural basis of this feedback process (Fig. 1). The actual neural circuit is, of course, much more complex, and one may doubt whether the cortex plays the primary role in the

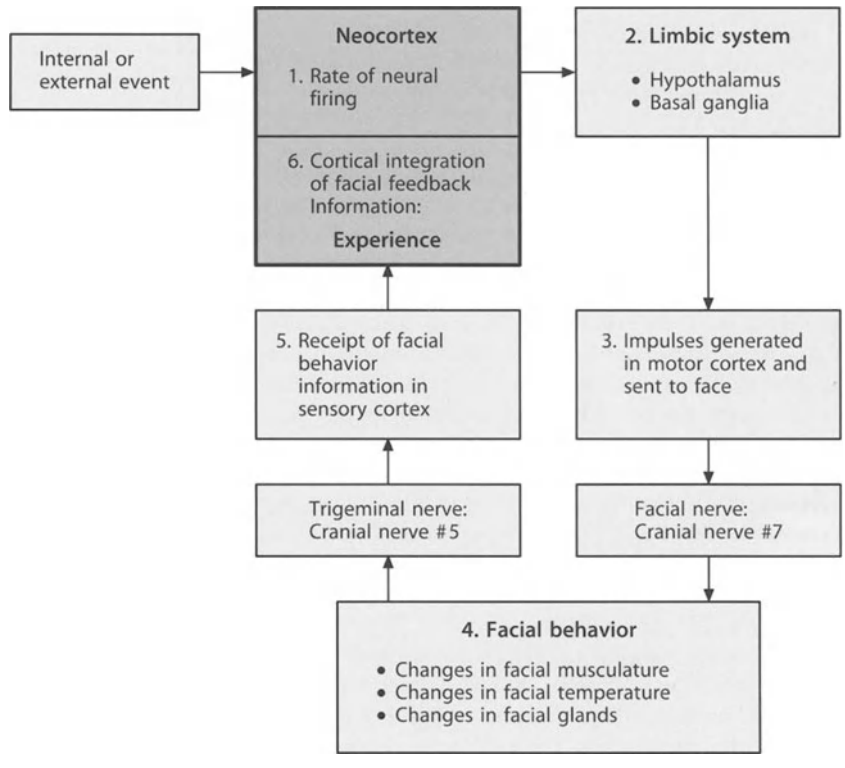


Fig. 1. Sequence of neural events in the facial feedback hypothesis. (From Reeve 1992)

internal processing of the facial feedback expressions (box at upper left). Recent experiments involving functional magnetic resonance imaging (MRI) have clearly demonstrated that empathic observation of pictures of sad or happy facial expressions can induce the corresponding emotions and that this occurrence correlates with neural activation in the amygdala, among other structures (Morris et al. 1996; F. Schneider et al. 1998a,b). These and similar studies provide verification of the facial feedback hypothesis.

In the field of motivational theory, it was again Tomkins (1970) and also Frijda (1986) who identified not physiologic drives, such as hunger and thirst, but rather the emotions as driving forces of motivated behavior. Not drives, but emotions are the determining factors of the primary system of motivation (see below). This conclusion is supported mainly by three arguments.

1. Emotions influence drives: hunger fades in the presence of fear or disappointment; thirst is driven away by disgust when the water is foul; excitement and joy promote sexual behavior, while fear and anger inhibit it.
2. Drives are periodic and temporally restricted, while emotions (in the sense we are now using) have no time limit: one can be afraid momentarily or for a long time. The function of drives is limited to homeostasis and the regulation of sexual behavior, while the motivational function of the emotions pervades all areas of life.
3. Unlike drives, emotions can always be consciously experienced, whatever their quality or intensity, while drives signify homeostatic imbalance (e.g. hunger, thirst, fatigue, sexual disposition).

The emotions constituting the primary system of motivation are also called fundamental or basal, because it is generally assumed – even by cognitive psychologists – that they have an inborn basis, are common to all human beings, and are of phylogenetic origin (Eibl-Eibesfeldt 1984). The recognition that emotions have an evolutionary history leads directly to the question of their function and thus also to that of their adaptive value. How does an emotion contribute to the adaptation of a living creature to its physical and social environment? An evolutionary theory of emotion should apply not only to humans, but to all vertebrates. Darwin's assumption that emotion increases the likelihood of an individual's survival, while simultaneously, in its expression, signifying behavioral intent remains the point of departure for all evolutionary considerations of emotion today.

The names we give to the (subjective) emotions on the basis of common human experience, the behavior corresponding to each of these emotions, and the functions they perform for the organism

Table 1. Three languages that may be used to describe emotional states. (From Plutchik 1984)

Subjective language	Behavioral language	Functional language
Fear, terror	Withdrawing, escaping	Protection
Anger, rage	Attacking, biting	Destruction
Joy, ecstasy	Mating, possessing	Reproduction
Sadness, grief	Crying for help	Reintegration
Acceptance, trust	Pair bonding, grooming	Incorporation or affiliation
Disgust, loathing	Vomiting, defecating	Rejection
Expectancy, anticipation	Examining, mapping	Exploration
Surprise, astonishment	Stopping, freezing	Orientation

are listed in corresponding columns in Table 1 (Plutchik 1984). Double naming in the first column indicates that emotions vary in their intensity. Fear and terror, in subjective language, correspond to withdrawing and escaping as the indicated behaviors, and both serve to protect the individual (functional description). Sadness and crying evoke offers of help, and so on.

When an ethologist speaks of the "cause" of a particular phylogenetic development, the word is being used in two senses. One concerns the molecular, cellular, and neural mechanisms underlying the behavior (or behavioral disturbance) in question; another concerns its purpose. The former are the so-called proximate, the latter the so-called ultimate causes. Both are required for the explanation of an evolutionary process; the questions "How?" and "Why?" necessarily go together. Both questions have been discussed repeatedly in earlier articles (e.g. Ploog 1964, 1980a, 1988); the second is of primary concern to us here. The answers offered will never be more than merely plausible, because teleological questions cannot be answered empirically.

A word on the concept of adaptation is in order before we embark on a discussion of the function of the emotional system. Both Darwin's expression "the survival of the fittest" and his term "adaptation," as well as their German translations, have been liable to misunderstanding for a long time and remain so today. It is not the strongest organism that preferentially survives, but rather the one best adapted to the prevailing social and environmental conditions, i.e. the one whose particular abilities and strengths give it the greatest chance of surviving and reproducing. The individual achieves these goals by deriving the maximum possible gain from its physical and social

environment. This also implies optimization of reproductive potential or, in other words, the greatest possible propagation of the individual's own genes. The business of reproduction thus not only serves to maintain the species in general, as classical Darwinism holds, but also gives each individual the means to propagate its own genes as effectively as possible (Dawkins 1976, 1988).

The ability of the organism to propagate its genes to the largest possible number of descendants is known as its "inclusive fitness." Since the late 1960s, sociobiologists have studied the behavioral strategies used to achieve this goal (Hamilton 1964, 1975; Wilson 1975). The methods of sociobiology allow the collection of empirical data about the ultimate causes of behavior. Discoveries have been made in this way about the behavior of social animals of several different phyla. This fact alone implies that the adaptive behavioral strategies observed (which correspond to the ultimate causes of behavior) cannot be considered conscious processes. Even in humans, adaptive behavior is a conscious process only in very small part. It is certain, however, that adaptive behavior is continually demanded of all participants in a hierarchical society – and this is, of course, the universal mode of human social organization. The behavioral analysis of a depressive female patient, shown in Fig. 2, concerns an extreme example of an adaptive process leading to illness.

Adaptive behavior is always based on emotion and is not only found among primates; it has, of course, a long evolutionary history. Prototypical behavioral sequences in which primary emotions manifest themselves are listed in Table 2 (Plutchik 1984). The first column lists typical releasing stimuli that activate behavioral sequences, and the second column lists the inferred cognition of these stimuli. The releasing stimulus either is recognized by an innate mechanism or else operates as a conditioned stimulus. The third column lists the primary emotions corresponding to the adaptive behaviors (fourth column), which have an innate basis and serve a particular purpose (last column).

3 Canon of Emotions

Which emotions underlie the primary or fundamental system of motivation? Carroll Izard (1977), in a further development of Tomkins' ideas, proposed a

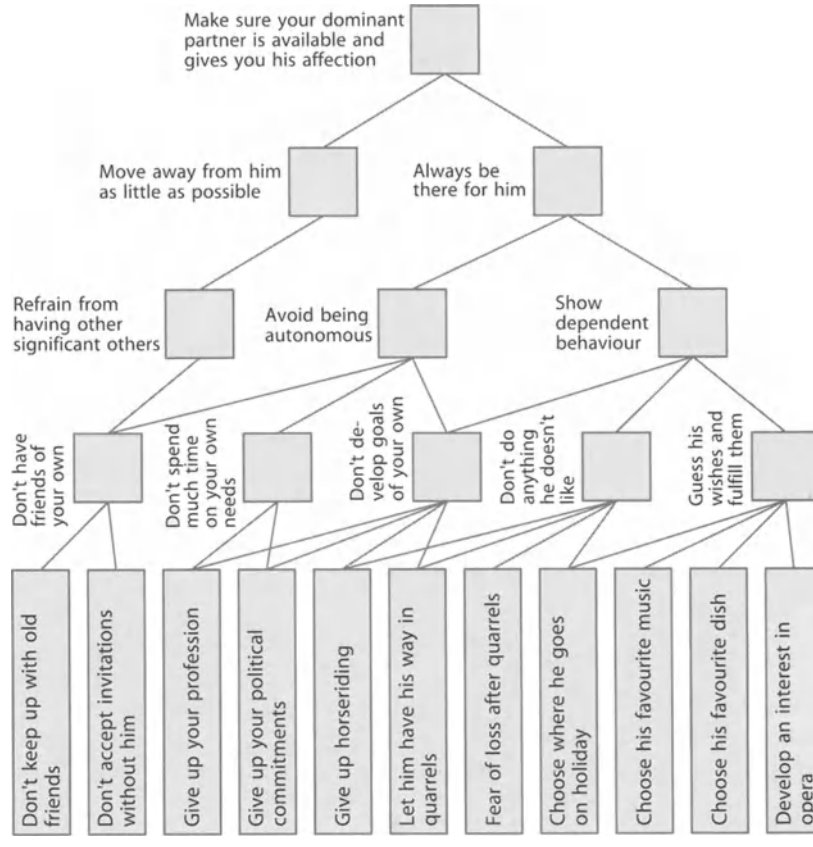


Fig. 2. Behavioral analysis of a depressive female patient. (From Sulz 1993)

Table 2. The complex, probabilistic sequence of events involved in the development of an emotion

Stimulus event	Inferred cognition	Feeling	Behavior	Effect
Threat	"Danger"	Fear, terror	Running, or flying, away	Protection
Obstacle	"Enemy"	Anger, rage	Biting, hitting	Destruction
Potential mate	"Possess"	Joy, ecstasy	Courting, mating	Reproduction
Loss of valued individual	"Abandonment"	Sadness, grief	Crying to be reunited	Reintegration
Group member	"Friend"	Acceptance, trust	Grooming, sharing	Affiliation
Gruesome object	"Poison"	Disgust, loathing	Vomiting, pushing away	Rejection
New territory	"What's out there?"	Anticipation	Examining, mapping, organizing	Exploration
Unexpected object	"What is it?"	Surprise	Stopping, alerting	Orientation

differential theory of emotion according to which the fundamental emotions are those that give rise to differential motivations. Each of these fundamental emotions has a specific subjective quality and a specific corresponding facial expression. Each emotion is assumed to be brought into consciousness through a specific pattern of neural discharges, and each one leads to specific behavioral dispositions. There is, however, no consensus on the number of basic emotions.

Panksepp's psychobiological theory (1982) limits the number of "primitive" emotions to four, which are assumed to arise during the development of the infant brain, arise in at least four visceral–limbic circuits, and activate corresponding modes of behavior. These are expectation (curiosity), anger, fear, and panic. Tomkins (1962, 1963) lists eight basal emotions, three of them positive – interest, surprise, and joy – and five of them negative – anger, fear, shame, disgust, and rage. Izard (1977) posits ten basal emotions.

Since the time of Spinoza and Descartes, there have been many lists and classifications of human emotions, most of them without justification. The earliest empirical studies of emotional expression, including some experimental studies, are those of Darwin (1872). Experimental studies of facial expression in psychiatry (e.g. those of Ellgring 1989; Ellgring and Smith 1998; Flack and Laird 1998) commonly make use of the list given by Ekman and Friesen (1978):

- Surprise
- Anger
- Fear
- Happiness (joy)
- Sadness (distress)
- Disgust
- Contempt (a later addition)

Plutchik (1980, 1994) adds two further emotions with an evolutionary and ethological basis, namely expectancy and acceptance. Expectant (i.e. curious) behavior is an important "drive" for juvenile animals and facilitates coping with the environment. Acceptance is striven for in all hierarchically oriented societies, especially among primates (including humans), and is one of the most important forms of reward. There is controversy over whether shame and guilt should be counted among the fundamental human emotions, mainly because cross-cultural study reveals no facial expressions that unambiguously express them in all cultures (Ekman et al. 1969).

The likely objection that the human emotions, at least, are not exhausted by a list of four, seven, eight, or ten primary ones can be answered by positing the existence of secondary emotions that are mixtures of the primary ones, in much the same way that the color spectrum is generated by mixing the primary colors (Plutchik 1980). Such mixtures or superpositions of emotions have been described by ethologists with regard to emotional expression, e.g. in wolves by Schenkel (1947), in dogs by Lorenz (1963), in cats by Leyhausen (1973), and in subhuman primates by Chevalier-Skolnikoff (1973). In these species, a spectrum of facial expression is generated by superposition of expressions signifying varying intensities of different emotions, e.g. readiness to attack or to flee. Similar mixtures have been found in humans; the best-known example is the mimetic superposition of surprise and fear (Ekman 1973; Fig. 3).

From an ethological and evolutionary viewpoint, the existence of such phenomena leads to certain conclusions concerning the cerebral organization of emotion (MacLean 1949, 1990; Ploog 1989). It is a basic principle that every peripheral structure of the body has its representation in the central nervous system. Thus, if a peripheral structure (such as the facial musculature) develops into a progressively more complex apparatus, then a corresponding organization of the central nervous system must take place to subserve the function of this apparatus. This principle is no longer open to doubt (Ciompi 1991).

Over the course of mammalian evolution, the facial apparatus has undergone a remarkable differentiation, as may be observed by comparing the variety of

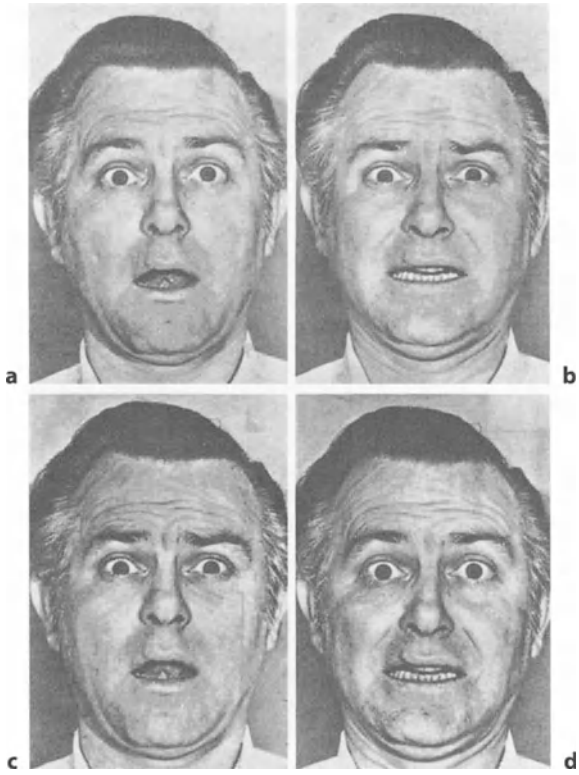


Fig. 3. Superposition of emotions in facial expression. *Upper left*, pure expression of surprise; *upper right*, pure expression of fear; *lower left*, this expression represents a superposition of surprise (mouth) and fear (eyes, eyebrows, forehead); *lower right*, superposition of surprise (eyebrows, forehead) and fear (mouth). (From Ekman 1973)

facial expressions of a rat to that of a dog or chimpanzee. Emotional expression in primates is almost exclusively concentrated in the face and the voice (Ploog 1986a). The facial musculature and the larynx have correspondingly undergone considerable structural and functional change over the course of human evolution, which has its reflection in the central nervous system. A good illustration of this is provided by the progressive increase in central representation of the facial and glossal musculature as one considers first the cat, then the monkey, and finally humans. If the premise is correct that a distinct emotion corresponds to each facial expression – even though higher primates may voluntarily suppress the expression itself – then there can be no doubt that the canon of emotions has evolved to its greatest extent, and greatest degree of differentiation, in humans. This is not to say that the total number of emotions that can be experienced cannot exceed the number of possible facial expressions (Leonhard 1997).

In view of the increasing complexity of emotion and the increasing complexity of the morphology and physiology of facial expressions over the course of evolution, it is questionable whether the notion of primary and secondary emotions accords well with the current understanding of nervous system function. A branching model, such as is found in many different biological systems, seems a likelier candidate to explain the processes underlying emotion. In such a branching system, the idea of a specific, and specifically localized, cerebral representation for each emotion would be better supported than that of primary and secondary systems of emotion. This idea was merely theoretical until recently, but has now gained experimental support with the aid of functional imaging techniques (e.g. Morris et al. 1996).

4
Emotion and Communication

The remarkable development of the facial and vocal apparatus for the expression of emotion serves an obvious purpose: each evolutionary step forward improves the ability of members of the species to communicate with each other (Jürgens and Ploog 1976). Especially in social mammals, including the subhuman primates, the extraordinary differentiation of communicative processes facilitates communal living and thereby promotes the survival of the species. It cannot be determined whether this differentiation of communication was a cause or an effect of the increasing complexity of social systems; the two developments probably went hand in hand, in small increments. It may be concluded, in any case, that there must have been constant selection pressure promoting the development of modes of communication (Ploog 1997).

Modes of communication among primates overwhelmingly serve social functions, while those devoted to the outside world are relatively few. The latter are mainly limited to warning and alarm cries and perhaps a few indicative sounds as well (Ploog 1972; Winter et al. 1973; Hopf et al. 1985). Facial and vocal expression, and also to a lesser extent bodily posture and some autonomically driven functions (hair position, glandular secretions, urine odor), are not only expressions of emotion, but also messages to other members of the species. Each expression is a signal which may occur at varying levels of intensity, ranging from a brief and (to humans) barely perceptible intentional movement to a long, repeatedly displayed complete movement. The best-known illustration is the intensity scale of threatening movements, ranging from a slight raising of the head, to facial and vocal expressions of varying intensities, to an imminent attack.

Even if the nonverbal aspects of human communication seem relatively insignificant in comparison to speech and may sometimes be manufactured or suppressed at will (particularly facial expressions), these phylogenetically old modes of communication are still very much in action. They unconsciously guide dialogue between individuals (as already in subhuman primates) (Ploog 1995) and are important indicators of defensiveness, approach, and other behavioral tendencies. As psychiatrists, we make diagnostic use of these nonverbal modes of expression. It may be questioned how well we succeed in decoding all of the messages they contain (Ploog 1992, 1993). Perhaps the greater or lesser ability to recognize such messages underlies the once decried and now rather derided "clinical intuition."

5 Hierarchy and Adaptation

As we remarked above, an evolved, differentiated system of communication is advantageous for communal life. A closer look at this communal life, with a sidelong glance at the human situation, will yield evidence for this assertion. Van Hooff (1973) used factor analysis to identify five categories, or systems, of social behavior in chimpanzees (our closest relatives): play, aggression, submission, bonding, and excitement. Only the last system mentioned partly concerns nonsocial, environmentally oriented behavior, while all of the other systems are partner or group oriented. No one observing chimpanzees, even superficially, could deny that these animals manifest emotions of each of these types that influence the behavior of their partners or of the group. The repertoire of communicative signals is selectively distributed and differentially applied within the different systems.

It is true of all primates, though to a variable extent depending on the species, that the use of social signals creates a hierarchical social organization, and a large amount of social activity is used to determine which animal takes which position in the rank ordering, and who plays what role. The maintenance of hierarchy is a dynamic process over the entire life cycle and is continually in flux because of aging, changes in group composition, and numerous environmental factors.

We described this process in squirrel monkeys in the early 1960s using sociometric methods (Ploog et al. 1963; Hopf 1967; Castell and Ploog 1967); it applies equally well to all old and new world monkeys studied thereafter. We shall see below that our primate inheritance plays a major role in human emotion. In any case, hierarchical organization seems to confer a reproductive advantage despite the large investment made by each member of the group. The sociobiologist's cost-benefit calculation comes out positive as long as the physical environment is favorable.

Human emotion is expressed mainly through facial expression and the voice. Animals, particularly subhuman primates, have innate expressive movements, homologous to human ones, which are generated by "preprogrammed" cerebral structures. The functional relationship between subjective (experienced) emotions and their expression is bidirectional: the neural programs generate expressive movements, and these movements activate patterns of central nervous activity (Ploog 1989). Expressive movements have a function. They occur mainly, but not exclusively, in the social context and act as communicative signals, which inform other members of the species of the individual's emotional state, thereby influencing their behavior. Social signals thus have an adaptive function with manifold effects in a hierarchically organized society.

Expressive movements highlight the adaptive function of the emotions. Emotions are at the root of every adaptive behavior. The ultimate causes of the emotions lie in the effects of adaptive behavior, e.g. self-protection, the appeasement or defeat of an opponent, reproduction (see Table 2). Just as appetite motivates the search for food, emotions motivate adaptive behavior in general. The emotions act as springs of activity.

6 Ethology of Affect

Up to this point, we have been discussing emotion as a general concept and have largely avoided more highly differentiated terms, such as mood, emotional state, feelings, and affect, that may involve the temporal dimension and imply a shorter or longer duration of the inner experience. We shall now discuss the origin and adaptive function of affect, by which is meant an acute, temporally limited, intense emotional occurrence.

According to Salzen's theory (1991), negative affect results from the prevention or frustration, by whatever means (often by a social partner), of an initiated or intended instinctive behavior (e.g. capture of prey). Positive affect, in contrast, is evoked when the removal of an obstacle enables the performance of instinctive behavior. The motor and autonomic phenomena appearing during this process serve as signals for the social partner and may evolve into ritualized movements (see examples below). The individual's own perception of the preprogrammed movement as it is performed, and of the accompanying autonomic changes, form a part of the experience (perception)

of the emotion, which in turn motivates further behavior and learning. Lorenz (1937) described this process, which he named "Instinkt-Dressur-Verschränkung" ("instinct-training interlock"), and thereby provided an ethological explanation for conditioned and learned behavior. This process may also be exploited in behavioral therapy (Ploog 1969).

Breland and Breland (1966) have shown experimentally that the execution of species-specific preformed movements ("fixed action patterns") is a necessary reinforcement in the learning process. The reward, or reinforcement, at the end of a behavioral chain with a desired objective need not always be food; it might be the sight of a long-absent member of the species, for example (Hupfer and Maurus 1975). Many mammals have developed species-specific sounds that signal separation from the group and, in turn, are answered by the group and bring about reintegration into it. We exploited this separation effect in learning experiments with juvenile squirrel monkeys. The isolated animal expresses its "separation distress" through an innate sound that we named the "isolation peep" (Winter et al. 1966). Production of this cry stops immediately upon the animal's rejoining the group.

Experimental attempts to get the animals to discriminate species-specific from synthetic sounds uniformly failed when food or water was used as the reward; the animals appeared to learn nothing. However, when another animal of the same species was shown to the isolated animal through a window, even for as little as 10 s, the learning curve climbed rapidly to a highly significant level. The sight of another species member is thus the reinforcement (reward) that guides this learning process (Hupfer et al. 1977). The innate behavior (isolation peep) is both an expression of emotion and a signal to the group, with the adaptive function of causing the reintegration of the isolated animal. Reintegration then leads to further emotional behavior in the form of a greeting ritual, consisting of mutual stereotypic movements and "positive" vocalizations.

The separation-and-reintegration model (Bowlby 1969, 1973) describes two situations that evoke affect: separation prevents fulfillment of the innate desire for affiliation (bonding), while reintegration satisfies this desire. The isolation peep is a rare instance of a behavior whose function (ultimate cause) and neural mechanism (proximal cause) can both be empirically demonstrated (Ploog 1979).

7 Bonding and Separation

Deprivation and separation experiments in monkeys, especially macaques, provide a well-known model for the human depressive syndrome (Harlow and Harlow 1962; Harlow and Suomi 1974). When juvenile monkeys aged 4–7 months are separated from their mothers, they let out loud cries, look for their mothers, and run around restlessly (phase of protest and despair). After 1–2 days, a state described as depression sets in. The animals cower, take no notice of their surroundings, and are apathetic. A similar picture results if one separates 3-month-old monkeys raised together, even after multiply repeated separations. The "friends" pass through the phase of protest and despair after each separation, and each reunification leads to new bonding. Ultimately, behavioral development is arrested. The animals cling to each other, lack curiosity, do not play, and show no interest in their surroundings (Suomi et al. 1970). Even 3-year-old adolescent monkeys develop a depressive behavioral syndrome after separation.

Considering the accumulated results of experiments on primates of various ages, with varying lengths of separation and varying lengths of prior social experience, it can no longer be doubted that the separation of tightly bonded animals leads to severe disturbances of affect and social behavior (Kaufman and Rosenblum 1969). It is, however, usually not mentioned that there are large individual differences in the severity and duration of the apathetic behavioral disturbance (Ploog 1980b).

Genetic differences play a role as well. It has been shown in two closely related macaque species that small interspecies differences in social structure and mother–child behavior produce different results in separation experiments. Bonnet macaques cling closely to one another, while pigtailed macaques maintain a greater distance. The bonnet macaques and their newborn infants remain in close contact with the group, while pigtailed macaque mothers and their infants keep away from the group. When juvenile bonnet macaques are separated from their mothers, they are "mothered" by the group as a whole and "adopted" by an "aunt," while the juvenile pigtailed macaques are left to themselves and are not uncommonly chased away by other group members. They develop a severe depressive syndrome, while the juvenile bonnet macaques, though they are upset and whimper for a time, nevertheless continue to show interest in their inanimate and social environment (Kaufman 1973). In humans, too, there are undoubtedly major individual differences in the realm of bonding, separation, and grief behavior. Twin studies reveal that such differences are, to a significant extent, genetically determined (Schepank 1982; Plomin et al. 1997; Segal and Bouchard 1993).

Rather than attempting to analyze the causes of depressive syndromes, we are discussing here the changes in emotionality resulting from interference

with the adaptive function of an emotion – in this case, the prevention or reinstatement of bonding (Ploog 1986b; see Table 2). The protest-and-despair reaction and the subsequent emotional disturbances are affected by multiple factors:

- The change of social environment after the loss of a tightly bonded member of the species
- The time of life at which the loss occurs
- The individual's own predisposition to the reaction
- Genetic determinants of social behavior

Attempts to explain human depressive disorders, invoking each of these factors in turn, have included the life-event, ontogenetic, vulnerability, personality, and genetic hypotheses. All of these are probably valid, depending on the circumstances. Only a few investigators have favored an evolutionary hypothesis (e.g. Price 1969; Sloman 1976; Gardner 1982; Price and Sloman 1987; McGuire and Essock-Vitale 1982; Pedersen et al. 1988; McGuire et al. 1992; Gilbert 1992); this will be discussed below.

We shall maintain that emotions, from the evolutionary point of view, represent basic phylogenetic adaptive processes that promote individual fitness and, thereby, the maintenance of the species. They not only inform the organism about internal bodily processes and external events, but also provide feedback regarding the effectiveness of its own behavior. Unpleasant emotions cause a change of goal-directed behavior, while pleasant emotions reinforce the organism's goal-directed behavior and thus enhance its fitness. Emotions of the former type include feelings of disappointment, of sadness, of fear, of powerlessness, of anger, of disgust; emotions of the latter type include feelings of being recognized, of being loved, of love, of happiness, of power, of triumph. These and all other emotions determine our behavior to a much greater extent than reason. In regard to human depressive disorders, we must ask what adaptive value is associated with depressive affect (Nesse 1998).

8 Grief and Empathy

In our search for the adaptive value of depressive affect in a sense relevant to psychopathology, we shall direct our attention to grief and its associated affect. These are evoked by the loss of a loved person and are certainly closely related to pathological depressive affect, though not identical with it.

A component of the Minnesota Twin Study (Segal et al. 1993) was devoted to the intensity of grief after the loss of a monozygotic twin or other near relative. Grief affect was expected to be stronger the more closely related the person who had died. A total of 49 mono- and 19 dizygotic surviving twins, as well as mothers, fathers, grandparents, and other relatives, were questioned about their grief and its (quantified) intensity in a standardized interview. The results were statistically significant and confirmed the hypothesis: the surviving monozygotic twins grieved more intensely than the dizygotic ones, and all twins grieved more intensely than mothers, fathers, and grandparents. Although the authors were cautious in their multidimensional evaluation of these results, it seems definitively demonstrated that there are genetically determined individual differences in the grief reaction, which, according to this study, are highly correlated with the closeness of the genetic relationship.

Empathy affect is closely related to grief affect (Brothers 1989). Darwin (1872) wrote: "When we witness any deep emotion, our sympathy is so strongly excited that close observation is forgotten or rendered almost impossible." It might seem that Darwin was concerned here merely with the "contagiousness," or transferal, of emotion from one individual to another. Other passages reveal, however, that he meant to describe the communication of an emotional state. Many developmental and experimental findings from human ontogeny show that the capacity for empathy matures in a number of developmental stages (Grossmann 1989; Grossmann and Grossmann 1991; Papoušek and Papoušek 1974).

The studies carried out by Bischof-Köhler (1989) conclusively reveal that receptive empathic understanding arises in the 20th to 24th month of life, in close association with the new ability to recognize oneself in a mirror. Empathic understanding contrasts sharply with the empathy defect of autistic children, who perceive social signals wrongly or not all, and with the alexithymia of adult patients (Krystal 1979; von Rad 1983), which presumably reflects a primary inability to perceive their own emotions.

9 Neuroethology of the Recognition of Social (Emotional) Signals

Up to this point, we have not explicitly distinguished the transmission of social signals from their reception. The fascinating and still unsolved problem of biological communication resides precisely in the coevolution of the transmitting and receiving apparatus for social signals. How can a human infant distinguish certain universal phonemes of speech as early as the first week of life and react to them with autonomically mediated responses (sucking frequency, heart rate) (Eimas et al. 1971)? How can it perceive and imitate

certain facial movements (sticking the tongue out; Meltzoff and Moone 1983) or modulations of the voice (Papoušek and Papoušek 1986)? Certain groups of neurons in the simian forebrain fire not only when the animal intends to move a limb (e.g. hand), but also when it sees the experimenter making such a movement (Rizzolatti et al. 1996).

The neural basis of the recognition of facial expression is particularly important to research on emotion. In the mid-1980s, workers in several laboratories independently discovered neurons in the midportion of the superior temporal sulcus and the amygdala of monkeys that displayed various responses only upon the presentation of pictures of monkey faces, with a preference for identical faces. Certain cortical cells in the superior temporal sulcus also responded to isolated portions of faces, such as the eyes or hairline, but fired more strongly when the entire face was displayed. Selectivity for facial features was found: some cells responded more strongly to the eyes, others to the mouth or the hairline. Finally, and most importantly for our purposes, there were significant differences in neuronal response depending on different facial expressions, directions of gaze, and facial orientations. There is no longer any doubt that the neuron populations studied are components of a module for the analysis of facial expression (Baylis et al. 1985; Leonard et al. 1985; Rolls and Baylis 1986; Perrett et al. 1982, 1984, 1985; Hasselmo et al. 1989).

More recently, it has also been demonstrated that injury to the amygdala leads to selective deficits in the perception of facial expression. A remarkable patient was studied by Adolphs and colleagues (1994). The total destruction of this phylogenetically ancient structure by calcium deposition (Urbach-Wiethe disease) made the patient unable to perceive facial expressions of fear and other emotions that normal individuals can distinguish effortlessly. A female patient with bilateral partial amygdalar injury, similarly impaired in the recognition of facial expressions, had great difficulty determining whether someone was looking at her or not (Young et al. 1995).

Experiments on the question of "being looked at," carried out on paranoid schizophrenics and controls, have revealed that normal individuals can reliably tell whether someone is looking them in the eye up to a distance of 80 cm, while paranoid patients imagined they were being looked at considerably larger distances (Ploog 1970). In social communication, whether among human beings or among subhuman primates, the direction of gaze signifies the individual concerned, while the facial expression signifies the intention, or readiness to act, of the "sender." Eye contact creates a reciprocal loop between the sender and the receiver. The amygdala has multisensory cortical inputs as well as outputs to the hypothalamus and brain stem (Aggleton and Mishkin 1986; LeDoux 1998) and is doubtless an important nodal point in the complex anatomy of emotion and social communication.

An indication that the perception of social signals (facial expressions) can be differentially impaired comes from the studies carried out by Berndl et al. (1986), Harrington et al. (1989), and Grüsser et al. (1990) in schizophrenics and by Sprengelmeyer et al. (1996) in patients with Huntington's disease. These patients had a mild impairment of the perception of direction of gaze, could not differentiate the expression of anger from that of fear, and were totally unable to recognize the expression of disgust. The authors rightly propose that certain primary emotions are subserved by specific neural substrates. Our own studies of simian vocal communication have demonstrated such functional relationships with certainty (Jürgens 1979; Ploog 1981).

10 Depressive Syndromes

After these preliminaries, we can now return to the question of the adaptive value of depressive affect (Price 1991). The likelihood that a depressive syndrome will develop is greater in the presence of excessive life stress, major material losses, loss of a loved one, divorce or loss of a partner, loss in conflict situations (e.g. sports), or loss of social position and power (i.e. loss of control over other persons and property). The common features of all of these life circumstances are that they dissolve bonds, make the achievement of expected goals questionable, and threaten the individual's position in the social rank ordering. Loss of fitness results. Depressive emotion might thus be seen as a suboptimal adaptation to negative life circumstances, an adaptation that no longer serves the purpose of optimal fitness.

In fact, however, depressive affect and its symptoms – be they exogenous, endogenous, or neurotic – evoke an empathic reaction in other individuals. They act as signals for help, lead to a temporary deferment of the material and social demands placed on the involved person, keep him or her in the family group, and preserve his or her social position. Depressive affect thus provides a grace period in which deficits can be made good and temporarily shifts the balance of material, personal, and social give-and-take toward the side of taking, i.e. toward a reduction of demands on the involved person.

It might be objected that any form of suffering or illness similarly evokes compassion and help from others, and that depressive affect therefore has no *specific* adaptive effect. This objection can be answered

as follows: The tremendous, organized effort against illness that modern culture has developed has biological roots and is likely a product of the coevolution of expressions of suffering and innate empathic reactions to them.

Further hypotheses are needed, however, to account for depression of moderate or prolonged duration, and for bipolar disorders, on an evolutionary basis. It remains to be explained how the primal fears that originated over the course of evolution – the fears of failing to meet excessively high demands, losing one's property, becoming powerless, feeling guilty and worthless, losing persons close to oneself, losing the respect of society – may arise and grow spontaneously, entirely without apparent cause. Biochemical hypotheses of proximate causation will have to be introduced (Nemeroff 1998).

If proximate causes are assumed to operate in the framework of ultimate (evolutionary) causes, then experimental findings relating to proximate causes must be reinterpreted as revealing not deficient neural or biochemical functioning, but rather attempts at compensation by the central nervous system, leading to suboptimal adaptation.

The temporal dimension of the depressive condition also requires explanation. Emotions have a limited temporal course. They usually arise suddenly and then subside. The depressive affect occurring during a depressive illness, however, may be moderately or extremely prolonged; abnormal duration is, indeed, the sign that this is a pathologic rather than a normal process. When depression becomes subacute or chronic, the benevolent attention of others may, not infrequently, give way to abandonment. The depressive signals that at first evoked empathy later habituate.

Price et al. (1994) developed an evolutionary hypothesis to explain human depression, while taking account of its temporal course. This hypothesis, too, rests on the fact that humans – like most terrestrial vertebrates – are born into a hierarchically organized society and have to compete for their place in it over the entire course of their life. Adaptive behavior is continually demanded. We have already shown (see Fig. 2) an extreme example of this. It is described how a wife's subordination, on the basis of a strong emotional bond, progressed ultimately to self-sacrifice (Grossmann 1993).

This case reveals the nature of the "social contest hypothesis." The contest consists not only of the attempt to be "better" or more successful than the next higher individual in the rank ordering, but also the attempt to gain confederates for mutual support in the upward struggle, as well as the attempt to obtain recognition and confirmation. Simultaneously, and (as we are about to discuss) paradoxically, close bonds may be formed at any time of life, primarily within the family group (to parents, children, siblings and their children, and so forth), but also outside it. The paradoxical aspect of this social network is that bonding behavior can either promote or impede the process of social adaptation.

The description provided here of the individual's immediate social environment applies not only to human society, but also to that of our nearest primate relatives (de Waal 1989). Only in the human case, however, can we appreciate the importance of recognition, and the intensity of the associated emotions. The recognition or admiration of those of equal or higher social rank or prestige raises an individual's mood and self-esteem more than the recognition or admiration of those considered of lower status (Gilbert 1992). Nonetheless, the admiration of many (or even of "the masses," as one might say) adds strongly, sometimes even excessively, to the individual's self-esteem.

The unfortunate wife we have mentioned gradually maneuvered herself into the position of a hopeless loser who can no longer offer any resistance. Her submissive posture manifested itself in chronic depressive symptoms. Alongside this model of depression is the analogous one of a persistently "agonistic" position. Such patients, despite generally being the loser in conflict situations, can never free themselves from the interpersonal power struggle with those close to them and therefore remain permanently at odds with their spouses, families, or other persons, are routinely "defeated," and therefore develop feelings of worthlessness, failure, incompetence, and helplessness.

In the normal course of development, a child acquires a realistic understanding of its own place in the family hierarchy, in kindergarten, and in school and learns to accept occasional defeats. In adulthood, too, there are good and poor losers. The latter are repeatedly beset by internal conflict over whether to keep fighting or give up. When this conflict becomes protracted and intense, the individual becomes depressed and is thereby relieved of the need to resolve the conflict. In a smoothly functioning hierarchical system, the winners cease their aggression and the losers their submissive behavior once the conflict is over. Individuals prone to depressive behavior, however, fail to recognize the signals that indicate the end of the conflict. They continue to behave as if they could be attacked again at any moment, and they thus fall, more readily than others, into a vicious circle in which attack and appeasement maneuvers follow close upon one another.

This model of depression also includes the idea of evolutionary selection mechanisms resulting in a hereditary predisposition to depression (Halliday 1983). Such arguments date back to Darwin's principles of sexual selection (1871) and have to do with the

attractiveness of the individual as a potential partner (Sloman 1977). Repeated success in social competition leads to increased self-confidence and improves the chances of future success, while repeated failure diminishes these same chances.

The effect is felt in partner selection as elsewhere: winners radiate a more powerful attractiveness than losers. A differential amplification of traits and status behavior results, so that winners are more likely to pair with winners, and losers with losers. The predisposition to depressive behavior patterns could thus be explained on a familial or population-genetic basis, considering that human history has now been in progress for hundreds of thousands of years. Indeed, recent studies have shown that attractiveness and social success are important parameters for partner selection even in our modern society (Grammer 1995; Grammer and Thornhill 1994).

Years ago, in ethological studies of the green iguana, we were literally eyewitnesses to the fact that the individual's position in the hierarchy affects the function of its autonomic nervous and endocrine systems. These animals live according to a linear pecking order. The alpha animal's head is bright white, an emblem of its rank, as it were. One such "boss," splendidly green with a white head, had increasingly to defend himself against a challenger one level down in the hierarchy. One day, it lost a major duel and immediately thereafter became grayish brown, and its head lost its white color. It crept away into a corner and remained submissive for the remaining months of the period of observation (Ploog 1970).

Sapolsky (1990) described hypercortisolism in subordinate free-ranging baboons, and many other investigators have described similar associations of rank ordering and central nervous function in mammals. Ethological and neuroethological studies thus provide a significant contribution to both the "state concept" and the "trait concept" of depression.

11 Anxiety

In this concluding section, we shall consider anxiety, the most common accompaniment of depressive affect, in its evolutionary aspect. The neuroanatomical, neurophysiological, and neurochemical bases of anxiety have come under particularly intense study in the past decade (Charney et al. 1998; LeDoux 1998). Anxious feelings are among the most common psychopathologic symptoms and undoubtedly arise from many different causes (Strian 1983, 1996). Nevertheless, it is easy to see that anxiety is an affect of enormous importance to the survival of the individual.

The development of a warning system that signals danger to the organism, and consequently induces fight-or-flight behavior, is an early evolutionary achievement of high adaptive value. In the ongoing concert of emotions, anxious affect drowns out all others and thus ensures the fastest possible reaction to danger. The releasers of anxious affect and its behavioral consequences – avoidance or flight – in lower vertebrates are exclusively innate; thus the behavioral reaction to the releasing stimulus does not have to be learned (Liebsch et al. 1998). Even among subhuman primates, species-specific warning sounds release innate behavioral patterns of searching for cover (Winter et al. 1966; Hopf et al. 1985). Countless experiments in rats and many other species have shown, however, that anxiety reactions can be conditioned, and then released, by practically any arbitrary stimulus. The original, innate releasers of anxiety may be replaced by others through experience and learning, but the anxious affect itself, and its behavioral consequences, remain the same.

In the case of pathological human anxiety released by a known stimulus – such as phobia of various types, or social anxiety – the methods of behavioral therapy may be effectively applied, as is well known, but the treatment of indefinable anxious states is more difficult. Post-traumatic stress syndrome, however, is an exception to this rule (Strian and Ploog 1992; Bronisch 1997). The cause of this disorder – the traumatic event or series of traumatic events – is known, retains its intensity and vividness over a long period, and evokes the same mental and autonomic symptoms of anxiety each time it is recollected. Nonetheless, attempts to neutralize the anxiety-producing event or events by behavior therapy appear to be ineffective.

Among patients with anxiety of one or more types, those with social anxieties are most common. From the evolutionary viewpoint, even these anxieties, which often seem easily explicable in terms of the sufferer's life history, have a phylogenetic origin. It lies in the establishment of a hierarchical rank order, a form of organization common to most social vertebrates, including all diurnal primates. Clearly, this social structure results in optimization of the probability of survival of the individual and of the species.

From childhood onward, human beings learn to fit into the hierarchical order through dominant and submissive modes of behavior, through coalitions and behavioral strategies, and thus to achieve the highest possible rank. Every day so lived is a new source of anxiety. In any case, the forms of expression of anxious affect through vocal and facial expression and autonomically regulated functions may be observed as early as the first days of life, and there is scarcely any

doubt that these are an important part of a primary, innate system of affect that directs our social behavior (Ploog 1997). All human beings have social anxieties, but only some experience a major limitation of their freedom to act because of them and thus become socially dysfunctional.

12 Conclusions

The emotions acquired over the course of evolution – pleasure, displeasure, surprise, anger, rage, anxiety, fear, disappointment, worry, and grief; feelings of love, of belonging, of hatred, of disgust; and all other emotions – are the common possession of mankind. They are the aspect of phylogenetically based, species-specific adaptive behavior that is open to subjective experience, and they serve the purpose of sustaining the life of the individual and of his or her progeny. The means by which they are expressed by various bodily organs, mainly through facial expressions and the voice, have evolved over millions of years and are subserved by the central and autonomic nervous systems. Facial expressions are primarily innate and are among the instinctive movements of the species ("fixed action patterns"). Each emotional expression creates an impression in other people ("receivers"), acts as a signal to them, and affects their behavior in turn; this process gives expressive behavior its adaptive function. The reciprocal exchange of signals constitutes a process of communication that guides and regulates the behavior of all individuals concerned. Humans have brought their expressive behavior under voluntary control and can apply it manipulatively in both verbal and nonverbal communication.

The function of emotion is not limited to expressive behavior and communication; indeed, emotion is the spring of all adaptive behavior in general. Human beings live in a social system and are equipped with species-specific behavior, which enables them to pursue biologically relevant objectives. Any behavior that brings them closer to these goals is adaptive. Despite the extraordinary variety of goal-directed human behavior, most of which is learned, the emotions that drive it and those that accompany its failure or successful accomplishment have remained remarkably universal over the entire documented course of human history. These emotions are undoubtedly anchored in the human genome; the canon of emotions capable of being experienced by humans is innate. Emotions cannot be learned, whether by imitation or by any other means. What can be learned, i.e. stored in memory, are those objects, events, persons, and social constellations that have evoked certain emotions and remain associated with them. The experience of emotion can be conveyed to others only by human speech, although, as with the recounting of dreams, what is conveyed is probably no more than a pale shadow of the experiential reality. I understand the feelings of others only because I have had similar feelings myself; I understand primarily by empathy, not by experience. I shall never know whether the anxiety, sadness, or joy of another person is exactly the same as mine.

In a hierarchically organized society, like those of subhuman primates and of humans, a large part of social activity is devoted to attaining the highest possible rank and to defending one's rank and role as long as possible. The goal of all adaptive behavior is the securing of the greatest possible gain from the physical and social environment with which the individual is confronted. This fact of life also lies at the root of human psychopathology. Emotions underlie all motivated activity and are the motor of all adaptive behavior. As the human population grows, the universal human objective of maximizing personal gain, which is anchored in the human genome, is gradually becoming the greatest threat to the survival of our species. The current appeals for the limitation of population growth are thus justly founded in human biological nature.

13 References

Adolphs R, Tranel D, Damasio H, Damasio A (1994) Impaired recognition of emotion in facial expressions following bilateral damage to the human amygdala. Nature 372: 669–672

Aggleton JP, Mishkin M (1986) The amygdala: sensory gateway to the emotions. In: Plutchik R, Kellerman H (eds) Emotion. Theory, research, and experience. 3. Biological foundations of emotions. Academic, London, pp 281–299

Baylis GC, Rolls ET, Leonard CM (1985) Selectivity between faces in the responses of a population of neurons in the cortex in the superior temporal sulcus of the monkey. Brain Res 342: 91–102

Berndl K, Cranach MV, Grüsser O-J (1986) Impairment of perception and recognition of faces, mimic expression and gestures of schizophrenic patients. Eur Arch Psychiatr Neurol Sci 235: 282–291

Bischof-Köhler D (1989) Spiegelbild und Empathie. Die Anfänge der sozialen Kognition. Huber, Bern

Bowlby J (1969) Attachment and loss. I. Attachment. International psycho-analytical library no 79. Hogarth, London

**Bowlby J (1973) Attachment and loss. II. Separation: anxiety and anger. International psycho-analytical library no. 95. Hogarth, London

Breland K, Breland M (1966) Animal behavior. Macmillan, New York

Bronisch T (1997) Posttraumatic stress disorder – posttraumatische Belastungsstörung. Neuere Forschungsergebnisse. Fortschr Neurol Psychiatr 65: 195–207

Brothers L (1989) A biological perspective on empathy. Am J Psychiatry 146: 1–19

Buck R (1988) Human motivation and emotion. Wiley, New York

Cannon W (1927) The James-Lange theory of emotion: a critical examination and an alternative theory. Am J Psychol 39: 106–124

Castell R, Ploog D (1967) Zum Sozialverhalten der Totenkopf-Affen (Saimiri sciureus): Auseinandersetzung zwischen zwei Kolonien. Z Tierpsychol 24: 625–641

Charney DS, Grillon CCG, Bremner JD (1998) The neurobiological basis of anxiety and fear: circuits, mechanisms, and neurochemical interactions (part II). Neuroscientist 4: 122–132

Chevalier-Skolnikoff S (1973) Facial expression of emotion in non-human primates. In: Ekman P (ed) Darwin and facial expression. Academic, New York

Ciompi L (1991) Affects as central organizing and integrating factors. A new psychosocial/biological model of the psyche. Br J Psychiatry 159: 97–105

Darwin C (1871) The descent of man and selection in relation to sex. Murray, London

Darwin C (1872) The expression of the emotions in man and animals. Murray, London

Dawkins R (1976) The selfish gene. Oxford University Press, Oxford

Dawkins R (1988) Auf welche Einheiten richtet sich die natürliche Selektion? In: Meier H (ed) Die Herausforderung der Evolutionsbiologie. Piper, Munich, pp 53–78

de Waal F (1989) Peacemaking among primates. Harvard University Press, Cambridge, MA

**Eibl-Eibesfeldt I (1995, [1]1984) Die Biologie des menschlichen Verhaltens. Grundriß der Humanethologie, 3rd edn. Piper, Munich

Eimas PD, Siqueland ER, Jusczyk P, Vigorito J (1971) Speech perception in infants. Science 171: 303–306

**Ekman P (ed) (1973) Darwin and facial expression: a century of research in review. Academic, New York

Ekman P, Friesen WV (1978) Facial action coding system. Consulting Psychologists, Palo Alto

Ekman P, Sorenson ER, Friesen WV (1969) Pan-cultural elements in facial display of emotion. Science 164: 86–88

Ekman P, Levenson RW, Friesen WV (1983) Autonomous nervous system activity distinguishes between emotions. Science 221: 1208–1210

Ellgring H (1989) Nonverbal communication in depression. Cambridge University Press, Cambridge

Ellgring H, Smith M (1998) Affect regulation during psychosis. In: Flack WF Jr, Laird JD (eds) Emotions in psychopathology. Theory and research. Oxford University Press, Oxford, pp 323–335

Emde RN, Gaensbauer TJ, Harmon RJ (1976) Emotional expressions in infancy. International Universities Press, New York

**Flack WF Jr, Laird JD (1998) Emotions in psychopathology. Theory and research. Oxford University Press, Oxford

*Frijda NH (1986) The emotions. Cambridge University Press, Cambridge

Gardner R (1982) Mechanisms of manic-depressive disorder, an evolutionary model. Arch Gen Psychiatry 39: 1436–1441

Gilbert P (1992) Depression: the evolution of powerlessness. Erlbaum, Hillsdale

Grammer K (1995) Signale der Liebe, 3rd edn. dtv-Wissenschaft, Munich

Grammer K, Thornhill R (1994) Human facial attractiveness and sexual selection: the roles of averageness and symmetry. J Comp Psychol 108: 233–242

Grossmann KE (1989) Die Bindungstheorie: Modell und entwicklungspsychologische Forschung. In: Keller H (ed) Handbuch der Kleinkindforschung. Springer, Berlin Heidelberg New York, pp 31–61

*Grossmann KE (1993) Bindungsverhalten und Depression. In: Hell D (ed) Ethologie der Depression. Fischer, Stuttgart, pp 65–79

Grossmann KE, Grossmann K (1991) Attachment quality as an organizer of emotional and behavioral responses in a longitudinal perspective. In: Parkes CM, Stevenson-Hinde J, Marris P (eds) Attachment across the life cycle. Tavistock/Routledge, London, pp 93–114

Grüsser O-J, Kirchhoff N, Naumann A (1990) Brain mechanisms for recognition of faces, facial expression, and gestures: neuropsychological and electroencephalographic studies in normals, brain-lesioned patients, and schizophrenics. Res Publ Assoc Res Nerv Ment Dis 67: 165–193

Halliday JR (1983) The study of mate choice. In: Bateson P (ed) Mate choice. Cambridge University Press, Cambridge, pp 3–32

Hamilton WD (1964) The genetical evolution of social behaviour. J Theoret Biol 7: 1–32

Hamilton WD (1975) Innate social aptitudes of man: an approach from evolutionary genetics. In: Fox R (ed) Biosocial anthropology. Malaby, London, pp 133–135

Harlow HF, Harlow K (1962) Social deprivation in monkeys. Sci Am 207: 137–146

Harlow HF, Suomi SJ (1974) Induced depression in monkeys. Behav Biol 12: 273–296

Harrington A, Oepen G, Spitzer M (1989) Disordered recognition and perception of human faces in acute schizophrenia and experimental psychosis. Compr Psychiatry 30: 376–384

Hasselmo ME, Rolls ET, Baylis GC (1989) The role of expression and identity in the face-selective responses of neurons in the temporal visual cortex of the monkey. Behav Brain Res 32: 203–218

Hopf S (1967) Ontogeny of social behavior in the squirrel monkey. In: Starck D, Schneider R, Kuhn H-J (eds) Neue Ergebnisse der Primatologie. Progress in primatology. Fischer, Stuttgart, pp 255–262

Hopf S, Herzog M, Ploog D (1985) Development of attachment and exploratory behavior in infant squirrel monkeys under controlled rearing conditions. Int J Behav Dev 8: 55–74

Hupfer K, Maurus M (1975) Operant conditioning of the squirrel monkey with social reinforcement. Naturwissenschaften 62: 42–43

Hupfer K, Jürgens U, Ploog D (1977) The effect of superior temporal lesions on the recognition of species-specific calls in the squirrel monkey. Exp Brain Res 30: 75–87

Izard CE (1977) Human emotions. Plenum, New York

James W (1890/1950) The principles of psychology, vol II, chap xxv. The emotions. Dover, New York, pp 442–485

Jürgens U (1979) Vocalization as an emotional indicator, a neuroethological study in the squirrel monkey. Behaviour 69: 88–117

Jürgens U, Ploog D (1976) Zur Evolution der Stimme. Arch Psychiatr Nervenkr 222: 117–137

Kaufman IC (1973) The role of ontogeny in the establishment of species-specific patterns. Res Publ Assoc Res Nerv Ment Dis 51: 381–397

Kaufman IC, Rosenblum LA (1969) Effects of separation from mother on the emotional behavior of infant monkeys. Ann NY Acad Sci 159: 681–695

Kraepelin E (1916) Einführung in die psychiatrische Klinik, 3rd edn. Barth, Leipzig, p 206

Kraepelin E (1920) Die Erscheinungsformen des Irreseins. Z Ges Neurol Psychiatr 62: 1–29

Kretschmer E (1953) Der Begriff der motorischen Schablonen und ihre Rolle in normalen und pathologischen Lebensvorgängen. Arch Psychiatr Nervenkr 190: 1–3

Krystal H (1979) Alexithymia and psychotherapy. Am Psychother 33: 17–31

**LeDoux J (1998) The emotional brain: the mysterious underpinnings of emotional life. Touchstone/Simon and Schuster, New York

Leonard CM, Rolls ET, Wilson FAW, Baylis GC (1985) Neurons in the amygdala of the monkey with responses selective for faces. Behav Brain Res 15: 159–176

Leonhard K (1997) Der menschliche Ausdruck in Mimik, Gestik und Phonik, 3rd edn. Wernicke-Kleist-Leonhard-Schriftenreihe, Würzburg

Leyhausen P (1956/1973) Verhaltensstudien an Katzen, 3rd edn. Parey, Berlin

Liebsch G, Montkowski A, Holsboer F, Landgraf R (1998) Behavioural profiles of two Wistar rat lines selectively bred for high or low anxiety-related behaviour. Behav Brain Res 94: 301–310

Lorenz K (1937) Über die Bildung des Instinktbegriffes. Naturwissenschaften 25: 289–300, 307–318, 324–331

Lorenz K (1953) Über angeborene Instinktformen beim Menschen. Dtsch Med Wochenschr 78: 1566–1569, 1600–1604

Lorenz K (1963) Das sogenannte Böse. Borotha-Schoeler, Vienna

*Lorenz K (1992) Die Naturwissenschaft vom Menschen. Einführung in die vergleichende Verhaltensforschung. Das "Russische Manuskript". Piper, Munich

MacLean PD (1949) Psychosomatic disease and the "visceral brain". Psychosom Med 11: 338–353

*MacLean PD (1990) The triune brain in evolution. Plenum, New York

McGuire MT, Essock-Vitale SM (1982) Psychiatric disorders in the context of evolutionary biology. The impairment of adaptive behavior during exacerbation and remission of psychiatric illness. J Nerv Ment Dis 170: 9–20

**McGuire M, Troisi A (1998) Darwinian Psychiatry. Oxford University Press, New York

McGuire MT, Marks I, Nesse RM, Troisi A (1992) Evolutionary biology: a basic science for psychiatry? Acta Psychiatr Scand 86: 89–96

Meltzoff AN, Moone MK (1983) Newborn infants imitate adult facial gestures. Child Dev 54: 702–709

Morris JS, Frith CD, Perrett DI, Rowland D, Young AW, Calder AJ, Dolan RJ (1996) A differential neural response in the human amygdala to fearful and happy facial expressions. Nature 383: 812–815

Nemeroff CB (1998) The neurobiology of depression. Sci Am 278(6): 28–35

Nesse R (1998) Emotional disorders in evolutionary perspective. Br J Med Psychol 71: 397–415

Oatley K, Jenkins JM (1996) Understanding emotions. Blackwell, Cambridge, MA

Panksepp J (1982) Toward a general psychobiological theory of emotions. Behav Brain Sci 5: 407–467

Papoušek H, Papoušek M (1974) Mirror image and self-recognition in young infants. I. A new method of experimental analysis. Dev Psychobiol 7: 149–157

*Papoušek H, Papoušek M (1986) Structure and dynamics of human communication at the beginning of life. Eur Arch Psychiatr Neurol Sci 236: 21–25

Pedersen J, Schelde JTM, Hannibal E, Behnke K, Nielsen BM, Hertz M (1988) An ethological description of depression. Acta Psychiatr Scand 78: 320–330

Perrett DI, Rolls ET, Caan W (1982) Visual neurons responsive to faces in the monkey temporal cortex. Exp Brain Res 47: 329–342

Perrett DI, Smith PAJ, Potter DD, Mistlin AJ, Head AS, Milner AD, Jeeves MA (1984) Neurones responsive to faces in the temporal cortex: studies of functional organization, sensitivity to identity and relation to perception. Hum Neurobiol 3: 197–208

Perrett DI, Smith PAJ, Potter DD, Mistlin AJ, Head AS, Milner AD, Jeeves MA (1985) Visual cells in the temporal cortex sensitive to face view and gaze direction. Proc R Soc London (Biol Sci) 223: 293–317

Plomin R, DeFries JC, McClearn GE, Rutter M (1997) Behavioral genetics, 3rd edn. Freeman, New York

Ploog D (1957) Motorische Stereotypien als Verhaltensweisen. Nervenarzt 28: 18–22

Ploog D (1958) Endogene Psychosen und Instinktverhalten. Fortschr Neurol 26: 83–98

*Ploog D (1964) Verhaltensforschung und Psychiatrie. In: Gruhle HW, Jung R, Mayer-Gross W, Müller M (eds) Psychiatrie der Gegenwart, vol I/1B. Springer, Berlin Göttingen Heidelberg, pp 291–443

Ploog D (1969) Die Trieb-Dressur-Verschränkung in der Verhaltenstherapie. Prax Psychother 14: 167–170

Ploog D (1970) Social communication among animals. In: Schmitt FO (ed) The neurosciences. Rockefeller University Press, New York, pp 349–361

Ploog D (1972) Kommunikation in Affengesellschaften und deren Bedeutung für die Verständigungsweisen des Menschen. In: Gadamer H-G, Vogler P (eds) Neue Anthropologie. Thieme, Stuttgart, pp 98–178

Ploog D (1979) Phonation, emotion, cognition, with reference to the brain mechanisms involved. In: Brain and mind. Excerpta Medica, Amsterdam, pp 79–98 (Ciba Foundation series 69)

*Ploog D (1980a) Soziobiologie der Primaten. In: Kisker KP, Meyer J-E, Müller C, Strömgren E (eds) Psychiatrie der Gegenwart, vol I/2, 2nd edn. Springer, Berlin Heidelberg New York, pp 379–544

Ploog D (1980b) Verhaltensbiologische Ansätze zur Depressionsforschung. In: Heimann H, Giedke H (ed) Neue Perspektiven in der Depressionsforschung. Huber, Bern, pp 18–26

Ploog D (1981) Neurobiology of primate audio-vocal behavior. Brain Res Rev 3: 35–61

Ploog D (1986a) Biological foundations of the vocal expressions of emotions. In: Plutchik R, Kellerman H (eds) Emotion: theory, research and experience. III. Biological foundations of emotions. Academic, New York, pp 173–197

Ploog D (1986b) Zur Psychopathologie der Emotionen unter neuroethologischem Aspekt. In: Heimann H, Gaertner HJ (eds) Das Verhältnis der Psychiatrie zu ihren Nachbardisziplinen. Springer, Berlin Heidelberg New York, pp 15–31

Ploog D (1988) An outline of human neuroethology. Hum Neurobiol 6: 227–238
Ploog D (1989) Human neuroethology of emotion. Progr Neuropsychopharmacol Biol Psychiatr 13[Suppl]:S15–S22
Ploog D (1992) Ethological foundations of biological psychiatry. In: Emrich HM, Wiegand M (eds) Integrative biological psychiatry. Springer, Berlin Heidelberg New York, pp 3–35
Ploog D (1993) Psychopathologische Prozesse in neuroethologischer Sicht. In: Schüttler R (ed) Organische Psychosyndrome. Springer, Berlin Heidelberg New York, pp 1–28
Ploog DW (1995) Mutuality and dialogue in subhuman primate communication. In: Marková I, Graumann CF, Foppa K (eds) Mutualities in dialogue. Cambridge University Press, Cambridge, pp 27–57
Ploog D (1997) Das soziale Gehirn des Menschen. In: Meier H, Ploog D (eds) Der Mensch und sein Gehirn. Die Folgen der Evolution. Serie Piper 2457. Piper, Munich, pp 235–252
Ploog D, Blitz J, Ploog F (1963) Studies on social and sexual behavior of the squirrel monkey (Saimiri sciureus). Folia Primatol 1: 29–66
Plutchik R (1962) The emotions: facts, theories and a new model. Random House, New York
Plutchik R (1980) Emotion: a psycho-evolutionary synthesis. Harper and Row, New York
Plutchik R (1984) Emotions: a general psychoevolutionary theory. In: Scherer KR, Ekman P (eds) Approaches to emotions. Erlbaum, Hillsdale, pp 197–219
Plutchik R (1985) On emotion: the chicken-and-egg problem revisited. Motivation Emotion 9: 197–200
*Plutchik R (1994) The psychology and biology of emotion. HarperCollins, New York
Prechtl HFR (1974) The behavioral states of the newborn infant (a review). Brain Res 76: 184–212
Price JS (1969) Neurotic and endogenous depression: A phylogenetic view. Br J Psychiatry 114: 119–120
Price JS (1998) The adaptive function of mood change. Br J Med Psychol 71: 465-477
Price JS, Sloman L (1987) Depression as yielding behavior: An animal model based on Schjelderup-Ebbe's pecking order. Ethol Sociobiol 8: 85S-98S
Price J, Sloman L, Gardener R Jr, Gilbert P, Rohde P (1994) The social competition hypothesis of depression. Br J Psychiatry 164: 309–315
*Reeve J (1992) Understanding motivation and emotion. Holt, Rinehart and Winston, Fort Worth
Rizzolatti G, Fadiga L, Gallese V, Fogassi L (1996) Premotor cortex and the recognition of motor actions. Cogn Brain Res 3: 131–141
Rolls ET, Baylis GC (1986) Size and contrast have only small effects on the responses to faces of neurons in the cortex of the superior temporal sulcus of the monkey. Exp Brain Res 65: 38–48
Salzen EA (1991) On the nature of emotion. Int J Comp Psychol 5: 47–88
Sapolsky RM (1990) Stress in the wild. Sci Am 262: 106–113
Schachter S, Singer JE (1962) Cognitive, social, and physiological determinants of emotional state. Psychol Rev 69: 379–399
Schenkel R (1947) Ausdrucksstudien an Wölfen. Behaviour 1: 81–129
Schepank H (1982) Ergebnisse zur Frage der Erblichkeit. 1. Berliner Sample. In: Heigl-Evers A, Schepank H (eds) Ursprünge seelisch bedingter Krankheiten. Eine Untersuchung an 100+9 Zwillingspaaren mit Neurosen und psychosomatischen Erkrankungen. II. Ergebnisse. Verlag für Med Psychol im Verlag Vandenhoeck & Ruprecht, Göttingen, pp 377–403
*Schneider F, Weiss U, Kessler C (1998a) Funktionelle Kernspintomographie von Emotionen bei psychiatrischen Patienten. In: Gaebel W, Falkai P (eds) Zwischen Spezialisierung und Integration – Perspektiven der Psychiatrie und Psychotherapie. Springer, Berlin Heidelberg New York, pp 50–54
Schneider F, Weiss U, Kessler C, Salloum JB, Posse S, Grodd W, Müller-Gärtner HW (1998b) Differential amygdala activation in schizophrenia during sadness. Schizophr Res 35: 133–142
Schneider K, Scherer KR (1988) Motivation und Emotion. In: Immelmann K, Scherer KR, Vogel C, Schmoock P (eds) Psychobiologie. Grundlagen des Verhaltens. Fischer, Stuttgart/Psychologische Verlagsunion, Weinheim, pp 257–288
Segal NL, Bouchard TJ Jr (1993) Grief intensity following the loss of a twin and other relatives: test of kinship genetic hypotheses. Human Biol 65: 87–105
Sloman L (1976) The role of neurosis in phylogenetic adaptation, with particular reference to early man. Am J Psychiatry 133: 543–547
Sloman L (1977) The role of attractiveness and mate selection in phylogenetic adaptation with particular reference to early man. Biol Psychiatry 12: 487–494
Sprengelmeyer R, Young AW, Calder AJ, Karmat A, Lange H, Hömberg V, Perrett DI, Rowland D (1996) Loss of disgust. Perception of faces and emotions in Huntington's disease. Brain 119: 1647–1665
Sroufe LA (1979) Socioemotional development. In: Osofsky J (ed) Handbook of infant development. Wiley, New York, pp 462–513
Steiner JE (1974) Innate, discriminative human facial expressions to taste and smell stimulation. Ann NY Acad Sci 237: 229–233
Strian F (1983) Angst – Grundlagen und Klinik. Ein Handbuch zur Psychiatrie und Medizinischen Psychologie. Springer, Berlin Heidelberg New York
Strian F (1996) Angst und Angstkrankheiten, 2nd edn. Beck, Munich
Strian F, Ploog D (1992) Post-traumatic stress disorder – neuronal damage from catastrophic events? In: Burrows GD, Roth Sir M, Noyes R Jr (eds) Handbook of anxiety, vol 5. Elsevier, Amsterdam, pp 365–386
Sulz SKD (1993) Verhaltenstherapie. In: Möller HJ (ed) Therapie psychiatrischer Erkrankungen. Klinische Psychologie und Psychopathologie, vol 58. Enke, Stuttgart, pp 63–73
Suomi SJ, Harlow HF, Domek CJ (1970) Effect of repetitive infant-infant separation of young monkeys. J Abnorm Soc Psychol 76: 161–172
Tomkins SS (1962) Affect, imagery, and consciousness: the positive affects, vol 1. Springer, Berlin Heidelberg New York
Tomkins SS (1963) Affect, imagery, and consciousness: the negative affects, vol 2. Springer, Berlin Heidelberg New York
Tomkins SS (1970) Affect as the primary motivational system. In: Arnold MB (ed) Feelings and emotions. Academic, New York, pp 101–110
Trivers RL (1985) Social evolution. Cummings, Menlo Park
Van Hooff JARAM (1973) A structural analysis of the social behavior of a semicaptive group of chimpanzees. In: von Cranach M, Vine I (eds) Social communication and movement. Academic, New York, pp 75–162
von Rad M (1983) Alexithymie. Empirische Untersuchungen zur Diagnostik und Therapie psychosomatisch Kranker. Mono-

graphien aus dem Gesamtgebiet der Psychiatrie, vol 30. Springer, Berlin Heidelberg New York

Wilson EO (1975) Sociobiology: the new synthesis. Harvard University Press, Cambridge, MA

Winter P, Ploog D, Latta J (1966) Vocal repertoire of the squirrel monkey (Saimiri sciureus), its analysis and significance. Exp Brain Res 1: 359–384

Winter P, Handley P, Ploog D, Schott D (1973) Ontogeny of squirrel monkey calls under normal conditions and under acoustic isolation. Behaviour 47: 230–239

Wolff PH (1987) The development of behavioral states and the expression of emotions in early infancy. University of Chicago Press, Chicago

Wolff PH (1993) Behavioral and emotional states in infancy: a dynamic perspective. In: Smith IB, Thelen E (eds) A dynamic systems approach to development: applications. MIT, Cambridge, MA, pp 189–208

Young AW, Aggleton JP, Hellawell DJ, Johnson M, Broks P, Hanley JR (1995) Face processing impairments after amygdalotomy. Brain 118: 15–24

Zivin G (1989) Some basic considerations in the field of expressive behavior development. Monogr Soc Res Child Dev 54: 114–124

CHAPTER

20

H.M. Emrich, W. Schiefenhövel

Philosophical Anthropology: Basic Science of Psychiatry

Translator: S. Henn

1 Introduction: Significance of Philosophical Anthropology for Psychiatry

The basic question in philosophical anthropology concerning the nature, the characteristic features, of human beings is, in a sense, the starting-point of all science, because the acquisition of knowledge, the understanding of "what in actual fact is" (Hegel 1807; see Spaemann and Löw 1981), cannot completely disregard the essence that acquires this knowledge. In this respect at least, all scientifically sound and lasting cultural epochs of humanity were epochs with questions of philosophical anthropology. However, anthropology has an additional, an even more specific significance for psychiatry, because anthropology is irrefutably a fundamental science for psychiatry inasmuch as psychiatric illnesses, viewed as departures from healthy psychological life, always both refer to and are counterpoint to what it means to be human. At the same time, the fundamental problem of all psychiatric anthropology arises, namely that, whether we like it or not, it is always concerned with ideal typology, with establishing criteria for normality, even though this might be problematic and might open up the way to both misunderstandings and abuse (see Weber 1993; Mitscherlich and Mitscherlich 1987; Lifton 1988).

On the other hand, psychiatric anthropology can, if it has learned to handle the problems of understanding normality appropriately, enrich general anthropology and many specialized branches of it (e.g. philosophical, phenomenological, and cognitive anthropology). The problematic situations particular to psychiatric illnesses can be understood as "nuanced phenomena" to a certain extent. Many different ways of explaining and many different explanatory contexts that have already been formulated in general anthropology can be clarified here, e.g. in the anthropology of perception by Sartre (1943) and Merleau-Ponty (1966) or in the anthropology of time by Theunissen (1992).

Viewed in this way, psychiatric illnesses are not "defects," "injuries" to the state of health, or – as Binding and Hoche (1920) irresponsibly argue according to social Darwinism, an argument fraught with consequences (see Lifton 1988) – in the extreme case "naked existences"; with the exception of profound mental disorders, psychiatric illnesses are often simply normal psychological variations, out of which an extremely rich mental life can arise. Especially in the connection between creativity and psychosis, there are indications that distinctive psychiatric features can represent an integral part of a person's creative mental essence (see Jaspers 1922).

Blankenburg (1971) has presented a survey of the question of "normality" in psychology, according to which Kurt Schneider (1923) asserted that the "average norm" for psychopathology as the "statistic norm of being" is inadequate. The same has been said by Spaemann and Löw (1981, p. 284): "Health is itself a normative, teleological notion." It cannot mean the suitability of an organism for anything specific, rather "suitability in and of itself," judged by a standard of normality which is neither a statistical nor a utilitarian measure. If 99% of all people had headaches, headaches would still not be "normal." Accordingly, Blankenburg refers to the idea developed by Müller-Suur of the "actual individual norm" as the "individual norm of becoming which is restricted by the collective norm of becoming," an individual norm at which therapy is aimed (cited in Blankenburg 1971).

Philosophical anthropology in psychiatry – and psychiatry in general – cannot, therefore, exclude the problem of "normality" and its criteria, but rather must face the difficult challenge of noting a deviant cognitive and mental event in a manner which acknowledges and does not condemn, includes it in a scientific context, and takes this as a basis to suggest and develop further therapeutic treatment methods in collaboration with the patient.

In this respect, psychiatric anthropology is both a danger zone and an ethical challenge for the development of a nonjudgmental systematization and classification of fundamental anthropological conditions of cognitive psychology.

Thus the essential and very difficult task of providing a meta-theoretical background for the theory and practice of psychiatry falls to philosophical anthropology. Consequently, Sullivan states (1953) that there are only a few psychiatrists with a solid theoretical frame of reference within which to think about life problems, their origins, their reliable manifestations or relatively reliable improvements. What matters to him is the level of all common human existence, which is necessary for an explanation of mental disorders. The difficulty of such a concept, as Sullivan says, has its origin in the fact that psychiatry deals with life (Sullivan 1953), and the fact that this is not purely an academic matter is expressed by Karl Jaspers in the following way: "We must free ourselves from the idea that to philosophize is in itself essentially a matter for professors. It is a concern for humankind, clearly, under all conditions and circumstances" (Jaspers 1978, p. 112).

However, psychiatry as "a concern for humankind," in comparison to other medical disciplines, is more indebted to a philosophy such as this, because philosophical reflection frequently shows a close relationship with psychologically exceptional circumstances.

Philosophizing appears naturally among children and among the mentally ill. It is at times (seldom) as if the ties of universal disguises come off and a touching truth is spoken. Deeply distressing metaphysical revelations occur at the beginning of many mental illnesses. In their form and language, their quality is admittedly not without exception such that their expression gains objective significance, outside of cases such as the poet Hölderlin or the painter van Gogh. But anyone present cannot deny the impression that a covering is tearing open here, under which we usually lead our life. Many healthy people also experience uncannily deep meanings upon awakening from sleep, which are then lost again when fully awake (Jaspers 1978, p. 35f).

An understanding of the naturalness of human philosophical reflection, the breaking through of fundamental questions of human self-interpretation in psychiatry, is only possible in the context of our knowledge concerning the exceptional features of human selfhood, as developed by philosophical anthropology (von Weizsäcker 1982; Russell 1959).

2 History of Philosophical Anthropology

The whole 2500-year history of philosophy is like a single, large moment of growing human self-awareness.
JASPERS (1978, p. 112)

The account of the "moment" of self-discovery of the human mind characterized by Jaspers (see Snell 1986) would require an entire reference book of the history of philosophy (see, e.g. Windelband 1976; Krings 1973). Only a few key ideas can be set forth here to serve as guides.

Jaspers refers to philosophy as that "which concentrates, through which the person becomes himself, and in which he partakes of reality" (1978, p. 37). Spaemann and Löw (1981, p. 23) describe philosophy as "human self-understanding within the totality of reality." The history of philosophy – and with it the development of philosophical anthropology – describes the considerable transformations that human self-understanding has undergone in the past approximately 2500 years. Aristoteles' work *De anima* (384–322 B.C.; Aristoteles 1995) presents an important starting-point in this respect in thinking about the human psyche, as mental life is seen here as so lively that it is an irreducible phenomena which cannot be reduced to yet another basic category of knowledge.

Modern human self-interpretation, oriented to the development of consciousness, is related mainly to the *Meditations* of Descartes (1596–1650). Here Descartes (1641), in the pressing question of the certainty of reality, arrived at the irreducibility of the *cogito*, which was able to put an end to the infinite regress of doubt about possible fictional realities in that moment of coming to the inability to question doubt itself. Thus *cogito*, self-consciousness, is at the crux of knowledge in modern "subjectivistic" philosophical reflection and is certainly associated with the constitutive difficulty, reaching into the present, of the split between *cogitatio* and *extensio*, between consciousness and complex, organic unity, between soul and body, between mental and physical. For psychiatry, in particular, this dualism, which has governed thinking since Descartes in the form of the mind-body problem, represents a central challenge which is difficult to overcome (see Metzinger 1995).

Thereafter, a new philosophical foundation emerged with the method of "transcendental apperception" developed by Kant (1724–1804) in his *Critique of Pure Reason.* Here, the a priori conditions of the possibility of experience are questioned and a formal categorical analysis of the mental sphere is carried out, a structure that is decisive for the entire subsequent formation of theory about subjectivity and consciousness. Hersch (1989, p. 190) writes as follows: "Through Kant, the philosophical climate has decisively changed, not only in Germany, but quite generally. One can no longer reflect philosophically after Kant as one did before him." German idealism thus represents one of the possible reactions to Kantian philosophy.

With his *Phenomenology of Mind* in 1807 and his *Science of Logic* in 1912 (Hegel 1986), Hegel attempted in a highly structured way to reconcile intellectually all elements of existence with one another, namely, in the sense of a systematic structure in which all contradictory moments of reality appear as dialectical mediations of the universal mind; in contrast, the philosophical psychologist Franz Brentano proposed a rather modestly argued conception of the difference between mental and physical, albeit one which was fundamental for the further development of philosophical anthropology, namely the idea of "intentionality." In his work *Psychology from the Empirical Point of View,* published in 1874, he stated:

Each mental phenomena is characterized by what the scholars of the Middle Ages called the intentional (also probably mental) inner existence of an object, and which we, although without completely unambiguous

Authors' note: the original German word *Inexistenz* means *innere Existenz,* Engl.: "inner existence."

expressions, would call reference to content, direction toward an object (where no reality is to be understood) or immanent objectivity. Each (mental phenomenon) includes an object intentionally within itself, although not everything in a similar manner (Brentano 1874, p. 124f.).

Jaspers describes intentionality in the following way: "We are conscious of objects, intending them, directed toward them. It is a unique opposition, not comparable to any relationship between objects" (Jaspers 1978).

The Danish scholar, theologian, and philosopher Søren Kierkegaard (1813–1855), both referring to Hegel's systems construction of the absolute and very strongly rejecting it, developed a concept of the subject, decisive for the later philosophical development of existentialism, as a "synthesis of soul and body, carried by spirit" (Kierkegaard 1843). As Rohde (1992) emphasizes in his biography of the philosopher, Kierkegaard actively tried to oppose the disappearance of subjectivity in the perspectives of world history in the dominant Hegelian philosophy. His pioneering conception of subjective freedom is described in 1844 in his work *The Concept of Dread*. The fundamental ideas to understand here are "existence" and "moment" (see Hersch 1989), whereby the subject works his or her way through to the moment of the "reality of freedom," namely to the condition of "dread." Dread is not really to be understood as an emotional state, but rather as an existential condition, a form of existence, as an increased form of existence of the subject who is autonomous ("choosing the self"), in the paradoxical situation of the "leap" (see Sect. 3.3).

In addition to Franz Brentano, Kierkegaard can also be described as one of the founders of modern (existential) philosophical psychology. Whereas Kierkegaard mainly influenced the French school (Sartre, Merleau-Ponty, Camus), Bretano had an enormous effect on his student Husserl (1859–1938), who founded the school of "phenomenology," which is essential to the history of psychiatry (Husserl 1950).

"Phenomenological reduction" can be understood as attention to the "nature of the condition of things" and thus "things as phenomena": "the world is excluded by the method of reduction. The eye of the philosopher is directed back to consciousness, in which every sense develops (Hersch 1989, p. 190). The consequence of this, which the phenomenological analysis of the act of consciousness tried to describe, is how the "corresponding intentional object (that is, the consciousness correlate) is found" (Hersch 1989, p. 298).

An additional concept that is essential in philosophical psychology comes from the philosophy of language of Ludwig Wittgenstein (1889–1951), at the center of which is the philosophy of the "language game" (Wittgenstein 1984). The psychological and depth psychology aspects of the late philosophy of Wittgenstein were developed by Fischer (1991).

The significance of the philosophy of language for philosophical anthropology becomes clear if we acknowledge, as Jaspers did, the "universal character of speech," although speech appears as the "work of individuals": "Speech is one work among other human works, but of a unique character, because it is universal" (Jaspers 1978, p. 299). This means that the coming-to-one's-self of the person, described above as "philosophy," is only possible on the basis of this self-creating work of speech:

> It is the expression, the tool, the foundation of community, an independent work (even if only incidentally pursued in language art). These can all be called a language, but language itself is none of them in particular. In its essence it is not a peculiarity. Language is rather a work of the whole person. But if it is omnipresent as this work, where the person is effective and conscious of him- or herself, it is at the same time still vanishing, because it cannot itself be an appropriate object of attention. In its universality, it has not taken on a determined content as language, but everywhere the person is directed in and through language to the other. Where language itself becomes an object, it may be used superstitiously for supposedly magical effect, or used aesthetically as specific art material, or positivistically to be analyzed as one human product among others in words and grammar, in the study of motor and sensory functions of speech (Jaspers 1978, p. 300).

Very important contributions to philosophical anthropology were made by Arnold Gehlen (1904–1976), taking particular account of the components of action. For Gehlen (1966), a person – following the formulation of Friedrich Nietzsche – is a "non-determined animal," whereby Gehlen interprets human nature as "acting" insofar as it creates its world, precisely by means of the characteristic of openness to the world. This is possible for Gehlen on the basis of the problematic assumption for evolutionary biology of human beings as "defective" creatures (see Wetz 1994).

A fundamental concept of human self-interpretation for philosophical anthropology is also the work of Helmuth Plessner (1892–1985), who developed a graduated order of living things in his central concept of "positionality." In contrast to animals, which show a "centered positionality," humans live "positionally eccentrically" (Plessner 1970). As Wetz (1994, p. 249) emphasizes, "the individual lives as the center, insofar as he knows himself as the center. The animal rests in

itself, without knowing about itself. It lives and experiences, but it is not aware of its experiencing." The subjective competence of the individual for self-relating, accompanied by human beings' eccentric positionality, plays such a decisive role in psychiatry because disorders of this competence can lead to many and diverse psychiatric illnesses, such as persecution mania, sensitive psychiatric delusion of reference, and affective disorders (see Emrich 1992).

Gadamer (1972, p. 25) emphasizes the cultural openness in Plessner's conception:

From the eccentric conception of human liveliness then arises the differentiated ways in which he works out his eccentricity and which we call culture. ... Plessner summarizes all this in the expression that the person 'is embodied.' Here that other spring of human knowledge has its source and pours itself out in advance of natural science, and it has given and shaped for the naturalist the manifold contributions to knowledge about the person as a topic. For, thanks to this knowledge of the person by himself, 'science,' which seeks to understand everything accessible to it by methodological means, is confronted in a particular way with the topic, the human person. Its task of discovery is constantly brought home as one which is open into infinity.

For present-day conceptions of philosophical anthropology, French existentialism, as reflected in the work of Jean-Paul Sartre (1905–1980) and particularly in his major philosophical work *Being and Nothingness*, is especially meaningful. An immediate connection to Kierkegaard is unmistakable, especially with regard to the concept of freedom, as developed by Kierkegaard in *The Concept of Dread*, as well as the idea of "choosing the self," which is taken as the general theme by Kierkegaard in the work *Either – Or*. As Bertrand Russell (1959) emphasizes, in Sartre the existentialist view of human freedom is driven to the extreme; we are constantly choosing our own fate, and every new decision brings with it total obligation.

An additional aspect in Sartre's *Being and Nothingness*, important for psychiatric anthropology, is the phenomenological analysis of the "glance." As Sartre shows, the process of "looking at," which belongs to "my distances," means first objectifying the other. However, it is then shown in phenomenological analysis that indications of effectiveness arise with regard to Plessner's concept of eccentric positionality, insofar as the world view of the observer is "decentered" by the observed person being a subject: "The appearance of the other in the world thus corresponds to a motionless slipping of the entire microcosmos, a decentering of the world" (Sartre 1943). What is more, the observed person can, by the spontaneity of looking back, move the observer into being "shamed": "But shame is ... shame about oneself; it is the recognition of the fact that I am really this object which the other sees and judges. I can only be ashamed of my freedom, insofar as it escapes me and becomes a given object" (Sartre 1943).

A strengthening and radicalization of this analysis by Sartre of the interpersonal relationship of the mutual exchange of glances ensues in the philosophy of Emmanuel Lévinas (1905–1995), in which interpersonal relationships are represented as the philosophy of capturing the face. According to Lévinas' analysis in his major work *Totality and Infinity*, "looking at" means an act of "totalizing" in the sense of an "ontology of war." It is – in contrast to Sartre's position – only able to be transcended by a radical ontological change in the view of reality, namely by the appearance of the conception of "infinity." The dimension deduced with "face," with "the otherness of the other," is ultimately "language" for Lévinas, connected to truthfulness: "The epiphany of the face is itself a word of honor in a certain way. All language as exchange of verbal signs already relates to this original word of honor" (Lévinas 1987, p. 291). This view of responsibility in the interpersonal encounter plays a decisive role in psychiatry in for understanding the empathic aspect of the doctor–patient relationship (see Theunissen 1965, 1992, 1993).

3 General Cognitive Anthropology

3.1 Intentionality, Understanding, and Explanation

A basic problem in philosophy is the question of the difference between two possible questions: first, that of "why?," of the principle of causality, i.e. the question of the relationship between cause and effect; and second, the question "for what?," directed at the principle of purposefulness, of orientation toward purpose, of teleology. Natural science research has restricted itself most generally to developing causal mechanistic principles to explain natural processes, and this has resulted in decisive progress. According to Spaemann and Löw (1981, p. 284):

Thus medicine, for example, owes its enormous progress in modern times above all to the mechanistic examination of nature, to the discovery of causal connections. The question remains, however, whether the radical implementation of the reification of the human body does not in the end cause health once again to suffer.

Precisely in those scientific areas which have humanistic and therapeutic implications, teleological methods of explanation cannot be excluded.

Two explanatory models are given for the basic goal of asking the "why" question, as Spaemann and Löw (1981) emphasize. Both are ultimately aimed at increasing the degree of familiarity with outer reality which has been reduced by unexpected events (misunderstandings). One type of explanation is the mode of "understanding," and the other, the mode of "explanation." Karl Jaspers, in his *General Psychopathology* (Jaspers 1913), described understanding and explanation as the two fundamental ways of accessing the psychological course of events, standing opposite one another, one being the hermeneutic, empathic way and the other the scientific, causal-mechanistic process of explanation. It is then emphasized that both these methods have to be used in psychiatry to achieve a balanced relationship. According to Spaemann and Löw (1981), the "process of understanding" describes the "reestablishment of intimacy through the reconstruction of an intentional structure," i.e. the "intentionality structure" of the other, the acting person, is comprehended psychologically. The "process of explanation" on the other hand is interpreted as the "reestablishment of intimacy by specification of a conformity to natural law" (Spaemann and Löw, pp. 17, 19).

Thus the idea of intentionality has played a decisive role in the development of philosophical psychology since the end of the nineteenth century up to the present. Mimetic processes of psychological interaction, as they will be described in detail below, are based on the reconstruction of the psychological intentionality of the other and are connected with it through the mode of "empathic understanding."

One of the main difficulties for present-day cognitive anthropology is the question of the connection between understanding and explanation. The psychoanalytic project of Sigmund Freud aimed to reduce psychological understanding and causal explanation to a common denominator, i.e. psychoanalysis should simultaneously understand and explain. It should be both hermeneutics and neurophysiology. However, Wittgenstein had already objected that Freud confuses cause and motive. His reasoning goes along the same line as the strict separation between understanding and explanation made by Jaspers, which can be formulated in line with Fischer (1991, p. 114) in Wittgenstein's sense as follows: "Physiological or neurological causes are completely irrelevant to the question of the meaning of an action." In agreement with this, we can see that current basic research in neuroscience has still not been able to develop a convincing neurophysiological model for the origin of semantics. Only the elucidation of a "meaning generating code" (see Emrich 1990) would also be able to lead to the standardization of understanding and explanation in the realm of the psychological, because the theory of psychosis (see below) means that, at present, a dual way of consideration is still necessary, in which understanding and explanation are used in parallel. From the concept of intentionality, it can be said, as Wulff (1993) has done, that a psychotic event represents a "form of intentionality frustrating itself." This means that the establishment of meaningful performance is disturbed in psychosis, a fact which, according to Fischer (1991), can be represented in a psychotic event as "altered grammar."

3.2 Constructivity and Consciousness

As the twentieth century draws to an end, the question of human existence has become essentially that of human *consciousness*. Thus, as explained in the introduction to the anthology *Theory of Subjectivity* by Horstmann et al. (1987, p. 9):

> The theme of subjectivity in the philosophy of this century is one of the areas of greatest differences of opinion and most severe conflict. While individual thinkers in the course of modern tradition still tried to defend subjectivity as the guarantor of absolute transparency, freedom, and foundation of all possible discovery, other philosophers proposed the complete renunciation of the acceptance of any entity with this name. The simplified label 'philosophy of consciousness' represents what remains of the history of this dispute.

Cognitive anthropology is at the center of the question about the individual, and the contributions it makes to our understanding of psychiatric illness can be verified. The characteristic of "constructivity" is generally attributed to consciousness-generating systems (see Roth 1992). According to Roth, the internal reality generated through the constructivity of the central nervous system (CNS) is judged by processes of assessment and, of course, on the basis of previous experiences. This internal process of assessment seems to be the source of consciousness: "Sensory perceptions only become conscious if they have been assessed in the light of earlier experiences" (Roth 1994, p. 310).

The conception giving rise to neurobiological constructivism here is part of that realm of a collective notion now common which is specially designed to emphasize a perspective connected with two alternative models. Either the human subject is seen as more of a basin, i.e. as a receptive system for processing information contained in signals entering from out-

side, or the subject is interpreted more strongly in the "constructivist" sense as a source, precisely in the way that a "world design" is generated, which can be corrected or modified in accordance with all the incoming data. Thus we can differentiate between a more psychological, a physiological, and a philosophical form of constructivism.

Psychological constructivism, e.g. that of Watzlawick, turns out to be particularly the projective character of subjective reality. For Watzlawick (1985), the decisive point is the "reality-creating fiction," also called "as-if fiction" following the *Philosophy of the As-If* by Hans Vaihinger (1927). Philosophical constructivism is traced back by von Glasersfeld (1985) to the Italian Renaissance philosopher Giambattista Vico, who writes in his work *De antiquissima Italorum sapientia* (1710) as follows: "If the senses are 'active' abilities, then it follows that we make colors by seeing, tastes by tasting, sounds by hearing, cold and hot by touching" (Vico 1710; see also von Glasersfeld 1985).

For physiological constructivism, perception is a "confirmation of previously dreamed reality," whereby this hypothesis of prestructured reality is once again based each time on previous experimental materials (Singer 1994; Roth 1992).

As Kurthen (1990, p. 23) emphasizes, what is called 'consciousness' is "nothing uniform, but rather a complex of heterogeneous phenomena such as awareness, carefulness, self-reflection, cognitive capability, and other things." From the point of view of evolutionary biology, the question of the origin of this complex event "consciousness" is interpreted as a reconstructible process of development, explainable by paleoanthropological and cultural-anthropological studies.

According to Donald (1991), the basis of the development, as Wolfgang Köhler's chimpanzee experiments show, is episodic memory, with the help of which these animals also show problem-solving behavior, e.g. using boxes to reach bananas. The next evolutionary step assumed by Donald is the transition from "episodic to mimetic culture." By this he understands an archaic, but already human culture in real ways in the sense of the transition from apes to humans. Donald applies this to *Homo erectus*, a species that appeared about 1.5 million years ago and that was already able to make a wide variety of complicated tools and to move far away from Africa, its place of origin. *Homo erectus* was able to use fire, cook food, and maintain complex, stable societies, within which cooperation was highly significant, especially in communal hunting, teaching the use of tools, etc. It is accepted that a cognitive ability had developed in *Homo erectus* that could not yet really be described as "language." It is hypothesized that "mimesis" represents a cognitive ability which has led from episodic memory to higher forms of communication and representation of cognitive meanings.

Mimesis makes it possible to use symbolic gestures in social communication or artistic expression. Mimesis is still used today, especially to transmit social and practical abilities (see Webb 1995).

The third stage, which first appeared in *Homo sapiens*, is the mythical stage, and the fourth, the theoretical stage. In the mythical stage, language is developed and used to build up narrative structures (with the goal of emphasizing topics taken from individual episodes and conveying their historical connections, topics which are good for interpreting human life). Donald assumes by this that these myths play an important role in the formation and stabilization of group identity (see Webb 1995).

In contrast to this, the modern, "theoretical" structure of consciousness is bound up with external symbolic representation. Writing, painting, the use of graphics, etc. are part of this. Voegelin (1988) speaks in this sense of a noetic differentiation of consciousness.

3.3 Anthropology of Cognition–Emotion Coupling

A basic experience of human life consists in the fact that humans not only have perceptions and thoughts, but that they also have feelings that are reducible neither to thoughts nor to perceptions. What is noteworthy in the world of emotions is that it is beyond the conceptual, i.e. it can only be described incompletely by thinking. Insofar as quite important emotional states are conceptually transcendent phenomena for psychiatry, Wittgenstein (1984, p. 107) analyzed pain and regarded it not as an "incident in the world," but described it as a subjective quantity in the sense of "there are toothaches if one has them" ("the question 'Are you sure that *you* are the one who has pain?' would be nonsense"). Thus emotional states belong to the subjective universal realm.

With computer simulations of cognitive processes and perceptive capabilities, much has been learned in recent years about the way in which information processing functions in the brain. In particular, it has been shown that information is not really generated and processed in the brain, but rather significance is, i.e. biological information is highly dependent on context and, finally, it is a matter of internal assessments. However, these assessments are significantly released in the form of emotions. Thus, for example, the cognition researcher and neurobiologist Edelman (1987) showed in robotics experiments that a connectionistically built sensorimotor robot, which he named "Darwin II," was only functional if an "internal value-world" was incorporated into it, i.e. internal context-related

assessment quantities that had a quasi emotional, simulating character. The problem of perception–emotion coupling is thus a central question of cognitive anthropology.

To explain transmissions between perception and emotion, McLean's model of the "triune brain," the three-part brain, is generally referred to; this model assumes that, as a consequence of the mammalian revolution, three structures in the CNS have settled down more or less on top of each another: the first, the "reptilian brain," is sited in the brain stem and is responsible for vegetative functions; the second, the limbic system, is situated above the reptilian brain and is related to the primeval mammalian brain; and the third are the neocortical structures (McLean 1982). The connections between the cortical association structures and the limbic system are thus fundamental to the question of the nature of perception–emotion coupling. While McLean still emphasized the hermetic character of the boundaries between the systems, greater value is certainly now placed on research into the transmissions between the cortical and limbic structures (see, e.g. Roth 1992, 1994).

The discovery that the olfactory brain, the rhinencephalon, probably played a role in leading the way in cortical evolution in higher mammals is especially meaningful from the biological evolutionary perspective. Here, there seems to be a direct relationship between olfactory brain functions and the "reward system" in the basal forebrain, which is decisively significant in addictive behavior, for instance, as experiments with microinjections and electrode self-stimulation have shown. In addition, it performs important functions in the regulation of food supply.

There are two other connections between cortical structures and the limbic system that neurophysiology has been able to characterize well in recent years. The first, the so-called hippocampal comparator structures, facilitate important context-dependent internal assessments. The second, which is connected to these structures, is the nucleus amygdala, which is sited in the temporal lobe and whose significance in the regulation of affect has only become clear in the last few years.

According to neuropsychological investigations, which have been pushed ahead by the English psychologists Gray and Rawlins (1986), in particular, an important function of hippocampal structures is generated with the particular situation of a compatible internal "model which runs along with the world" and particular comparator functions carried out between "expected reality" and the incoming sensory data. This comparator function leads to the generation of an internal "alarm signal" when a critical threshold is exceeded, which is transmitted to the limbic system as an anxiety-triggering signal. On the basis of this idea, which is well supported by animal experiments and neuropsychology, it is assumed that anxiety-relieving medications raise the threshold at which the internal alarm signal appears or reduce the alarm signal amplitude, so that context-dependent anxiety stimuli are less likely to appear.

From this concept, a self-consistent neuropsychological and biochemical theory of anxiety not only can be developed, but the high number of γ-aminobutyric acid (GABA)-benzodiazepine receptor complexes in the hippocampus is plausible. In agreement with this theory, it can be shown that patients with benzodiazepine dependency show high anxiety and emotion scores in assessment scales by self and others after withdrawal and an overintensity of perceptions, e.g. oversensitivity to light and noise and hyperarousal (see Apelt and Emrich 1990).

These hippocampal comparator functions are closely related to the function of the nucleus amygdala lying in the temporal cortex. In particular, the American author Aggleton (1992) was able to show that an important function of the nucleus amygdala is that it exercises a "gate function" between the cortical association structures and the various perception systems. Disorders of this function can bring about abrupt emotional changes and even a crazed rampage, etc.

The neuropsychological idea of cognition–emotion coupling, represented in this way, can also be compared to a similar concept based on the anthropology of language. Language, understood in a universal sense not as something lexical and grammatical, but as a general expressive ability, is the area in which cognition–emotion coupling takes place mentally. A possible interpretation of this arises on the basis of the "mimetic" coupling of emotion developed by Girard (1992), within which the "language" of the mimetic processes of interaction within socioculturally conditioned hierarchies of values brings out the internal fine-tuning assignment of emotions to cognitive achievements. Since this fine-tuning process is of considerable importance for the appearance or non-appearance of psychiatric illnesses, these events play a central role in psychiatric anthropology.

4 Evolutionary Anthropology

The way in which human beings are seen in German intellectual history is greatly influenced by the concept of the defective human being. Plessner (1970), following Herder (1772), Gehlen (1966), and other authors, bases his view of *Homo sapiens* on old models as a being who is deficient not only morally, but also

biologically. Bischof (1985, p. 513) suspects that philosophy and the social sciences are trying, by means of the myth of the defective being, "to escape the biologically inherited burden, in which one mourns its alleged loss." The Persian physician Avicenna, originally Ibn Sina, one of the great Arabic-speaking interpreters of classical Greek medicine, characterizes human beings as biologically deficient beings; hence physicians constantly have to take countermeasures (Schipperges 1990). In those cultures influenced by Christianity, the concept of "original sin," the fundamental moral inferiority of Adam, Eve, and their descendants, corresponds to that of physical inferiority. Similar themes are found in the traditions of other religions.

Especially in modern medicine, the concept of the deficient being continues to have an effect. Clinical textbooks are necessarily concerned with disorders of the human organism, which is obviously inclined to malfunction. It is also beginning to become increasingly clear which individual gene defects have which pathological consequences, and where precisely described disorders have somatic or psychological effects. Indeed, the wonder of organic life, developed by evolution and maintained in a fine balance of functioning by an incredible diversity of extremely complex biocybernetic control processes, is not usually part of medicine's field of vision.

Further development, above all based on the evolutionary biology of Darwin (1859, 1872, 1871), has yielded advances in recent decades, especially in the precise formulation of theoretical positions and development of hypotheses and models. Their inherent predictive power makes predictions about even human behavior possible or at least can make other good suggestions for empirical investigations.

Alexander (1974) and some other authors (see vol. 8, special edition of *Ethology and Sociobiology*, 1987; also Ploog 1993; Voland and Voland 1989; Nesse 1994; Schelde 1994; Schelde and Hertz 1994; Marks and Nesse 1994) have drawn attention to a common point of intersection between psychology, psychiatry, and evolution, i.e. there are biologically related perspectives where the observation and evaluation of basic conflicts and disorders in the phylogenetic or ontogenetic dimensions of relational development are concerned. The current debate about the observable and real basis of adaptation is also significant for psychiatry.

A mark of the evolutionary and biological interpretation of the phenomenon of life is the search for functionality, for adaptive ability brought about (in the classical manner) through mutation and selection. That the past explains the present is a central statement of evolutionary biology, which, interestingly enough, like history or archeology, is a discipline in which events that occurred in the distant past have explanatory power.

In recent years, the term "environment of evolutionary adaptedness" (EEA) has gained acceptance for this way of observing. This refers to a scenario established in the Pleistocene age, during which, 1.5 million years ago, the first representative of the genus *Homo* arose, whose descendants were our Stone-Age ancestors. The fact that we exist proves that they not only survived the climatic, ecological, social, psychological, and cognitive kinds of challenges of the time, but also had sufficient descendants. It is also quite likely that this "evolutionary adaptation to the environment" in the form of abilities to adapt, which are typical for our species, continues in us. This is something that is generally accepted for the morphological, anatomical features of our phenotype, even by critics of human evolutionary biology. These widely uncontentious areas include the biped with its hand permanently freed (measured by subhuman primates) from the requirements of locomotion, the (possibly associated) further development of the neocortex, and the complete picture of *Homo sapiens* and the externally rather similar characteristics of human beings.

The argument reaching far back into Western history about "nature versus nurture," about genetically conditioned structure versus adaptation to the environment achieved by learning, is explosive where psychosocial and/or cognitive domains are concerned, i.e. the area which was traditionally investigated and interpreted in the humanistic approach. Following advances in knowledge in general evolutionary biology and in human ethology (Eibl-Eibesfeldt 1984), human behavioral biology (Hassenstein 1973), and evolutionary anthropology (Durham 1992) and psychology (Tooby and Cosmides 1989, 1990), in particular, it can be assumed that this kind of research will lead to future discoveries of additional ways of perceiving, thinking, and behaving based on primarily biological mechanisms. This in no diminishes the humanistic approach to understanding the expression of human life. On the contrary, the more knowledge we obtain based on biology and evolutionary biology, the clearer the puzzles of our existence will become, puzzles that are unsolvable or not yet solvable (Chalmers 1996; on the synthesis of evolutionary biology and the social and cultural sciences, see Schiefenhövel et al. 1994).

In addition to wide agreement in the fundamental accuracy of the Darwinian hypothesis, contradictory positions are also taken up. For example, at present the real basis of adaptive ways of human perception and behavior is being discussed. Tooby and Cosmides (1989, 1990) and other authors assume that a very large number of specific, genetically mediated programs are effective, which can be described using a Darwinian type of psychology and must have been found as

universal characteristics in all (or at least most) members of our species, independent of geography and culture.

Other evolutionary biologists (see MacDonald 1991) assume that a considerably limited set of genetically mediated perceptual and behavioral tendencies suffice in order for humans to behave "adaptively" in particular situations. MacDonald postulates that, instead of delicate domain-specific psychological mechanisms for coping with the many and diverse daily problems, "evolved motive dispositions," which emerge evolutionarily as motivational dispositions, guide our perception and behavior in more general ways. He and other authors also point out that the evolution of *Homo sapiens* is not over and that, because of the separate development of various populations on earth – which according to Erikson (1966) takes place as a kind of co-evolution between biology and culture, as "cultural pseudospeciation" – genetically explainable differences in behavioral repertory can also occur between populations.

The search for particular adaptations and basic underlying mechanisms inspired by evolutionary biology has also led psychiatry to interesting new perspectives, including the idea that psychiatric illnesses can be interpreted as excessive forms of evolutionarily adaptive biological achievements. Thus philosophical anthropology also gains in significance precisely from the evolutionary biological point of view as a culturally related and intercultural science.

5 References

Aggleton JP (1992) The functional effects of amygdala lesions in humans: a comparison with findings from monkeys. In: Aggleton JP (ed) The amygdala: neurobiological aspects of emotion, memory, and mental dysfunction. Wiley, New York, pp 485–503

Alexander RD (1974) The evolution of social behavior. Annu Rev Ecol System 5: 325–383

Apelt S, Emrich HM (1990) Sodium valproate in benzodiazepine withdrawal. Am J Psychiatry 147: 950–951

Aristoteles (1995) Über die Seele. Meiner, Hamburg

Binding K, Hoche A (1920) Freigabe der Vernichtung unwerten Lebens. Leipzig

*Bischof N (1985) Das Rätsel dipus. Piper, Munich

Blankenburg W (1971) Der Verlust der natürlichen Selbstverständlichkeit. Ein Beitrag zur Psychopathologie symptomarmer Schizophrenien. Enke, Stuttgart

*Brentano F (1874) Psychologie vom empirischen Standpunkt, vol 1. Duncker and Humblot, Leipzig

Chalmers DJ (1996) Das Rätsel des bewußten Erlebens. Spektrum Wissenschaft 2: 40–47

Darwin C (1859) On the origin of species by means of natural selection, or the preservation of favoured races in the struggle of life. Murray, London

Darwin C (1871) The descent of man and selection in relation to sex. Murray, London

Darwin C (1872) The expression of emotions in man and animals. Murray, London

Descartes R (1641) Meditationes de prima philosophia. Soly, Paris

Donald M (1991) Origins of the modern mind: three stages in the evolution of culture and cognition. Harvard University Press, Cambridge/MA

Durham WH (1992) Coevolution: genes, culture and human diversity. Stanford University Press, Stanford

Edelman GM (1987) Neural Darwinism. The theory of neuronal group selection. Basic, New York

Eibl-Eibesfeldt I (1984) Die Biologie des menschlichen Verhaltens. Grundriß der Humanethologie. Piper, Munich

Emrich HM (1990) Psychiatrische Anthropologie: Therapeutische Bedeutung von Phantasiesystemen/Wirklichkeitsrelativismus und Sinnfrage. Pfeiffer, Munich

Emrich HM (1992) Subjective error correction capacity, and the pathogenesis of delusions of reference. In: Spitzer M, Uehlein F, Schwartz MA, Mundt C (eds) Phenomenology, language and schizophrenia. Springer, Berlin Heidelberg New York, pp 372–377

Erikson EH (1966) Ontogeny of ritualization in man. Philosoph Transact R Soc 251: 337–349

Fischer HR (1991) Sprache und Lebensform. Auer, Heidelberg

*Gadamer HG (1972) Theorie, Technik, Praxis – die Aufgabe einer neuen Anthropologie. In: Gadamer HG, Vogler P (eds) Biologische Anthropologie, vol 1. Thieme, Stuttgart

Gehlen A (1966) Der Mensch. Seine Natur und seine Stellung in der Welt. Athenäum, Frankfurt am Main

**Girard R (1992) Das Heilige und die Gewalt. Fischer, Frankfurt am Main

Gray JA, Rawlins JNP (1986) Comparator and buffer memory: an attempt to integrate two models of hippocampal functions. In: Isaacson RL, Pribram KH (eds) The hippocampus. Plenum, New York, pp 151–201

Hassenstein B (1973) Verhaltensbiologie des Kindes. Piper, Munich

Hegel GWF (1807) Phänomenologie des Geistes. Goebhardt, Bamberg

Hegel GWF (1986) Wissenschaft der Logik. Suhrkamp, Frankfurt am Main (1st edn: 1812)

Herder JG (1772) Abhandlungen über den Ursprung der Sprache. Berlin

Hersch J (1989) Das philosophische Staunen. Piper, Munich

Horstmann RP, Cramer K, Fulda HF, Posthast U (eds) (1987) Theorie der Subjektivität. Suhrkamp, Frankfurt am Main

Husserl E (1950) Ideen zu einer reinen Phänomenologie und phänomenologischen Philosophie I. Nijhoff, The Hague (Husserliana, vol III)

Husserl E (1952) Ideen zu einer reinen Phänomenologie und phänomenologischen Philosophie II. Nijhoff, The Hague (Husserliana, vol IV)

*Jaspers K (1913) Allgemeine Psychopathologie. Springer, Berlin

Jaspers K (1922) Strindbergh und van Gogh. Barth, Leipzig

Jaspers K (1978) Was ist Philosophie? Piper, Munich

Kant I (1781) Kritik der reinen Vernunft. Hartknoch, Riga

Kierkegaard S (1843) Entweder – Oder. Reitzel, Copenhagen

Kierkegaard S (1992) Der Begriff Angst. Reclam, Stuttgart (1st edn: 1844)
Krings (1973) Handbuch philosophischer Grundbegriffe. Koesel, Munich
Kurthen M (1990) Das Problem des Bewußtseins in den Kognitionswissenschaften. Enke, Stuttgart
**Lévinas E (1987) Totalität und Unendlichkeit. Versuch über die Exteriorität. Alber, Munich
Lifton RJ (1988) Ärzte im Dritten Reich. Klett-Cotta, Stuttgart
MacDonald K (1991) A perspective on Darwinian psychology: the importance of domain-general mechanisms, plasticity, and individual differences. Ethol Sociobiol 12: 449–480
McLean PD (1982) On the origin and progressive evolution of the triune brain. In: Armstrong E, Falk D (eds) Primate brain evolution. Plenum, New York, pp 291–316
Marks IM, Nesse RM (1994) Fear and fitness: an evolutionary analysis of anxiety disorders. Ethol Sociobiol 15: 247–261
*Merleau-Ponty M (1966) Phänomenologie der Wahrnehmung. de Gruyter, Berlin
Metzinger T (ed) (1995) Bewußtsein – Beiträge aus der Gegenwartsphilosophie. Schöningh, Paderborn
Mitscherlich M, Mitscherlich A (1987) Medizin ohne Menschlichkeit. Fischer, Frankfurt
Nesse R (1994) An evolutionary perspective on substance abuse. Ethol Sociobiol 15: 339–348
*Plessner H (1970) Philosophische Anthropologie. Fischer, Frankfurt
Ploog D (1993) Psychopathologische Prozesse in neuroethologischer Sicht. In: Schüttler R (ed) Organische Psychosyndrome. Springer, Berlin Heidelberg New York, pp 1–28 (Tropon-Symposium, vol 8)
Rohde PP (1992) Kierkegaard. Rowohlt, Hamburg
Roth G (1992) Das konstruktive Gehirn: Neurobiologische Grundlagen von Wahrnehmung und Erkenntnis. In: Schmidt SJ (ed) Kognition und Gesellschaft – Der Diskurs des radikalen Konstruktivismus, vol 2. Suhrkamp, Frankfurt am Main, pp 277–326
Roth G (1994) Gehirn und Geist. In: Schiefenhövel W, Vogel C, Vollmer G, Opolka U (eds) Vom Affen zum Halbgott. Thieme, Stuttgart
Russell B (1959) Wisdom of the West. Rathbone, London
**Sartre JP (1943) L'être et le néant. Gallimard, Paris
Schelde T (1994) Ethological research in psychiatry. Ethol Sociobiol 15: 349–368
Schelde T, Hertz M (1994) Ethology and psychotherapy. Ethol Sociobiol 15: 383–392
Schiefenhövel W (1994) Krankheit, Altern, Tod. In: Schiefenhövel, W, Vogel C, Vollmer G, Opolka U (eds) Der Mensch in seiner Welt. Trias, Stuttgart, pp 217–244
Schipperges H (1990) Die Kranken im Mittelalter. Beck, Munich
*Schneider K (1923) Die psychopathischen Persönlichkeiten. In: Aschaffenburg G (ed) Handbuch der Psychiatrie, Sect 7, part 1. Deuticke, Leipzig, pp 1–96
Singer W (1994) Gehirn und Bewußtsein. Spektrum, Heidelberg
Snell B (1986) Die Entdeckung des Geistes: Studien zur Entstehung des europäischen Denkens bei den Griechen. Vandenhoeck and Ruprecht, Göttingen
Spaemann R, Löw R (1981) Die Frage Wozu? Piper, Munich
Sullivan HS (1953) The interpersonal theory of psychiatry. Norton, New York
**Theunissen M (1965) Der Andere. de Gruyter, Berlin
Theunissen M (1992) Negative Theologie der Zeit. Suhrkamp Frankfurt am Main
Theunissen M (1993) Der Begriff Verzweiflung. Suhrkamp, Frankfurt am Main
Tooby J, Cosmides L (1989) Evolutionary psychology and the generation of culture, part I. Ethol Sociobiol 10: 29–49
Tooby J, Cosmides L (1990) The past explains the present. Emotional adaptations and the structure of ancestral environments. Ethol Sociobiol 11: 375–424
Vaihinger H (1927) Die Philosophie des Als Ob und das Leben. Scientia, Aalen
Vico G (1710) De antiquissima Italorum sapientia. Neapel
Voegelin E (1988) Ordnung - Bewußtsein - Geschichte. Klett-Cotta, Stuttgart
Voland E, Voland R (1989) Evolutionary biology and psychiatry: the case of anorexia nervosa. Ethol Sociobiol 10: 223–240
von Glaserfeld E (1985) Konstruktion der Wirklichkeit. In: Gumin H, Mohler A (eds) Einführung in den Konstruktivismus. Oldenbourg, Munich, pp 1–26 (Schriften der C.F. v. Siemens-Stiftung, vol 10)
von Weizsäcker CF (1982) Der Garten des Menschlichen. Hanser, Munich
Watzlawick P (1985) Die erfundene Wirklichkeit. Piper, Munich
Webb E (1995) Mimesis, evolution, and differentiation of consciousness. Paragrana 4/2: 151–165
Weber MM (1993) Ernst Rüdin – Eine kritische Biographie. Springer, Berlin Heidelberg New York
Wetz J (1994) Der Mensch in der neuzeitlichen Philosophie. In: Schiefenhövel W, Vogel C, Vollmer G, Opolka U (eds) Gemachte und gedachte Welten. Thieme, Stuttgart
Windelband W (1976) Lehrbuch der Geschichte der Philosophie. Mohr, Tübingen
Wittgenstein L (1984) Tractatus logico-philosophicus. Suhrkamp, Frankfurt am Main (1st edn: 1921)
Wulff E (1993) Sich-selbst-durchkreuzende Intentionalität. Wahnsinn als Aushebelungsversuch aus Gesellschaft und Geschichte. Argument 197: 91–104

CHAPTER

21

A. Kraus

Phenomenological–Anthropological Psychiatry

Translator: O.P. Wiggins, Jr.

1 Introduction

A number of overviews have been published in periodicals and as book chapters on phenomenological-anthropological psychiatry (Boss 1951; Kuhn 1963; Lanteri-Laura 1963; Natanson 1963; Straus 1963; Zutt 1963a,b; Edie 1966; Wyss 1976; Blankenburg 1977, 1980a,b, 1991b; Rovaletti 1994; Mooij 1995; Figueroa-Cave 1996a,b, 1997a,b; Kimura 1996; Naudin et al. 1997; Schmidt-Degenhard 1997). Presentations of the field can also be found in several monographs (Spiegelberg 1972; Tatossian 1979; Kruger 1981; De Koning and Jenner 1982; Tellenbach 1987; Herzog and Graumann 1991; Kimura 1997; Dörr-Zegers 1995; Lopes 1996; Fédida and Schotte 1991; Braun 1992-1994; Holzhey-Kunz 1994a; Passie 1995). The sources of these selected overviews confirm the impression that, although the approaches subsumed under the title "phenomenological-anthropological psychiatry" had their origins largely in German-speaking areas, they have now gained a foothold in France and Italy, in Spanish-speaking countries, in Japan, and most recently in the United States. Moreover, the number of works in this field has grown steadily during the course of the century.

In this chapter, various approaches will be considered, ranging from the purely phenomenological, through the phenomenological-anthropological, to the purely anthropological. Daseinsanalytic and existential approaches can be understood as phenomenological, but not always anthropological, in their methods.

2 Anthropological Approaches

Like the concept of "phenomenology," the concept of "anthropology" can have very different meanings. The concept of anthropology, in particular, is extremely broad, and in English it is only seldom used in the sense in which it is used here. In the natural sciences, anthropology is concerned with the natural history of human beings as a species, while in the field of ethnology and cultural anthropology, it relies mainly on sociological and socio-psychological methods.

For anthropological medicine and above all for phenomenological anthropological psychiatry, the philosophical anthropology of M. Scheler (1976), H. Plessner (1928), A. Gehlen (1966), F.J.J. Buytendijk (1967), the life-philosophy of H. Bergson, W. Dilthey, L. Klages, and O.F. Bollnow, and the phenomenology of E. Husserl (1950), the fundamental ontology of M. Heidegger (1963), and the existential philosophies of J.-P. Sartre (1943), M. Merleau-Ponty (1966), E. Levinas, and P. Ricoeur were of primary significance. Until quite recently, the phenomenological-anthropological orientation in psychiatry was most strongly inspired by M. Heidegger (1963) and E. Husserl (1950) and by the French phenomenologists J.-P. Sartre (1943) and M. Merleau-Ponty (1966). It was established primarily by L. Binswanger (1955a,b, 1958, 1960, 1965, 1992, 1994), E. Straus (1963, 1978a,b), and V. von Gebsattel (1954). Other chief representatives of the older generation include J. Zutt (1963a,b, 1958), R. Kuhn (1963), von Baeyer (1978), W. Blankenburg (1969, 1970, 1971, 1972, 1977, 1979a,b, 1980a,b, 1981, 1982, 1983, 1987, 1989, 1980, 1991a,b, 1995, 1997), H. Tellenbach (1968, 1983), Bräutigam (1961), M. Boss (1951), G. Condrau (1963, 1990, 1992), D. Wyss (1973, 1976, 1990), Callieri and Castellani (1981), B. Kimura (1975, 1980, 1982, 1995, 1996, 1997), L. Lanteri-Laura (1963, 1982), H. Kunz (1962), von Uslar (1989a,b, 1991), and Tatossian (1979, 1996).

Anthropological approaches were introduced in internal medicine by V. von Weizsäcker (1986) and others and are to be found in the existential analysis of V.E. Frankl (1959, 1984), in the psychoanalysis of P. Schilder (1968) and J. Lacan, and in the humanistic psychology and therapeutic methods of Rogers (1973). Influences of phenomenological-anthropological psychiatry are also recognizable in the structural-dynamic approach of Janzarik (1959, 1988) and in the interactionist psychopathology of J. Glatzel (1981a,b, 1998). Heimann (1994) pleads for a neurobiological anthropology.

Binswanger called his (at that time) new methodological approach "phenomenological anthropology." However, he also understood his later "daseinsanalysis" as anthropological and thus sought to distinguish it as a form of existential anthropology from Heidegger's daseinsanalytic approach with its philosophical intention as a fundamental ontology.

In attempting to point out the most important goals of an anthropological approach in psychiatry, it must be emphasized that, with regard to method, this should be understood as almost exclusively phenomenological. The great interest in present ethical problems not only in psychiatry but also in medicine in general entails the danger that the anthropological approach will be understood primarily as simply an ethical entreaty or appeal to deal with patients in a humane way, sympathizing with their affliction. That, however, would not require us to move along a new path of psychopathological knowledge. The humanization of psychiatry is certainly the ultimate goal of all anthropological endeavors in psychiatry, but this ought to occur by way of a methodology that is appropriate for the essence of human beings and a corresponding

understanding of psychiatric disorders. Because psychiatry cannot disregard the human being as a whole, i.e., in his or her individuality, subjectivity, freedom, and historicalness, it poses the question of how psychiatry as a science of mentally ill human beings is possible. In this way, the anthropological approach is connected with the general problem of securing a high-quality scientific method that is adequate to deal with its subject matter.

Phenomenological–anthropological psychiatry makes the essence of human existence the locus of interpretation and that which is human in us the organon of experience (Blankenburg 1979b). However, even then, we are not concerned with human existence in its entirety in the sense of defining all meaningful content. According to Jaspers (1965), such an exhaustive definition cannot be achieved, nor can it be an object of scientific psychopathology. Instead, we are concerned with the significance of formal determinations of human existence for an understanding of the essence of psychiatric disorders. Only in this way can we prevent an anthropological approach from deteriorating into a *Weltanschauung*.

The issue of an appropriate methodology for psychiatric patients and their disorders continually creates a methodological openness that allows us to achieve two goals: it allows us to steer clear of a one-sided anthropological or ontological reductionism, and it makes it possible for us to recontextualize the objects of the individual objectivating sciences, which out of methodological necessity have to reduce their objects. It is also possible in this way to provide new impetus for research in the field of phenomenological–anthropological psychiatry and to initiate new therapeutic approaches.

Since this approach can allocate the objectivating sciences their epistemic places in psychiatry and can thus correlate them with one another, it is the indispensable field of basic principles in psychiatry, as Kisker (1964), Blankenburg (1991b), and Tatossian (1996) have pointed out. Moreover, outside the realm of psychiatric illnesses, it is hardly possible to introduce existential border experiences into our understanding of the essence of human beings. By doing this, the approach also has important repercussions on philosophical anthropology, a field from which it also draws inspiration.

At the center of all anthropological approaches in medicine stands the concept of person and hence of his or her world. The person is understood as the subject in the relation of the human being to the world and as the irreducible center of intentional acts. Hegel's statement that "individuality is that which is its world as its own" (Hegel 1952) thus serves as a leitmotiv in the research by Binswanger and in other phenomenological–anthropological approaches. If we wished to specify the central themes of more recent phenomenological–anthropological approaches, we would have to mention above all that of the self or identity, role (Blankenburg and Haltenhoff 1994; Kraus 1977, 1980a, 1982a,b, 1996c,d, 1997b), and the body (Blankenburg and Haltenhoff 1994; Blankenburg 1983, 1995; Fuchs 1998; Northoff et al. 1992; Schmoll and Koch 1989; Schmitz 1987; Schmoll 1992) ("embodiment") as well as time and space (see below) in their significance for various psychiatric illnesses and for various therapeutic approaches.

Phenomenological–anthropological approaches, in particular, including that of an existential–anthropological orientation, focus on the attempt to understand and describe schizophrenic, melancholic, phobic, anankastic, manic, and other patients on the basis of a priori structures of human existence. Here, two distinct directions can be seen. The first attempts to understand psychopathological phenomena by comparing them with normal psychological phenomena, e.g., by comparing delusion with the imagination of dreams and oneiric phenomena with creative imagination (Schmidt-Degenhard 1986, 1991, 1994, 1995) or with archaic or paleological encounters with reality. Both neurotic and psychotic signs are thus usually conceived merely as intensifications of normal forms of experience.

The second direction taken by phenomenological–anthropological research was developed by Binswanger, in particular. He conceived forms of delusion, schizophrenia, mania, depression, neurosis, psychopathy, etc., as possible modifications (grounded in the essence of human world- and self-relationships) or as "factual" (qualitative) modifications of the "a priori uncovered" structure of being-in-the-world or of *Dasein*. This procedure disregards for the time being the fact that what is at issue here are forms of mental or psychic illness.

Although here the distinction between health and illness is undercut, these analyses, as I will try to point out later, are not without significance for psychiatric classification, because this second direction shows how patients with various disorders differ from healthy individuals and from one another. In other words, on the basis of an understanding of the essence of "psychiatric disorders" such as this, the issue of psychopathological entities can be addressed anew (see below).

Binswanger's approach (1955b) has proved to be especially fruitful by delineating the anthropological proportions of a height and breadth dimension in human existence. Other authors (e.g., Blankenburg 1972, 1980a,b; Kraus 1977), following his example, have expanded this to include additional proportions such as self-realization and world-realization, individuation and community-relationship, tolerance of

ambiguity and intolerance of ambiguity, over-identification and under-identification, etc. The particular advantage of this way of looking at things is, as Blankenburg (1980a,b) has shown, that what is psychopathologically abnormal is not measured only by one norm, be it a real, ideal, or mean norm, by which the abnormal manifests itself as such. Instead, both the "abnormal" and the "normal" are located in a polar field based on ideal norms that oppose one another, a field in which "deviations" can always occur from at least two sides. Thus no absolute essential norm is posited, but rather different proportions are possible, even in the domain of the nonpathological. A possible variability in the measure therefore exists which is significant for transcultural investigations with regard to the culture-specific relativity of psychiatric abnormality, among other things.

However, as Blankenburg has shown (1981, 1987), a really new way of viewing psychopathological phenomena results not from merely seeing these phenomena under the aspect of a deficiency of particular structures and functions characteristic of human existence, but rather from a dialectical perspective. From a dialectical point of view, delusion, for example, is conceived positively as what happens when one of the integrating roots of human reality becomes independent, while in healthy individuals these roots are integrated. Thus, according to Blankenburg (1987), a theory of delusion can be referred to as "anthropological" only if it lies in the essence of reality-founding intentional acts to be able to go off the rails into delusion. In other words, the normal formation of representations and judgments must be able to be portrayed as a continually negated *in status nascendi* formation of delusion.

Compared with this, it is evident that a completely different understanding of anthropological aspects in psychiatry is envisioned when the view is advocated that structures that are in themselves alien to meaning are filled with anthropological content in psychosis simply because of the "anthropological matrix" (Weitbrecht 1971). The problem of all anthropological approaches is whether, with regard to anthropological bases (*Radikale*), only a quantitative abnormality or also (and particularly) a qualitative abnormality can be conceived and described.

3 Phenomenological Methods

Anthropologically oriented psychiatry employs chiefly phenomenological methods (Strasser 1964; Spiegelberg 1972; Herzog and Graumann 1991; Waldenfels 1992; Azorin 1996; Embree et al. 1997; Gros-Azorin 1997). "Phenomenological" and "phenomenology" are concepts which are often used in psychiatry in different ways. In phenomenological-anthropological psychiatry, the emphasis is on the methodological aspects of these concepts in the sense of a label for a definite theory of methods in psychiatry. The ambiguity of the concept of phenomenon goes back to Greek antiquity, where a phenomenon is, on the one hand, that which is open to view and, on the other hand, that which appears to be something which it is not. A different understanding of phenomenon is found when, as in medicine, people talk about phenomena of illnesses or symptoms of illnesses. Here the phenomena, which are called symptoms, indicate something, namely, an illness, which does not manifest itself as such. In contrast, phenomena in phenomenological-anthropological and daseinsanalytic-existenzanalytic psychiatry always means that which shows itself in itself. The usually very broad, promiscuous use of the terms phenomena and symptoms is therefore not permissible in this form of psychiatry.

For Heidegger (1963, p. 34), phenomenology means "to let that which shows itself as it shows itself from itself be seen from itself." That "which shows itself in itself" can be understood in a twofold sense - on the one hand as the manifest, the indubitably observed, and on the other hand as that which has yet to appear and therefore must still be elicited.

Blankenburg (1991b) pointed out this double meaning of the concept of phenomenon, which has evidently led to different phenomenological approaches but also to many misunderstandings. Phenomenology in the first sense is usually understood as descriptive phenomenology, e.g., that of Jaspers. In this sense, "phenomenological" is equated with "descriptive." On the other hand, phenomenology as the science that aims to make the logos of phenomena visible and to demonstrate the essence of experiencing and structures of experience has primarily become significant in the various approaches of hermeneutic psychiatry.

In the following, in carefully considering the details of different phenomenological approaches - which can be classified as descriptive, eidetic-essentially scientific, constitutive-phenomenological, daseins- and existenzanalytic - further distinctions in the meaning of the concept of phenomena are indicated. The important discovery of the meaning-establishing significance of the subject in phenomenology has led to a polarization of the concept of phenomena such that the focus is sometimes on the subjective meaning of phenomena, i.e., reception by the subject (e.g., Husserl 1950; Sartre 1943), and at other times on the objective meaning of phenomena, as the "self-showing" of the object (e.g., Heidegger).

4 Phenomenological Approaches

4.1 Descriptive Phenomenology

In common with all the other approaches that have been mentioned, Jaspers' descriptive phenomenology (1965) refrains from all theoretical presuppositions. Jaspers' phenomenological approach became significant for psychiatry because in it "the subjective appearances of the sick mental life" (p. 47), i.e., the subjective symptoms as opposed to the objective ones, achieved due importance in psychiatry. Compared with the currently prevailing "externalized approach," i.e., a line of research oriented chiefly to external observations of data as these are deposited in diagnostic glossaries and classifications of psychiatric pictures of disorders, Jaspers' phenomenology seems to present a challenge even now and increasingly so (Schwartz and Wiggins 1987; Wiggins et al. 1992).

Descriptive phenomenology in Jaspers' sense is based methodologically on the intuitive presentation (*Vergegenwärtigung*) of the other person's mental life through a self-transposal into the patient in the sense of an empathic understanding. In this manner, Jaspers arrives at the forms in which that which the patient experiences is given.

Since, unlike my originary experience of my own mental life, patients' mental lives are always mediated through their interpretations, a problem arises which Jaspers overlooks or does not recognize: namely, how the other person's mind can be made accessible to the investigator (Schäfer 1996). For this reason, he is unable to recognize the importance of the processes of mutual understanding for diagnostic inquiry as this has been worked out in the constitution-phenomenological approaches and also in Glatzels' interactional psychopathology (1978, 1981a).

The influence of Husserl's phenomenology on that of Jaspers has been surveyed by Wiggins et al. (1992) and Wiggins and Schwartz (1997). While Jaspers wanted to contrast his phenomenology with eidetic and constitution-phenomenological approaches in Husserl's sense, his writings also contain essence-phenomenological insights and a methodological procedure which follows Husserl by varying individual psychopathological phenomena in order to work out what is identical in the pattern of the eidetic (Schäfer 1996). This is shown clearly in his analysis of the essence of the hysterical, for example. The emphasis on the form as opposed to the content, i.e., a formal taxonomy of modalities of the experienced independent of any nosology, which Jaspers inaugurated and which is so important for psychiatry is a further agreement with Husserl's phenomenology. Differences between the two phenomenologies have been set out by Blankenburg (1980b, 1991b), among others.

4.2 Eidetic-Essence-Phenomenological and Constitution-Phenomenological Approaches

The significance of Husserl's phenomenology for psychiatry lies in its particular empirical access to the experiences of the world-situated subject. According to Husserl, among the empirical "objective sciences there is also a science of subjectivity, but this is a science of the objective subjectivity which belongs to the world" (1950, p. 68). In this statement it already becomes clear that, for Husserl, consciousness is not, as it is in the natural sciences understanding of it, a kind of container, but rather consciousness constantly goes beyond itself toward things or toward the world. Consciousness is characterized by a being outside of itself with things. When it goes beyond itself, consciousness also goes beyond things toward the unity of a meaning. This going beyond Husserl calls "transcendence" (1950). Consciousness is therefore transcendence. The going beyond occurs in intentional acts which are at the same time constitutive acts because, through them, objects are constituted.

Because consciousness is always directed, intentionality (which is not to be confused with purpose) is characterized as the mode of being of consciousness. Intentionality forms a unity with what is produced by it. Because "seeing" and "the seen" cannot be separated from one another, phenomena, for Husserl, are nothing other than "consciousness of," the appearance of the thing. Husserl's phenomenology draws its empirical claim from the evidence of what is immediately given in consciousness. Like the evidence of the *ego cogito* for Descartes, every originary perception, for Husserl, is evident, i.e., "unquestionable" ("*selbstverständlich*"). Evidence does not, according to Szilasi's interpretation of Husserl (1959), supervene on the content of experience. Rather to be "evident" and to be "experienced" (*erfahren*) are identical designations.

What in psychiatry and psychology are designated as eidetic- or essential-phenomenological and also qualitative methods are often only loosely connected to what Husserl understood by "eidetic" (1950). Husserl's methodological process of free imaginative variation allows us to get from a fact to the essence of a fact, from a sensuous to a categorial intuition. In playing through various conditions of experience and contexts of experience, we can advance to the essence of a thing, i.e., to what asserts itself as invariant (see Waldenfels 1992). The goal of eidetic variation is to arrive at the

unity which encompasses all possible factual divergences and all individual cases.

Thus a great number of hysterical phenomena can be thrown into relief in a unified essence as a "seeming instead of being" (to express it in a formula) (Jaspers 1965). Through the comparison of psychopathological phenomena with one another and with comparable normal phenomena, a differential phenomenology of the pathological arises. In this phenomenology, it is less a question of the "what" than of the "how" of an experience, i.e., of the modes of givenness of a hallucinatory experience, for example (Merleau-Ponty 1966; van den Berg 1982; Naudin 1997; Kraus 1994a; Silva and Silva 1975).

Constitutive phenomenology, when it applies the so-called transcendental or phenomenological epoché (bracketing), goes beyond the previously mentioned descriptive and essential-phenomenological approaches by not only thematizing the "what" of that which appears, but also analyzing the occurrence of the "something as something" (Waldenfels 1992, p. 31). The eidetic reduction to the essence of a thing (*Sache*) takes place in a believing in the being of the real existence of the world and things (*Dinge*) independently of an intending subject; in the transcendental reduction, however, one refrains from this belief in being. "Transcendental" means precisely this disregarding of the mundane contents of acts of consciousness. Through the bracketing (epoché) of the objectives intended in the intentional acts, the conscious acts as such, i.e., the synthetic accomplishments of constituting consciousness in its peculiarity, should be able to be investigated. The object of investigation here is not the factual experiencing but rather the transcendental conditions for the possibility of experiencing.

In this manner, it becomes clear that the phenomenological epoché, the analysis of the constitution of psychopathological phenomena, is of decisive importance as an organ for the understanding of the "ununderstandable" experiences of any psychotic person. This presupposes that we also to some extent bracket our own world-behavior and that we make it a theme of constitutional analysis in order to attain that free space that makes it possible for us to turn to and elucidate without prejudice a world-relationship judged to be pathological (Blankenburg 1979a).

If we speak of the constituting accomplishments of consciousness, it seems obvious to think only of an active constitution. However, Husserl - especially in his late work - thematized a passive constitution, i.e., passive syntheses, alongside active constitution. In this context, "constitution" signifies that for a specific person something appears or exists as something.

In the constitution-phenomenological approach, it is no longer a question, as it is in descriptive or eidetic approaches, of only describing the world of a manic or delusional person, for example, but rather a question of coming to know the elements of its construction. This results in a completely different way of viewing manic behavior, for example. Thus the objectively undistanced behavior in mania can prove to be "adequate" in relation to a changed intersubjectivity to which a different distance between fellow human beings corresponds (Kraus 1998a). Mania and depression then no longer appear as they usually do, as disturbances of mood, but rather as disorders of temporal constitution, for example (Binswanger 1960). Specific psychopathological phenomena, especially psychoses, point to a modification of "transcendental organization," as Blankenburg (1971) described this in schizophrenia with few symptoms. Binswanger (1960) saw the investigation of transcendental accomplishments as so important for psychiatry that he equated the importance of Husserl's phenomenology for psychiatry with the importance of biology for the medicine of the body. He sought to show the bearing of this not indisputable thesis not only for melancholia and mania (1960), but also for delusion (1965).

An especially difficult problem for phenomenology is the knowledge of other minds or the constitution of others. The contribution of phenomenology to the question of communication and interaction with patients and also to pathological modes of communication and interaction is crucial for psychiatry and consists in the following: each specific encounter with a patient already presupposes and is made possible by an intersubjective relatedness to others and to community with them. This set of problems is addressed by Husserl (1950, vols. XIII–XV, summary in vol. I) under the title of transcendental intersubjectivity. A similar set of problems is intended by Heidegger (1963) with the concept of "being-with" as an existential.

Such transcendental intersubjectivity is to be distinguished from a factual relatedness to personality. Pathological modes of communication and interaction of psychotic individuals are mostly related to a changed intersubjectivity. The importance of the problem of intersubjectivity becomes recognizable in the fact that not only the common lifeworld but also the self and the body are intersubjectively constituted. Transcendental intersubjectivity as the condition for the possibility of factual-concrete being related to others and of encounter with others are related to each other like intimacy and trusting (Blankenburg 1991a). This is a distinction that has become particularly important for Zutt (1963a,b), Zutt and Kulenkampf (1956), and others in understanding the paranoid syndrome.

Relatedness to intersubjectivity makes it possible not only to see ourselves and the world with the eyes of others and to widen our field of vision, but also to relativize our own ways of seeing. An impairment in

the relatedness to intersubjectivity can be supposed in connection with many delusional phenomena and in schizophrenic disorders in general (Kimura 1975, 1980, 1997). In relation to others and to society, I experience an alteration of myself. This happens because I become an object for others and of their normative (role) expectations. There accrues to me an identity (a role-identity) which is designated through others' decree, but which at the same time makes it possible for me to comport myself socially and in relation to others.

Here a connection ensues between phenomenology and socio-psychological role theory, which has also become important in the realm of phenomenological-anthropological psychiatry. Role theory aspects have been highlighted for manic-depressive illness (Kraus 1982a, 1996c,d, 1997b), schizophrenia, personality disorders, and alcoholism (for an overview, see Kraus 1980b, 1982b). Intentional (Mundt 1984, 1985, 1988) and preintentional disorders (Parnas and Bovet 1991) have been observed in schizophrenia and other disorders. Attempts have been made to render these accessible to empirical research (Mundt et al. 1985).

4.3 Lifeworld

The problem of intersubjectivity was widened in Husserl's later work to form that of the intersubjectively constituted lifeworld. This concept, which had already been taken up by Jaspers (1963), was developed further by Heidegger (1963) as "the world of everyday being" and in the work of Merleau-Ponty (1966) as "being-in-the-world" ("*être-au-monde*") and by Schütz (1962–1966) and by Schütz and Luckmann (1979) for the social sciences. The concept of lifeworld was introduced into psychiatry by Natanson (see Wiggins 1995), Callieri and Castellani (1981), and Blankenburg (1979a), among others. This results in the recognition that the "objects" of consciousness are constituted not as individual objects, but as objects of the world as the open total horizon.

The world in which we constantly live with a naive-natural attitude is referred to by Husserl (1950, 1954) as "lifeworld." This pre-given world is defined by cultural-historical concerns and by the practical occupation with everyday concerns. By way of our body, our consciousness is always already originally and inseparably bound to this world. Through the body, consciousness is always already united with its world, bound to its situation at the time and its horizon. Since we are always already passively-pathically related to it, there arises a tension to active-spontaneous intentional relating-oneself-to-something (Blankenburg 1987).

Schütz (1962–1966) has shown that this unreflective living in the everyday world can be characterized by an epoché II, which inverts the phenomenological epoché I (see above). While the epoché I deliberately puts out of action the naivety of our "belief in the world" (*doxa*) – all the way from our absorption in unexamined suppositions, prejudices, and self-evidences to believing in reality generally – in epoché II, which characterizes our carrying on of everyday life, we comport ourselves in precisely the opposite manner. Here, in contrast, not only reality, but also the meaning that we give the world is taken for granted just as it is. In epoché II, in contrast to epoché I, all doubt in the lifeworld in which we live at any given time, all doubt in the validity of its norms and consciousness of its intersubjectively constituted being, is bracketed. This bracketing characteristic of the "natural attitude" makes possible our everyday actions which inevitably presuppose self-evidences.

The self- and world-relationship of the schizophrenic individual, according to Blankenburg (1979a), resembles the phenomenological epoché I. This occurs through the loss of natural self-evidence in schizophrenic experience. While the schizophrenic individual undergoes or suffers this negation of self-evidence, the phenomenologist carries it out deliberately. Through a lack of rootedness in the lifeworld, the contexts and relevances become muddled up, which is manifest not only in delusion but also in the extravagance and oddity of schizophrenic patients (Binswanger 1992).

A special expression of rootedness in the lifeworld, indeed precisely the logic of this rootedness, is common sense (Blankenburg 1969). While individuals with schizophrenia, especially those with hebephrenia, frequently appear to lack an inner standard which makes it possible for them to behave in a way that fulfills social norms, people with melancholia and manic-depressive illness appear to be too strongly dominated by common sense (Blankenburg 1969; Stanghellini 1997). This is already shown premorbidly by their hypernomic behavior, which is characterized by a lack of distance in their relationship to normative expectations (Kraus 1977, 1980b, 1991b). This is found in the typus melancholicus (Tellenbach 1983) in patients' over-identification with their social roles at any given time, i.e., a predominance of role-identification over ego-identification (Kraus 1977, 1991c, 1996c,d).

From this perspective, role-identity and ego-identity are distinguished by the degree of absorption or nonabsorption in the lifeworld. In this regard, individuals with schizophrenia appear to have special difficulties in taking on social roles: in such roles they frequently experience themselves as alienated. This is

obviously connected with the fact that each role performance, on account of the reciprocity of each role with other roles, contains the capacity to imaginatively transpose the self into others, to assume their point of view (Kraus 1982b).

Through the epoché II of the lifeworld with its typifications and its predominant "oneself" ("*Man-Selbst*"), the existential situation of human beings, which is characterized by "being-toward-death" and therefore by threat, anxiety, unavoidable guilt, and loneliness, becomes concealed.

We can thus ask, as Wiggins (1995) does, closely following Natanson (1963), whether the anxiety syndrome, neuroses, and psychoses represent something new as against health or rather, like the primordial anxieties of Kurt Schneider (1950), they become revealed simply through a breakdown of epoché II. Analyses of the lifeworld are in principle possible with all mental disorders. Callieri and Castellani (1981), for example, have very vividly described the different lifeworld relationships of senile, phobic, depressive, phobic-compulsive, manic, and schizophrenic patients (on the latter, see Corin 1990).

4.4 Daseinsanalytic and Existenzanalytic Approaches

Kuhn (1963), Condrau (1992), Blankenburg (1977), and Holzhey-Kunz (1990, 1994a-c), among others, have provided introductory presentations on daseinsanalytic and existenzanalytic approaches.

Binswanger's psychiatric daseinsanalysis as an empirical phenomenological method differs from Heidegger's daseinsanalytics as a philosophical-phenomenological ontology, even though the former is grounded on the latter. Heidegger's point of departure is a fundamental distinction between beings which have the being of *Dasein* and beings which do not. The former are characterized by so-called existentials which strictly depend on one another, such as being-in-the-world, oneself (*Man*), historicity, attunement, etc., while determinations of the being of the latter are designated by categories. Being-in-the-world is the leading concept of Binswanger's daseinsanalysis (Holzhey-Kunz 1990). This concept rejects the being of human beings in the world as in a container. Being-in-the-world is rather the fundamental dimension of this being itself. The dichotomy of subject/object and of conscious/unconscious is thus circumvented.

Although in his later work Husserl (1954, 1986), with his notion of the lifeworld, his emphasis on the primordial, prereflective world of everyday experience, and his concept of passive syntheses, paved the way for Heidegger's fundamental ontology, the latter represents a certain reaction against Husserl's philosophy of consciousness. By turning the focus to the prepredicative or prepopositional determinations of *Dasein*, i.e., to its projecting-thrownness (which is not experienced but rather lived) or to its being-in, being-with, being-body, "being-to-the-world" (Merleau-Ponty 1966), "being-for-oneself" and "being-for-the-other" (Sartre 1943), more is kept of the whole human being, in his or her existence, situation, and freedom, in the bodily, interpersonal, social, and historical conditions of his or her existence. In this lies the advantage of psychiatric daseinsanalysis – as well as of existenzanalytic approaches – over constitution-phenomenological approaches.

For the most part, these approaches mark a greater closeness to individual patients and therefore also to therapy, while constitution-phenomenological approaches, as we shall see, become increasingly more closely connected with empirical-experimental and cognitive research approaches. Even if with the concept of *Dasein* a "transanthropological matrix" (Blankenburg 1967) is ontologically intended, Binswanger's earlier anthropological procedure, which is methodologically oriented in "natural-ontological experience" (Szilasi 1961), appears to entitle us to classify his daseinsanalysis as a phenomenological-anthropological approach. We cannot go into the set of problems involved in the difference between ontic and ontological here.

Daseinsanalysis seeks to use various psychiatric concepts of disorders to describe the modification of the fundamental condition of *Dasein*, i.e., being-in-the-world in the modification of the components of its structure. These components of its structure are, for example, its temporalization and spatialization showing event-character, its self-being, and being-with. Through "hermeneutic communication" with patients (Binswanger 1958), daseinsanalysis seeks to investigate the mode of being or condition of being of patients with specific clinical disorders, or it seeks to uncover which possibilities of *Dasein* are opened up at any given time; it does this against the background of the general, i.e., a priori, fundamental structure of human *Dasein*, which it uses as a guiding rule or norm. Daseinsanalysis is not satisfied with describing in a manic patient, e.g., his or her flight of ideas and the flow of mental processes connected with it; it rather seeks to define that form of being human which the flight of ideas, and also delusion, hallucination, etc., first makes possible.

With the description of different forms of being-in-the-world of a unified form of *Dasein*, something like the recognition of a principle of order occurs (Kuhn 1963). According to Binswanger (1958), this principle of order is not a logical principle, but rather an eidos or intuitable schema. Insofar as daseinsanalysis refers to the life-history of a patient, it is a matter of uncovering

the style of life of the patient; in others words, it is a matter of the connection of a whole conduct of life (Kuhn 1963) out of which the particular modes of experience and motive become intelligible.

With psychiatric daseinsanalysis, the horizon of understanding, according to Binswanger (1958), should be transcended to its a priori horizon of understanding, i.e., to the a priori framework structure and the a priori movement of *Dasein*. Daseinsanalysis should be understood, however, as being basically a science based on experience "with its own methods and its own ideal of exactitude" (Binswanger 1961, p. 191; Szilazi 1961). We cannot discuss daseinsanalysis in the sense of Boss (1951) here, for whom phenomenology is a "method-free intuition of the essence of things" (Holzhey-Kunz 1990, p. 94).

Among the manifold influences on psychiatry of French phenomenologists such as Merleau-Ponty, Sartre, Ricoeur, and Levinas, I can only refer to Sartre's "existential psychoanalysis" here (Sartre 1943). In addition to his problematization of the unconscious, shame, guilt, the emotions, and other important psychiatric themes, Sartre's existential psychoanalysis has become significant for phenomenological–anthropological psychiatry as an empirical method for the illumination of the psychic constitution of an individual *Dasein* or of a type of *Dasein*. For this reason, it can certainly be compared with daseinsanalysis, even if it is less well known. "Existential psychoanalysis" begins with the fact that the individual, despite all the heterogeneity of his or her inherited and acquired biological and psychic traits and dispositions, projects and comprehends him- or herself at any moment as a personal unity. There is correspondingly an original choice ("*choix originelle*") of the individual which, in all of his or her modes of behavior, inclination, and striving, comes to expression as a choice of a totality, i.e., in a definite kind of existing. According to Sartre (1943), this choice is not thetic but rather prereflective. And because it enters into the fundamental relation of being, this choice is, as choice of being, much more fundamental than, for example, sexuality or the will to power.

The significance of "existential psychoanalysis" for psychiatry lies, among other things, in discovering eidetic or essential entities in the sense of existential types in the traditional entities of classification and in going beyond these (Kraus 1977, 1996c). The multitude of factual findings, which can be reduced to a common denominator under the hypothesis of a determinate being-relation, serves as a criterion for such a type. The issue of the conditions or causes of such a type can be left aside. Like the forms of being-in-the-world of daseinsanalysis with regard to self-being (authenticity), being-with-others, body-being, temporality, etc., "existential types" can be read purely structurally as definite modes of being-in-the-world in the sense of an existential enacting of being. In the latter sense, they can provide immediate indications for rehabilitation and therapy.

Kraus (1996c) has portrayed people with melancholia and people with hysteria with regard to their relation to social roles, to their value stances, in relation to their bodies, to their emotions, etc. as contrary existential types of identity formation: a tendency to over-identification in the individual with melancholia contrasts with a tendency to non-identification in the individual with hysteria. In melancholia, increasing over-identification leads into depersonalization, which is conceived as the core of melancholic disorder (Kraus 1991a, 1992).

In connection with Bachelard's "*psychoanalyse des choses*" (1987) and the existential analysis of thing metaphors of dirt in dirt phobia (Kraus 1996a) and metaphors of specific qualities of space in agoraphobia and claustrophobia (Kraus 1997a) and also of the technical in technical delusion (Kraus 1994a), pictures of these disorders can likewise be characterized as existential types of specific self-world relations and relations of being-with-others.

5 Psychotherapy and Rehabilitation Therapy

From early on right up to today, phenomenological–anthropological psychiatry has intensively concerned itself with problems of psychotherapy and rehabilitation, or rehabilitation therapy (von Gebsattel 1954; Bräutigam 1961; Condrau 1963, 1990, 1992; Blankenburg 1982, 1987, 1990; Csef 1990; Wyss 1990; Lang 1990, 1997; Holzhey-Kunz 1994c; Holm-Hadulla 1997). The focus is generally on reflections on the doctor–patient relationship and therapeutic possibilities rather than the elaboration of therapeutic techniques. The therapist is a *Dasein* partner. Psychotherapy ought to work through establishing a "real" relationship, through empathy and experience of existential truth (Wyss 1990). This has created a special culture of intimacy with patients (Mundt 1989), which has also had an impact outside this psychiatric approach.

Beginning with the a priori fundamental structure of being human, Binswanger's anthropological proportions as polar structures in the development of *Dasein* make it possible for the therapist to partially identify him- or herself with patients (Blankenburg 1980a) and hence to enter into a special kind of "hermeneutic communication" with them (Binswanger 1958).

In phenomenological analyses of constitution, Blankenburg (1982) sees the possibility of a "constitutive psychotherapy." The former addresses the issue

of how something is constituted as something for someone and of how someone is constituted as someone for someone. It is a matter of performing an epoché (bracketing) of the self-relation and world-relation of patients in order to realize processes of the formation of reality for them and to work with them; it is also a matter of eventually achieving help with constitution. What is crucial, however, is that therapists should also be ready to bracket their own self- and world-relationships in order to produce that special space which makes it possible for them to "translate" themselves into a different form of being-in-the-world, e.g., the psychotic (Blankenburg 1979a, 1982). With the phenomenological approach, Blankenburg (1987, 1991a) also sees possibilities of developing procedures for training in intentionality and of provoking a flexibility of perspectives which can be started with hallucinatory and delusional patients with their fixation on a definite theme or with their "rigid perception" (Matussek, see Blankenburg 1987).

Phenomenologically-anthropologically oriented forms of psychotherapy are better aligned to helping patients exercise their own responsibility than are psychotherapies which analyze conditions. They assert what is at least a relative autonomy on the part of patients by requiring of them and offering them new possibilities of being (Blankenburg 1997b). Paradox intervention, which was introduced by Frankl (1959) and propagated in a modified form by Blankenburg (1990), therefore appeals to the autonomy of patients for a quasi-active bringing forth of symptoms which were previously only passively suffered. The therapist sees him- or herself as someone who makes it possible to point out the future. These therapeutic approaches are therefore less oriented to the past than to the future (Blankenburg 1989), corresponding to the great significance which is accorded by phenomenological-anthropological authors to the experiencing of time and temporalization (Alonso-Fernandez 1982; Bollnow 1963; Bühler 1986; Kobayashi 1989; Kraus 1985, 1998a; Minkowski 1971, 1972; Mooij 1995; Mundt et al. 1998; Kimura 1982; Pauleikhoff 1979; Csef 1985) and to the experiencing of space and spatialization (Tellenbach 1987; Fuchs 1994, 1998) in people with neuroses and psychoses.

Pretherapy was established by Prouty and further developed by Van Werde and Portner (Prouty et al. 1998). It can be regarded as standing entirely in the tradition of phenomenological psychology and psychiatry. The therapy is seen as indicated for patients with chronic schizophrenia who have been hospitalized for years. This form of therapy starts from Minkowski's thesis (1971, 1972) of the lack of vital contact between the person and reality as the chief characteristic of schizophrenia, and it seeks, through so-called "reflections of contact" (responding to the situation or facial expression, reproducing body posture, verbatim repetition of what the patient has said, etc.) to bring patients into contact with reality again, i.e., to bring them into contact with themselves or with others and then to strengthen these functions of contact.

Sartre's "existential psychoanalysis" (1943) provides possibilities previously little used before in order to illuminate the "original choice" (see above) of the grounding project of the individual or of people with specific neuroses (Kraus 1996c, 1997a, 1998b), personality disorders, and psychoses (Kraus 1996c, 1977, 1998a). On that basis, it seeks to understand specific, typical ways of behaving and experiencing and hence to find a therapeutic clue for a new possibility of being. On the basis of "existential psychoanalysis," Kraus (1997b) has developed an identity therapy for people with melancholia and manic-depressive illness which starts from the above-mentioned over-identification of these patients with their social roles, spatio-temporal givens (dwelling), etc. This also expresses itself in the social behavior of these patients, namely, in their hypernomic, i.e., overly precise, fulfillment of social norms or role expectations and in their emotional and cognitive intolerance of ambiguity (Kraus 1988). Their intolerance of ambiguity helps to maintain the identity they find in particular role relationships. On the basis of this concept, which was designed to be empirical, it is possible to undertake a psychotherapeutic management of these patients which is guided by theory and to conceive of definite measures for the prevention of phases and for rehabilitation.

There are analyses of transcendental-intersubjective constitution of the lifeworld or of identity formation in people with endogenous psychoses which have led to practical implications for their rehabilitation. The concept of "accomplishing life," whereby Husserl in his later work (1954, 1986) understood the entire stratified structure of meaning-bestowing acts which constitute a world or the "having of a world," and Heidegger's (1963) being-in-the-world as accomplishment caused Blankenburg (1970) to plead for a more comprehensive concept of accomplishment that includes categorial accomplishments. In this way, it is possible to characterize the loss of accomplishment and the condition of exhaustion of the schizophrenic deficit syndrome, e.g., the syndrome of apathy (Mundt 1985) under the aspect of a deficiency of categorial accomplishment. By characterizing them in this manner, they can be more precisely distinguished from an insufficiency of accomplishment that is purely bodily conditioned, as can occur, for example, under the aspect of a loss of energetic potential (Conrad 1958) or a dynamic insufficiency (Janzarik 1959).

For the healthy individual, the coherence of ego and world is given in a basic way, namely, through the natural self-evidence of the embeddedness of the

accomplishment in the common lifeworld. However, according to Blankenburg (1970), people with schizophrenia have to "co-accomplish" this coherence of ego and world, the access to reality and to everydayness, as their own specific categorial accomplishment, and doing this frequently demands too much of them. This leads to an interweaving of bodily and categorial exhaustion such that categorial failure is experienced as bodily frailty and bodily exertion as ego loss.

6 Phenomenological–Anthropological and Daseinsanalytic Approaches and the Empirical-Objectivating Sciences

In a certain sense, a one-sided relationship exists between sciences of essence and sciences of fact insofar as, according to Husserl (1954), any eidetic science can in principle be independent of factual sciences, while no empirical science can be completely free of eidetic knowledge. The empirical sciences cannot be completely free from the eidetic sciences because any factual science essentially has theoretical foundations in eidetic ontology (e.g., the natural sciences in the ontology of nature) (Husserl 1950, vol. I, p. 111). Not only do the empirical sciences draw their fundamental categories directly or indirectly from the lifeworld (Husserl 1954), but also the interpretation of empirical data often relies on a connection with the lifeworld.

Empirical–objectivating approaches are frequently characterized entirely as reductionistic, and phenomenological–anthropological ones as holistic. However, a distinction needs to be made between methodological and ontological reductionism. Methodologically, the empirical sciences deal with human beings as objects of investigation in an unavoidably reductive manner insofar as they abstract from the full essential reality of human beings, from their subjectivity and historicalness, for example. However, essential sciences no less than empirical sciences are exposed to the danger of an ontological reductionism.

Through their primarily heuristic alignment with the essence of human beings, the phenomenological–anthropological and daseinsanalytic forms of psychiatry attain a special place in the realm of scientific concerns in psychiatry. They attain this special place insofar as it is one of their tasks to display the reductionism of empirical–objectivating approaches in order to open up for those sciences, through referring to the more comprehensive "essence" of human reality, new questions and dimensions, but also to promote the adequacy of their methods for the object. As a circular relationship exists between problem-oriented (e.g., phenomenological–anthropological) and method-centered empirical investigations, Mundt (1989) speaks of a relationship of alternation between hermeneutic–holistic seeing and empirical–distancing objectivating. Qualitative investigation frequently precedes quantitative investigation, but qualitative research can also follow quantitative research in the interpretation of results.

The psychiatric subdisciplines have become increasingly more anthropological, i.e., more adequate for the essence of human beings. This is associated not only with the influence of phenomenological–anthropological or other "humanistic" approaches. It has also frequently developed out of the aporias of the findings in these subdisciplines. As an example, we can point to life-event research, which saw itself increasingly compelled to consider personality-specific sensitivity to traumatic events.

Human beings cannot be exclusively natural objects, and, on principle, not all of mental life can be objectivated (Herzog 1994) (although, insofar as the mental and the psychic can be distinguished, this is only valid for the psychic in the narrow sense of the term). Because of this and because of their orientation to subjectivity and human freedom, phenomenological–anthropological and daseinsanalytic psychiatry can claim not only a methodological, but also a systematic autonomy. And they can claim this autonomy in addition to their auxiliary function for the empirical sciences.

In contrast to the fixing of human beings in an objectivity which is definable in various ways in the empirical or, more adequately characterized, objectivating sciences, phenomenological–anthropological and daseinsanalytic psychiatry allow human beings to be what they are in all their transcending possibilities. These disciplines thus enter into an openness of questioning which is not only their strength but also their weakness, because of the greater difficulties involved in planning and designing, e.g., in the domain of research projects. Among the manifold fruitful relationships between phenomenological–anthropological and objectivating research approaches, I shall refer here only to the personality research concerning manic-depressive illness (see Mundt et al. 1996; Kraus 1991c, 1996d) and to empirical investigations of time experience and temporality in melancholia (Mundt et al. 1998) and mania.

7 Phenomenology, Cognitive Theory, and Cognitive Neuroscience

Between phenomenological orientations and cognitive approaches there exist manifold but as yet insufficiently

disclosed historical and systematic connections. Thus the relatively earlier emphasis on the subject's individual experiences and meaning-bestowing by Husserl (1950), Jaspers (1965), Binswanger (1955a, 1961), Minkowski (1971, 1972), Straus (1963, 1978a), and other representatives of phenomenological-anthropological orientations found a parallel later in Kelly's "personal construct theory" (1955) and the cognitive theory of Beck (1970), among others. As a representative example, I shall refer only to Straus. In his early monograph, *Event and Experience* (1978a, first edition 1930), and in his chief work, *On the Sense of the Senses* (1978b, first edition 1956), Straus conceived an individually and culturally constituted world of experience and developed the notion of a "representative meaning" of traumatic events. In this respect, he can be viewed as one of the predecessors of contemporary cognitive psychology.

Reciprocally fruitful relationships have been noted between phenomenology, in particular the phenomenology of constitution in Husserl's sense, in its original form plainly a science of cognition, and the cognitive neurosciences by Baumgartner (1996), Wiggins and Spitzer (1997), Gallagher (1997), Moss (1981), Naudin et al. (1997), and Schwartz et al. (1997).

Phenomenology, especially existential-anthropological approaches, continue to present a great challenge for the cognitive sciences. A chief difference exists in the dual status of the mental and the world. Phenomenology does not adopt this dualism because, in a certain sense, consciousness is always already embodied and socialized "in the world." In Gallagher's words (1997), the cognitive sciences tend, at least implicitly, toward an "internalism." Phenomenology, however, could be characterized rather as an "externalism," as Heidegger's concept of being-in-the-world (1963) and that of embodied existence (Merleau-Ponty 1966) make especially clear. Cognitive behavior is not primary, but rather secondary in relation to a primary involvement in an already precognitively meaningful world.

According to a number of phenomenological-anthropological authors, analyses of precognitive, prepredicative world-relatedness and social world-relatedness result in a critical revision of cognitive approaches to delusion (Sass 1992; Kraus 1991a, 1994a, 1996b; Lopez-Ibor 1982; Schwartz and Wiggins 1992), hallucinations (Blankenburg 1987; van den Berg 1982; Naudin 1997), and mania (Kraus 1998a), for example. If in the sciences of cognition there exists a tendency toward scientific explanation on the subpersonal level, then, according to Gallagher (1997), subpersonal and subintentional processes are frequently treated as isomorphic, i.e., as if they were directed intentionally, and this is littered with the danger of anthropomorphizing or intentionalizing them.

Varela (1996) proposes a neurophenomenology, and we can certainly agree with him if, as he says, it is a matter of showing not only correspondences and correlations, but also oppositions between phenomenological and cognitive approaches, including the cognitive neurosciences. In this manner, both research approaches can enrich and reciprocally illuminate one another, but also be distinguished from one another.

8 Diagnostics and Classification

The diagnostic glossary of DSM-IV (APA 1994) and analogously that of ICD-10 (Dilling et al. 1991) were drawn up on the basis of the logical empiricism of the philosopher Carl Hempel (1965). Through the application of the procedures of multivariate statistics, such as factor and cluster analysis, but also through consensus and compromise, these glossaries have led to entirely new diagnostic entities, which, in relation to the traditional entities, have led to a growing crisis in taxonomy.

The diagnostic glossary largely replaces intuitive-eidological acts with the generation of so-called diagnostic criteria and the introduction of algorithmic decision trees. However, this ignores the originally "hermeneutic nature of diagnostic processes" (Spitzer 1994), with their acts which are anticipatory-proleptic and retrospective with regard to part and whole (Kraus 1996b).

At the beginning of the diagnostic situation in psychiatry - if we view it phenomenologically - we are primarily given phenomena of a particular being-in-the-world and being-ill. With reference to medical models of illness, we reduce these phenomena to symptoms or criteria of an illness. This initial intuition of a totality of a particular being-in-the-world is more than merely a characteristic combination of symptoms, as Wyrsch (1946), Kraus (1996b), and others have shown. This totality expresses itself not only in so-called expressive symptoms, but also in a definite mode of *Dasein.* As the being-schizophrenic, being-manic, being-depressive of the person, it can be contrasted with schizophrenia, mania, and depression as illnesses.

There are therefore two forms of diagnosis, a phenomenological one and a symptomological-criteriological one (Kraus 1991b,d, 1994a,b; Ballerini 1997). Both forms of diagnostics supplement one another, but they also stand in a certain relation of tension with one another. What intuition as such apprehends only vaguely can be typologically differentiated with the aid of the phenomenologial categories of temporalization, spatialization, embodiment, relation to self and to others, etc. In this way, the oneness apprehended through intuition is definable, e.g., as an existential

type (Kraus 1991b; Mishara 1994). The significance of this phenomenological diagnostic method lies primarily in the fact that, through reference to the form of being-in-the-world, a higher degree of phenomenological specificity can be achieved, as Rümke (1958), Wyrsch (1946), Müller-Suur (1958), and Dörr-Zegers and Tellenbach (1980) have pointed out. This can be contrasted with a vast lack of specificity in symptomological diagnostics (Weitbrecht 1957). Such existential types and essential entities can, like the ideal types in Jaspers' sense (1965), contribute to the discovery of real, i.e., etiologically based illnesses (Wiggins and Schwartz 1994). On account of their reference to the lifeworld of the patient, they are especially important for rehabilitative and therapeutic measures.

In summary, phenomenological-anthropological approaches in psychiatry make the following contributions to diagnostics and classification:

1. They analyze more precisely the procedure of diagnosis itself, to which too little attention was previously paid.
2. They generate ideal-typical and essential-typological entities.
3. They carry out critical reflection on previous descriptive categories with regard to evaluative elements.
4. They advance a phenomenological form of diagnostics which takes account of prepredicative world-relationships and social world-relations.
5. They contribute to the development of psychiatry's own conception of illness.

The methods of psychiatry may change, but the issue of the conditions in the essence of human beings for the possibility of being mentally ill and of the methodology that is right for these conditions will always continue to exist.

9 References

Alonso-Fernandez F (1982) Space and time for the manic person. In: De Koning AJJ, Jenner FA (eds) Phenomenology and psychiatry. Academic, London, pp 165–172

APA (1994) Diagnostic and statistical manual of mental disorders, 4th edn (DSM-IV). APA, Washington DC

Azorin JM (1996) Position de la psychiatrie phénoménologiste par rapport aux psychoses et aux névroses. In: Confrontations psychiatriques: epistémiologie et psychiatrie. Specia Rhône-Poulenc Rorer

Bachelard G (1987) Poetik des Raumes. Fischer, Frankfurt am Main

Ballerini A (1997) La diagnosi in psichiatria. La nuova Italia Scientifica, Roma

Baumgartner E (ed) (1996) Phenomenology and cognitive science. Röll, Dettelbach

Beck A (1970) The core problem in depression: the cognitive trial. In: Massermann JH (ed) Depression: theories and therapies. Grune and Stratton, New York, pp 47–55

Binswanger L (1955a) Ausgewählte Vorträge und Aufsätze. Francke, Bern

Binswanger L (1955b) Vom anthropologischen Sinn der Verstiegenheit. Ausgewählte Vorträge und Aufsätze. Francke, Bern

Binswanger L (1958) Psychiatrisches Denken der Gegenwart in der Schweiz. J Psychol Psychother 6: 175–192

Binswanger L (1960) Melancholie und Manie. Neske, Pfullingen

Binswanger L (1961) Ausgewählte Vorträge und Aufsätze, 2nd edn. Francke, Bern

Binswanger L (1965) Wahn. Beträge zu seiner phänomenologischen und daseinsanalytischen Erforschung. Neske, Pfullingen

Binswanger L (1992) Formen mißglückten Daseins. In: Herzog M (ed) Ludwig Binswanger. Ausgewählte Werke, vol 1. Asanger, Heidelberg, pp 1–443

Binswanger L (1994) Der Mensch in der Psychiatrie. In: Holzhey-Kunz A (ed) Ludwig Binswanger. Ausgewählte Werke, vol 4. Asanger, Heidelberg, pp 57–72

Blankenburg W (1967) Die anthropologische und daseinsanalytische Sicht des Wahns. Stud Gener 20: 639–650

Blankenburg W (1969) Ansätze zu einer Psychopathologie des 'common sense'. Confinia Psychiatr 12: 144–163

Blankenburg W (1970) Zur Leistungsstruktur bei chronischen endogenen Psychosen. Nervenarzt 41/12: 577–587

**Blankenburg W (1971) Der Verlust der natürlichen Selbstverständlichkeit. Enke, Stuttgart

Blankenburg W (1972) Grundsätzliches zur Konzeption einer 'anthropologischen Proportion'. Z Klin Psychol Psychother 22: 322–333

Blankenburg W (1977) Die Daseinsanalyse. In: Eicke D (ed) Freud und die Folgen. Kindler, Zürich, pp 942–964 (Die Psychologie des 20. Jahrhunderts, vol III/2)

Blankenburg W (1979a) Phänomenologische Epoche und Psychopathologie. In: Sprondel WM, Grathoff R (eds) Alfred Schütz und die Idee des Alltags in den Sozialwissenschaften. Enke, Stuttgart, pp 125–139

Blankenburg W (1979b) Towards a more man-centered psychiatry. In: Schäfer KE (ed) A new image of man in medicine, vol 3. Futura, Mt Kisko

**Blankenburg W (1980a) Anthropologisch orientierte Psychiatrie. In: Peters UW (ed) Ergebnisse für die Medizin. 2. Psychiatrie. Kindler, Zürich, pp 182–197 (Die Psychologie des 20. Jahrhunderts, vol X)

Blankenburg W (1980b) Phenomenology and psychopathology. J Phenomenol Psychol 1: 50–78

Blankenburg W (1981) Wie weit reicht die dialektische Betrachtungsweise in der Psychiatrie. Z Klin Psychol Psychopathol Psychother 29: 45–66

Blankenburg W (1982) Zur Indikation hermeneutischer Methoden in der Psychotherapie am Paradigma der Daseinsanalyse. In: Helmchen H, Linden M, Rüger U (eds) Psychotherapie in der Psychiatrie. Springer, Berlin Heidelberg New York, pp 41–46

Blankenburg W (1983) Der Leib als Partner. Psychother Psychosom Med Psychol 33: 206–212

Blankenburg W (1987) Phänomenologisch-anthropologische Aspekte von Wahn und Halluzination. In: Olbrich HM (ed) Halluzination und Wahn. Springer, Berlin Heidelberg New York, pp 77–101

Blankenburg W (1989) Futur-II-Perspektive in ihrer Bedeutung für die Erschliessung der Lebensgeschichte des Patienten. In:

Blankenburg W, Heinrich K, Peters UH, Neundörfer B (eds) Biographie und Krankheit. Thieme, Stuttgart, pp 76–84
Blankenburg W (1990) Wirkfaktoren paradoxen Vorgehens in der Psychotherapie. In: Lang H (ed) Wirkfaktoren der Psychotherapie. Springer, Berlin Heidelberg New York, pp 122–138
Blankenburg W (1991a) Perspektivität und Wahn. In: Blankenburg W (ed) Wahn und Perspektivität. Enke, Stuttgart, pp 4–28
Blankenburg W (1991b) Phänomenologie als Grundlagendisziplin der Psychiatrie. Fundam Psychiatr 5: 92–101
Blankenburg W (1995) Das Sich-Befinden zwischen Leiblichkeit und Gefühl. In: Grossheim M (ed) Leib und Gefühl. Beiträge zur Anthropologie. Akademieverlag, Berlin, pp 201–214
Blankenburg W (1997) "Zumuten" und "Zumutbarkeit" als Kategorien psychiatrischer Praxis. In: Krisor M (ed) Was du nicht willst, dass man dir tut ... Roderer, Regensburg, pp 21–48
Blankenburg W, Haltenhoff H (1994) Selbst und Leib. Psycho 20: 48–51
Bollnow OF (1963) Mensch und Raum. Kohlhammer, Stuttgart
Boss M (1951) Beitrag zur daseinsanalytischen Fundierung des psychiatrischen Denkens. Schweiz Arch Neurol Psychiatr 67/1: 15–19
Braun HJ (ed) (1992–1994) Ludwig Binswanger. Ausgewählte Werke, vols 1–4. Asanger, Heidelberg
Bräutigam W (1961) Psychotherapie in anthropologischer Sicht. Enke, Stuttgart
Bühler KE (1986) Zeitlichkeit als psychologisches Prinzip. Über Grundfragen der Biographie-Forschung. Janus, Cologne
Buytendijk FJJ (1967) Prolegomena einer anthropologischen Physiologie. Müller, Salzburg
Callieri B, Castellani A (1981) On the psychopathology of the lifeworld. In: Bello AA (ed) Analecta Husserliana. Reidel, London, pp 172–202
Condrau G (1963) Daseinsanalytische Psychotherapie. Huber, Bern
Condrau G (1990) Heilfaktoren in der daseinsanalytischen Psychotherapie. In: Lang H (ed) Wirkfaktoren der Psychotherapie. Springer, Berlin Heidelberg New York, pp 150–155
Condrau G (1992) Daseinsanalyse. In: Battegay R (ed) Handwörterbuch der Psychiatrie. Enke, Stuttgart, pp 101–105
Conrad K (1958) Die beginnende Schizophrenie. Thieme, Stuttgart
Corin EE (1990) Facts and meaning in psychiatry. An anthropological approach to the lifeworld of schizophrenics. Cult Med Psychiatry 14: 153–188
Csef H (1985) Zum Zeiterleben des Zwangskranken. In: Bühler KH, Weiss H (eds) Kommunikation und Perspektivität. Königshausen and Neumann, Würzburg, pp 127–138
Csef H (1990) Anthropologisch-integrative Psychotherapie. Fundam Psychiatr 4: 18–26
De Koning AJJ, Jenner FA (eds) (1982) Phenomenology and psychiatry. Academic, London
Dilling H, Mombour W, Schmidt MH (1991) Internationale Klassifikation psychischer Störungen: ICD-10. Huber, Göttingen
Dörr-Zegers O (1995) Psychiatría antropológica. Editorial Universitaria, Santiago de Chile
Dörr-Zegers O, Tellenbach H (1980) Differentialphänomenologie des depressiven Syndroms. Nervenarzt 51: 113–118
Edie JM (1966) Phenomenology and psychiatry: the need for a "subjective method" in the scientific study of human behavior. In: von Baeyer W, Griffith RM (eds) Conditio humana. Springer, Berlin Heidelberg New York, pp 55–73
Embree L, Behnke EA, Carr D et al (1997) Encyclopedia of phenomenology. Kluwer, Dordrecht
Fédida P, Schotte J (1991) Psychiatrie et existence, 2nd edn. Millon, Grenoble
Figueroa-Cave G (1996a) Hacia una antropología psiquiátrica. Rev Chil Neuropsiquiatr 34: 131–137
Figueroa-Cave G (1996b) Los fundamentos filosophicos. Rev Chil Neuropsiquiatr 34: 381–390
Figueroa-Cave G (1997a) Hacia una antropologia psiquiátrica. III. Los hallazgos psiquiátricon. Rev Chil Neuropsiquiatr 35: 7–15
Figueroa-Cave G (1997b) Towards a psychiatric anthropology. III. The psychiatric findings. Rev Chil Neuropsiquiatr 35: 7–15
Frankl VE (1959) Grundriß der Existenzialanalyse und Logotherapie. In: Frankl VE, von Gebsattel VE, Schultz JH (eds) Handbuch der Neurosenlehre und Psychotherapie. Urban and Schwarzenberg, Munich
Frankl VE (1984) Der leidende Mensch, 2nd edn. Huber, Bern
Fuchs T (1994) Die Welt als Innenraum. Kafkas "Bau" als Paradigma paranoider Räumlichkeit. Nervenarzt 65: 470–477
Fuchs T (1998) Phänomenologie des Leib- und Raumerlebens. Post-doctoral dissertation, University of Heidelberg
Gallagher S (1997) Mutual enlightment: recent phenomenology in cognitive science. J Consciousn Stud 4/3: 195–214
Gehlen A (1966) Der Mensch. Athenäum, Frankfurt am Main Bonn
Glatzel J (1978) Allgemeine Psychopathologie. Enke, Stuttgart
Glatzel J (1981a) Spezielle Psychopathologie. Enke, Stuttgart
Glatzel J (1981b) Zum Problem der Krankheitseinsicht in der Psychiatrie seit Jaspers. In: Burchard JM (ed) Psychopathologie. Schattauer, Stuttgart, pp 133–145
Gros Azorin C (1997) Phénomenologie et expérience psychiatrique chez Ludwig Binswanger. Doctoral thesis, University of Paris XII
Hegel GWF (1952) Phänomenologie des Geistes (Sämtliche Werke, vol 5). Hoffmeister, Hamburg
Heidegger M (1963, [1]1927) Sein und Zeit, 10th edn. Niemeyer, Tübingen
Heimann H (1994) Psychiatrie und Anthropologie in Geschichte und Gegenwart. Fundam Psychiatr 8: 60–64
Hempel CG (1965) Fundamentals of taxonomy. In: Neurath O, Carnap R, Morris C (eds) Aspects of scientific explanation and other essays in the philosophy of science. Free Press, New York, pp 137–154
**Herzog M (1994) Weltentwürfe. Ludwig Binswangers phänomenologische Psychologie. de Gruyter, Berlin
Herzog M, Graumann CF (eds) (1991) Sinn und Erfahrung. Phänomenologische Methoden in den Humanwissenschaften. Asanger, Heidelberg
Holm-Hadulla R (1997) Die psychotherapeutische Kunst. Vandenhoeck and Ruprecht, Göttingen
Holzhey-Kunz A (1990) Ludwig Binswanger: Daseinsanalyse als wissenschaftlich exakte Untersuchung von Weltentwürfen. Daseinsanalyse 7: 81–101
**Holzhey-Kunz A (1994a) Die Daseinsanalyse und ihre Aufgabe einer Hermeneutik psychopathologischer Phänomene. Passagen, Vienna
Holzhey-Kunz A (1994b) Einleitung der Herausgeberin. In: Holzhey-Kunz A (ed) Ludwig Binswanger. Ausgewählte Werke. Der Mensch in der Psychiatrie. Ansanger, Heidelberg, pp 1–55

Holzhey-Kunz A (1994c) Leiden am Dasein. Die Daseinsanalyse und die Aufgabe einer Hermeneutik psychopathologischer Phänomene. Passagen, Vienna

Husserl E (1950–1996) Husserliana, vols I–XXIV. Nijhoff, The Hague

Husserl E (1954) Die Krisis der europäischen Wissenschaften und die transzendentale Phänomenologie. Eine Einleitung in die Phänomenologie. (Husserliana, vol VI). Nijhoff, The Hague

Husserl E (1986) Phänomenologie der Lebenswelt. (Ausgewählte Texte, vol II). Reclam, Stuttgart

Janzarik W (1959) Dynamische Grundkonstellationen in endogenen Psychosen. Springer, Berlin Göttingen Heidelberg

Janzarik W (1988) Strukturdynamische Grundlagen der Psychiatrie. Enke, Stuttgart

Jaspers K (1965, [1]1913) Allgemeine Psychopathologie, 8th edn. Springer, Berlin Heidelberg New York

Kelly GA (1955) The psychology of personal constructs. Norton, New York

Kimura B (1975) Schizophrenie als Geschehen des Zwischenseins. Nervenarzt 46: 434–439

Kimura B (1980) Phänomenologie des Zwischen – Zum Problem der Grundstörung der Schizophrenie. Z Klin Psychol Psychopathol Psychother 28: 34–42

Kimura B (1982) Zeit und Psychose. In: Janzarik W (ed) Psychopathologische Konzepte der Gegenwart. Enke, Stuttgart, pp 117–126

Kimura B (1995) Zwischen Mensch und Mensch. Strukturen japanischer Subjektivität. Wissenschaftliche Buchgesellschaft, Darmstadt

Kimura B (1996) Mögliche Beiträge des japanischen Denkens zur phänomenologischen Psychopathologie. In: Peters UH, Schifferdecker M, Krahl A (eds) 150 Jahre Psychiatrie. Martini, Cologne, pp 98–100

Kimura B (1997) Ecrits de psychopathologie phénomenologique. Presse Universitaires de France, Paris

Kisker KP (1964) Kernschizophrenie und Egopathien. Nervenarzt 35: 286–294

Kobayashi T (1989) Melancholie und Zeit. Stroemfeld, Basel Frankfurt am Main

*Kraus A (1977) Sozialverhalten und Psychose Manisch-Depressiver. Enke, Stuttgart

Kraus A (1980a) Bedeutung und Rezeption der Rollentheorie in der Psychiatrie. In: Peters UW (ed) Ergebnisse für die Medizin. 2. Psychiatrie. Kindler, Zürich, pp 125–148 (Die Psychologie des 20. Jahrhunderts, vol X)

Kraus A (1980b) Vom Umgang Manisch-Depressiver mit sozialen Normen. Med Mensch Gesellsch 5: 250–255

Kraus A (1982a) Identity and psychosis of the manic depressive. In: De Koning AJJ, Jenner FA (eds) Phenomenology and psychiatry. Academic, London, pp 201–216

Kraus A (1982b) Rollenkonzepte in der Psychiatrie. In: Janzarik W (ed) Psychopathologische Konzepte der Gegenwart. Enke, Stuttgart, pp 107–116

Kraus A (1985) Zeitlichkeit in der prämorbiden Verfassung Melancholischer. In: Bühler KE, Weiss H (eds) Kommunikation und Perspektivität. Königshausen and Neumann, Würzburg, pp 183–191

Kraus A (1988) Ambiguitätsintoleranz als Persönlichkeitsvariable und Strukturmerkmal der Krankheitsphänomene Manisch-Depressiver. In: Janzarik W (ed) Persönlichkeit und Psychose. Enke, Stuttgart, pp 140–149

Kraus A (1991a) Der melancholische Wahn in identitätstheoretischer Sicht. In: Blankenburg W (ed) Forum der Psychiatrie. Wahn und Perspektivität. Enke, Stuttgart, pp 68–80

Kraus A (1991b) Metholodogical problems with the classification of personality disorders: the significance of existential types. J Pers Disord 5/1: 82–92

Kraus A (1991c) Neuere psychopathologische Konzepte zur Persönlichkeit Manisch-Depressiver. In: Mundt C, Fiedler P, Lang H, Kraus A (eds) Depressionskonzepte heute: Psychopathologie oder Pathopsychologie? Springer, Berlin Heidelberg New York, pp 42–54

Kraus A (1991d) Phänomenologische und symptomatologisch-kriteriologische Diagnostik. Fundam Psychiatr 5: 102–109

Kraus A (1992) Lügenmotiv und Depersonalisation. In: Schmitt W, Hofmann W (eds) Phänomen – Struktur – Psychose. Roderer, Regensburg

Kraus A (1994a) Phenomenology of the technical delusion in schizophrenics. J Phenomenol Psychol 25/1: 51–69

Kraus A (1994b) Phenomenological and criteriological diagnosis. Different or complementary? In: Sadler JZ, Wiggins OP, Schwartz MA (eds) Philosophical perspectives on psychiatric diagnosis and classification. Johns Hopkins University Press, Baltimore

Kraus A (1996a) Das Problem der Freiheit im Zwang. In: Nissen G (ed) Zwangserkrankungen. Prävention und Therapie. Huber, Göttingen, pp 75–82

Kraus A (1996b) Die Bedeutung der Intuition für die psychiatrische Diagnostik und Klassifikation. In: Sass H (ed) Psychopathologische Methoden und psychiatrische Forschung. Fischer, Jena, pp 156–169

Kraus A (1996c) Identitätsbildung Melancholischer und Hysterischer. In: Seidler GH (ed) Hysterie heute. Enke, Stuttgart, pp 103–110

*Kraus A (1996d) Role performance, identity structure and psychosis in melancholic and manic-depressive patients. In: Mundt C, Goldstein MJ, Hahlweg K, Fiedler P (eds) Interpersonal factors in the origin and course of affective disorders. Gaskell, London, pp 31–47

Kraus A (1997a) Existential and differential aspects of anxiety. In: Boer JA de (ed) Clinical management of anxiety. Dekker, New York, pp 63–78

Kraus A (1997b) Identitätstherapie Melancholischer. In: Mundt C, Linden M, Barnett W (eds) Psychotherapie in der Psychiatrie. Springer, Berlin Heidelberg New York, pp 111–115

Kraus A (1998a) Der Sinn der Manie in identitätstheoretischer Sicht. In: Csef H (ed) Sinnverlust und Sinnfindung in Gesundheit und Krankheit. Königshausen and Neumann, Würzburg

Kraus A (1998b) Das Problem der Freiheit im Zwang. In: Nissen G (ed) Zwangserkrankungen. Huber, Göttingen

Kruger D (1981) An introduction to phenomenological psychology. Duquesne University Press, Pittsburgh

Kuhn R (1963) Daseinsanalyse und Psychiatrie. In: Gruhle HW, Jung R, Mayer-Gross W, Müller M (eds) Psychiatrie der Gegenwart, vol I/2: Grundlagen und Methoden der Klinischen Psychiatrie. Springer, Berlin Göttingen Heidelberg, pp 853–902

Kunz H (1962) Die eine Welt und die Weisen des In-der-Welt-Seins. Psyche 16: 58–221, 378–464, 544–560, 705–720

Lang H (1973) Die Sprache und das Unbewußte. Suhrkamp, Frankfurt am Main

Lang H (1990) Beziehung und Gespräch als psychotherapeutische Wirkfaktoren. In: Lang H (ed) Wirkfaktoren der Psychotherapie. Springer, Berlin Heidelberg New York, pp 36–48

Lang H (1997) Hat die Hermeneutik noch eine Chance? In: Mundt C, Linden M, Barnett W (eds) Psychotherapie in der Psychiatrie. Springer, Berlin Heidelberg New York, pp 33–48

Lanteri-Laura G (1963) La psychiatrie phénoménologique. PUF, Paris

Lanteri-Laura G (1982) Phenomenology and a critique of the foundations of psychiatry. In: De Koning AJJ, Jenner FA (eds) Phenomenology and psychiatry. Academic, London, pp 51–62

Lopes RG (1996) A escolha de si-proprío. II. Encontro de antropologia fenomenógica e existencial. Hospital do Conde de Ferreira, Porto

Lopez-Ibor J (1982) Delusional perception and delusional mood. A phenomenological and existential analysis. In: De Koning AJJ, Jenner FA (eds) Phenomenology and psychiatry. Academic, London, pp 135–154

Merleau-Ponty M (1966) Phänomenologie der Wahrnehmung. de Gruyter, Berlin

Minkowski E (1971) Die gelebte Zeit. 1. Über den zeitlichen Aspekt des Lebens. Müller, Salzburg

Minkowski E (1972) Die gelebte Zeit. 2. Über den zeitlichen Aspekt psychopathologischer Phänomene. Müller, Salzburg

Mishara AL (1994) A phenomenological critique of commonsensical assumption in DSM III-R: the avoidance of the patient's subjectivity. In: Sadler JZ, Wiggins OP, Schwatz MA (eds) Philosophical perspectives on psychiatric diagnostic classification. John Hopkins University Press, Baltimore, pp 148–162

Mooij A (1995) Towards an anthropological psychiatry. Theor Med 16: 73–91

Moss D (1981) Phenomenology and neuropsychology: two approaches to consciousness. In: Valle R, Eckartsberg R (eds) The metaphors of consciousness. Plenum, New York, pp 153–166

Müller-Suur H (1958) Die schizophrenen Symptome und der Eindruck des Schizophrenen. Fortschr Neurol Psychiatr 26: 140–150

Mundt C (1984) Der Begriff der Intentionalität und die Defizienzlehre von den Schizophrenien. Nervenarzt 55: 582–588

Mundt C (1985) Das Apathiesyndrom der Schizophrenen. Springer, Berlin Heidelberg New York

Mundt C (1988) Zur intentionalen Struktur einer schizophasischen Selbstdarstellung. In: Spitzer M, Uehlein FA, Oepen G (eds) Psychopathology and philosophy. Springer, Berlin Heidelberg New York, pp 85–95

Mundt C (1989) Psychopathologie heute. In: Kisker KP, Lauter H, Meyer JE, Müller C (eds) Psychiatrie der Gegenwart, 3rd edn, vol 9. Brennpunkte der Psychiatrie. Springer, Berlin Heidelberg New York, pp 148–180

Mundt C, Fiedler P, Pracht B, Rettig R (1985) InSka (Intentionalitätsskala) – ein neues psychopathometrisches Instrument zur quantitativen Erfassung der schizophrenen Residualsymptomatik. Nervenarzt 56: 146–149

Mundt C, Goldstein MJ, Hahlweg K, Fiedler P (1996) Interpersonal factors in the origin and course of affective disorders. Dorset, Dorchester

Mundt C, Richter P, Hees H van, Stumpf T (1998) Zeiterleben und Zeitschätzung depressiver Patienten. Nervenarzt 69: 38–45

Natanson M (1963) Philosophische Grundfragen der Psychiatrie. In: Gruhle HW, Jung R, Mayer-Gross W, Müller M (eds) Psychiatrie der Gegenwart, vol I/2: Grundlagen und Methoden der Klinischen Psychiatrie. Springer, Berlin Göttingen Heidelberg, pp 903–925

Naudin J (1997) Phenomenology and psychiatry. The voices and the thing. PUM, Toulouse

Naudin J, Azorin JM, Stanghellini G et al (1997) An international perspective on the history and philosophy of psychiatry: the present day influence of Jaspers and Husserl. Curr Opin Psychiatry 10: 390–394

Northoff G, Schwartz MA, Wiggins OP (1992) Psychosomatics, the lived body, and anthropological medicine: concerning a case of atopic dermatitis. In: Drew Leder (ed) The body in medical thought and practice. Kluwer, Dordrecht, pp 139–154

Parnas J, Bovet P (1991) Autism in schizophrenia revisited. Compr Psychiatry 32(1): 7–21

Passie T (1995) Phänomenologisch-anthropologische Psychiatrie und Psychologie. Pressler, Hürtgenwald

Pauleikhoff B (1979) Person und Zeit. Im Brennpunkt seelischer Störungen. Hüthig, Heidelberg

Plessner H (1928) Die Stufen des Organischen und der Mensch. de Gruyter, Berlin

Prouty G, Van Werde D, Pörtner M (1998) Prätherapie. Klett-Cotta, Stuttgart

Rogers CR (1973) Die klientenbezogene Gesprächspsychotherapie. Kindler, Munich

Rovaletti ML (ed) (1994) Psicologia y psiquiatria fenomenologica. Hipólito Irigoyen, Buenos Aires

Rümke HC (1958) Die klinische Differenzierung innerhalb der Gruppe der Schizophrenien. Nervenarzt 29: 49–53

Sartre JP (1943) L'être et le néant. Gallimard, Paris

Sass H (1992) Phenomenological aspects on "Zerfahrenheit" and incoherence. In: Spitzer M, Uehlein F, Schwartz MA, Mundt C (eds) Phenomenology, language and schizophrenia. Springer, Berlin Heidelberg New York, pp 147–159

Schäfer ML (1996) Philosophie und Psychiatrie. Die Gegebenheit des Fremdseelischen – Versuch einer epistemischen Analyse der Einfühlung. In: Peters UH, Schifferdecker M, Krahl A (eds) 150 Jahre Psychiatrie. Martini, Cologne, pp 315–322

Scheler M (1976) Die Stellung des Menschen im Kosmos. Nymphenburger, Munich

Schilder P (1968) Deskriptiv-psychologische Analyse der Depersonalisation. In: Meyer JE (ed) Depersonalisation. Wissenschaftliche Buchgesellschaft, Darmstadt, pp 46–141

Schmidt-Degenhard M (1986) Oneiroides Erleben bei intensivbehandelten panplegischen Polyradikulitis-Patienten. Nervenarzt 57: 712–718

Schmidt-Degenhard M (1991) Zum Problem der oneiroiden Erlebnisform. Fundam Psychiatr 5: 165–171

Schmidt-Degenhard M (1994) Das Imaginäre in den phantastischen Erlebniszusammenhängen. Überlegungen zu einem wenig beachteten Aspekt des Wahnproblems. Nervenarzt 65: 293–295

Schmidt-Degenhard M (1995) Wahn und Imagination. Fortschr Neurol Psychiatr 63: 350–357

Schmidt-Degenhard M (1997) Zur Standortbestimmung einer anthropologischen Psychiatrie. Fortschr Neurol Psychiatr 65: 473–480

Schmitz H (1987) Der vergessene Leib. Z Klin Psychol Psychopathol Psychother 35: 270–275

Schmoll D (1992) Vom symbiotischen Leib zum zerschnittenen Körper. Fundam Psychiatr 6: 180–189

Schmoll D, Koch T (1989) Leibgefühlsstörungen in der schizophrenen Psychose. Eine Kasuistik. Nervenarzt 60: 619–627

Schneider K (1950) Aufdeckung des Daseins durch die cyclothyme Depression. Nervenarzt 21: 192–194
Schütz A (1962–1966) Collected papers I-III. Nijhoff, The Hague
Schütz A, Luckmann T (1979) Strukturen der Lebenswelt, vol 1. Suhrkamp, Frankfurt am Main
Schwartz MA, Wiggins OP (1987) Diagnosis and ideal types: a contribution to psychiatric classification. Compr Psychiatry 28/4: 277–291
Schwartz MA, Wiggins OP (1992) The phenomenology of schizophrenic delusions. In: Spitzer M, Uehlein F, Schwartz MA, Mundt C (eds) Phenomenology, language and schizophrenia. Springer, Berlin Heidelberg New York, pp 305–318
Schwartz MA, Wiggins OP, Spitzer M (1997) Psychotic experience and disordered thinking: a reappraisal from new perspectives. J Nerv Ment Dis 185(3): 176–187
Silva F, Silva MC (1975) Die Theorie der Halluzination bei Merleau-Ponti. Z Klin Psychol Psychother 23: 106–137
*Spiegelberg H (1972) Phenomenology in psychology and psychiatry. Evanston, Illinois
Spitzer M (1994) The basis of psychiatric diagnosis. In: Sadler JZ, Wiggins OP, Schwarzt M (eds) Philosophical perspectives on psychiatric diagnostic classification. John Hopkins University Press, Baltimore, pp 163–177
Stanghellini G (1997) For an anthropology of vulnerability. Psychopathology 30: 1–11
Storch A (1965) Wege zur Welt und Existenz des Geisteskranken. Hippokrates, Stuttgart
Strasser S (1964) Phänomenologie und Erfahrungswissenschaft vom Menschen. de Gruyter, Berlin
Straus E (1963) Philosophische Grundfragen der Psychiatrie. In: Gruhle HW, Jung R, Mayer-Gross W, Müller M (eds) Psychiatrie der Gegenwart, vol I/2: Grundlagen und Methoden der Klinischen Psychiatrie. Springer, Berlin Göttingen Heidelberg, pp 926–1036
Straus E (1978a) Geschehnis und Erlebnis. Zugleich eine historiologische Deutung des psychischen Traumas und der Renten-Neurose. Springer, Berlin Heidelberg New York
Straus E (1978b) Vom Sinn der Sinne. Ein Beitrag zur Grundlegung der Psychologie. Springer, Berlin Heidelberg New York
Szilasi W (1959) Einführung in die Phänomenologie Edmund Husserls. Niemeyer, Tübingen
Szilasi W (1961) Die Erfahrungsgrundlage der Daseinsanalyse Binswangers. Schweiz Arch Neurol Psychiatr 67: 74–82
Tatossian A (1979) Phénoménologie des psychoses. Masson, Paris New York Barcelona Mailand
Tatossian A (1996) La phénoménologie: une épistémologie pour la psychiatrie. Epistemol Psychiatr 37: 177–196
Tellenbach H (1968) Geschmack und Atmosphäre. Müller, Salzburg
Tellenbach H (1983) Melancholie. Problemgeschichte, Endogenität, Typologie, Pathogenese, Klinik, 4th edn. Springer, Berlin Heidelberg New York
**Tellenbach H (1987) Psychiatrie als geistige Medizin. Verlag für angewandte Wissenschaften, Munich
van den Berg JH (1982) On hallucinating: critical-historical overview. In: De Koning AJJ, Jenner A (eds) Phenomenology and psychiatry. Academic, London, pp 97–110
Varela F (1996) Neurophenomenology. A methodological remedy for the hard problem. J Conscious Stud 4: 330–349
von Baeyer W (1978) Über die Bedeutung psychiatrischer Schlüsselwörter moderner Psychiatrie. In: Kraus A (ed) Leib, Geist, Geschichte. Brennpunkte anthropologischer Psychiatrie. Hüthig, Heidelberg, pp 29–44
von Gebsattel V (1954) Prolegomena einer medizinischen Anthropologie. Springer, Berlin Göttingen Heidelberg
von Uslar D (1989a) Sein und Deutung. 1. Grundfragen der Psychologie, 2nd edn. Hirzel, Stuttgart
von Uslar D (1989b) Sein und Deutung. 2. Das Bild des Menschen in der Psychologie. Hirzel, Stuttgart
von Uslar D (1991) Sein und Deutung. 3. Mensch und Sein. Hirzel, Stuttgart
von Weizsäcker V (1986) Gesammelte Schriften, vols 1–10. Suhrkamp, Frankfurt am Main
Waldenfels B (1992) Einführung in die Phänomenologie. Fink, Munich
Weitbrecht JH (1957) Zur Frage der Spezifität psychopathologischer Symptome. Fortschr Neurol Psychiatr 25: 41–56
Weitbrecht HJ (1971) Was heißt multikonditionale Betrachtungsweise bei den Schizophrenien? In: Huber G (ed) Ätiologie der Schizophrenien. Schattauer, Stuttgart, pp 181–200
Wiggins OP (1995) Natanson on phenomenology in psychiatry. In: Crowell SG (ed) The prism of the self. Kluwer, Dordrecht, pp 31–41
Wiggins OP, Schwartz MA (1997) Edmund Husserl's influence on Karl Jaspers' phenomenology. Philosophy, Psychiatry, and Psychology. Johns Hopkins University Press, Baltimore
Wiggins OP, Spitzer M (1997) Cognitive science. In: Embree L, Behnke EA, Carr D et al (eds) Encyclopedia of phenomenology. Kluwer, Dordrecht, pp 101–104
Wiggins OP, Schwartz MA, Spitzer M (1992) Phenomenological/descriptive psychiatry: the method of E. Husserl and K. Jaspers. In: Spitzer M, Uehlein F, Schwartz MA, Mundt C (eds) Phenomenology, language and schizophrenia. Springer, Berlin Heidelberg New York, pp 46–69
Wiggins OP, Schwartz MA (1994) The limits of psychiatric knowledge and the problem of classification. In: Salder JZ, Schwartz MA, Wiggins OP (eds) Philosophical perspectives on psychiatric diagnostic classification. Johns Hopkins University Press, Baltimore
Wyrsch J (1946) Über die Intuition bei der Erkennung des Schizophrenen. Schweiz Med Wochenschr 46: 1173–1176
Wyss D (1973) Beziehung und Gestalt. Entwurf einer anthropologischen Psychologie und Psychopathologie. Vandenhoeck and Ruprecht, Göttingen
Wyss D (1976) Die anthropologisch-existenzialontologische Psychologie und ihre Auswirkungen insbesondere auf die Psychiatrie und Psychotherapie. In: Balmer H (ed) Die europäische Tradition: Tendenzen, Schulen, Entwicklungslinien. Kindler, Zürich, pp 460–569 (Die Psychologie des 20. Jahrhunderts, vol I)
Wyss D (1990) Heilfaktoren in der anthropologisch-integrativen Psychotherapie. In: Lang H (ed) Wirkfaktoren der Psychotherapie. Springer, Berlin Heidelberg New York, pp 156–163
Zutt J (1963a) Auf dem Weg zu einer anthropologischen Psychiatrie. Springer, Berlin Göttingen Heidelberg
Zutt J (1963b) Über verstehende Anthropologie. In: Gruhle HW, Jung R, Mayer-Gross W, Müller M (ed) Psychiatrie der Gegenwart, vol I/2: Grundlagen und Methoden der Klinischen Psychiatrie. Springer, Berlin Göttingen Heidelberg, pp 763–852
Zutt J, Kulenkampff C (eds) (1958) Das paranoide Syndrom aus anthropologischer Sicht. Springer, Berlin Göttingen Heidelberg

CHAPTER

22

H. Kächele, A. Buchheim,
G. Schmücker, K.H. Brisch

Development, Attachment and Relationship: New Psychoanalytic Concepts

Translator: G. Schmücker

1 Introduction

In the last decade, the term "depth psychotherapy" has lost much of its fascination. This is reflected by its omission in the index of recent textbooks on psychoanalytic therapies (Thomä and Kächele 1987, 1991; Heigl-Evers et al. 1993). In therapeutic settings, the terms "psychodynamic" and "psychoanalytic psychotherapy" have won much ground, especially in the empirical world.

Three manuals published in 1984 (Luborsky 1984; Strupp and Binder 1984; Klerman et al. 1984) were all akin in different ways to the interpersonal and dynamic way of thinking. They demonstrated that it was obsolete to contrast psychodynamic procedures with procedures that were empirically supported.

The systematic pathology of conflict in Freud's paper (1917) characterised the scientific paradigm of psychoanalysis, which Freud summarised by stating: "We seek not merely to describe and to classify phenomena, but to understand them as signs of an interplay of forces in the mind" (Freud 1917, p. 67).

A significant assumption of psychoanalytic theory is the role which conflict plays during the lifespan of a person – starting from birth, manifest in interpersonal contact and personal well-being. If the role that conflict plays in the emergence of a psychic or psychosomatic illness is considered purely as intrapsychic and not also as interpersonal, the implications of the theory, as well as the technique, would be limited. Psychoanalysis and both the Jungian and Adlerian schools are based on the concept of development. Traditional psychoanalytic understanding of a symptom almost requires a search of its origin in the life-history of the person. This genetic point of view is not incompatible with Kurt Lewin's belief that only forces and conditions which are present in the here and now can induce an effect in the here and now. He would say that much of what is "presently seen" in the individual (in the here and now) can only be identified by the genetic discovery of what came before (Rapaport 1960).

The main tool with which the developmental dimension is captured is reconstruction. Freud based his psychoanalytic theory on the treatment and observation of adults. It was only incidental that he also observed young children. One of these observations is the well-known so-called reel of string game, which can be found in Freud's *Beyond the Pleasure Principle* (1920). Here, Freud recounts how his 1½-year-old nephew comes to terms with the absence of his mother by playing with a reel of string. Emde (1992) sees this contribution as being ahead of its time. Freud attributes the small child with the ability to actively reduce tension by repeating his experience of separation and reunion in the game. In contrast, in traditional psychoanalytically based developmental psychology, infants were seen as passive, undifferentiated beings who were governed by their instincts (Dornes 1993).

In the last 30 years, there has been a great change in the understanding of the various developmental processes that take place in early childhood. Empirically based research into the early mother–child relationship was initiated by Spitz. As early as 1935, he was able to observe hundreds of infants growing up in orphanages and described the "hospitalism" he saw, which he attributed to emotional deprivation. Research into the evolution of the mother–child relationship in the first year of life, as reported by Spitz (1965), provided the psychoanalytic world of the infant or child with new ideas, a world which until that point had largely been constructed or reconstructed. Mahler et al. (1975) followed the tradition of Spitz and developed his work further in her groundbreaking monograph entitled *The Psychological Birth of the Human Infant.*

In addition to observing psychotic children and working with them (Mahler 1958, 1969), she focused on the effect that being separated from the main caregiver had on children. Based on her work, Mahler conceptualised a developmental theory which was mainly pathomorphic. Bowlby (1969) was the first psychoanalyst of his generation to use ethological terms to describe the infant's biologically predisposed availability of attachment to a main caregiver. He saw relatedness in early childhood as a primary and independent developmental goal that is not subservient to a physiological need such as hunger (see also Künzler 1969). These theoretical developments perceive the infant more from an interactional point of view and also focus on relationship aspects. The concept of dyadic interplay replaces the drive-conflict model. A dynamic and conflict-oriented psychology which describes psychic processes has been extended by more current concepts of development, attachment and relationship. The new theories of child development have had two main effects. They have promoted an integration of ethology and of theories of communication and action and they have also had a marked impact on psychoanalysis and other psychodynamic schools (Stern 1995; Dornes 1997; Krause 1998). In the following chapter, these three concepts will be explored further in order to demonstrate the ways in which each has made a contribution to new psychoanalytic thought.

2 Development

The diversity of methods used to observe infants has contributed to a change in the perception of

children's development and hence led to an increased base of knowledge.

The function attributed by psychoanalysts to direct observation was initially to correct the retrospective information obtained in therapeutic analyses. Today, psychoanalysts are required to also use empirical and experimental information gained from direct observation and to reflect on the consequences for clinical and retrospectively obtained knowledge. Daniel Stern (1985) captured this element of tension by talking about the "observed infant" and the "clinically reconstructed infant". His work criticises traditional psychoanalytic concepts such as "normal autism" (Mahler 1958), the process of splitting into "good" and "bad" (Klein et al. 1952; Kernberg, 1968), orality, undifferentiatedness or "normal symbiosis" (Mahler 1969) and manages to demonstrate the absurdity of some of the clinical constructs. Lichtenberg (1983) sums up the infant's world in the first year of life: Firstly, infants are capable of much more than was previously thought possible and secondly, they are not as able in some matters as had previously been assumed. They can master complex developmental steps when their activities are guided and enhanced by affect. However, the infant does not have or even need the ability to imagine symbolically.

Clinical reconstruction primarily focused on the revelation of subjective experience, but the interest in direct observation focuses on identifying what really happens in childhood, in so far as it can be observed from the outside. However, therapists are always in danger of reifying experiences recounted by patients and of taking them as the real picture of the event. Freud took great pains with the retrospective attribution of meaning when it concerned the conception of psychological causality:

> I admit that this is the most delicate question in the whole domain of psycho-analysis. I did not require the contributions of Adler or Jung to induce me to consider the matter with a critical eye, and to bear in mind the possibility that what analysis puts forward as being forgotten experience of childhood (and of an improbably early childhood) may on the contrary be based upon fantasies created on occasions occurring late in life (Freud 1918, p. 137).

However, the path "inexorably led analysts to trace the etiological conditions of psychic and psychosomatic illnesses back to the first hour and even earlier" (Thomä and Kächele 1991, p. 100). The analysis of pathological development provided the basis for an uncritical view of normal psychological development (Peterfreund 1978). This peculiarity of psychoanalytic theory formation so far is reflected by the fact that infants' characteristics are described as deficient modi of the adult world. In addition to this so-called adultomorphism, pathomorphism can also be found and is equally abundant. Here, the infant is described using pathological categories. This is based on the assumption that formative processes that constitute development can be derived from the observation of pathological states. The key to discovering early phases of psychic life lies in the data concerned with fixation and regression (e.g. Tustin 1994).

In the following sections, some concepts which have modified the psychoanalytic view of development from infancy to childhood will be described further.

2.1 Complex Abilities of the Infant

Freud's assumption that the tension/relief principle representing the pleasure/unpleasure principle is the basis of developmental processes can no longer be maintained.

Since the 1980s, developmental psychologists have emphasised that a newborn infant is provided with basic activity which has a tendency to stimulate the organism to increase psychological complexity (Brazelton et al. 1974; Stern 1974; Emde 1992). Newborns have a substantial repertoire of possible behaviours that prepare them for interactive relations in a care-taking environment.

The drive/discharge model viewed development through the eyes of the entropy model. However, today's developmental psychobiologists take into account the fact that the neurobiologically determined complexity due to 10^{10} neurones with thousands of interconnections leads to uncertainty and a limited ability to predict behaviour (Spitzer 1997). This degree of complexity guarantees individuality and assures self-determination (Emde 1992). Complexity grows in the course of development. Humans are attributed with the ability to socialise themselves into the animate and inanimate world. Activity generated endogenously represents a fundamental principle and has taken the place of the drive/discharge hypothesis. Similarly, we should be critical of the perception that the infant is born as a psychological "blank slate" and is only formed by parental socialisation. Schaffer (1982) states that, from the beginning, a baby's behaviour shows organisation and order. It is only due to our inadequate recording methods that the small child seems as if it is in a state of bubbling confusion. The discovery of this complexity is attributable to the detailed work that has been done on different types of behaviours, each demonstrating its own intricacy.

Finally, the revolution seen in research into infancy was based on methodological innovations (Stern

1985). Today, we ask what reactions an infant shows that can be taken as an answer to the questions put by the researcher. It has been demonstrated that, from the beginning, infants are able to show a preference, habituate or let themselves be taken by surprise. These signs, which point to the complex abilities of infants, have led to the development of research paradigms.

Research has shown that infants have excellent olfactory abilities. They are able to distinguish their mother from others by their scent from the eighth day of life (Brazelton and Cramer 1991).

In addition, the visual abilities of infants have long been underestimated. Experiments with intrauterine ultrasound have shown that the fetus turns towards a moderate source of light, but turns away if this source is intense (Brazelton 1981).

The postnatal auditory abilities of infants are also remarkable. Even during the first hours after birth, newborns are more likely to turn their head and eyes towards a tone, noise or a voice than away from it.

Infants also have a capacity for intersensory co-ordination or cross-modal perception. Hence, an infant is able to co-ordinate sensory perceptions with different sensory canals (seeing, hearing and touching). This incredibly complex ability has been investigated repeatedly (e.g. Meltzoff and Borton 1979). It can be concluded that infants perceive objects as a gestalt and do not live in a world of separate sensations (Dornes 1993). Psychoanalytically based developmental psychology assumes that, at the beginning, self and object perception is fragmented. However, the findings concerning cross-modal competence seem to show that from early on the separate aspects of sensory information are put in relation to each other. It is not the case that separate part objects are perceived next to each other.

Newborns are organised in such a way that, after birth, they can start interacting in a complex way with the animate and inanimate world. The regulation inherent in these interactions imprints the patterns of sleeping and waking cycles, feeding and social exchange. This regulation is established in the first 2 months of life. It is manifest in various phases replacing one another, such as alert attention, calm wakefulness, arousal, crying, rapid eye movement (REM) sleep, non-REM sleep and the infant's search for different stimuli (Greenspan 1989). The concept of self-regulation acts as a basic motive of development and is related to the organism's ability to smooth out deficits which arise out of challenges or disturbances (Clarke and Clarke 1976). Another strong motive in the developmental plan of the small child is the innate willingness to adapt socially. Surprising to some, research in developmental psychology has shown the great degree to which taking part in social interactions is pre-programmed. Many of these abilities are already present at birth and include a preference for eye contact or a state-dependent susceptibility for the activating and pacifying effect induced by being held and touched by one's mother.

Social pre-adaptation is manifest in a multitude of communicative channels. According to Papoušek (1981), social pre-adaptation is based on an ability to discover and master contingencies in stimuli. This would imply a biological basis. In addition to children's behaviour as described above, parental response to children's communicative offers should also be mentioned, which Papoušek and Papoušek (1983) defined as "intuitive parenting." Such parental behaviour seems to be characteristic of the species, not conscious and not a product of individual experience. Many microscopic interactions between mother and child may be subsumed under the terms synchronicity and reciprocity (Jörg et al. 1994).

The psychoanalytic pleasure/unpleasure principle has lost its economic quality. Today, it is conceived of as affective monitoring. This is a basic motivational system that evaluates affective experiences according to the quality of "pleasurable" or "unpleasurable" (Emde 1981). Infants do not divide the world into two, but instead abstract daily experiences into different levels of pleasurable and unpleasurable experiences. This will eventually lead to the formation of schemata as conceived by Piaget, where cognitive elements play as important a role as emotional quality. This principle will govern both the mother's and the child's actions. At the age of 3 months, a stable emotional organisation can already be seen that includes three dimensions, namely, hedonic quality, activation and internal/external orientation. Early coherent emotional experience forms the affective core of self (Emde 1983). This emphasises the significance attributed to the emotional attention provided by the caretaker in early childhood.

In this process of emotional exchange, attunement is attributed with a special meaning (Stern 1985). A series of dialogic sequences in different communicative channels provide this exchange, which is seen in the ninth month. Stern assumes that around this time the subjective self is formed and joint affective experience becomes prominent.

A mother interacts with an infant who is motorically active (e.g. kicking the legs rhythmically) by responding in the verbal channel (e.g. with "lalala"), hence she does not change the rhythm but adds variation by verbalising.

Kohut's concept of mirroring (Kohut and Wolf 1978) comes closest to the process of attunement. In the clinical world the use of the term encompasses other different affective processes. Empathy is more closely linked to cognitive processes than the

unconscious process of attunement (Moser and von Zeppelin 1991; Basch 1983). In all research approaches addressing early mother–child interaction, the processes of reciprocity, intersubjectivity, intentionality and a willingness to relate are emphasised. These represent signs of the early processes of communication.

From the beginning, children are equipped for social interaction. They partake in reciprocal exchange with the caretaker. We cannot see our fellow man as static targets of drive. From this point of view, terms such as object relationship are not suitable due to their implications (Emde 1983).

2.2 Implications for the Psychoanalytic Understanding of Development

These developments caused the fundamental position of the drive theory in classical psychoanalysis to be given up. The psychoanalytic object psychology of Balint and Winnicott, for example, prepared the way for this critique. The libido theory did not account for the process of affective reciprocity. Freud regarded the libidinous object from the point of view of the child (and his or her unconscious wishes) and not from the view of the reciprocal relationship between mother and child. This tradition was so deeply embedded that Kohut (1971) derived his "self-objects" from the hypothetical view and experience of the infant. Self-psychologists assume a significant phase of undifferentiation between the self and others (Stern 1985).

From today's point of view, the inner object is not seen as an isolated object, but rather as a memory framed by a context of activity. From birth, the object representations take place in a multiple context of acts of varying quality. By repeated communicative, acts unconscious schemata are created and these can become very stable.

Stern (1985) refers to this active process as the representations of interactions that have been generalised (RIG). He assumes that the infant divides the flow of an interaction into episodes (e.g. feeding), and from repeated similarities (invariances) a prototype or schema is built and generalised. This schema guides expectations and behaviour in the interactional sequences. More recently, Stern has extended his theory of the representation of interaction by starting from the subjective perspective of the infant.

Stern (1996) conceived a model of infant representations not so much concerned with behaviours, but much more with feelings of the infants. He refers to the generalised representations of interactions as "schemas-of-being-with". However, in each interactive episode, different feelings are experienced (e.g. subjective feeling of hunger, negative affect, tactile contact with the mother, kicking). These feelings are stored in the form of protonarrative envelopes (series of events such as in a narrative). In summary, the infant experiences the world of interactive events as if they had the structure of a story.

The main difference between Stern's concepts and those of psychoanalysis are that, in traditional psychoanalysis, unconscious fantasies have been attributed to the infant which originate in his or her instincts. Stern, on the other hand, assumes that the infant develops his or her representations from real interactions. These are not motivated by tension or a lack of it, but are created and processed continually (Dornes 1997).

It may be asked to what extent psychoanalytic clinical thought has changed theoretically and practically due to the rich data set on early parent–child interaction. It is feasible to maintain that these early processes are interesting, but do not have any significant influence on the complex process of symptom formation observed in neuroses and other disorders, as the psychological organisation of adults is fundamentally different. With the development of language and accompanying symbolic processes, early experiences would be transformed in such a way that the familiar ground of the psychotherapist would not be disturbed. Parallels can be drawn here with a phenomenon which is also familiar from developmental psychobiology. The developmental context of a child permanently changes and transforms along a developmental pathway, and later behaviour cannot be predicted from earlier events that have taken place at a predetermined sensitive phase.

The consequence would be that infants should not be seen as "pseudo-adults" by ascribing to them the ability to symbolise in the first year of life (Lichtenberg 1983). However, the Kleinian notion of fantasy and/or the theory of splitting as an early form of defence would imply this.

These new insights make Kernberg's (1984) concept of splitting questionable as an explanatory concept of early ontogeny. If infants split good and bad, they would be forced into "double bookkeeping" (Stern 1985), but this does not correspond to the infant's abilities at this developmental stage. Clinical use of the concept of splitting as a description of pathological states has also been affected by these research developments (Reich 1995). These states require a certain degree of symbolisation, such as a labelling of memories and cognitive reorganisation. Hence the process of splitting is most probably found at a later developmental phase when symbolic transformation of experience is possible.

Similarly, it is difficult to justify the concept of an undifferentiated phase of the id and ego in which the inner world of the infant consists of separate elements. This would equally be the case for Mahler's concepts of "normal autism" and "symbiosis" (Stern 1985;

Lichtenberg 1983). Even though Mahler does not conceive of symbiosis as biological, research into the abilities of infants indicate that the term "symbiotic merging" as an appropriate category of early experience needs to be reviewed. Terms such as "primary narcissism" equally become set terms of what are most probably outdated theoretical positions (Eagle 1984).

Psychological research emphasises the characteristic of openness and also limited predictability from one developmental phase to the next as being important for healthy development. This is especially the case if we focus on individual behaviours. With the term "transference", Freud tried to capture pathological development which is formed in accordance with psychodynamic theory through established motivational and relationship-regulating structures (Thomä and Kächele 1987).

Instead of aetiological assumptions concerning which phase of development may be the origin of a specific disorder, it is possible to establish that the different developmental steps are linear, even though they continue to interact in parallel as functional contexts. Erikson's epigenetic model (1950) is relativised by Stern's concept of four senses of the self. Disorders may develop at any time of life in one of the four senses of self: the emergent self, the core self, the subjective self and the verbal self. The simplified link of severe disorders with early development is undone, something which has substantial therapeutic implications (Stern 1985). Instead of seeing the development of disorders fixed at critical phases, the whole chain of interacting influences should be taken into account. The focus should not only be on the first or the last link in the chain. The formation of psychopathology can thus only be understood as an accumulation of pathological patterns of interaction (Blatt 1990).

The value of this object psychology approach, which is also developmentally oriented, can be seen in the reanalysis of the Menniger and the National Institute of Mental Health (NIMH) depression project (Blatt 1992; Blatt et al. 1995). Here, a differential effect of therapies with regard to specific developmental and psychoanalytic configurations (anaclitic vs. introjective) could be demonstrated.

2.3 Mother–Child Psychotherapy: A New Development in the Therapeutic World

The developmental theory of Sander assumes that, in the first 3 years of life, the mother–child system represents an interconnection of mutual regulation and self-regulation (Sander 1985). The negotiation of self-perception, self-determination and initiative are the core problems which an infant has to tackle with his or her caretaker. These configurations become the enduring adaptive strategies of an individual (Quinton and Rutter 1988), the implications of which may be found in the transference patterns (Luborsky and Crits-Christoph 1990). This process of fitting and experience of what is possible in a relationship becomes clinically relevant when the caregiver introduces neurotic aspects into early interactions, leading to maladaptive interaction patterns.

Cramer (1991) describes the first attempts at mother–baby psychotherapy, where such disturbances may be therapeutically resolved. In his psychoanalytically oriented therapy, Cramer assumes that the mother has her own unresolved conflicts, which become reactivated through the birth and the child's behaviour. The unresolved conflicts are then projected onto the child and distort and burden the interaction.

In his latest book Stern (1995) compares the different psychotherapeutic approaches in mother–child psychotherapy and critically discusses their theoretical foundation. He develops his own theoretical model of a so-called motherhood constellation, which should be the starting point for an all-encompassing view of mother–child psychotherapy.

He defines this constellation as a mother's fundamental psychological organisation, which should be acknowledged in the therapeutic relationship. The central themes are as follows: Can this mother provide the nourishment and care this baby needs to survive? Can she have an emotional relationship, build up a system of support and help the baby find his or her own identity? A new trilogy develops: the mother's mother, the mother herself and the baby. This motherhood trilogy is central in every therapeutic intervention. Stern pleads for a positive and supportive transference in mother–child psychotherapies to counteract any additional hurt and insecurity.

An all-encompassing concept of preventive intervention was conceptualised by us for parents of extremely premature infants. This intervention programme consists of four components (Brisch et al. 1996), which were conceived to help parents cope better with the accompanying insecurity and anxiety brought about by this situation. In addition, the intervention programme was designed to help establish a positive parent–child relationship.

Immediately after the birth, parents were offered individual attachment-focused therapy to enable reflection of reactivated experiences of loss and separation, as these could interfere in the establishment of close contact with the premature baby. The focus of a continual parent group was to provide the possibility for emotional exchange and support with other parents. Once the child had been discharged from hospital, a home visit from a neonatal nurse was offered to provide information of a medical nature.

The final component of the intervention programme consists of a video training to improve a parent's sensitive handling of the child.

According to attachment theory, a sensitive parent–child interaction allows secure development on the part of the child. This significant direction of research will be explored in the following section.

3 Attachment

John Bowlby, a psychiatrist and psychoanalyst, formulated his attachment theory in the 1960s (Bowlby 1969, 1973). He turned away from the traditional psychoanalytic view of fantasy life during childhood and focused on the implication of actual real events such as separation and loss on the emotional development of a child. Hence Bowlby's theory has long been the topic of heated debate among psychoanalysts (Bretherton and Waters 1985; Bretherton 1995).

The independent motivational aspect of attachment has now been widely accepted as safeguarding the establishment of social relationships independent of hunger and sexuality. Freud's view that social relationships are primarily formed because of the need for nourishment would not find support by the researchers working in the field of attachment (Grossmann et al. 1989). The theory of attachment has taken up aspects of psychoanalytic theory and also developed some aspects further (Diamond and Blatt 1994). Contrary to the psychoanalytic theory of development, attachment theory has also managed to establish itself empirically. Observation of important aspects of dyadic interactions (also prospective) are possible and are systematically described and operationalised.

3.1 Basic Concepts and Methods of Attachment Theory

Attachment theory sees the desire for close emotional relationships as specifically human. This desire, which is already present in the newborn, remains present until old age and is a basic element with a function for survival. In infancy and childhood, attachment to one's parents ensures shelter and care. Analogously, the task of the parents is to provide sensitive care for their child. These two systems are in delicate balance and develop in a specific sequence.

In the middle of the first year of life, using his or her attachment behaviour and reactions of the attachment figures, the child develops an inner representation of attachment, a so-called inner working model (Bowlby 1969, 1973, 1979). The child's daily interactions with his or her attachment figures are the basis of the inner working models. The experiences gained from the interaction of the attachment figures are integrated into a whole. For the child, this model is a basic organisation of expectations and corresponding feelings which accompany him or her in different situations. This basic organisation remains the same even if the child changes his or her behaviour in different situations (Fremmer-Bombik 1995).

The inner working model is seen as a construction that becomes increasingly complex over a life time; however, the focus is exclusively on relationships to attachment figures. To assume an internalisation of relationship experiences shows a parallel to other psychoanalytic theories, especially the object-relations theory. The difference is one of specificity, as attachment theory only concerns itself with experiences of attachment to specific attachment figures and operationalises these.

The term "working model" corresponds to the "basic assumptions" of Beck et al. (1979) and to the "representations of interactions that have been generalised (RIGs)" of Stern (1985), which are now referred to as "schemas-of-being-with" (Stern 1995, 1996), in addition to the "role relationship models" and "self-other schemas" of Horowitz (1991).

Attachment quality between mother and child finds its expression in the second half of the first year of life. The child experiences whether the attachment figure responds sensitively to his or her signals and needs and whether availability is shown. A sensitive mother is alert and notices her child's signals; she interprets them correctly and reacts promptly and appropriately to the needs of her child (Ainsworth et al. 1974). This dialogue enables the child to have inner security, to show flexibility in situations of conflict and to have confidence in his or her continually expanding competence, emotional reactivity, sensitivity and assertiveness (e.g. Grossmann et al. 1988).

The relationship between parental sensitivity and a secure attachment classification is moderate ($r = 0.32$; van IJzendoorn 1995). However, the relationship between parents' ability to reflect (see below) upon their own childhood and the development of secure attachment by the child is much stronger ($r = 0.47$, $\kappa = 0.49$; Main et al. 1985; Grossmann et al. 1989; Fonagy et al. 1991a). A so-called transmission gap exists, which is the "gap" of knowledge concerning the interactive ways in which attachment experiences are passed on (van IJzendoorn 1995).

Ainsworth and Witting (1969) developed the so-called strange situation, where the attachment quality of the child to the mother was operationalised. This standardised laboratory situation is made up of eight episodes, each lasting 3 min. The behaviour of 12- to 18-month-old children is observed while in contact with a "stranger", after two short separations from

their attachment figure and reunions with this person. The episodes of separation were conceptualised to activate the attachment system and to trigger attachment behaviour (e.g. clinging, seeking proximity, crying). Behaviours of attachment and exploration are the central paradigms of attachment research, which ideally should be balanced. How children react in the reunion episodes enables a reliable assessment of the quality of their interactional life so far (for a review, see Buchheim et al. 1998).

Four patterns of attachment have been identified to date, and data from other cultures have confirmed these:

1. Securely attached (B pattern). These are children who have made reliable attachment experiences. In the separation episode, they can show their distress openly; the attachment figure is usually able to comfort them easily, and the children are able to return to play and exploration.
2. Insecure-avoidant (A pattern). These children have usually had predictable but rejecting experiences with their attachment figure, especially concerning the expression of negative feelings. They avoid this rejection by concentrating on play in a seemingly unaffected manner. When the attachment figure leaves or returns, they show no distress or need of proximity. Spangler and Schieche (1995) measured the children's cortisol level and found these children to be very stressed, thus indicating maladaptive avoidance strategies.
3. Insecure-ambivalent (C pattern). These children have had unpredictable experiences with their attachment figure, who is sometimes very sensitive in responding to their needs and sometimes not sensitive at all. Upon separation, insecure-ambivalent children cry a lot and are characteristically difficult to comfort. They show anger or passive despair as an expression of their ambivalence. Their attention is mostly focused on their attachment behaviour.
4. Disorganised/disoriented (D pattern). This group was described as such in the 1980s (Main and Solomon 1986) and is assessed separately from the other classifications. After separation, the children have not developed an organised coping strategy. They are neither able to approach the attachment figure (as B or C does), nor can they distract themselves (avoid as A does). Upon reunion, they show unintegrated behaviours, such as stereotypic movements after seeking proximity, phases of rigidity, so-called "freezing" and an expression of fear towards their parent. This disorganised behaviour is especially seen in abused children (Carlson et al 1989), neglected children (Lyons-Ruth et al. 1993) or children whose parents have not worked through their own process of grieving (Main and Hesse 1990). Internationally, the distribution of attachment patterns has been shown to be 66% for the B pattern (secure), 20% for the A pattern (avoidant) and 12% for the C pattern (ambivalent) (e.g. Baltimore study, Ainsworth et al. 1978). In non-clinical samples, the proportion of the D pattern can be assumed to range from 15% to 35% (Main 1995). In clinical populations of abused children, the frequency of the D pattern is 80% (Main 1995).

So far, the results confirm a stability of the attachment quality of children from 1–10 years of age (Grossmann and Grossmann 1991). The prognostic value of attachment experiences or attachment deficits is high for the later social development of the child, his or her self-image, self-esteem, social competence and cognitive ability (Grossmann and Grossmann 1991).

An early secure attachment relationship can be a protective factor in the development of psychological disorders (Bowlby 1988, 1995). However, early avoidant or ambivalent attachment relationships may start negative mechanisms, which can establish themselves in inappropriate psychic structures (Fonagy 1993). At the same time, early experiences may be changed by subsequent stress. A straightforward stability of attachment security cannot be assumed. Bowlby never took a deterministic view of early attachment experiences. He saw the developmental path of attachment organisation as flexible and would not subscribe to the view that once an attachment relationship was secure it would always be secure (Bowlby 1988, 1995). Extreme emotional experience due to separation or loss may change attachment quality and may lead to a change in self-esteem (Zimmermann et al. 1995). Similarly, it may be assumed that the inner working model of an early insecure attachment experience may be reorganised. This might be the case after a new positive experience with a partner or psychoanalytic therapy (Fonagy et al. 1995).

The systematic description of childhood relationship experiences enables an attachment theory with a life-cycle perspective to be constructed (e.g. Ainsworth and Bowlby 1991). Since early relationship experiences seem to influence adult relationships, there has been a growing interest in the attachment representations of adults.[1] An essential step in this development was the so-called "move to the level of representation", which was taken by Main et al. (1985). She assessed the attachment representations of 6-year-olds (Strage and

[1] A critical review of alternative methods to assess adults' attachment representations can be found in Crowell and Treboux (1995) and Buchheim et al. (1998).

Main 1985; Main and Cassidy 1988; Grossmann and Grossmann 1991) and of adults (Main et al. 1985) by using language. The Adult Attachment Interview (AAI) (George et al., unpublished) was developed to capture the attachment representations of adults. The themes touched upon in the interview correspond to Bowlby's trilogy of relationship, separation and loss.

The semi-structured AAI assesses current attachment experiences with respect to the past and the present using 18 questions. The scoring method focuses not so much on the content of the narrative, but on the ways and means of linguistic organisation. The coherence of the discourse in the linguistic sense is essential (see Grice 1975).

Parental attachment representations have also been classified into four groups (Main et al. 1985; Main 1991; Grossmann et al. 1988; Ainsworth and Eichberg 1991) and correspond conceptually and empirically to the attachment qualities of children:

1. Autonomous secure adults recount their childhood memories/experiences in an open and coherent fashion. They provide positive and negative examples, can reflect on them and are able to integrate them into an appreciating whole.
2. Dismissing adults provide incomplete, incoherent examples and often have blanks, especially when concrete examples are required. Attachment figures are either idealised or devalued in order to avoid painful memories.
3 Preoccuppied adults relate in an angry and never-ending way the conflicts they experienced with their caregiver. They are entangled and give the impression that their experiences were very recent. These adults characteristically oscillate between positive and negative value judgements without being aware of the contradiction.
4. Specific passages in the AAI may show unresolved grief. These passages are rated separately and relate to traumatic events (loss or abuse) which have not been resolved. The linguistic presentation seems disorganised (confusing senses of time and space, long silences, unusual details), incoherent and at times irrational.

Longitudinal results show the clear relationship between a mother's attachment representations and the observable attachment quality of her child (Main 1991; Fonagy et al. 1991a). Hence the transgenerational aspect of attachment is a confirmed result.

The statistical proof of the transmission of attachment experiences is satisfactory. Agreement between the category of attachment representation in adults and the quality of attachment of their children was examined in 18 studies (854 dyads) (van IJzendoorn 1995). The agreement (secure vs. insecure) is 75% ($\kappa = 0.49$) (Main 1995). In a study by Fonagy et al. (1991a), the predictive validity of the AAI is clearly shown. The attachment quality of the child could be predicted from the attachment representations obtained during an interview conducted while the mothers ($n = 100$) were pregnant ($\kappa = 0.44$; 69%). These results have been replicated in several studies (Benoit and Parker 1994; Ward and Carlson 1995).

3.2 Clinical Relevance of Adult Attachment Research

It has also proved useful to apply the AAI clinically. Van IJzendoorn and Bakermans-Kraneburg (1996) demonstrated a higher distribution of insecure representations of attachment in clinical than in non-clinical samples. Hence a distinction of clinical and non-clinical groups could be made using the AAI, even though a differentiated correlation between insecure attachment and psychopathology is not yet possible. The coding system of the AAI takes into account the extent to which attachment-relevant information has been processed. The way the information is presented, especially regarding linguistic aspects, is central to the AAI, and the coding system therefore also incorporates defensive processes. This semi-structured interview can "surprise" the unconscious, and there have been discussions on whether to include the systematic application of this instrument in clinical training.

Psychoanalytic research into borderline personality has already profited from the results of attachment research (Clarkin et al. 1992). A lack of sympathetic understanding, being untouched by the feelings of others and an inability to have relationships are all phenomenological characteristics of dissociated and narcissistic personality disorder and of a borderline personality type. Among other factors, a pathological component of this disorder is discussed, where from the perspective of object psychology disorders of "containments" (Bion 1962) are assumed.

People with borderline disorders have a greater preponderance to have had unresolved traumatic experiences and have an attachment classification of "entangled" when compared to a control group (Patrick et al. 1994). They also seem to have an inadequate meta-cognitive ability for self-reflection (self-reflective function; Fonagy et al 1991b, 1995). Successful psychoanalytic therapy is able to improve the ability to self-reflect, i.e. to be able to identify with another person's state of mind. Patients are able to obtain an adequate representation of themselves and others through the continual and repeated evaluation of the therapist's and patient's consciousness in transference (Fonagy et al. 1995). The following section will deal with the core concept of transference

from an interpersonal perspective, also presenting methods of operationalisation.

4 Interpersonal Relationship Pattern

From the beginning of the 1970s, the shibboleth of psychoanalysis, namely transference, has been investigated theoretically and empirically in a differentiated manner. A multitude of different methods have been developed to capture this interactive process of regulation.

The methodology of structural analysis of social behaviour (SASB; Benjamin 1993, Tress et al. 1990), in which each speech act of a therapeutic interaction becomes the object of analysis, is distinct from approaches that draw systematic information from verbal exchange about subjectively relevant structures. Examples of the latter approaches are the Central Conflict of Relationship Theme (CCRT; e.g. Luborsky and Crits-Christoph 1990), the methods of cyclic maladaptive pattern (Strupp and Binder 1984), plan diagnosis (Weiss and Sampson 1986), Dahl's FRAME method (Dahl 1988) and the role-relationship-conflict constellation (Horowitz 1991). The observation focuses on identifying functional and dysfunctional, observable or experienced interactions which may be influenced therapeutically. Microanalytic investigations of individual speech acts (SASB) are found at one end of the spectrum, and the global instruments which capture complex psychological processes of conflict (e.g., plan diagnosis, FRAME) at the other. The description of individual components of the interaction (CCRT) to different, partly parallel intrapsychic and interpersonal schemata may be seen as situated somewhere in between the two.

These procedures of interactional analyses with which interpersonal relationship patterns were established correspond to biographical methods which have experienced a revival over the last decade (Jüttemann and Thomae 1987). The SASB is based on the interpersonal circumplex model. This enables the analysis of connections between interpersonal and intrapsychic processes through the introduction of three levels of foci: the transitive (active: cause something to happen in others), intransitive (reactive: tell others about oneself) and introjective (focused on the self) (Benjamin 1974). The systematic application of the SASB model for psychiatric diagnosis and classification (Benjamin 1993) shows the influence of the approach. Other approaches use narrative material.

Luborsky developed the best-known approach for the systematic analysis of individual transference disposition, the CCRT (Luborsky and Kächele 1988). This procedure is based on the assumption that the narrative of the patient transports and condenses "clotted", subjectively meaningful interpersonal relationship experiences. Poignant subject-object behavioural relations such as burnt-in clichés are made visible by this method.

This instrument, which judges a relationship event that is experienced, prepares narrative material so that the imprinted internalised relationship structures that can be seen in individual behaviour become transparent. The relationship world of an individual is represented with a type of enduring life-history "motto", "chiffre" or "schema". Even more differentiated insights into the "macromolecular" relationship structures can be obtained from a further development of the Central Relationship Pattern (CRP) method (Dahlbender et al. 1998). This method demonstrates the variable organisation with different objects and contexts, shows a lifetime regulation and manages to demonstrate a change in therapy.

So-called relationship episodes are filtered from accounts of interactions from which three components are extracted. These are presented as sequential schemas.

The subject's wish for something from an object leads to a satisfactory or unsatisfactory reaction from the object, which is subsequently followed by a corresponding reaction from the subject. We can remain at the idiographic level, so that the patient's formulations are in speech form, or the statements can be transformed into an abstract categorical level.

One remarkable result obtained in this area of research is that the more relationship episodes with a variety of objects are recounted from the past and present, the clearer the differentiated patterns of clotted relationship experiences are (Luborsky and Crits-Christoph 1990). Research into convergent and discriminant validity of therapy transcripts confirms the value of analysing interpersonal relationship patterns and allows consideration of clinical progress and diagnosis (Luborsky and Barber 1995; Kächele and Dahlbender 1993).

In addition, self-rating methods were developed which assess interpersonal behaviour and experience, such as the Inventory of Interpersonal Problems (IIP) presented by Horowitz et al. (1988). This instrument is theoretically based on Sullivan's interpersonal theories (1953) from which the circumplex models of interpersonal behaviour were derived.

The circumplex model is based on the assumption that all interpersonal behaviours can be represented using two orthogonal and bipolar dimensions. The dimension of control ranges from dominant/controlling to submissive behaviour, and the dimension of affiliation ranges from affectionate/oriented towards to hostile/distanced behaviour. In Leary's model (1957),

16 interpersonal categories or segments were defined which are arranged around two orthogonal dimensions. Starting from this model, different groups of researchers have developed other models for the taxonomy of interpersonal behaviour that partly differ from one another in the number of segments they use (e.g. Wiggins 1982; Kiesler 1983).

Empirical relationships of the attachment style and people schemata underline the interconnection of constructs based on two different theoretical traditions (Horowitz 1994; Strauß and Schmidt 1997).

5 Implications

Psychoanalytic therapies have been influenced by the concepts of development, attachment and relationship. It has been shown that psychopathology can begin anywhere on the developmental path (Stern 1985). Modern psychoanalysis takes account of the environmental aspect. Instead of the pleasure/unpleasure principle, the principle of safety is a prime regulating mechanism using such concepts as mirroring, communication, affective exchange and bodily contact. Research into pathological development supports these new concepts with impressive evidence. The results of attachment research show the overriding significance of attachment needs for personality development.

Even though stringent empirical confirmation has yet to be delivered, knowledge of the patterns of attachment imply differentiated therapeutic strategies. It is reasonable to assume that the development of disorder-specific, axis II-oriented therapies would prove useful.

Knowing the implication of loss and the influence of death on the development of children is helpful for the therapeutic process (Köhler 1995). Diffuse associations made by the patient when touching upon such a topic should not be interpreted as a defence, but are developmentally based deficits of attention and concentration.

The therapist should not treat his or her patients as infants and should not mother them. However, it can be an advantage if the process of understanding the patient's childhood is enriched by images provided by new developmental psychology. The current interaction during therapy can be understood with as much differentiation as the mother-child relationship has illustrated. This leads to a multitude of communicative and interactive processes which enrich clinical conceptualisation (Emde 1991).

The process of empathic agreement takes shape through preverbal processes which are manifest in eye contact, position of the body and verbal adjustment. The saying that with our own unconscious we decode the unconscious of the patient would not be much more than an empty metaphor without these microstructural processes of exchange (Krause 1998). We may assume that the empathic understanding and intuitive grasp of the therapist are based on consciously or subconsciously perceived affective and motor patterns. These may be rooted in the early mother-child, father-child and sibling-child interactions (Lichtenberg et al. 1992).

The great progress made in infant research demonstrates the necessity to decode the grammar of non-verbal interactions (Krause 1990). The results focus on the importance of situational factors that play a part in both dyadic and group therapy situations.

In order to attain a helpful relationship, which is a prerequisite for good therapy, many different verbal and non-verbal communicative subprocesses may be seen as important. Their significance has already been demonstrated in the mother-child relationship, and parallels may be drawn with the therapeutic relationship. Therapy research has shown (Henry et al. 1994) that mutual esteem is an important ingredient in a helpful therapeutic relationship.

To summarise the significance of new insights in early development, attachment and relationship, the following may be said: They provide us with relevant facts and plausible models that allow us to create and enrich (Bornstein and Masling 1998) a current relationship, while having access to the so-called present unconscious and past unconscious (Sandler and Sandler 1984).

6 References

Ainsworth M, Bowlby J (1991) An ethological approach to personality development. American Psychologist 46: 333–341

*Ainsworth MDS, Eichberg CG (1991) Effects on infant-mother attachment of mother's unresolved loss of an attachment figure, or other traumatic experience. In: Parkes CM, Stevenson-Hinde J, Marris P (eds) Attachment across life cycle. Tavistock/Routledge, London, pp 160–183

Ainsworth MDS, Witting B (1969) Attachment and the exploratory behavior of one-years-olds in a strange situation. In: Foss BM (eds) Determinants of infant behavior. Basic Books, New York, pp 113–136

Ainsworth MDS, Bell SM, Stayton DJ (1974) Infant-mother attachment and social development: 'socialisation' as a product of reciprocal responsivness to signals. In: Richards MPM (ed) The integration of a child into a social world. Cambridge University Press, New York, pp 99–135

**Ainsworth MDS, Blehar MC, Waters E, Wall S (1978) Patterns of attachment. A psychological study of the strange situation. Erlbaum, Hillsdale, New Jersey

Basch MF (1983) Empathic understanding. J Am Psychoanal Assoc 31: 101–126

Beck AT, Rush JA, Shaw BF, Emery G (1979) Cognitive therapy of depression. Guilford, New York

Benjamin LS (1974) Structural analyses of social behavior (SASB). Psychol Rev 81: 392–425

Benjamin LS (1993) Interpersonal diagnosis and treatment: the SASB approach. Guilford, New York

Benoit D, Parker KHC (1994) Stability and transmission of attachment across three generations. Child Dev 65: 1444–1456

Bion WR (1962) Learning from experience. Heinemann, London

Blatt S (1990) Interpersonal relatedness and self-definition. In: Singer J (ed) Repression and dissociation: implications for personality theory, psychopathology and health. University of Chicago Press, Chicago

Blatt S (1992) The differential effect of psychotherapy and psychoanalysis with anaclitic and introjective patients: the Menninger Psychotherapy Research Project revisited. J Am Psychoanal Assoc 40: 691–724

Blatt S, Quinlan D, Pilkonis P, Shea MT (1995) Impact of perfectionism and need for approval on the brief treatment of depression: the NIMH treatment of Depression Collaborative Research Program revisited. J Consult Clin Psychol 63: 125–132

Bornstein RF, Masling M (1998) Empirical perspectives on the psychoanalytic unconscious. American Psychological Association, Washington, DC

**Bowlby J (1969) Attachment and loss. 1. Attachment. Basic, New York

Bowlby J (1973) Attachment and loss. 2. Separation. Anxiety and anger. Basic, New York

**Bowlby J (1979) The making and breaking of affectional bonds. Tavistock, London

Bowlby J (1988) A secure base: parent-child attachment and healthy human development. Basic, London

Bowlby J (1995) Bindung: Historische Wurzeln, theoretische Konzepte und klinische Relevanz. In: Spangler G, Zimmermann P (eds) Die Bindungstheorie. Grundlagen, Forschung und Anwendung. Klett-Cotta, Stuttgart, pp 17–29

Brazelton TB (1981) Precursors for the development of emotions in early infancy. In: Plutchik R, Kellerman H (eds) Emotion, theory, research and experience. Academic, New York

Brazelton TB, Cramer B (1991) Les premiers liens. Calman-Levy, Paris

Brazelton TB, Koslowski B, Main M (1974) The origins of reciprocity: the early mother-infant interaction. In: Lewis M, Rosenblum LA (eds) The effect of the infant on its caregiver, vol 4. Wiley, New York, pp 49–76

Brisch KH, Buchheim A, Köhntop B, Kunzke D, Kächele H, Schmücker G, Pohlandt F (1996) Early preventive psychotherapeutic intervention program for parents after the delivery of a very small premature infant: the Ulm Study. Infant Behav Dev (special issue) 19: 356

Bretherton I (1995) Die Geschichte der Bindungstheorie. In: Spangler G, Zimmermann P (eds) Die Bindungstheorie. Grundlagen, Forschung und Anwendung. Klett-Cotta, Stuttgart, pp 27–49

Bretherton I, Waters E (1985) (eds) Growing points of attachment theory and research. Monogr Soc Res Child Dev 50: 3–35

Buchheim A, Brisch KH, Kächele H (1998) Einführung in die Bindungstheorie und ihre Bedeutung für die Psychotherapie. Psychother Psychosom Med Psychol 48: 128–138

Carlson V, Ciccheti D, Barnett D, Braunwald KG (1989) Finding order in disorganization: lessons from research on maltreated infants' attachments to their caregivers. In: Cicchetti D, Carlson V (eds) Child maltreatment. Cambridge University Press, Cambridge, MA, pp 494–528

Clarke AM, Clarke ADB (1976) Early experience, myth and evidence. Free Press, New York

Clarkin J, Marziali E, Monroe-Blum H (eds) (1992) Borderline personality disorder: clinical and empirical perspectives. Guilford, New York

Cramer B (1991) Frühe Erwartungen. Unsichtbare Bindungen zwischen Mutter und Kind. Kösel, Munich

Crowell J, Treboux D (1995) A review of adult attachment measures: implications for theory and research. Soc Dev 4: 294–327

Dahl H (1988) Frames of mind. In: Dahl H, Kächele H, Thomä H (eds) Psychoanalytic process research strategies. Springer, Berlin Heidelberg New York, pp 51–66

Dahlbender RW, Albani C, Pokorny D, Kächele H (1998) Central connected relationship themes CCRT: a structural version of the CCRT. Psychother Res 8: 408–425

Diamond D, Blatt SJ (1994) Internal working models and the representational world in attachment and psychoanalytic theories. In: Sperling MB, Berman WH (eds) Attachment in adults. Clinical and developmental perspectives. Guilford, New York, pp 72–97

*Dornes M (1993) Der kompetente Säugling. Fischer, Frankfurt am Main

*Dornes M (1997) Die frühe Kindheit: Entwicklungspsychologie der ersten Lebensjahre. Fischer, Frankfurt am Main

Eagle M (1984) Recent developments in psychoanalysis. A critical evaluation. McGraw-Hill, New York

Emde RN (1981) Changing models of infancy and the nature of early development. Remodeling the foundation. J Am Psychoanal Assoc 29: 179–219

*Emde RN (1983) The prerepresentational self and its affective core. Psychoanal Study Child 38: 165–192

Emde RN (1991) Positive emotions for psychoanalytic theory: surprises from infancy research and new directions. J Am Psychoanal Assoc 39: 5–44

Emde RN (1992) Individual meaning and increasing complexity: contributions of Sigmund Freund and Rene Spitz to developmental psychology. Dev Psychol 28: 347–359

Erikson EH (1950) Childhood and society. Norton, New York

Fonagy P (1993) Psychoanalytic and empirical approaches to developmental psychopathology: an object-relations perspective. In: Shapiro T, Emde R (eds) Research in psychoanalysis: process, development, outcome. International Universities Press, New York, pp 245–260

*Fonagy P, Steele H, Steele M (1991a) Maternal representations of attachment during pregnancy predict the organization of infant-mother attachment at one year of age. Child Dev 62: 891–905

Fonagy P, Steele M, Steele H, Moran GS, Higgitt AC (1991b) The capacity for understanding mental states: the reflective self in parent and child and its significance for security of attachment. Infant Mental Health J 12: 201–218

Fonagy P, Steele M, Steele H, Leigh T, Kennedy R, Mattoon G, Target M (1995) Attachment, the reflective self, and borderline states: the predictive specificity of the Adult Attachment Interview and pathological emotional development. In: Goldberg S, Muir S, Kerr J (eds) Attachment theory: social,

developmental, and clinical perspectives. Analytic Press, Hillsdale, pp 233–278

Fremmer-Bombik E (1995) Innere Arbeitsmodelle von Bindung. In: Spangler G, Zimmermann P (eds) Die Bindungstheorie. Grundlagen, Forschung und Anwendung. Klett-Cotta, Stuttgart, pp 109–119

Freud S (1917) Introductory lectures on psychoanalysis. Standard edition, vols 15/16. Hogarth, London

Freud S (1918) From the history of an infantile neurosis. Standard edition, vol 17. Hogarth, London, pp 1–122

Freud S (1920) Beyond the pleasure principle. Standard edition, vol 18. Hogarth, London, pp 1–64

Greenspan SI (1989) The development of the ego: implications for personality theory, and the psychotherapeutic process. International Universities Press, Madison

Grice H-P (1975) Logic and conversation. In: Cole P, Morgan J (eds) Syntax and semantics. Speech acts, vol 3. Academic, New York, pp 41–58

Grossmann K, Grossmann K (1991) Attachment quality as an organizer of emotional and behavioral responses in a longitudinal perspective. In: Parkes CM, Stevenson-Hinde J, Marris P (eds) Attachment across the life cycle. Tavistock/Routledge, London, pp 93–114

Grossmann, K, Fremmer-Bombik E, Rudolph J, Grossmann, KE (1988) Maternal attachment representations as related to child-mother attachment patterns and maternal sensitivity and acceptance of her infant. In: Hinde RA, Stevenson-Hinde J (eds) Relations within families. Oxford University Press, Oxford, pp 241–260

*Grossmann K, August P, Fremmer E et al (1989) Die Bindungstheorie: Modell und entwicklungspsychologische Forschung. In: Keller H (ed) Handbuch der Kleinkindforschung. Springer, Berlin Heidelberg New York, pp 31–55

Heigl-Evers A, Heigl F, Ott J (eds) (1993) Lehrbuch der Psychotherapie. Fischer, Stuttgart

*Henry W, Strupp HH, Schacht TE, Gaston L (1994) Psychodynamic approaches. In: Bergin AE, Garfield SL (eds) Handbook of psychotherapy and behavior change. Wiley, New York

*Horowitz MJ (1991) Person schemas. In: Horowitz MJ (ed) Person schemas and maladaptive interpersonal patterns. University of Chicago Press, Chicago, pp 13–31

Horowitz LM, Rosenberg SE, Baer AE, Ureno G, Villasenor VS (1988) Inventory of interpersonal problems: psychometric properties and clinical applications. J Consult Clin Psychol 56: 885–892

Horowitz LM, Strauß B, Kordy H (1994) Manual zum Inventar zur Erfassung interpersonaler Probleme (IIP-D). Beltz-Test-Gesellschaft, Weinheim

Jacobson E (1964) The self and the object world. International Universities Press, New York

Jörg M, Dinter R, Rose F, Villalba-Yantorno P, Esser G, Schmidt M, Laucht M (1994) Kategoriensystem zur Mikroanalyse der frühen Mutter-Kind-Interaktion. Z Kinder Jugendpsychiatrie 22: 97–106

Jüttemann G, Thomae H (eds) (1987) Biographie und Psychologie. Springer, Berlin Heidelberg New York

Kächele H, Dahlbender R (1993) Übertragung und zentrale Beziehungsmuster. In: Buchheim P, Cierpka M, Seifert T (eds) Lindauer Texte. Springer, Berlin Heidelberg New York

Kernberg O (1968) The treatment of patients with borderline personality organization. Int J Psychoanal 49: 600–619

**Kernberg OF (1984) Severe personality disorders. Psychotherapeutic strategies. Yale University Press, New Haven

Kiesler DJ (1983) The 1982 interpersonal circle: a taxonomy for complementarity in human transactions. Psychol Rev 90: 185–214

**Klein M, Heimann P, Isaacs S, Riviere J (1952) Developments in psychoanalysis. Hogarth, London

Klermann GL, Weissman MM, Rounsaville BJ (1984) Interpersonal psychotherapy of depression. Basic, New York

Köhler L (1995) Bindungsforschung und Bindungstheorie aus der Sicht der Psychoanalyse. In: Spangler G, Zimmermann P (eds) Die Bindungstheorie: Grundlagen, Forschung und Anwendung. Klett-Cotta, Stuttgart, pp 67–85

Kohut H (1971) The analysis of the self. A systematic approach to the psychoanalytic treatment of narcissistic personality disorders. International Universities Press, New York

Kohut H, Wolf ES (1978) The disorders of the self and their treatment: an outline. Int J Psychoanal 59: 413–425

Krause R (1990) Psychodynamik der Emotionsstörungen. In: Scherer K (ed) Psychologie der Emotion. Enzyklopädie der Psychologie. Hogrefe, Göttingen, pp 630–705

*Krause R (1998) Allgemeine Psychoanalytische Krankheitslehre. 2. Modelle. Kohlhammer, Stuttgart

Künzler E (1969) Zwei Hypothesen über die Natur der frühkindlichen Sozialbeziehungen. Psyche 23: 25–57

Leary T (1957) Interpersonal diagnosis of personality. Ronald, Chicago

*Lichtenberg J (1983) Psychoanalysis and infant research. Analytic Press, Hillsdale

Lichtenberg J, Lachmann F, Fosshage J (1992) Self and motivation systems. Analytic Press, Hillsdale

*Luborsky L (1984) Principles of psychoanalytic psychotherapy. A manual for supportive-expressive treatment. Basic, New York

Luborsky L, Barber J (1995) Perspectives on seven transference-related measures applied to the interview with Mr. Smithfield. Psychother Res 4: 152–155

*Luborsky L, Crits-Christoph P (1990) Understanding transference, 2nd edn. Basic, New York

Luborsky L, Kächele H (1988) Der zentrale Beziehungskonflikt. PSZ-Verlag, Ulm

Lyons-Ruth K, Alpern L, Repacholi B (1993) Disorganized infant attachment classification and maternal psychosocial problems as predictors of hostile-aggressive-behavior in pre-school classroom. Child Dev 64: 572–585

Mahler M (1958) Autism and psychosis: two extreme disturbances of identity. In: Mahler M (ed) The selected papers of Magret Mahler. Infantile psychoses and early contributions, vol 1. Aronson, New York, pp 169–181

Mahler MS (1969) On human symbiosis and the vicissitudes of individuation. Hogarth, London

Mahler M, Pine F, Bergmann A (1975) The psychological birth of the human infant. Basic, New York

Main M (1991) Metacognitive knowledge, metacognitive monitoring, and singular (coherent) vs. multiple (incoherent) model of attachment: findings and directions for future research. In: Parkes CM, Stevenson-Hinde J, Marris P (eds) Attachment across the life cycle. Routledge, London, pp 127–159

*Main M (1995) Recent studies in attachment: overview with selected implications for clinical work. In: Goldberg S, Muir R, Kerr J (eds) Attachment theory: social developmental and clinical perspectives. Erlbaum, New Jersey, pp 407–474

Main M, Cassidy J (1988) Categories of response to reunion with the parent at age six: predicted from attachment classifications and stable over a one-month period. Dev Psychol 24: 425–426

Main M, Hesse E (1990) Parents' unresolved traumatic experiences are related to disorganized attachment status: is frightened and/or frightening parental behavior the linking mechanism? In: Greenberg MT, Cicchetti D, Cummings EM (eds) Attachment in the preschool years. University of Chicago Press, Chicago, pp 161–182

Main M, Solomon J (1986) Discovery of an insecure disorganized/disoriented attachment pattern: procedures, findings and implications for the classification of behavior. In: Brazelton TB, Yogman M (eds) Affective development in infancy. Ablex, Norwood, pp 95–124

Main M, Kaplan N, Cassidy J (1985) Security in infancy, childhood, and adulthood: a move to the level of representation. In: Bretherton I, Waters E (eds) Growing points in attachment theory and research. Monogr Soc Res Child Dev 50: 66–106

Meltzoff A, Borton R (1979) Intermodal matching by human neonates. Nature 282: 403–404

Moser U, von Zeppelin I (1991) Cognitive-affective processes. Springer, Berlin Heidelberg New York

Papoušek H (1981) The common in the uncommon child: comments on the child's integrative capacities and on parenting. In: Lewis M, Rosenblum LA (eds) The uncommon child. Plenum, New York, pp 317–328

*Papoušek H, Papoušek M (1983) Interactional failures. Their origins and significance in infant psychiatry. In: Call JD, Galenson E, Tyson RL (eds) Frontiers of infant psychiatry. Basic, New York, pp 31–37

Patrick M, Hobson RP, Maughan B (1994) Personality disorder and the mental representation of early social experience. Dev Psychopathol 6: 375–388

Peterfreund E (1978) Some critical comments on psychoanalytic conceptualizations of infancy. Int J Psychoanal 59: 427–441

Quinton D, Rutter M (1988) Parenting breakdown: the making and breaking of intergenerational links. Gower, Brooksfield

Rapaport D (1960) The structure of psychoanalytic theory. A systematizing attempt. International Universities Press, New York

Reich G (1995) Eine Kritik des Konzeptes der "primitiven Abwehr" am Begriff der Spaltung. Forum Psychoanal 11: 99–118

Sander J (1985) Toward a logic of organization in psychobiological development. In: Klar K, Siever L (eds) Biologic response styles: clinical implications. American Psychiatric Press, Washington (Clinical Insights Monograph)

Sandler J, Sandler AM (1984) The past unconscious, the present unconscious and interpretation of the transference. Psychoanal Inquiry 4: 367–399

Schaffer R (1977) Mothering. Fontana, London

Spitz R (1965) The first year of life. A psychoanalytical study of normal and deviant development of object relations. International Universities Press, New York

*Spitzer M (1997) Geist im Netz. Spektrum der Wissenschaften, Heidelberg

Stern DN (1974) Mother and infant at play: the dyadic interaction involving facial, vocal, and gaze behaviors. In: Lewis M, Rosenblum LA (eds) The effect of the infant on its caregiver, vol 4. Wiley, New York, pp 187–213

**Stern DN (1985) The interpersonal world of the infant. Basic, New York

Stern D (1995) The motherhood constellation. Basic, New York

Stern DN (1996) Ein Modell der Säuglingsrepräsentation. Forum Psychoanal 12: 187–203

Spangler G, Schieche M (1995) Psychobiologie der Bindung. In: Spangler G, Zimmermann P (eds) Die Bindungstheorie. Grundlagen, Forschung und Anwendung. Klett-Cotta, Stuttgart, pp 297–310

Strage M, Main M (1985) Attachment and parent-child discourse patterns. Presented at the biennal meeting of the Society for Research in Child Development, Toronto, 25–28 April 1985

Strauß B, Schmidt S (1997) Die Bindungstheorie und ihreRelevanz für die Psychotherapie. Psychotherapeut 42: 1–16

Strupp HH, Binder J (1984) Psychotherapy in a new key. A guide to time-limited dynamic psychotherapy. Basic, New York

Sullivan HS (1953) The interpersonal theory of psychiatry. Norton, New York

**Thomä H, Kächele H (1987) Psychoanalytic practice. 1. Principles. Springer, Berlin Heidelberg New York

**Thomä H, Kächele H (1991) Psychoanalytic Practice. 2. Clinical studies. Springer, Berlin Heidelberg New York

Tress W, Henry P, Strupp H, Reister G, Junkert B (1990) Die strukturale Analyse sozialen Verhaltens (SASB) in Ausbildung und Forschung. Ein Beitrag zur "funktionellen Histologie" des psychotherapeutischen Prozesses. Z Psychosom Med Psychoanal 36: 240–257

Tustin F (1994) The perpetuation of an error. J Child Psychother 20: 3–23

van IJzendoorn MH (1995) Adult attachment representations, parental responsiveness and infant attachment: a meta-analysis on the predictive validity of the Adult Attachment Interview. Psychol Bull 117: 387–403

van IJzendoorn MH, Bakermans-Kranenburg MJ (1996) Attachment representations in mothers, fathers, adolescents and clinical groups: a meta-analytic search for normative data. J Consult Clin Psychol 64: 8–21

Ward MJ, Carlson EA (1995) Associations among adult attachment representations, maternal sensitivity, and infant-mother attachment in a sample of adolescent mothers. Child Dev 66: 69–79

*Weiss J, Sampson H (1986) The psychoanalytic process: theory, clinical observation, and empirical research. Guilford, New York

Wiggins JS (1982) Circumplex models of interpersonal behavior in clinical psychology. In: Kendall PC, Butcher JN (eds) Handbook of research methods in clinical psychology. Wiley, New York

Zimmermann P, Spangler G, Schieche M, Becker-Stoll F (1995) Bindung im Lebenslauf: Determinanten, Kontinuität, Konsequenzen und künftige Perspektiven. In: Spangler G, Zimmermann P (eds) Die Bindungstheorie: Grundlagen, Forschung und Anwendung. Klett-Cotta, Stuttgart, pp 311–334

CHAPTER

23

R. Michels

Psychoanalysis in Practice

Excerpts from Michels (1985, 1995, 1997) were used in this chapter and have been reprinted with the permission of the publishers.

1 Introduction

Psychoanalysis developed out of the clinical experience and theoretical speculations of Sigmund Freud, a Viennese physician born in 1856. In 1883, Freud's senior colleague, Joseph Breuer, told him of a fascinating patient, an intelligent young German woman, Bertha Pappenheim, who in later years was to become one of the founders of psychiatric social work. Breuer's clinical observations on this patient, who suffered from the symptoms of what was then called hysteria, led to a new theory about the nature of the disorder and some suggestions for a possible new treatment (Breuer and Freud 1895). Breuer himself stopped working in the area, perhaps in part because of his discomfort over the intense personal feelings stimulated by his relationship with his patient and his inability to understand those feelings. Freud explored Breuer's method further with other patients, continued to develop the theory, and in time even came to understand the meaning of Breuer's discomfort. The theory and the treatment based upon it were the beginning of psychoanalysis and psychoanalytic psychiatry (Freud 1925).

Bertha Pappenheim, like other patients diagnosed with hysteria, had multiple symptoms related to several organ systems, altered states of consciousness, and disturbances of thought, affect, and communication. Nineteenth-century psychiatry had understood these as reflecting abnormalities of the nervous system, i.e., as signs of an underlying neuropathology that had not yet been identified. First Breuer and then Freud began to employ a practice that grew out of the tradition of nineteenth-century medicine far more than that of nineteenth-century psychiatry: they spent many hours talking with their patients or, to be more precise, listening to them. As they did this, they noted a relationship between their patients' symptoms and what was on their patients' minds, and – particularly exciting – they discovered that talking about these issues actually influenced the symptoms. Freud and Breuer reconceptualized the symptoms as communications – symbolic expressions of thoughts and memories of which the patient might not be aware – rather than as signs of a disordered nervous system. In brief, the meaning of the symptoms and the cause and treatment of the disorder were to be discovered through a dialogue between the patient and the therapist rather than by a dissection of the brain and nervous system. Psychoanalytic psychiatry was to be a science of mental life. In Freud's famous formulation, his patients suffered not from brain disease, but from "reminiscences" (Breuer and Freud 1895).

Freud's background as a neurologist and neurobiologist shaped much of his early thinking; he initially attempted to formulate his theory as a neuropsychology, a project that quickly ran into difficulty because of the then primitive understanding of the nervous system (Freud 1895). Freud was also interested in evolutionary biology, cultural anthropology, and history, and one can trace each of these themes, as well as the ideas of late nineteenth-century positivist natural science, throughout his work. All of this is to say that psychoanalytic theory has been richly informed by the dominant ideas that influenced its creators and the intellectual climate in which it developed. In recent years, developmental psychology (Mahler et al. 1975; Stern 1985) has largely replaced neurobiology (Reiser 1984) and evolutionary biology as the source of its basic psychological concepts, and linguistics (Shapiro 1979), cognitive neuropsychology (Benjamin and Freidrich 1991; Stinson and Palmer 1991), group dynamics, and clinical psychopathology have contributed to its conceptual models as well.

According to Freud, psychoanalysis is three things: a psychology, a method of treatment, and a method of inquiry (Freud 1923). We will first discuss those principles of psychoanalytic psychology that are fundamental to the treatment (Michels 1995), then the clinical concepts that are central to contemporary psychoanalytic practice (Greenson 1967), and finally the scope of psychoanalytic psychotherapy and psychoanalysis today (Michels 1997).

2 Basic Principles

2.1 Mentalism

Psychoanalysis is about the psyche, i.e., the mind. It is about thoughts, feelings, experiences, wishes, fears, fantasies, and memories. It is not about the brain or the body, except insofar as these are seen as the substrate of mental activity having an impact on the mind or as being represented in mental imagery. It is also not about behavior viewed "objectively" from the outside, as by a behaviorist, seen as the activity of an organism without regard to the realm of subjective experience. Contemporary psychiatry often pushes neurobiological reductionism to the extreme, regarding mentalist concepts as referring to epiphenomena, with the "real" action occurring at the level of the nervous system (Kandel 1979). Psychoanalysis regards mental phenomena as its subject matter, and therefore it is of greatest value in understanding the world of inner experience. It is not an important tool in understanding brain dysfunction, seizure phenomena, or the psychological defects caused by the biologic

diatheses that underlie the major psychoses, but it is a major tool in understanding wishes, fears, and fantasies and the relationships among them – how these phenomena arise from earliest experiences, are transformed in the course of development, predispose to both adaptive and maladaptive behavior, and change in response to meaningful human relations. If we want to understand mental life and mental symptoms, we must employ mentalist models (Michels 1995).

2.2 Psychic Determinism

Common-sense psychology is, like psychoanalysis, mentalist; in other words, it explains people and their behavior in terms of thoughts, wishes, fears, and fantasies and the interplay among them. However, common-sense psychology, unlike scientific thinking, is not strictly determinist; although it explains many things, it accepts that other things might not be explainable – they just happen. For example, most people eat because they feel hungry and stop because they feel satisfied; their behavior reflects a mental experience. However, some people stop eating while still hungry, while others continue to eat although sated. These behaviors can be labeled as pathological – anorexic or bulimic – and attributed to neurobiological abnormality or unspecified "disease," but there is no common-sense psychological explanation for them. Even more familiar are so-called Freudian slips (psychoanalysts call them parapraxes). Most often we say what we are thinking, but sometimes we "just make a mistake." Other mental phenomena that seem to common-sense psychology to be unexplained by psychic determinism include dreams and most forms of neurotic symptomatology, such as obsessions or phobias.

Psychoanalytic theory is different from common-sense psychology in that it asserts a strict and all-embracing psychic determinism – all mental events can be understood as the result of antecedent mental events rather than as epiphenomena of brain events (except perhaps for the behavioral aspects of a seizure) and certainly never as random or without cause. An individual who does not eat may have the unconscious fantasy that food is poison, while one who overeats may – without being aware of it – see food as evidence of love. In short, dreams and symptoms are symbolic representations of disguised thoughts and fantasies. Psychoanalytic theory argues that mental determinism is pervasive and in doing so accepts that self-awareness or consciousness – a characteristic of mental life that common-sense psychology views as universal and constant – is in fact only partial and intermittent. The anorexic or bulimic individual, the dreamer, and the person with a neurosis all share a lack of awareness of the mental forces that govern their experience and, through it, their behavior. From a psychoanalytic perspective, these individuals' behavior is totally determined, although the determinants are unconscious. The psychoanalytic method allows us access to unconscious mental life and thereby permits us to fill in the gaps in the incomplete psychic determinism of common-sense psychology.

2.3 Mental Dynamics

The psychoanalytic approach is not only "psycho," i.e., mentalist, but also "dynamic," i.e., about forces, motives, and inevitably (because multiple forces have multiple and disparate goals) about conflict. Psychoanalytic thinking views mental life as the product of conflicting mental forces, wishes, fears, and emotions, each of which "presses" and all of which together direct thoughts and behavior. One of the ongoing theoretical dialogues in psychoanalysis relates to how such forces are best conceptualized. Freud thought of these mental dynamics as biologically rooted drives which unfold in a largely predetermined maturational sequence. Others have emphasized socially and culturally determined attitudes and desires (Kardiner 1945). The classification of mental forces, their origins in biology or social experience, their development over the life span, their plasticity, and their basic content – sexuality, aggression, mastery, curiosity, self-fulfillment, and others – have been ongoing themes of psychoanalytic discourse (Rapaport and Gill 1959). Much dispute in psychoanalytic theory relates to the basic nature of these forces. There is much more agreement as one moves from theoretical to clinical phenomena and observes the central role of the wishes and fears that stem from these forces in organizing patients' mental lives (Brenner 1982).

2.4 Unconscious Mental Life

Recognition of unconscious mental processes long antedates psychoanalysis. People frequently remember things they had previously forgotten and that must have been "stored" somewhere in the mind during the interval. Furthermore, they can be shown to know something they were not aware of knowing, by hypnosis or by less exotic means, such as behavioral responses to subliminal stimuli. However, unconscious mental processes assume a unique importance in psychoanalytic thinking because not only casual memories but also major dynamic forces – the wishes and fears that shape our lives and, particularly, our

neurotic symptoms and character traits – are believed to be unconscious (Arlow 1969). That mental processes could occur without our awareness was known to Aristotle; that many of the determinants of our major life choices are unknown to us, but nonetheless – and even because of this – exert powerful influence on us was one of Freud's great discoveries. Psychoanalysts further differentiate unconscious mental activity into "preconscious" (i.e., out of awareness, but easily accessible if attention is directed) and "unconscious" in the narrow sense, (i.e., inaccessible to the individual under normal circumstances). The clinical psychoanalytic method, with its free association (Kris 1982) and dream analysis (Altman 1978), is designed to facilitate the exploration of unconscious mental activity, making it preconscious and thus accessible to consciousness.

2.5 Repression and Defense

Why are mental themes unconscious? The most obvious answer – and the one largely accepted by common-sense psychology – is that they are likely to be forgotten if they are unimportant. However, this cannot explain why major dynamic forces that shape our lives are unconscious. Freud recognized that dynamic unconscious forces are unconscious for the very reason that they are important and would cause distress if they were to become conscious. The psychological force or mechanism that actively keeps powerful and potentially disturbing wishes or fears unconscious is called repression. First Freud himself, and then his daughter Anna and other psychoanalysts, came to recognize that repression was only one of the mental mechanisms available to avoid conscious distress engendered by the emergence of disturbing unconscious mental themes, and the term defense was introduced to encompass all such mechanisms, with repression as the prototype. In contemporary psychoanalytic thinking, defense is a function which may be served by almost any mental activity which is used to avoid conscious awareness of a disturbing unconscious theme. While early models of psychoanalytic therapy focused on helping people become more aware of forbidden unconscious wishes and fears, contemporary views emphasize helping people understand their defensive strategies and diminish the power that these exert in mental life.

2.6 The Present and the Past

Psychoanalysis is about the dynamic mental forces that shape behavior, forces that are very much alive in the present. It has no interest in the past except insofar as that past is preserved in those present dynamic forces. In their earliest inquiries into the unconscious forces that determined psychopathologic symptoms, Breuer and Freud had no special interest in childhood. However, within a few years, Freud discovered (and countless psychoanalysts since him have confirmed) that the persistent wishes and fears of childhood are the major themes of unconscious mental life. Thus the study of psychoanalysis, which is the study of unconscious mental forces, became of necessity the study of childhood psychology. The influence of unconscious forces on behavior is the influence of persisting themes of the mental life of the child on the adult. One of the important methodologic issues in contemporary psychoanalytic thinking is the relationship between the direct observation and scientific study of the psychology of children and the exploration of themes in adult mental life that can be traced back to childhood. The inner subjective record of the past is not the same as the past that would have been observed by a contemporary onlooker. The record is constructed, shaped, and reshaped during the course of development, and then is profoundly influenced by the circumstances in which it is eventually recounted. Thus, although the study of developmental psychology has relevance for psychoanalysis, it is only one source of data relevant to the clinical exploration of persisting infantile themes in unconscious mental life.

The concept of transference refers to a particularly important instance of the role of the past in shaping the present. Emotionally charged personal relationships, such as the relationship between a patient and doctor, are always shaped by the capacity for and the style of relatedness that the patient acquired in childhood. It is similar to the accent that is always present when speaking a language learned in adult life. The forms and characteristics of the first language shape the sounds of all later ones. All important relationships are to some extent repetitions of primary relationships, and a patient's relationship with his or her doctor reflects the patient's relationship with earlier caregivers, particularly parents. The appreciation of this process of transference is vital to our understanding of how therapy works, how it helps patients, and why they often struggle against it (Michels 1985).

2.7 Infantile Passions

Adults generally think of infants and small children as having bland and innocent thoughts, if indeed they think of them as having any thoughts at all. However, the truth appears to be quite otherwise, based on

observations made by experienced infant caregivers and developmental psychologists and, most importantly, on the persisting unconscious fantasies of adults. The child's mental life is driven by elemental passions and primitive terrors, unchecked by the understanding of reality available to the adult. The child's experience of the external world is organized in terms of the role it plays in the gratification and frustration of basic needs. Freud emphasized the sensual pleasures of bodily experience, which he labeled by the name of the bodily pleasure central to adult experience, sexuality. Other psychoanalysts have emphasized aggressive as well as sexual drives, and the psychological meaning of attachment to and separation from primary caregivers. These themes – sexual, aggressive, clinging, and autonomy-seeking wishes and fantasies persisting from the mental life of childhood as unconscious dynamic forces in adults – are the building blocks of psychoanalytic models of psychic conflict.

2.8
Mental Structures

Psychoanalytic models of mental life focus on dynamic forces. However, certain patterns of mental activity are stable over long periods of time, and the concept of structure has seemed more appropriate than that of force to describe such stable patterns. Motivational systems such as sexuality and aggression, believed by many to stem from the neurobiological organization of the nervous system, are structures. Patterns of defense resulting from innate styles and developmental experiences are also structures. Freud's so-called structural model organized mental activity into three overarching structures (Arlow and Brenner 1964). One was the id, which referred to the organismic and biologically rooted drives and their psychological representations. The second was the ego, which referred to the adaptive and external reality-oriented aspects of the mind, including perception, cognition, memory, motor control, and adaptive behavior as well as defensive processes. Finally, there was the superego, a specialized portion of the ego that tended to function as a coherent organized system and that was often in conflict with the rest of the ego as well as with the id. The superego encompasses values and standards, notions of good and bad, right and wrong, approval and disapproval, and the inner sources of guilt, shame, and pride. The common-sense psychology notion of "conscience" refers to that small aspect of superego functioning that is conscious. Psychoanalytic developmental theory views the primary origin of the superego as the child's internal psychological representation of the parent – approving and disapproving, loving and criticizing, rewarding and punishing. Some psychoanalytic writers separate the positive ideals from the critical injunctions of the superego and call the former the "ego-ideal."

2.9
Adaptive Function of Behavior

Common-sense psychology views most behavior as adaptive, but sees some behavior differently, as a mistake, wrong, stupid, or pathologic. In many ways, this view is a corollary of seeing some behavior as not fully determined. Psychopathology seen from a common-sense perspective is not adaptive; it is wrong, and the goal of treatment should be to expunge it and replace it with correct behavior.

Psychoanalytic psychology holds a different view. All behavior, including pathological behavior, is seen as adaptive. Instead of being a mistake, pathological behavior involves the effective pursuit of goals that may be concealed both from the patient and from the rest of the world. Such behavior constitutes an adaptive component of an unconscious strategy. The central goal of treatment is not to expunge the "wrong" behavior, but rather to identify its hidden goal and bring it into the open. When the goals of pathologic behavior become known to the patient, the patient is freer to integrate them with the rest of his or her life, either to purge them or to abandon them, but their maladaptive impact is diminished.

Modern clinical psychiatry classifies the phenomenology of psychopathology according to the familiar clinical diagnostic categories. Psychoanalytic psychiatry, by contrast, sees behavioral phenomenology as the product of underlying mental forces. Whereas a clinical psychiatric diagnosis is a summary description of pathological phenomena, a psychoanalytic formulation is a statement of the major mental themes, wishes, fears, conflicts, and compromises leading to that behavior and the history of their development (Perry et al. 1987). Diagnoses focus on the boundaries and discontinuities between health and pathology; in contrast, psychoanalytic formulations focus on the continuities and similarities that reveal the underlying integration of the patient's healthy and pathologic mental life.

2.10
Conflict and Compromise – Symptom and Character

Psychoanalytic psychology views behavior as the product of conflicting and often unconscious mental forces. Psychological conflicts lead to internal signals of potential dysphoria, anxiety, and depression, signals

that will lead to conscious distress and the disruption of functioning if other psychological strategies are not called into play. Most often, these signal affects elicit coping strategies, such as the so-called defense or mental mechanisms – strategies for constructing compromises among the conflicting dynamic themes, wishes, and fears. Typically, such compromises manage to keep the potentially most disturbing themes unconscious, while partially and symbolically gratifying the wishes and accommodating the fears. These compromise structures may diminish adaptation to reality in order to satisfy unconscious needs, in which case they can lead to pathology. If the compromise structures are integrated into the individual's experience of self, they are called character traits (Auchincloss and Michels 1983). If they remain psychologically sequestered and alien from the sense of self, they are called symptoms. Paradoxically, an early task in the psychoanalytic psychotherapy of symptoms is to demonstrate to the patient that his or her symptoms are meaningful products of inner psychological themes (i.e., they are linked to character), whereas an early task in the psychotherapy of pathological character traits is to establish some potential dystonicity from the patient's sense of self (i.e., to support the patient's image of a possible self without the undesirable pathology).

3 Clinical Principles

The initial focus of psychoanalytic study was on the exciting new discoveries in the psychology of the mind. Many of Freud's most famous early publications – on dreams, parapraxes, jokes, sexuality, art, culture, and even psychopathology – had little to do with psychoanalysis as a therapy. It was as though knowledge of the psychology would be sufficient to conduct the therapy, and no further concepts would be required. However, problems soon emerged, and the basic concepts of clinical psychoanalysis were formulated and developed in response to these. Today, psychoanalytic discourse and the psychoanalytic literature is more interested in issues of clinical therapy than in basic psychologic matters.

3.1 Relationship

The early clinical interest of psychoanalysts focused on the content of the therapeutic sessions – the dreams, fantasies, and memories of the patient – with the hope that their exploration would lead to the recovery of repressed childhood memories and the reconstruction of a coherent account of the forgotten past. However, it soon became apparent that the relationship between the patient and psychoanalyst within which this exploration occurred was important as well, perhaps even more important than the content of what was discussed. Today, it is widely recognized that the therapeutic effectiveness of psychoanalysis, as that of all other psychotherapies, is heavily dependent on the quality of the relationship between patient and therapist.

3.2 Alliance

Early formulations of the relationship emphasized two aspects: (1) the transferential relationship between the regressive infantile neurotic themes in the patient and the responsive aspects of the therapist's personality and (2) the relationship between the mature, rational adult, self-observing and help-seeking themes in the patient and the corresponding expert professional aspects of the therapist. This latter "mature" relationship was first differentiated from transference and referred to as the "therapeutic" or "working" alliance. It was seen as an essential frame for the psychoanalytic process to progress. Today, a broader notion of transference has become more prevalent, one that emphasizes the transferential origins of all relationships, both adaptive and maladaptive, and thus of the alliance as well.

3.3 Transference

Transference refers to the impact of unconscious mental themes that persist from childhood relationships on the experience and form of contemporary relationships.

While transference phenomena are important in understanding unconscious motivating factors in all human relationships, they are particularly prominent in those that are strongly charged emotionally and, like that between patient and analyst, recreate aspects of early parent–child relationships. In most situations in life, and in many therapeutic situations, we try to emphasize external reality and minimize or suppress the impact of transference. In psychoanalysis, we do the reverse – we try to control, limit, and minimize the determining effect of contemporary external reality so that the transferential determinants of behavior will be clear and vivid, both to the patient and the therapist.

Transference phenomena are essential in understanding both the motivating and resistive factors in

any type of treatment, biologic or psychologic, psychoanalytic or otherwise. However, in psychoanalytic treatments, they have additional significance. Because the treatment itself revolves around an exploration of the transference, an attempt is made to study and explore the persistent unconscious dynamic forces that shape the patient's relationship with the therapist rather than merely to exploit or control their impact, positive or negative, on the therapeutic process. The exploration of the expression of these themes in the transference relationship is the essence of psychoanalysis and psychoanalytic psychotherapy.

Transference responses are shaped by contemporary external reality as well as by persisting unconscious determinants from the past. Therefore, the transference in psychoanalysis evolves in part in response to the psychoanalytic process itself. It is the exploration of this aspect of the transference, the development and evolution of transference phenomena within the relationship between patient and therapist, rather than the exploration of the initial transferentially determined attitudes toward the therapist that is at the core of psychoanalytic treatment. The concept of "transference neurosis," considered essential by some and superfluous by others, refers to the integrated cohesive transference that develops in the course of psychoanalysis and that encompasses the core themes of the patient's psychopathology in the newly developed relationship between patient and analyst – a constellation that may in some cases provide a substitute for the patient's neurotic symptoms.

Any aspect of mental life may be involved in transference responses, e.g., wishes, fantasies, emotions, defenses, attitudes, patterns of relationship with others. Further, transferences may stem from any of the various developmental epochs. Transferences may be positively or negatively toned. They may motivate the patient or may stimulate resistance. They may be erotic or aggressive. At times, they may be traced back to specific early relationships; at other times, they may be seen as embodying mixtures of characteristics from several different relationships. Transferences have been classified according to each of these characteristics, and for specific purposes each type of classification may be useful.

Although it is common to think of psychotherapy as exploring the links between past and present, these links are already established in the transference, and it is the exploration of the transference that is central to psychoanalysis and much exploratory psychotherapy. A common misunderstanding of psychoanalysis would emphasize the discussion of the historic past and its relation to contemporary external reality as a core theme. However, if such discussion is conducted in arenas other than the transference, it rarely has the emotional power or conviction that accompanies the analysis of the transference. Some have argued that the only really effective interpretations are transference interpretations. The degree to which this is true, the extent to which transference interpretations are qualitatively exceptional, as opposed to merely frequently important, is a matter of contemporary controversy.

Exploration of the transference entails not only exploration of the patient's response, the unconscious themes that shape it, and their origins in the past, but also exploration of the current situation that has precipitated it. All transference responses are real responses to contemporary stimuli; they are all more or less appropriate. Some of the most important transference phenomena are thoroughly realistic and adaptive. That something is transferentially determined by unconscious mental factors does not mean that it is wrong, pathological, or maladaptive, and all human relationships are transferentially determined by unconscious factors (Michels 1985).

3.4 Countertransference

Just as the patient's response to the therapist reflects persisting unconscious themes, so does the therapist's response to the patient, or the countertransference. Indeed, our notion that transference is universal would require that this be so. Like transferences, countertransferences can be both motivating and resistive factors. Furthermore, like transferences, they are responses to contemporary determinants, i.e., to the patient. As a consequence, they may become valuable sources of information about the patient's mental life.

The first and most vivid indication of the patient's transference may be in the therapist rather than in the patient; hence the study of the therapist's countertransference is an important source of information about the patient's transference. The patient in effect expresses the transference by eliciting, provoking, or seducing responses in the therapist. The extent to which there are unique mental mechanisms involved in this type of phenomena, the extent to which they have central clinical significance, particularly in the treatment of more disturbed patients, and the close link of this idea and certain technical notions emanating from it to specific schools of psychoanalytic thought are areas currently being debated (Sandler et al. 1992).

3.5 Resistance

Resistance refers to the forces within the patient that oppose the treatment. It is a reflection of the patient's defenses and character structure in the treatment

situation and is manifested by a variety of behaviors, including not coming to treatment, coming late, being silent, avoiding emotions, ruminating about trivia, journalistic accounts of daily events, keeping secrets, forgetting, not paying, and focusing on obtaining transference gratifications from the analyst even at the risk of destroying the treatment.

Resistance appears early in every therapy and was recognized early in the history of psychoanalysis. The initial response to this recognition was the obvious one, the development of a variety of techniques for overcoming the resistance, e.g., rational persuasion, exploitation of the positive transference, hypnosis. With time, psychoanalysts came to realize that resistant behavior presented a therapeutic opportunity as well as an obstacle and that the dynamic structure of the patient's character could be explored as it emerged in the therapeutic setting. The focus shifted from what was being resisted, with the corollary of penetrating, dissolving, overcoming, undermining, or circumventing the resistance to the study of the resistance itself, with the goal of understanding its form and its history and exploiting its potential as a model of the patient's character structure in the therapeutic process.

3.6 Interpretation

Interpretation is the major specific activity of the psychoanalyst and is one of the essential ingredients of psychoanalytic psychotherapy. Interpretation has been defined in various ways. For some, any intervention that aims to increase the patient's insight and understanding or to expand his or her awareness is an interpretation. For others, interpretations must be differentiated from the preliminary steps of clarification and confrontation or from the more comprehensive process of construction and reconstruction. The term interpretation may refer to the process of discovery, decoding, and translation of latent or concealed meaning, or it may refer to communicating the results of that process to the patient.

Psychoanalysts discovered that the meanings that individuals create from events can be far more significant than the events that serve as stimuli for those meanings. This is as true of interpretations as other events. The analyst may believe that he or she is interpreting the here and now or the past, but the patient will hear the analyst's interpretations in terms of the psychological complexes that are active at that time in his or her own mental life and in terms of the phase of the treatment and the nature of the transference. As a result, if we wish to know what an interpretation is about, it may be more important to know what was going on in the analysis and in the patient's mind at the time that the interpretation was made than to know the words that the analyst said. Most of the time in a successful analytic process, the patient is involved in a complex mixture of concerns: concerns with his or her current relationship with the analyst, concerns with memories of earlier relationships reactivated by the current one, concerns with the impact of the current relationship on the memories, and so on in a never-ending process of reverberating resonance.

Individual interpretations may be aimed at one or another concern, at the past, at the here and now, or at the interaction between them. If they are successful, however, it is because they become part of the reverberating process and because they in turn influence it. An interpretation directed to the here and now must therefore be related to or at least open to the patient's memories and fantasies of the past if it is to have the possibility of enlarging the patient's experience of the here and now, of providing it with new meanings. Similarly, an interpretation directed to the past must be about the living past that is shaping and forming the patient's present if it is to have a therapeutic impact and alter the present. Some interpretations do both at once. Some poor interpretations may be "true," but they are truths which have no psychic relevance. Such statements about the patient's mental development or mental functioning may be of interest to a psychodynamicist or a psychogeneticist, but they will have no impact on the patient's inner world. The question is not whether a specific interpretation is directed at the present or at the past, but whether it is part of an interpretive net, a therapeutic process that involves both. It is the nature of this process, rather than the focus of a specific interpretation, that defines psychoanalysis (Michels 1983).

4 Psychoanalytic Psychotherapy

Many other treatments have been developed from various components of Freud's early experiments with the complex mixture of activities which occur when a patient and a therapist talk to each other. Some derivative treatments focus on the patient's symptoms; others on pathogenic cognitive strategies; still others on interpersonal relationships; some focus on the therapeutic relationship itself or on constructing a story about the past. Thus we have symptom-oriented treatments, cognitive therapies, interpersonal therapies, relationship and existential therapies, and reconstructive therapies that create stories about those early experiences which are thought to be of pathogenic significance.

All of these treatments have evolved in parallel with the continuing evolution of psychoanalytic psychotherapy and psychoanalysis (Michels 1997). Certain contemporary themes are of particular importance. A prominent one is the trend toward technical innovations that make the treatment more "user-friendly," with less emphasis on developing a painful awareness of disavowed aspects of oneself, and more on relieving the pain and suffering involved in coming to know oneself more deeply. Modern psychoanalytic psychotherapy is more concerned with making the process of therapy comfortable for patients.

A second trend is toward a sharper definition of goals. Today, few psychoanalytic psychotherapists or psychoanalysts believe that one should try to reconstruct a complete account of early life, resolve all psychological conflicts, or understand all dimensions of the transference. One of the goals of treatment – prominent early in the field's history, lost in the middle third of the century, and now recognized again – is to end the treatment.

A third trend is the widening of scope of the mental theories that constitute the content matter of psychoanalytic psychotherapy from their original focus on sexual motivations or wishes to a broader interest in other kinds of motives (e.g., aggression), in self and object representations, and in mental structures other than motives. The modern psychoanalytic therapist has a larger range of metaphors upon which to draw in making interpretive interventions.

The psychoanalytic psychotherapist listens with more directed or focused attention than does the psychoanalyst, with a goal in mind other than the analyst's goal of full comprehension and communication. Both therapist and analyst are guides on a journey, but the analyst is a guide in an uncharted territory, while the psychoanalytic psychotherapist is a guide with a map and a plan for the trip.

The goal of psychoanalysis is change, often understood as a continuation or resumption of an arrested course of development with overt behavioral change viewed as a product and sign of more fundamental, deeper change involving integration of previously unintegrated experiences. The goals of other psychotherapies may also include change, either intrapsychic or behavioral. They often emphasize relief or maintenance and support as well. When change is sought in other therapies, it may or may not be through insight (Wallerstein and Nemetz 1979).

In general, psychoanalytic psychotherapy tends to draw on the same array of techniques as psychoanalysis does. However, noninterpretive techniques are relatively more prominent in psychotherapy. Techniques are selected in regard to specific therapeutic goals. Insight is not the privileged route to change; what works best is preferred.

Psychoanalysis is more importantly differentiated from psychotherapy by the process than by the various techniques employed. The process of psychoanalysis involves the relatively spontaneous evolution of a regressive transference with the goal of its understanding and resolution. Most psychotherapies involve some activity designed to direct or influence the nature of the transference or to focus the therapeutic process on some other aspect of the patient's experience. This may be required by the patient's pathology, the therapist's capacity, or situational variables. The diagnostic assessment may lead the therapist to avoid analyzing certain infantile themes that contribute to a stable adult adjustment.

Psychoanalytic therapies have grown in significance so that they now constitute some of the major treatments of contemporary psychiatry. The range of disorders to which they have been applied has broadened, so that today a large number of individuals appear to be potential patients to some practitioner of psychoanalysis or psychoanalytic psychotherapy.

Much of the literature suggests that psychoanalysis is for healthier patients and psychoanalytic psychotherapy for sicker ones. This formulation is both misleading and impolitic. Not everyone who is analyzable is best analyzed. Positive indications for psychoanalytic psychotherapy would include the value of theories other than psychoanalytic theories in guiding the treatment, the value of noninterpretive and noninterpreted techniques, and goals not mediated by insight and integration.

Today, psychoanalytic psychotherapy is a valid treatment for a number of anxiety disorders, with the exception of obsessive–compulsive disorder. Psychoanalytic treatments are effective for nonpsychotic affective disorders and for mixed anxiety and depressive syndromes. They remain the dominant treatments for personality disorders, excluding antisocial personality.

Psychoanalytic psychotherapy is an important treatment for people who have problems in living, adjusting to stress, adapting to medical or psychiatric disabilities, complying with treatment regimens, rehabilitation, and other situations in which subthreshold personality characteristics limit happiness or successful adaptation. Psychoanalytic psychotherapy is probably at its best with patients who do not have disorganizing psychopathology but nonetheless have some psychological limitations, and therefore might be helped to function more effectively through treatment.

Finally, it is a powerful adjuvant combined with biological therapies for the chronic and persistently mentally ill, both by enhancing compliance with the other therapies and by alleviating secondary demoralization.

5 Psychoanalysis Today

Psychoanalysis proper, in the narrow sense, is only about 70 years old, emerging several decades into the "official" history of psychoanalysis. The treatment that was called psychoanalysis before that, from the late nineteenth century until the end of the second decade of the twentieth century, was more akin to contemporary short-term psychoanalytic psychotherapy than to contemporary psychoanalysis. It was not until 30 years into the history of the discipline, with the discovery of the central role of the analysis of transference and resistance rather than the reconstruction of childhood experiences, that patients routinely came back to a second year of treatment after their first summer vacation. Contemporary psychoanalysis is thus one of the offspring of the original treatment rather than the direct continuation of it (Michels 1994).

Psychoanalysis proper is far less prevalent than the various psychoanalytic and other psychotherapies that have been derived from it. It is more time consuming and more expensive, and therefore practical considerations often lead to the selection of a psychotherapy instead. However, the scope of psychoanalysis is broader, its potential greater, and, when the modification of pathologic character structure is the primary indication for treatment and the patient can tolerate the frustrations and deprivations of the process, it is often the treatment of choice. For some patients with serious character pathology, even though the prognosis is uncertain, it may be the only treatment that offers any real chance for change. Finally, it has remained the richest source of new concepts and ideas for the entire field of psychotherapy and for other areas of applied psychoanalysis as well (Cooper and Michels 1978; Michels 1988).

6 References

Altman L (1978) The dream in psychoanalysis. International Universities Press, New York

Arlow J (1969) Unconscious fantasy and disturbances of conscious experience. Psychoanal Q 38: 1–27

*Arlow J, Brenner C (1964) Psychoanalytic concepts and the structural theory. International Universities Press, New York

Auchincloss E, Michels R (1983) Psychoanalytic theory of character. In: Frosch J (ed) Current perspectives on personality disorders. American Psychiatric Press, Washington, DC, pp 2–17

Benjamin LS, Freidrich FJ (1991) Contributions of structural analysis of social behavior to the bridge between cognitive science and a science of object relations. In: Horowitz MJ (ed) Person schemas and maladaptive interpersonal patterns. University of Chicago Press, Chicago, pp 379–412

*Brenner C (1982) The mind in conflict. International Universities Press, New York

Breuer J, Freud S (1895) Studies on hysteria. In: Strachey J (ed) (1975) The standard edition of the complete psychological works of Sigmund Freud, vol 2. Hogarth, London, p 7

Cooper A, Michels R (1978) Psychoanalysis and future growth. In: Quen J, Carlson E (eds) American psychoanalysis: origins and development. Brunner/Mazel, New York, pp 189–209

Freud S (1895) Project for a scientific psychology. In: Strachey J (ed) (1975) The standard edition of the complete psychological works of Sigmund Freud, vol 1. Hogarth, London, pp 283–397

Freud S (1923) Two encyclopedia articles. In: Strachey J (ed) (1975) The standard edition of the complete works of Sigmund Freud, vol 18. Hogarth, London, p 235

Freud S (1925) An autobiographical study. In: Strachey J (ed) (1975) The standard edition of the complete psychological works of Sigmund Freud, vol 20. Hogarth, London, pp 3–70

Greenson RR (1967) The technique and practice of psychoanalysis, vol 1. International Universities Press, New York

Kandel E (1979) Psychotherapy and the single synapse: the impact of psychiatric thought on neurobiologic research. N Engl J Med 301: 1028–1037

Kardiner A (1945) The psychological frontiers of society. Columbia University Press, New York

Kris A (1982) Free association: method and process. Yale University Press, New Haven

Mahler MS, Pine F, Bergman A (1975) The psychological birth of the human infant: symbiosis and individuation. Basic Books, New York

Michels R (1983) Contemporary psychoanalytic views of interpretation. In: Grinspoon L (ed) Psychiatry update: the American Psychiatric Association annual review, vol II. American Psychiatric Press, Washington, DC, pp 3–11

Michels R (1985) Transference: an introduction to the concept. In: Schwaber E (ed) The transference in psychotherapy: clinical management. International Universities Press, New York, pp 13–19

Michels R (1988) The future of psychoanalysis. Psychoanal Q 57: 167–185

Michels R (1994) Psychoanalysis enters its second century. In: Winer J (ed) The annual of psychoanalysis, vol XXII. Analytic, Hillsdale, pp 37–45

*Michels R (1995) Basic principles of psychodynamic psychiatry. In: Schwartz H, Bleiberg E, Weissman S (eds) Psychodynamic concepts in general psychiatry. American Psychiatric Press, Washington, DC, pp 3–11

Michels R (1997) Psychodynamic psychotherapy in modern psychiatry. J Pract Psychiatry Behav Health 3: 95–98

Perry S, Cooper A, Michels R (1987) The psychodynamic formulation: its purpose, structure and clinical application. Am J Psychiatry 144: 543–550

Rapaport D, Gill M (1959) The points of view and assumptions of metapsychology. Int J Psychoanal 40: 153–162

Reiser MF (1984) Mind, brain, body: toward a convergence of psychoanalysis and neurobiology. Basic Books, New York

Sandler J, Dare C, Holder A (1992) Countertransference. In: Sandler J, Dare C, Holder A (eds) The patient and the analyst, 2nd edn. International Universities Press, Madison, pp 81–98

Shapiro T (1979) Clinical psycholinguistics. Plenum, New York
Stern D (1985) Interpersonal world of the infant. Basic Books, New York
Stinson CH, Palmer SE (1991) Parallel distributed processing models of person schemas and psychopathologies. In: Horowitz MJ (ed) Person schemas and maladaptive interpersonal patterns. University of Chicago Press, Chicago, pp 339–377
Wallerstein R, Nemetz S (1979) Conceptualizing the nature of the therapeutic action of psychoanalytic psychotherapy. J Am Psychoanal Assoc 27: 127–144

Part 2
General Psychiatry

CHAPTER 1

G.E. Berrios

The History of Psychiatric Concepts

1 Introduction

Like any other intellectual and practical discipline (or, indeed, institution or culture), psychiatry needs a stable conceptual scaffolding. Such a structure can be found only after much effort, well hidden from the view of ordinary users. All the latter can see is putative "empirical facts" and (of late) what is grandly called "evidence-based" guidelines for their diagnostic and therapeutic interventions. What is more, since the nineteenth century, decisions as to which "scaffolding" to accept and what "facts" and "rules" to follow have been taken by small, changing cliques and are well beyond the control of ordinary users. This vaticanisation of psychiatry started in continental Europe and has since moved across the English Channel.

The history of psychiatry itself began in similar social circumstances. Thus, basking by the fires of a weekend tryst, the princes of psychiatry might on some occasion have felt that the past of their subject ought to be sung. Knowing what was good for him, some shivering minstrel would have then spun a glorious yarn leaving much unsaid, particularly concerning the rights of self-styled princes to decide what scaffolding to use (for examples of such accounts, see Berrios 1994a). Mercifully, things have changed since and the psychiatric historian of today waits upon no prince. He or she wants to know about the conceptual and moral warrants of princely powers, about the sources of their money and about the real motivation of the meek narratives that extol their virtues.

This started in the 1960s, when Foucault (1972a) and his epigoni first introduced the idea that the historian should be the scourge of psychiatry. Their exaggerations and pseudo-historical productions, however, alienated for a time the psychiatric brotherhood. Things have improved of late, and the clinical historian once again attends the high table of science and is allowed to tell his or her *roman d'épouvante*. Through this, the profession has learned that the conceptual scaffolding of psychiatry is on occasions tinkered with for purposes other than the good of patients. Although by no means perfect, this new historical approach is imbued with the idea that history *must* also contribute to the development of clinical psychiatry.

2 Conceptual History

Of all the historiographic approaches available, the one referred to as "conceptual history" is the most apposite to deal with scaffoldings and the other serious matter treated in this chapter. Such a method will need to focus on four interacting frames: descriptive psychopathology (DP), aetiological theory, pathogenesis and taxonomy. DP refers to the language of description; aetiology to the causes of disease (Berrios 1984a); pathogenesis to the manner in which disrupted brain mechanisms generate mental symptoms; and taxonomy to the rules governing the classification of disease. Ideological forces within and without psychiatry have contributed to the generation and maintenance of these frames. DP (i.e., "semiology"), for example, owes much to eighteenth-century linguistics and the theory of signs (Landre-Beauvais 1813; Lanteri Laura 1966; Juliard 1970; Barthes 1972), and aetiological theory and pathogenesis to developments in general medicine, microscopy and nineteenth-century psychological theory (Canguilhem 1966; Laín Entralgo 1978; Albarracín Teulón 1983; Berrios and Porter 1995), while taxonomy is partly based on seventeenth- and eighteenth-century metaphors of order (Whewell 1857; Boyer 1873; Delasiauve 1861; Larson 1971; Georgin 1980; Slaughter 1982). Little else will be said on the latter, as it is covered in Chap. 2 (this volume, Part 1).

It goes without saying that these interacting frames only gain full meaning when placed on the canvas of the nineteenth-century practice of alienism. The resulting fabric includes explanatory elements ranging from the evolutionary and biological to the sociopolitical – as rightly required by social historians (e.g., Foucault 1972a,b; Dörner 1969; Blasius 1980; Scull 1979; Donelly 1983). Conceptual historians start from the premise that the "meaning" of mental disorder is as dependent upon knowledge of its biological origins (i.e., the source of the distorted biological signal) as it is upon knowledge of the psychosocial envelopes. In other words, most "psychiatric" phenomena are the final expression of a biological signal modulated by personal and cultural grammars (Marx 1970; Berrios 1984a).

It follows that the stability of DP and of the disease categories that populate psychiatry is a function of the rate at which biology and language change (Daumezón 1957; Berrios 1994a); in other words, descriptions and diagnoses are kept stable as much by symbols, myths and other constructs (Devereux 1980) as they are by actual biological invariants (Berrios 1994a). Indeed, psychiatrists have not yet developed accurate ways of deciding on the relative contribution of each. For example, while "manipulative behaviour" (Mackenzie et al. 1978) may be fully the result of human interaction – and hence be "social" in origin – "grand mal seizures" (Berrios 1984b), "delirium" (Berrios, 1981a) and "hallucination" (Berrios, 1982b) can be considered as fundamentally "biological" phenomena.

Pre-nineteenth-century literature is rich in descriptions of insanity (Postel 1984; R. Porter 1987;

MacDonald 1981), but little is known about the conceptual frames that propped them up. More is known about the nineteenth century, but the three great changes that transformed the nature of psychiatry remain only partially understood (Berrios and Porter 1995). These changes are the following:

1. The transformation of "insanities" into "psychoses" (on this, see History of Psychiatry 1996)
2. The narrowing (and eventual disappearance) of the "neuroses" as a general category (López-Piñero 1983)
3. The fragmentation of the old monolithic descriptions of insanity into what are now called mental "symptoms" (Berrios 1996)

This chapter addresses some of these issues and chronicles the interaction between theory, observation and the biological phenomena of madness. To avoid confusion, it will keep separate the history of words, behaviours and concepts. It will also take as proven that the protagonists of the tale were men with families, political interests, fears and ambitions and that many of their choices were determined by "non-cognitive" factors. However, it will also assume that they were also ethical beings who, when faced with real patients, exercised (as current psychiatrists would like to be thought of as doing) objectivity and a modicum of descriptive freedom; hence their writings can be considered as scientific documents.

3
Development of Descriptive Psychopathology

3.1
Definitions

DP is defined here as a "language" comprising a syntax, lexicon and rules of application (Berrios 1996). Because DP imposes order on a universe of complex behavioural forms, it is also a "cognitive system". For each term (purportedly dealing with a self-contained piece of behaviour or "symptom"), DP is expected to contain "caseness" rules, i.e., ways of determining whether a given "symptom" is present or absent (Berrios and Chen 1993). Symptoms (conceived of as referents or signifiers) are defined by means of decision-making routines profitably analysed in terms of signal-detection theory (Macmillan and Creelman 1991). At a basic level, symptoms are assumed to result from a "fracturing" of insane behaviour, although things may be more complex than that (on symptom formation, see Berrios et al. 1995).

Consequently, observers may differ in the way in which this task is done; indeed, before estimations of inter-rater reliability (e.g., the kappa values; Shrout et al. 1987) were available, nineteenth-century alienists used consensual (qualitative) rules to determine *when* a symptom was present. For example, they might appeal to the higher tribunal of "common sense", to the "obvious" nature of some disordered behaviours and occasionally to intuition and the "clinical eye". When such aids failed, as not infrequently happened in court, particularly in relation to the predication of intentionality (Smith 1979), impasses might occur as to how symptom recognition could be achieved (Helmchen 1985).

3.2
Problems

Absence of a recognisable DP is a striking feature of psychiatric discourse *before* the nineteenth century. However rich in literary detail, earlier references to insanity (or germane terms such as dementia) were made in terms of "holistic" categories (Berrios 1987a,b). One explanation for this may be that detailed descriptions were unnecessary or inconvenient, because "insanity" in those days fulfilled a different social or legal function (Beaugrand 1865). For example, any assumption that there might be a continuity between mad and normal behaviour – often made by DP – would have threatened the "all-or-none" concept of "total insanity" which was so important before the nineteenth century. Furthermore, since Greek times, psychiatric categories were founded on descriptions of polar "overt" behaviours (Berrios 1987b) and of social competence (Platt and Diamond 1965) and left little room for nuances and transitions.

The creation of DP took about 100 years. It started during the second decade of the nineteenth century and was completed just before the Great War. It has changed little since. This means that the success of current clinical and research endeavours depend to no small extent on the quality of a conceptual machinery tooled during the nineteenth century (Berrios 1983, 1995; Berrios et al. 1995). The twentieth century has no doubt refined the psychiatric discourse by the introduction of techniques of statistical calibration and decision-making, but the historical question remains: how did nineteenth-century alienists manage to extract, based on longitudinal observation of what were often institutionalised patient cohorts, stable descriptions and classifications? Five factors will be explored in this regard:

1. The descriptive and medico-legal obligations of medical officers gradually introduced in nineteenth-century asylums
2. The availability of psychological theories

3. The changing importance of the notion of "sign" and "symptom" in medicine
4. The introduction of *subjective* symptomatology
5. The adoption of a *temporal* dimension in the description of abnormal behaviour

3.3 New Descriptive Needs

During the early years of the nineteenth century, a drive to build asylums for the insane simultaneously appeared in various European countries (Walk 1964). Once built, these institutions created social and scientific consequences of their own. First of all, they encouraged the accumulation of the mentally ill within confined physical spaces. Overcrowding and lack of medical care led to decimation through intermittent infections and highlighted the need for a regular medical presence. In Great Britain, this was made good by the 1828 Asylums Act (Jones 1972). The incorporation of medical practitioners into asylums generated, in turn, new changes. They brought with them the habit (and the medico-legal obligation) of monitoring and documenting clinical change. As long as this need related to the physical state of patients there was no problem, as, during the early nineteenth century, recognised methods of history-taking already existed (see Laín Entralgo 1961).

This was not so with regards to mental state. Perusal of pre-1840 clinical log-books shows a poverty of description consonant with the absence of "symptom lists". Early asylum doctors were thus forced to improvise and borrow, and their activity can be said to be an important factor in the creation of a "semiology" of mental illness. After 1850, however, a change is noticed in the quality of descriptions.

In this regard, it is important to point out that, although on occasions accounts of madness can be found *before the nineteenth century* that include elegant descriptions of mental states (e.g., Battie 1758; Burton 1883; Diethelm and Heffernan 1965; Hunter and Macalpine 1963; MacDonald 1981), they cannot be said to amount to a *common* language of description (i.e., shared by all physicians), nor were they intended to be. What emerged from the nineteenth-century descriptive enterprise is totally different, namely a common language based on an *analytical and pictorial epistemology* dealing with symptoms *independently* and assuming that the *same* symptom could be seen in *different* forms of madness. The creation of such language of description (DP) led to a shift in the *perception* of madness. (It could be claimed, of course, that it was the other way around – see Foucault 1972b – namely, that changes in the perception of madness, e.g., its medicalisation, led to treating these phenomena as if they were brain lesions expressed in signs and symptoms. This might have been so at a general level, but the point made here is that, once the old monolithic notion of insanity was broken up, semantic interpretation concentrated on individual symptoms and on the way they clustered together. Consequently, the general semantics of insanity became unimportant.) Changes in the "semiology" of medicine are no doubt important in this process. But the sine qua non for this change was the availability of psychological theories in terms of which behavioural profiles could be constructed.

3.4 Psychological Theories

DP started to emerge in France during the second decade of the nineteenth century. It did so in the way in which specialised books began to use illustrative case histories and to include sections on "elementary" symptoms. This is the main difference between the earlier books by Pargeter, Arnold, Crichton, Haslam, Rush, Heinroth or Pinel (predominantly holistic and taxonomic) and those published after the 1930s by Prichard (1835), von Feuchtersleben (1847), Bucknill and Tuke (1858), Falret (1864), Griesinger (1867), Krafft-Ebing (1893), Séglas (1895) or Chaslin (1912).

Melancholia, mania, phrenesis, delirium, paranoia, lethargy, carus and dementia were the main diagnostic categories inherited by the nineteenth century. By the 1850s, these clinical notions had undergone fractionation (Ey 1952; Berrios 1977), and recombinations of the fragments (i.e., of the new mental symptoms) gave rise to new clinical entities, many of which have survived to this day. Delirium was one of the few old categories to escape unscathed. Carus, phrenesis and catalepsy were not so lucky and disappeared altogether. Many old terms survived, but lost their old contents and were filled with new ones (e.g., melancholia and mania; Berrios 1988a,b). The fractionation of the old clinical categories of madness occurred along planes of cleavage contained in the templates of the mind that faculty psychology and (less obviously associationism) made available.

3.4.1 *Faculty Psychology*

Faculty psychology, an ancient (and recurrent) view of the structure of the mind (Blakey 1850) underwent a revival towards the end of the eighteenth century. In fact, the early development of DP in France can be partially explained by the fact that, in their reaction against associationism, alienists in this country opted for a form of faculty psychology (e.g., Damiron 1828;

Dwelshauvers 1920; Ravaisson 1885). Influenced by the Scottish philosophy of common sense (Boutroux 1908; Grave 1960) and under the leadership of Maine de Biran (Drevet 1968; Moore 1970), Royer-Collard (Swain 1978), Cousin, Jouffrey and Garnier, the French philosophical establishment accepted a "functionalist" view of the mind. This led to a shift from the reigning Condillacean sensationalism to a new essentialist view emphasising a form of Biranean "inner experience" (Royer-Collard 1843; Losserand 1967).

Phrenology was one of the nineteenth-century intellectual disciplines inspired by faculty psychology (Lanteri Laura 1970). Indeed, it has not been sufficiently emphasised that the conceptual basis of craneology (later called phrenology by Spurzheim) was but an "anatomised" form of faculty psychology. By suggesting "trait profiles", the phrenological approach allowed typological theories of character (later called "personality"; Spoerl 1936). These profiles were de facto conglomerates of faculties, and long after phrenology had been discarded, faculty psychology remained a conceptual matrix for nineteenth- and twentieth-century views on psychiatric taxonomy and localisation (Young 1970; Clarke and Jacyna 1987; Radden 1996).

3.4.2 Kant's Version of Faculty Psychology and the Nineteenth Century

Kant's tripartite concept of the mind was almost contemporary to that of the Scottish philosophers (Hilgard 1980; Radden 1996). Influenced by von Wolff via Knutzen, his teacher at Königsberg, and via Tetens, whose version of the "tripartite view" he was to follow (Windelband 1948), Kant rebelled against the "dogmatic rationalism" of his teachers (Brett 1953). Following the views of Baumgarten (Buchner 1897), Kant suggested that affect (and semantic appurtenances) constituted an independent faculty. In the *Critique of Judgement*, Kant (1914) put forward the view that the "three faculties are irreducible and cannot be derived from a common root" and believed that emotions had a causal role in mental disease (*Krankheiten des Gemüts*; Kant 1974). As Mora (1975) has remarked, "at the end of the 18th century, Kant came to think of mental illness as a form of weakness of the faculties, in line with his belief in Faculty Psychology. . ." (p. lix).

In addition to the *Anthropologie*, another early tract dating back to 1764 carried some of Kant's ideas on mental disorder (see Jalley et al. 1977). Influenced by Locke, Kant differentiated in this work between delusions originating from faulty perception and faulty thinking and suggested a separation between perceptual and thinking functions. The impairment of the former he called "hallucination" and of the later "delusion": "so far (in the hallucinations) the thinking faculty is not involved at least not necessarily, and the failure occurs in the empirical notions. . . as opposed to this, when thinking is impaired the conclusions from experience are wrong; the first degree of this disturbance is called delusion" (p. 225, Jalley et al. 1977). Sauri (1969) noted that the "ideal schema" of the faculties allowed Kant to classify delusional thinking no longer in terms of content, but as a manifestation of a disorder of intellectual function.

Little has been written on the influence of Kant on psychiatric thinking (Leary 1982). Leibbrand and Wettley (1961) remarked upon the fact that Kant's contribution was theoretical and was made in spite of the fact that he had no first-hand knowledge of clinical disorders. Evidence of the long-term influence of Kant's views in nineteenth-century psychiatry can be found, however, in an acknowledgement by Magnan and Serieux (1911), who referred to him as a "precursor" of the concept of *délire chronique à evolution systematique* (p. 607). Jaspers was also influenced by Kant (Kauffmann 1957; Stierlin 1974; C. Walker 1988; Berrios 1992a): "Kant became for me, and remained for me, the philosopher par excellence" (Kauffman 1957, p. 407). In his autobiography, on the other hand, Jaspers (1957) admitted that, in his university years, he found the German philosopher "hard to understand". Likewise, in *General Psychopathology*, Kant is only alluded to in relation to the role played by alienists in deciding on the legal responsibility of the criminal offender. The Kantian view has not been without critics; for example, from a Marxist perspective, Merani (1976) commented upon his over-descriptivism and anti-experimentalism.

3.5 Associationism

3.5.1 Before the Nineteenth Century

The atomistic model of associationism provided the epistemological basis for the development of seventeenth- and eighteenth-century science (Schofield 1970; Hoeldtke 1967). The notion of "simple idea" (the psychological counterpart of the Newtonian "atom") was used by Locke (1959) as a "unit of analysis" and is at the basis of the "laws of association", a combinatorial algebra in terms of which the mind reconstructed the world from simple experiences. This view found its ideal model in perception; in later years, this was to favour certain functions (e.g., the intellect) to the detriment of others, such as the emotions (Berrios 1985b; Gardiner et al. 1937). The intellectualistic definition was challenged during the

early nineteenth century. Thus, when describing cases with "lesions of the function of the will. . . whose symptoms appeared enigmatic upon the definitions of mania given by Locke and Condillac" (p. 102), Pinel (1809) was forced to write: "one can admire the writings of Locke, and agree nonetheless that his views on mania are incomplete in that he considers it to be always associated with a delusion. . . I thought likewise until resuming my research at Bicêtre; and I was not a little surprised to find many maniacs who at no period gave evidence of any lesion of the intellect, but who were under the control of a sort of instinctive fury, as if the affective faculties alone were impaired" (Pinel 1809, p. 156).

3.5.2 During the Nineteenth Century

For all its influence, pre-nineteenth-century associationism was more epistemolo ;ical than psychological. However, the books by Brown (1828) and Mill (1829) marked a change of emphasis: in spite of their heavy philosophising, these works also undertook to explain behaviour. John Stuart Mill (Warren 1921) and Alexander Bain (Bain 1859; Greenway 1973) developed this further.

A similar situation arose in France, where associationism was confronted, early in the nineteenth century, by faculty psychology and by the anti-analytic views imported from Scotland. The associationism of Condillac and Bonnet had been epistemological in intention, and development of its psychological aspects had to wait until Destutt de Tracy and the "ideologues" arrived (Destutt de Tracy 1818; Mora 1981). Upon becoming "psychological", associationism entered into conflict with faculty psychology, whose direct psychological usefulness (as in its application to phrenology and in the classification of mental disease; e.g., Esquirol 1838 inspired by his teacher Laromiguière 1820) had been obvious from the start.

The situation was also similar in Germany (Ribot 1885), a good illustration of which is found in the work of Herbart (1884), with his emphasis on education and psychology (Boring 1950; Watson 1963; Fritzsch 1932). These views influenced Griesinger (Ackerknecht 1957; Wahrig-Smith 1985; Verwey 1985), and through him the analytical tradition was passed on to Krafft-Ebing, Meynert and Wernicke. For example, it is reflected in the latter's "connectionist" view of aphasia. During the second half of the nineteenth century and influenced by Fechner's views on the correlations between stimuli and intensity of sensation (and the underlying metaphysics of the mind–body relationship) (Marshall 1982), German psychiatrists sought to establish objective experimental descriptions of some of the symptoms of insanity. Kraepelin (1896, 1983), a disciple of Wundt and follower of his particular brand of associationism, carried over this tradition into the twentieth century.

Most alienists accepted the analytical epistemology of associationism and adopted the important concept of "unit of analysis" as applied to behaviour and experience. Symptoms such as obsessions, delusions and hallucinations became the indivisible units of madness (Berrios 1996). This was consolidated in the work of Chaslin (1912, 1914) and Jaspers (1913). However, taxonomy remained based on faculty psychology, and this created a tension in the evolution of DP.

3.6 Surface Markers of Disease

As mentioned above, psychiatric "semiology" was born out of the observation of asylum patients, all of whom suffered from severe functional or organic psychoses. During the nineteenth century, the "neuroses" did not yet fall within the purview of alienists (López-Piñero 1983; Drinka 1984), so their "symptoms" and other clinical features contributed little to psychiatric semiology. There is also evidence that the symptomatology of the psychoses was modelled upon that of delirium (Calmette 1874; Roubinovitch 1896; Berrios 1981a).

The notion of "sign" is not free from ambiguity (Martinet 1973; Land 1974; Malberg 1977; Manetti 1993), particularly in the field of medicine (King 1968; Barthes 1972) and psychopathology. In some cases, it might relate to an underlying dysfunction rather directly in the same way in which smoke relates to fire (e.g., disorientation) (Berrios 1982a), i.e., analogous to Pierce's notion of (1931–1935) "index"; in others, it "signifies" a behavioural form (e.g., manipulative behaviour; what Pierce called "symbols"). It is plausible to believe that "indices" are more likely to reflect a specific neurobiological disorder than "symbols".

3.6.1 Assumptions and Concepts

Since descriptive psychopathology has changed little since the nineteenth century, a historical analysis of its conceptual bases should help to understand the enduring quality of some of the symptoms it generated (e.g., delusions, hallucinations). The following issues will be briefly touched upon in this section:

- "Form and content" of symptoms
- The role of numerical descriptions
- conographic representations
- The relationship between mental disease and "time"

- The incorporation of subjective information into the definition of mental illness

3.6.2 Form and Content

The distinction between "form" and "content" of a symptom is one of the enduring contributions of nineteenth-century DP. The Aristotelian *eidos* referred to the essence or common character of objects and is one of the origins of the current concept of "form" (Emerton 1984; Ferrater Mora 1958). With some modifications, the Aristotelian notion of form lasted well into the seventeenth century, when Bacon proposed that "form" might be simply considered as a synonym for "figure" (see Bacon 1858, Book II, Para 17, p. 474). Kant in turn suggested that the sense modality in which a perception took place, in conjunction with its cognitive network, should be called "form" (Abbagnano 1961).

Nineteenth-century DP, and indeed Jaspers (1963, pp. 58–59) at the beginning of the twentieth century, followed the Kantian definition: "form must be kept distinct from content which may change from time to time, e.g., the fact of a hallucination is to be distinguished from its content, whether this is a man or a tree . . . Perceptions, ideas, judgements, feelings, drives, self-awareness, are all *forms* of psychic phenomena; they denote the particular *mode of existence* in which content is presented to us. It is true, in describing concrete psychic events we take into account the particular content of the individual psyche, but from the phenomenological point of view it is only the form that interests us. . ." (for an excellent discussion of form and content in Jaspers' work, see C. Walker 1993).

To this day, "form" refers to structures that act as warrants of symptom stability, i.e., "constancy" elements which render mental symptoms recognisable in time and space. The notion of "form" is easier to understand in physical medicine. Colour, sound, surface, solidity, smell and temperature provide the natural media in which "form" achieves expression and stability (Laín Entralgo 1982). Inspired by medical semiology, alienists also expected to identify signs of insanity which were stable, public and observable. To do so, they committed themselves to the epistemology of "natural kinds" (Mill 1898; Markman 1994), i.e., to the view that the solutions of continuity between symptom and symptom (whether mental or not) existed in re. This led to a loss of interest in "contents": symptoms were just signposts to a brain lesion. This neglect of the semantic aspects of symptoms impeded the development of a comprehensive model; as a result, a theory that catered exclusively for "contents" was to develop with a vengeance at the end of the century (Ellenberger 1970).

However, in clinical *practice*, the "contents" of symptoms were never neglected altogether, and alienists made use of such information to establish aetiological connections between the subject's illness and his or her past. By the second half of the century, and before Janet or Freud had come on the scene, associations between content and past history had been already conceptualised as cause–effect chains. For example, it was felt that the content of a delusion or a hallucination, or the form of a hysterical conversion, might tell us something about the circumstances in which the symptom was first acquired (e.g., trauma, financial loss, infection; Bucknill and Tuke 1858). These putative cause–effect chains acted as veritable second-order "psychological" explanations (Billod 1861; Dagonet 1881a,b; Despine 1876), and their ubiquitous presence in nineteenth-century psychiatric practice calls into question the view that during this period alienists only entertained "somatic" aetiologies (e.g., Jacyna 1982). Not surprisingly, these psychological accounts matched beliefs in popular psychology. When the neuroses, particularly hysteria, came within the purview of alienists (something which only occurred towards the end of the century), they found that the content of a sign could reveal a great deal about the circumstances of its acquisition (e.g., Charcot's "idea" expressing itself in the symptom; Charcot 1971; Owen 1971; Bannour 1992).

The emphasis on "form" also led to changes in the type of explanation; for example, the "form" of a hallucination highlighted the informational value of the sense modality in which it occurred, and this in turn suggested a brain address (Tamburini 1881).

3.7 Numerical Representation and Measurement

The mathematisation of the natural world started in Europe during the seventeenth century (Dijksterhuis 1961). However, the Newtonian paradigm did not greatly affect psychological thinking during this period, as both Cartesian and Lockean psychology coincided on the view that numerical descriptions did not apply to behaviour (Moravia 1983).

The suggestion that "psychometry" (i.e., the measurement of psychological experience) was possible and desirable is attributed to Christian von Wolff. While describing ways of assessing the magnitude of pleasure and displeasure, he stated in a footnote: "these theorems belong to 'psychometry' which conveys a mathematical knowledge of the human mind and continues to remain a desideratum" (Ramul 1960). Ramsay, Baumgarten, Crusius, de Maupertuis, Buck, Mendelssohn and Ploucquet are mentioned among other eighteenth-century writers who prepared the

conceptual terrain for the advent of measurement in psychology; no one, however, seems to have carried out actual experimental work.

The path to numerical description, originally suggested by von Wolff (and opposed by Kant and Comte), was continued by Herbart, who suggested the development of a "statistics" of the soul (Ribot 1885). This conceptual change made the work of Müller and Du Bois Reymond easier (Rothschuh 1973), and the instruments (Sokal et al. 1976) they designed facilitated in turn the ideas of Weber and Fechner.

The introduction of quantification in medicine follows a different path (Underwood 1951; Shryock 1961; Murphy 1981; Matthews 1995). Numerical management of data was already familiar to pre-nineteenth-century epidemiologists and administrators (e.g., the bills of mortality), but inferential interpretations were scanty (Perrot and Woolf 1984). During the nineteenth century, statistical analysis of data, based on probability theory (Hilts 1981; T.M. Porter 1986; E.S. Pearson 1978) was started in earnest (Esquirol 1838). This is clear in the work of Gavaret (Wulff et al. 1986), Louis (Ackerknecht 1967), Radicke (1861), Renaudin (1856) (in this paper Renaudin reported on the positive view on the role of statistics taken during the 1856 Statistics Symposium in Paris) and Esquirol (1838), who also made use of inferential percentages.

Numerical descriptions extended only gradually to other areas of psychopathology, and this occurred around the middle of the century (Parchappe 1856). There is little historical evidence to suggest that any serious effort had been made before the 1850s to measure personality traits (Boring 1961; Zupan 1976; Bondy 1974; for a classical example of the use of mathematical and statistical reasoning in memory research, see Ebbinghaus 1964). This is surprising, as the ideas of Gall and Spurzheim had made available to psychology a conception of individual differences susceptible to quantification (Damiron 1828; Spoerl 1936; Lesky 1970), and phrenology had sought to establish correlations between anatomical and psychological magnitudes (Lanteri Laura 1970). After the 1830s, growing opposition to phrenology appeared in Europe (Cantor 1975; Cooter 1976). This might have discouraged alienists from espousing, at least publicly, Gall's interesting "modular" view of the mind (Fodor 1983; Shallice 1988).

3.8 Psychopathology of Non-verbal Behaviour

The great diagnostic categories of the past (mania, melancholia, phrensy, lethargy, catalepsy) relied on the observation of what the individual did, looked like and thought rather than on what he or she "felt". This is particularly so with regards to mania and melancholia (Berrios 1988a,b). The idea that these two notions are forerunners of the clinical categories currently bearing the same name is wrong. Before the nineteenth century, there was very limited historical evidence that "elation" or "sadness" (i.e., pathological mood) were the central "criteria" for the *medical* definition of mania and melancholia, respectively (Couchoud 1913; Heiberg 1927; Drabkin 1955; Flashar 1966; Siegel 1973; Simon 1978; for an illustration of the view that "melancholia" has more or less retained its meaning since ancient times, see Jackson 1986). The fact that literary uses of melancholia included a reference to "low spirits" (Babb 1951) is, of course, no argument in favour of a continuity in medical meaning.

The use of overt behaviour as the métier of psychopathological description seems to have been started by the Greeks (Simon 1978; Roccatagliata 1973). Symptom-mapping was influenced by their views on what constituted harmonious behaviour. The categories they created became the archetypal forms of insanity which, with little change, lasted well into the eighteenth century. Interest in the description of overt behaviour was renewed during this period (Bühler 1968), particularly in the study of facial expression in normal subjects, in what came to be known as the science of physiognomy (Lavater 1891; Mantegazza 1878; Caro Baroja 1988). During the late eighteenth century, Parsons (1747) attempted to establish correlations between emotions and gestures. The application of these techniques to pathological states gave rise in the fullness of time to an iconography of madness. This in turn influenced the way in which the mentally ill were perceived (Gilman 1982). For example, exaggerated or distorted facial expressions were thought to indicate the intensity of the underlying derangement. During the nineteenth century, a change occurred in the way in which the insane were depicted. The old stereotyped ways of Hogarth and Tardieu gave way to a more "realistic" approach; after 1839, this was made possible by the invention of the daguerreotype. This introduced another bias in the type of photographic records kept for posterity, as the long exposure times required by this technique encouraged its use in "static" conditions (e.g., stupor).

Likewise, the view that there was a one-to-one correlation between inner states and gestures became less acceptable; indeed, it was believed that the two factors could be dissociated. This in turn led to the idea that insanity could be either concealed or simulated. Morison, Laurent and the great Pierret, for example, developed a complex theory of "mimia" and "paramimia" during the second half of the century (Régis 1906). Charles Darwin's (1904) interest in this issue is also well known (Browne 1985).

3.9
Disease and the Time Dimension

Until the early nineteenth century, the descriptions of insanity had been, in a real sense, atemporal and identification of symptoms in cross-sectional analysis sufficed to make a diagnosis (Arnold 1806).

This resulted from the belief, based on the ontological view, that madness was an irreversible process. This view reflected similar notions in the field of physical disease. Referring to this period, Charcot (1881) wrote: "disease was formerly considered as being independent of the organism, a kind of parasite attached to the economy" (p. 4). Disease, therefore, was not considered as contained within the temporal and spatial constraints of the body (for a discussion of this problem, see King 1982, pp. 131–183; Haas 1864). The fact that subjects might occasionally show "normal behaviour" (which was then called the "lucid interval") was not considered as contradictory evidence, as patients could "stifle their disorder" (Haslam 1809).

Asylum psychiatry allowed for the first time the longitudinal observation of groups of patients. This forced a change in the observational framework, and as a result a *temporal dimension* was gradually introduced in the 1850s (Lanteri Laura 1972, 1986; Del Pistoia 1971). The longitudinal approach encouraged major changes in the concept of mental illness. Thus information obtained from longitudinal observation could be used to correct or modify earlier diagnosis. Kahlbaum (1863) used this to good advantage in his definition of mental disease, and for the first time a distinction between acute and chronic insanity was made (Berrios 1987b; Beer 1996a,b). By the end of the century, "duration" had become the central category in the analysis of disease. For example, in Kraepelin's work, the evolution and outcome of a condition were crucial to making a diagnosis (Hoff 1985; Berrios and Hauser 1988); indeed, it has been claimed that he purposely went to Dorpat to gain experience with chronic hospitalised patients (Marx 1980). It has also been suggested that the fact that the Dorpat patients spoke no German (Kraepelin did not speak Russian) forced him to choose "more objective" signs to define "dementia praecox" (Berrios and Hauser 1988).

3.10
Incorporation of Subjectivity

The incorporation of subjectivity is perhaps the single most important contribution of the nineteenth century to DP. As was mentioned above, pre-nineteenth-century descriptions of insanity relied heavily on the observation of overt behaviour, psychosocial competence and cognition. During the early nineteenth century, changes in psychological thinking, particularly in France, led to the acceptance of the "contents of consciousness" as a legitimate field of inquiry (Dwelshauvers 1920; Royer-Collard 1843). This opportunity was seized by alienists, who were at the time searching for additional sources of clinical information. Methods of eliciting and recording data were sought (Laín Entralgo 1961; Helmchen 1985), and the mental state assessment, in its dialogical form, appeared during this period.

Moreau de Tours can be considered as one of the protagonists in this development (Bollote 1973; Ey and Mignot 1947). In his *Psychologie Morbide*, Moreau (1859) attempted to legitimate the value of subjective information (Pigeaud 1986). At the same time as French psychiatry was paying great attention to mental contents, British psychiatrists were worrying about "morbid introspections" as a potential cause of mental disorder (Clark 1988). In addition to the analysis of hallucinatory images and delusional contents, this new source of symptoms also included a rich gamut of emotional, affective and volitional experiences (Lanteri Laura 1984).

During the early stages of this process, much emphasis was put on the "form" of the newly discovered symptoms. For example, efforts were made to decide whether hallucinatory voices were bilateral or unilateral, recognisable, single or multiple, etc. (Parish 1897; Gurney 1885). Towards the second half of the century, the influence of Brentano (Fancher 1977) redirected attention towards the "content" of the symptom. The psychodynamic doctrines can be said to be an extreme illustration of this trend (Bercherie 1983, 1988). This is not the place to discuss in any detail the changes in the history of psychology that legitimated the use of this new experiential source (Berrios 1987c). Suffice it to say that they pertain to the appearance of a *psychological* notion of consciousness (Burt 1962; Bastian 1870; Viziolo and Bietti 1966; Ey 1966) and to the acceptance of the epistemological value of introspection (Boring 1953; Danziger 1980).

The acceptance that purely subjective experiences could be mental symptoms encouraged the re-definition of some mental diseases. For example, the newer notions of melancholia and mania were made possible by the availability of direct experiential information with regards to mood states and emotions (Berrios 1988a,b). Likewise, the concept of paranoia was to reappear in the 1860s, this time based on the presence of delusional experiences (Lewis 1970a). The various types of stupor, until then lumped together, were

also classified according to whether or not there was amnesia for the episode (Berrios 1981b). Subjective data were also used to identify "subtypes" of insanity, and "symptomatic classifications" proliferated (e.g., religious, metaphysical and erotic mania; Berrios 1994b).

4 Development of the Concept of Mental Disease

Historians of psychiatry are struck by two major clinical changes that occurred during the nineteenth century: one concerns the metamorphosis of the insanities into psychoses, and the other the "cross-over" of meaning and aetiology that took place between the psychoses and neuroses (Beer 1996b). For reasons of space, only the first will be dealt with in this chapter. The change from insanity to psychosis was made possible by the development of a new concept of disease in general medicine (Ackerknecht 1967), by the availability of new ways of defining behaviour (Berrios 1987b) and by the appearance of new taxonomic principles for the classification of biological entities (Georgin 1980; Berrios 1987b).

4.1 Clinico-anatomical View

Already present in embryonic form in the work of Sydenham (Laín Entralgo 1978), the clinico-anatomical view reached full development during the early nineteenth century. By then, the view that symptoms were signs of an underlying anatomical lesion had been fully adopted (López-Piñero 1983; Ackerknecht 1967). As the century wore on, the concept of lesion was successively re-defined as inhabiting organs, tissues, and finally cells (Albarracín Teulón 1983; Canguilhem 1966; Foucault 1972b). The failure to identify *anatomical* lesions in many diseases led, during the second half of the nineteenth century, to envisaging "lesion" in *physiological* terms, and this led to concepts such as "irritation" and "inhibition" (López-Piñero 1983; Smith 1992). In this way, the concept of "functional" lesion was born. A logical continuation of this was the development of the notion of "psychological" lesion, which took place by the 1890s. Janet and Freud helped this view to be incorporated into psychiatry (López-Piñero and Morales Meseguer 1970).

4.2 Psychological Definitions of Behaviour

Associationism and faculty psychology helped alienists to map behaviour, to seek new sources of symptoms (e.g., subjectivity) and to develop new classifications. They also suggested ways of fragmenting behaviour. Consciousness was metaphorically described as a "theatre" with its "contents" captured by means of introspection. The acceptance of a view of the mind as a set of functionally autonomous modules provided a natural framework of classification. For example, and as mentioned above, Pinel (1809) and Prichard (1835) abandoned Locke's "intellectualistic view" of madness (still extant in Arnold's distinction between ideal and notional insanity; Arnold 1806) and turned to faculty psychology. By the end of the century, symptoms were considered as either exaggerations of normal behaviour (continuity view) or as new forms (discontinuity view) (Dumas 1908; Deshaies 1967; Griesinger 1865; Delasiauve 1861).

4.3 Taxonomic Changes

The philosophy of medical taxonomy also changed during the early nineteenth century. The botanical principles implemented by Linné, Sauvage, Cullen and others (Bowman 1975; Temkin 1965; Faber 1923) were replaced by an empirical approach based on the following:

1. Frequential analysis of symptoms (Griesinger 1865)
2. Aetiological speculation (Morel 1860)
3. Knowledge, later in the century, of the natural history of the disease (Remond and Lagriffe 1902; Baillarger 1853; Foville 1872; Vié 1940; Desruelles et al. 1934; Menninger 1964; de Boor 1954; Goldstein 1988)

The early nineteenth-century view that intellectual, emotional and volitional mental functions could be separately affected by disease offered a solid taxonomic principle. Thus the "intellectual insanities" became the nucleus around which the notions of schizophrenia and paranoia were to crystallise; the emotional insanities served the same function in relation to mania and depression (Berrios 1988a), and the volitional insanities with regards to the psychopathic disorders (Werlinder 1978; Ey 1978; Berrios 1993).

While the eighteenth-century view of the insanities was cross-sectional and related to specific life events (Postel 1984), the nineteenth-century view became longitudinal, particularly after Kahlbaum, Wernicke and Kraepelin. In the event, the latter two were to

developed rival systems of classification (Donalies 1971). Had Wernicke lived longer, the classification of the functional psychoses would be very different today, partly because his was a model based on a veritable physio- and psychopathology. However, Kraepelin was the victor, and in his scheme of things the number of insanities was drastically reduced to two psychoses characterised by stable and overlapping symptom clusters (Hoff 1994). Organic aetiology and recognisable natural history and prognosis became the final diagnostic criterion (Berrios and Hauser 1988). From the start, the finding of clinical anomalies ("intermediate" cases) challenged the Kraepelinian dichotomy, and Kraepelin was to abandon it in later years (Kraepelin 1920).

4.4
The "Combined Psychoses" Debate

Current preoccupations with the notion of co-morbidity (generated by the peculiar diagnostic structure of DSM IV; e.g., Wittchen 1996) are not new, and a similar debate took place during the early twentieth century, when the question was asked whether two independent psychoses could simultaneously affect the same individual. Jaspers (1963) clearly stated that "the idea of disease entity leads one to expect that no more than one illness can be diagnosed in any one person. Where a schizophrenic process is present we duly suppose that we should hold it responsible for all the symptoms, but that is a presupposition" (p. 612). However, if the two psychoses are truly independent, what prevents their simultaneous occurrence? It would be tempting (but historically inaccurate) to blame the "taboo of incompatibility" on the psychodynamic view in spite of the fact that, according to this doctrine, it is nonsensical for two such "diseases" to occupy the same "psychological space". This very issue was debated under the general heading of the "combined psychoses" (Meeus 1908), but no agreement was ever reached. Indeed, no answer to this question seems available yet in current psychiatry.

4.5
Heritability of Mental Illness

The view that mental illness was passed on from generation to generation was well accepted during the nineteenth century, but Morel, Magnan and others re-conceptualised this belief in terms of the so-called degeneration theory. However, the application of this doctrine led to a number of blind classificatory alleys and to the search for somatic stigmata and other genetic markers (Morel 1857; Saury 1886; Walter 1956; Friedlander 1973; Danion et al. 1985; Dowbiggin 1985; Pick 1989).

5
Transformation of Insanity into "Psychoses" During the Nineteenth Century

5.1
Adoption of the Word "Psychosis"

The word "psychosis" was used in the first half of the nineteenth century to refer to the subjective states accompanying insanity (Jastrow and Baldwin 1901). According to von Feuchtersleben (1847) (who, it must be remembered, used the terms "psychosis" and "neurosis" in their late eighteenth-century sense), the term "used with regard to normal processes is equivalent to the mental or psychical element in a psychophysical process, just as neurosis refers to that aspect of the process which belongs to the nervous system" (p. 392). In most cases, however, these experiences relate to "conditions which we usually call, in a more restricted sense, mental derangement" (p. 241). "Every psychoses (must be), at the same time, a neuroses; as without the intervention of nervous action, no change in the psychical realm can occur, (however) every neuroses is not a psychoses; of this convulsions afford a sufficient example. . ." (p. 246). A similar view was taken by another commentator, who chose to emphasise "the neural act corresponding to the mental phenomena" (Warner 1892, p. 1025).

These two usages ("normal" and "pathological") became official in German psychiatry by the device of using the singular form (*Psychose*) to name the former, and the plural (*Psychosen*) the latter (Tuke 1892a). Meyer (1901) attempted to clarify the issue: "used pathologically (and in this sense the usage is rapidly gaining ground both in foreign and in the English literature) the term designates an abnormal mental condition, specially inasmuch as it is correlated with a specific disease-process (a disease entity, if the term be allowed) with characteristic origin, course and symptoms. The typical forms of insanity which can be scientifically differentiated would rank as psychoses in this sense." Maudsley (1885) made use of the "pathological meaning": "no wonder that the criminal psychosis, which is the mental side of the neurosis, is for the most part an intractable malady. . ." (p. 33).

After the Great War, the term "psychosis" gained popularity and soon replaced "insanity"; indeed, Sir George Savage was one of the last to use it in the title of a textbook (Savage 1898). The fact that its original meaning referred to the *experiential* aspects of behaviour no doubt helped in that it dovetailed well with the

growing trend towards including subjective symptoms in the description of insanity. The word "insanity", on the other hand, was too laden with pre-nineteenth-century notions to be acceptable, and only survived in legal language. Another reason for the adoption of the term might have been that it lent itself to easy adjectival derivations (e.g., psychotic) (Sauri 1972).

5.2 The Defining Dichotomies

During the late nineteenth century, the "psychoses" became defined in relation to five dichotomies: psychoses versus neuroses, functional versus organic, exogenous versus endogenous, total versus partial and unitary versus multiple.

5.2.1 Psychoses Versus Neuroses

As has been mentioned above, during the first half of the nineteenth century, "psychoses" referred to subjective states and "neuroses" to the underlying neurological processes. By 1900, meanings had been exchanged: psychoses were the official name for all the "organic states" (whether exogenous or endogenous), and the neuroses had become psychologised (López-Piñero and Morales Meseguer 1970).

As mentioned earlier, the practice of alienism was very different before the incorporation of the neuroses into psychiatry. Asylum alienists only dealt with severe functional or organic psychoses and were (mercifully) free from worrying about the forms of deviancy and psychological incompetence that have since been thrown their way. Insanity therefore provided the descriptive, taxonomic and aetiological templates on which all future conceptions of mental illness were to be based.

This view demanded that a "lesion" be sought in all cases. Failure to identify such changes in the case of the neuroses led to a gradual watering down of the concept of lesion. The basis of the neurotic symptoms was successively thought of as anatomical (disorder of sense and motion), physiological (irritability and inhibition) and psychological (compromise between instinct and reality demands). This transformation was helped by the gradual expurgation from the class of the neuroses of real "organic" states. (The last conditions to be excluded were the vasomotor disorders, i.e., Raynaud's syndrome. A similar mechanism attended the development of neurology, which after the 1860s only encompassed motor and sensory deficits resulting from *localised* brain and spinal lesions.)

However, the mechanism did not work well, and a number of clinical states such as catatonia, motor stereotypes, sleep disorders, hallucinations and deficits of higher cortical functions remained in the borderland between neurology and psychiatry. During the late nineteenth century, the transient popularity of hypnosis and putative "functional mechanisms" (Barrucand 1967; Bercherie 1980), together with psychoanalysis, led to considering many of these states as "psychological". Soon enough, however, the epidemics of encephalitis lethargica pushed views in the opposite direction (for a discussion, see Rogers 1988; Berrios and Dening 1990).

5.2.2 Functional Versus Organic

At the dawn of the twentieth century, the distinction between functional and organic psychoses became a fundamental aspect of their classification.

The functional group included dementia praecox, manic depressive insanity, paraphrenia, the paranoid states and paranoia; the organic psychoses comprised delirium, dementia and a panoply of "symptomatic" psychoses. Since all "psychoses" were supposed to have, at some level, an "organic" basis, it is at first sight unclear why the dichotomy was necessary at all.

Mendel's work (1907) contains one of the earliest references to "functional" psychoses. He offered a negative definition: "on the other hand, there is a great difference of opinion amongst authors as to how to divide those mental diseases in which no anatomical findings have hitherto been met and which do not belong under any of the forms named. They are designated as *functional psychoses*, by which it is not said that anatomical changes do not exist, but only that we have so far been unable to verify them . . . in this respect (absence of a recognizable lesion) they resemble the *functional* peripheral neuroses" (p. 160).

Mendel considered the following as functional psychoses: delirium hallucinatorium, mania, melancholia, circular psychosis and paranoia and acute dementia (pp. 175–213). The first and last categories included conditions which would currently be diagnosed as schizophrenia. Mendel also recognised a separate group of "organic psychoses" which included progressive paralysis of the insane, senile dementia and atherosclerotic and syphilitic psychoses. He also distinguished between "psychoses which are called forth by focal diseases of the brain" (e.g., apoplectic attacks, brain tumours, traumata) and "psychoses arising from central neuroses" (e.g., epileptic, hysteric and choreic psychoses).

By the end of the Great War, the dichotomy between functional and organic had become sharper. For example, Jaspers (1963) listed three functional psychoses, genuine epilepsy, schizophrenia and manic depressive illness: "these three . . . have four points in common. In the *first* place their study gave rise to the concept of disease entity. . . in the *second* place the

cases which belong to this group cannot be subsumed under the disorders of group I and III (i.e., organic). One must however assume that many of these psychoses have a somatic base ... in the *third* place (they) are not exogenous but endogenous psychoses. Heredity is an important cause ... in the *fourth* place they all lack anatomical cerebral pathology..." (pp. 607–608; see also Beer 1996a).

5.2.3 Exogenous Versus Endogenous

Nineteenth-century neurobiological beliefs on the aetiology of mental illness are enshrined in the exogenous/endogenous distinction. Already controversial and unclear in meaning at the time of its inception, it is now but a noble archaism that has, nonetheless, survived its obituarists (Lewis 1971; Heron 1965; Gaston and Tatarelli 1984).

First of all, it must be stated that the term "endogenous", coined by Candolle in 1813 (Berrios 1987b), was never meant to signify "genetic" in the modern sense, nor "exogenous" to signify "environmental". As Kraepelin (1924) pointed out, it was the German neurologist Möbius who first introduced these terms into medicine in 1893. In his short neurology textbook, Möbius described endogenous disorders as those in which "the principal condition must lie in the individual, in a congenital disposition (*Anlage*), other factors being merely contingent and quantitative". Examples were neurasthenia, hysteria, epilepsy, migraine, Huntington's chorea and Friedreich's disease. On the other hand, "exogenous" diseases were toxic and infectious conditions, thought to be "engendered from without" such as trigeminal neuralgia, thyroid disease, multiple sclerosis and Parkinson's disease (Schiller 1982).

The value and meaning of the dichotomy hinged on the feasibility of drawing an operational boundary between "without" and "within". Möbius did not make his criteria explicit, but it would seem that he did not consider the skin as the natural boundary (hence for him exogenous is not necessarily environmental), nor did he mean the neck (hence for him exogenous is not necessarily non-cerebral). "Endogenous" in Möbius' work relates to the *Anlage*, to the intrinsic form or essence, and it is therefore not a spatial but a metaphysical concept.

The origin of this metaphysical view is to be found in nineteenth-century degeneration theory, a doctrine that, as has been hinted at before, dominated European psychiatric and social thinking during the second half of the nineteenth century (Mechler 1963). Phenomena as wide apart as Lamarck's inheritance of acquired features, somatic stigmata and the fact that disease could suddenly disappear after one generation were all explained by degeneration theory (Huertas 1987; Ribot 1871; Talbot 1898; Mairet and Ardin-Delteil 1907; Genil-Perrin 1913; Apert 1919; Salas 1920; Wettley 1959; Peset 1983; Hermle 1986). During the 1890s, the over-religious and fatalistic Morelian model of degeneration (Morel 1857) was made flexible by Magnan (Genil-Perrin 1913), and after the 1910s, it gradual merged with the new genetic views on mental disease (Saury 1886; Huertas 1985).

However, during the 1890s, the concept of "endogenous" was still a shorthand for degeneration theory and postulated that the cause of mental illness lay deep in the metaphysical being (Mechler 1963). Hence "endogenicity" was not just about genetic control or personality or constitution. Kraepelin liked the concept precisely because of these features. The terms "endogenous" and "exogenous", no longer current in neurology (where they started their career), might have also sunk into oblivion in psychiatry had it not been for Kraepelin, who had adopted them by the time the 1896 edition of his textbook appeared.

The dichotomies functional/organic and endogenous/exogenous therefore have a different historical provenance and overlap only partially. To a late nineteenth-century alienist, exogenous did not necessarily mean organic, and functional did not necessarily mean endogenous. Further obscurities have accrued during the twentieth century due to their well-nigh untranslatability into modern biological terminology.

5.2.4 Total Versus Partial Insanity

Before the nineteenth century, the clinical concept of insanity was thought of as representing an involvement of *all* mental faculties, most particularly the intellectual functions (i.e., presence of delusions) (Postel 1984). Furthermore, insanity was thought to be irreversible (Foucault 1972a; Quetel and Morel 1979); mania, for example, signified a behavioural, physical and metaphysical state (Middleton et al. 1780; Calmeil 1839). Insanity, in general, was defined as a genus whose species were each defined in terms of a salient behaviour (Arnold 1806). The categorical nature of "total insanity" is even clearer in legal usage. Bracton, for example, defined mania as "totally lacking in discern", and 400 years later, Coke and Hale were still defining it as "absolute madness" (N. Walker 1968).

The concept of "partial insanity", i.e., the view that insanity may be less than total, originated to some extent in the interaction between medicine and the law in relation to cases where the "all-or-none" concept seem to break down (Eigen 1995). For example, in 1736, Hale defined it thus: "there is partial insanity of mind...; some persons that have a competent use of reason in respect of some subjects, are yet under a

particular dementia in respect of some particular...; or else it is partial in respect of degrees..." (N. Walker 1968). During the early nineteenth century, however, partial insanity had more than one clinical meaning (Berrios 1991b):

1. Involvement of one faculty (from the view point of faculty psychology)
2. Intermittent insanity, i.e., presence of "lucid intervals" (from the longitudinal viewpoint) (Haslam 1809)
3. Mild or moderate (from the point of view of intensity)
4. Presence of a "partial" delusional system (from the point of view of the extension of the delusions) (Berrios 1994c)

The concept of "monomania" (fallen into desuetude after 1850; Reports of the Sessions of the Société Médico-Psychologique 1854) was formed in the 1820s around the first and last definitions given above (Kageyama 1984). In his translator's introduction to Hoffbauer's treatise, Chamberyon (1827) presented the simplest classification: "insanity is divided into mania and dementia, according to whether the mental faculties are overactive or weakened; mania is subdivided in turn into polymania (or total insanity) and monomania (partial insanity)" (pp. 10–11). The final acceptance of the tripartite classification (intellectual, emotional and volitional), in addition to the gradual ascent of the natural sciences view of disease, led after the 1850s to the quiet disappearance (in medicine) of the concept of total insanity. In the legal field, it never disappeared altogether and enjoyed a period of revival after the McNaughton rules were passed in Great Britain in 1843 (West and Walk 1977), because the narrowness of the criteria made it difficult to find a case of "partial insanity".

5.2.5 Unitary Versus Multiple Psychoses

One of the interesting offshoots of the nineteenth-century taxonomic controversy was the development of the unitary and multiple views of insanity. Reacting against the proliferation of classifications and inspired by the principle of the indivisibility of the mind, some alienists put forward the view that there was only one insanity which could take many forms and that its multiple clinical presentations were epiphenomenic, i.e., the result of pathoplastic factors (Llópis 1954; Menninger et al. 1958; Rennert 1968; Vliegen 1980; Janzarik 1969; Berrios and Beer 1994). These included factors such as idiosyncratic responses (i.e., individual styles of coping with the ravages of insanity), proximate and remote aetiologies (e.g., emotional precipitants), intensity (over- or under-activity of the mental faculties) and duration of the condition. It was suggested, for example, that mania, melancholia, delusional insanity and vesanic dementia might just be successive stages of the same disease.

This is an important point to remember when, for example, the history of circular insanity (bipolar disorder) is studied (Sedler and Dessain 1983; Pichot 1995). Thus the suggestion made by Baillarger and Falret (Pichot 1995) that mania and melancholia were related should not be considered as solely based on clinical evidence (as the observation that some patients might intermittently suffer from both had previously often been made); indeed, these two great alienists were also influenced by the "unitary psychosis" concept (Berrios and Beer 1994). A new factor was added to the factors mentioned above after 1857, namely degeneration theory (Morel 1857). According to this view, the degeneration taint was transmitted from parents to offspring and on each occasion gave rise to ever worse forms of insanity until dementia appeared.

5.3 Three Modules of the Mind and Their Insanities

Up to the beginning of the nineteenth century, the intellectualistic view of insanity reigned supreme (Berrios 1985b). The growth of faculty psychology led to the division of the mind into functional modules (Fodor 1983) and provided a new model both to phrenology and to later studies on brain localisation (Bentley 1916). As suggested by Kant (Hilgard 1980) and the Scottish philosophers (Seth 1890; Albrecht 1970; Brooks 1976), three modules were distinguishable: intellectual, emotional and volitional. The old forms of madness were re-defined as "intellectual insanities" (including both total and partial). This change in the original meaning of insanity (until then only considered as intellectual) opened a clinical space for purely "emotional" and "volitional" insanities. It also freed mania and melancholia from their secular duty as forms of delusional insanity (Linas 1871). The emotional insanities were to evolve into the modern concept of melancholia, mania and circular insanity (Berrios 1988a; Pichot 1995), and the volitional disorders played a role in the explanation of psychopathy (Werlinder 1978) and aboulia (Ribot 1904; Lapie 1902; Berrios and Gili 1995).

The decline in popularity by the turn of the century of "will" and "volition" undercut this neat arrangement (Herzen 1880; Paulhan 1903; Daston 1982; Keller 1954; O'Shaughnessy 1980). If "will" had no explanatory power, then it made little sense to ponder over its disorders; the question of what psychological functions (if any) are dysfunctional in psychopathy and aboulia was, however, left unanswered. Likewise, the shift in

the meaning of "emotional insanity" occurred during the 1850s. No clearer statement can be found than Bucknill and Tuke's remark: mania, "perhaps the most interesting and best recognized form of mental disease, has been usually treated of by writers, as essentially a disorder of the reasoning faculties. Dr. Prichard classed it under intellectual insanity. We are disposed however to regard it as belonging primarily to the affective group" (Bucknill and Tuke 1858, p. 221).

5.4 Separation of the Organic States

Ever since the times of ancient Greece, the symptomatology of delirium (phrensy) had included fever, fleeting hallucinations, delusions and behavioural disorder (Berrios and Freeman 1991a). Other "organic" psychiatric disorders, however, such as dementia, were not distinguished during this period from the rest of the insanities. Analysis of pre-nineteenth-century case reports confirms the view that, during this long period, mania and melancholia included "organic" states such as encephalitis, neuro-intoxications, brain tumours and certainly schizophrenia.

The modern concept of "organic disorder" only appeared after 1822, when Bayle described chronic arachnoiditis in patients with psychiatric manifestations of the type that were to be later called "general paralysis of the insane" (Bulbena and Berrios 1986). Battie (1758) had already differentiated between "original" and "consequential" madness, but the latter category cannot be said to refer to "organic states" in the modern sense. It was in fact Willis, in his description of dementia, who came closer than any other before the nineteenth century to separating dementia from insanity (Berrios 1987a).

The interesting aspect of Bayle's work (1826) was not so much that he identified anatomical lesions in relation to a given neurobehavioural syndrome, but that he made possible the view that kaleidoscopic psychopathology (ranging from typical mania to melancholia, hallucinatory states and dementia) could be associated with the same brain lesion (Bercherie 1980) and vice versa, i.e., that the same syndrome could be caused by many types of lesions. Griesinger (1865), in his perceptive lecture on the occasion of the opening of the Zürich Psychiatric Clinic, stated: "One can suffer of melancholia on account of eight or ten different brain diseases; and of dementia on account of twenty. . ." (p. 11).

It took the rest of the century for the "organic disorders" to be aggregated into a self-contained group. As expected, some of these clinical states had been included for centuries under the old "mania" concept, and hence their gradual separation contributed to the narrowing down of the latter (Couchoud 1913).

5.5 Narrowing Down of the Concept of Mania

The concept of mania underwent radical transformation between 1800 and 1900 (Berrios 1981c). At the end of the eighteenth century, it was tantamount to insanity (Linas 1871) or madness (Middleton et al. 1780; Chiarugi 1793), but by the end of the nineteenth century it was used exclusively to refer to "elated hyperactivity" with or without psychotic symptoms (e.g., Mendel 1907).

Gauchet and Swain (1980) have stated that the reconstruction of the concept of "mental alienation" during the early nineteenth century revolved around a repeated analysis of the concept of mania and point out that Esquirol himself used it as a synonym of "mental alienation" (p. viii). Likewise, Erasmus Darwin (1796), Charles Darwin's grandfather, described in *Zoonomia* three forms of mania, one of which included ecstatic states, despair and melancholia. Arnold (1806) put it thus: "maniacal insanity, properly so called, as a species, is of all others, perhaps, the most *comprehensive*; since it extends its dominion over the whole internal world of ideas, and comprehends every possible combination of sensible images which can enter into, and delude, a distempered brain. To enumerate all its varieties would not only be difficult, but impossible. . .".

The concept began to change in the work of Pinel (1809), who tellingly deleted the term "mania" (used in this general sense) from the subtitle of the second edition of his *Traité*. He classified "manie" in the *Nosographie* as a genre of the "vesanies" (insanities) characterised by "disorder of one or more faculties with sad, gay, extravagant or raging affect, occasionally free from disordered thought but always blind aggression". He also recognised delusional and non-delusional forms of mania (Pinel 1818, p. 592). Couchoud (1913) rightly stated that it was from the large group of the manias that melancholia, dementia and idiocy were eventually separated off.

In German psychiatry, situation is similar. In the historical introduction to his book, Heinroth (1818) wrote that "mania was a *general* insanity, accompanied by rages and by a bold execution of the demands of the will" (p. 65). He described four types of "rage" (mania): simplex (pure rage), ecstatica (insane), ecnoa (rage accompanied by folly) and catholica (common rage) and differentiated mania from the following: daemonomania, erotomania, raging melancholia, lycanthropia, mania cum tristitia (with sadness), continua acuta, chronica, periodica, nynphomania, satyriasis and melancholia saltans (savage impulse to jump).

At the other end of the century, Mendel (1907), perhaps the greatest specialist on mania in the late

nineteenth century (his work on mania appeared in 1881), defined it as: "a functional psychosis which is characterized by a) pathological acceleration of the efflux of ideas, b) motor unrest, and c) absence of symptoms which confirm an organic disease of the brain". In its symptomatology, Mendel included hallucinations, anomaly of thought, delusions, confusion, hypermnesia, heightening of motility and loss of weight. He also recognised four stages of mania, "initial, exaltation, frenzy and decline", and four subtypes, "hypomania, recurrent, gravis and periodic". Kraepelin's (1921) own definition was influenced by Mendel's.

Between Pinel and Mendel, "mania" had not only become narrower, but also more syndromatic (and defined in terms of exclusively affective symptomatology). How was this change possible? It is unlikely that it resulted simply from the gradual erosion of the old concept of mania. Historical analysis shows that the word "mania" almost dropped out of circulation around the 1830s, when some of its clinical functions were taken over by the concept of monomania (Kageyama 1984; Calmeil 1839; Goldstein 1988). "Monomania" itself lost popularity (Falret 1864), and its decline brought back the word "mania". On its return, however, it was furnished with a new meaning.

I should like to suggest that six factors (inter alia) contributed to the metamorphosis of the old into the new notion. First of all, "mania" was too general a category to be acceptable to the analytical view of mental disease popular during the nineteenth century (Lanteri Laura 1982, 1983). Secondly, the concept of "total insanity" (on which it was based) had been replaced by one of partial insanity (Kageyama 1984). Thirdly, faculty psychology led to the acceptance of emotional insanities as a separate group of disorders (Berrios 1987c). Fourthly, organic disorders such as "general paralysis of the insane", in the past a regular source of "manic states", were now well recognised and excluded (Berrios 1985a). Fifthly, the language of description became more rigorous, as alienists created stable definitions of the "elementary symptoms" (Griesinger 1867). Lastly, subjective symptomatology became incorporated into the definition of insanity, thus making it possible to include "elation" as a central symptom.

6 The Twentieth Century

Official time divisions have little meaning in history. For example, nothing special suddenly happened in the year 1900 (as nothing will in 2000) to alter the meaning of psychiatric concepts or the seamless way in which they had been flowing since their creation during the nineteenth century. Mental symptoms and diseases are born out of a historical mechanism or process which can be called a convergence, and it goes something like this: a particular observer reports a particular clinical phenomenon as discrete and stable enough to represent a mental symptom, syndrome or disease; the behaviour in question is described and examples (a casuistry) are usually reported. The observer may often give it a name (a neologism or a refurbished old term) and, if possible, an explanation (a concept) in terms of current medical doctrine. What converges in the work of an author are therefore a form of behaviour, a word and a concept. When enduring, convergences can be called successful or felicitous. The reasons for this are unclear. To say that success depends on the fact that the convergence has captured a true biological fact is not sufficient, as many convergences have been kept alive for social and political reasons.

Lasting convergences tend to produce collective mirages which contribute to their perpetuation. Psychiatrists cannot help but see things in that particular way (or dare not see them in any other way). Social and financial investment are also strong factors in the survival of a convergence. Once pharmaceutical companies and the reputation of pundits are committed to the truth of a particular convergence, it becomes very difficult for common mortals to call it into question. The success and stability of ICD-10 and DSM IV are, in a way, based on this social apparatus.

This can be said to constitute a triumph for psychiatry, but also its tragedy. It is a triumph because it is possible for researchers to claim that there is scientific continuity in what they are doing, that they are not standing on the shoulders of giants as it were. However, it can also be the tragedy of psychiatry if the convergence so sedulously defended turns out to be totally wrong. What, for example, if the current cognitive model of dementia were to prove infelicitous (what would happen to the enormous financial investment in a miracle drug) or if the view that delusions and hallucinations are crucial "positive" symptoms of schizophrenia were to turn out to be misconceived (as it is likely to)? What will happen to those whose reputation has been built on selling the idea of "positive symptoms", scales, neuro-images and specific neuroleptics?

This is why psychiatry must remain a conceptual broad church and why current pundits (particularly those locked within the Anglo-Saxon tradition) must look beyond the narrow confines of their language to listen to what is being said in French, Polish, Russian, German or Chinese psychiatry. For the same reason, marginalising psychopathologists and phenomenologists as fuddy-duddies who have not yet heard about neuro-imaging would seriously harm the future of

psychiatry. The main lesson of history, as iterated in this chapter, is that psychopathology and its attending concepts remain the very foundations of psychiatry. Without their help, it would not be possible to recalibrate the descriptions in response to secular genetic changes in the expression of disease, to social warrants and to the development of new investigative techniques. It is of the essence to understand that there will never be a stable set of psychopathological descriptions and concepts. Re-adjustments will have to be made every so often in order to do justice to our patients.

It is beyond the scope of this chapter to map out the twentieth-century fate of the ideas and concepts discussed in previous sections. However, something must be said to give younger colleagues a sense of continuity, emphasising French and British psychiatry.[1]

6.1 France

Writing on the history of French psychiatry during the twentieth century is a complex task, not so much because it is conceptually difficult (although on occasions it can be), but because there is too much to say. This embarrassment of riches forces selectivity, and the latter risks distortion.

Since the nineteenth century, French psychiatry has been characterised by (a) a fruitful interaction between concepts and clinical practice, (b) an interest in the creation of a second-order language and (c) the strong presence, after the 1850s, of a historicist, longitudinal (diachronic) perspective. Psychoanalysis (Mordier 1981) and the First World War had a negative influence on this epistemological stance in spite of the efforts of writers such as Semelaigne (1932), who tried in his work to reconcile the inter-bellum generation with their great psychiatric past. The 1930s also saw the development of a society of scholars which came to be known as *L'Evolution Psychiatrique* and which started publishing a quarterly under the same name. One of its members was Henri Ey, then in his early 30s and already intellectually productive. This group was characterised by a keen historical sense and what could be described as an ambiguous attitude towards their own psychiatry which included both pride and rejection. Thus, on the one hand, they referred to the great nineteenth-century debates at the *Société Médico-Psychologique* as the great source of concepts; on the other, influenced by psychoanalysis and a version of the Jacksonian model, they criticised French psychiatry as being excessively descriptivist. Their publications include classical articles on hallucinations, obsessions, etc.

Once again, the Second World War affected intellectual life in France, including academic psychiatry. Great men such as Dide (Mangin-Lazarus 1994) and Halbwachs (Coser 1992) were killed, and specialised publications slowed down or disappeared. This sombre mood can be noticed, for example, in the centenary issue of *Annales Médico-Psychologiques* (1943), in which all review articles were clearly written in a hurry and without the back-up of libraries, which were shut or hidden. Notable among these is the paper on the history of dementia written by Guiraud (1943), then one of the great figures of French psychiatry. Things started improving by 1950, when the First World Congress of Psychiatry was held in Paris, with Henri Ey as its president and with the supporting presence of world figures who gave a fillip to French psychiatry. Ey's explanatory views, whether of the origin of mental disorder or of the evolution of psychiatric thought, were based on a longitudinal, diachronic view. On account of the influence of Hughlings Jackson on his work, Ey saw disease and the historical process in the same way, namely, as resulting from dissolution and circular motions (Berrios 1977).

Another good example of a psychiatrist transversing three generations is Henri Baruk, who started in 1926 with a magnificent book on the psychiatry of brain tumours (Baruk 1926). In 1965, Coury, then the head of the Department of the History of Medicine at the University of Paris, organised a seminar in which Baruk gave five lectures on the history of psychiatry in the nineteenth and twentieth centuries. Published under the title "French Psychiatry, from Pinel to the present day" (Baruk 1967), it provides a useful division into what Baruk called the philanthropic, clinical, anatomo-clinical, psychopathological and medico-legal periods of French psychiatry.

From this historical sketch, it can be surmised that, during the twentieth century, French psychiatry tried to continue along furrows traced by earlier psychiatrists. Thus, on occasions, it resisted new concepts, and a good example is the debate at the beginning of the century on whether Kraepelin's concept of dementia praecox had clinical validity. On other occasions, this defensiveness became xenophobic, as it is the case with Chaslin's (1914) anti-German paper. More permeability was shown to other ideas, for example psychoanalysis, which since the time of the early writings of Hesnard (1971) found a special place in many a French hospital.

[1]Editor's note: On the development of psychiatry in the German-speaking countries in the twentieth century, see also Ackerknecht (1985), Finzen (1996), Hoff (1994), Kolle (1956, 1959, 1963), Kreuter (1996), Schliack and Hippius (1998).

On the other hand, some of the great nineteenth-century ideas were developed further. De Clérambault modified the old notion of *automatisme psychologique* and made it central to his rich, but idiosyncratic psychopathology (Marchais 1995). Inspired by Bergson and Durkheim, Charles Blondel produced one of the more original books on mental symptoms and pathological changes in consciousness ever written (Fuentenebro and Berrios 1997), and Henry Ey tried to reconcile psychoanalysis, Jacksonian ideas and the old French descriptive tradition (Berrios 1977). During the 1930s, Guiraud developed a subtle neurobiological model of mental disease, Mourgue helped to develop a fascinating diachronic model of mental illness (Monakow and Mourgue 1928), Lacan (1977) gave a hint of things to come with his powerful thesis on paranoia and Ajuriaguerra, after training in Paris, took up a chair in Geneva to undertake his legendary neuropsychiatric research (Berrios 1992b).

All these very French approaches had two things in common: one was the view that mental symptoms are not just units of analysis clustering together and occurring in clear consciousness, and the other that they are manifestations of deep structural changes in the psyche, of dissolutions and regressions, which also cause major changes in the frame of consciousness and which make mental illness a complex and dynamic phenomenon (similar in a way to the suggestive ideas developed by Janzarik 1959).

The historian can still sense the major effort made by French psychiatry after the Second World War to remain true to its nosology and philosophy of care. During the 1950s and 1960s, the impact of left-wing ideas and the intellectual and practical work of talented immigrants such as Fanon and Tosquelles led to the development of a new social psychiatry and the growth of the concept of sector and continuity of psychiatric care (Raynier and Beaudouin 1961; Fourquet and Murard 1980). Psychoanalysis, neurobiologism, descriptivism and phenomenology are allowed to run parallel, and little effort is made to find common ground. The great French journals continued to be published in French, and with few exceptions (e.g., Pierre Pichot) little effort is made inform the outside world of French ideas.

This landscape has began to change. Descriptive psychopathology and the clinical tradition – à la Séglas, Chaslin, De Clérambault, Guiraud or Ey – is in danger of being eclipsed by the growing acceptance of DSM IV. Important journals (e.g., *L'Encéphale*) have switched to English, and the results of French research projects are increasingly appearing in Anglo-Saxon publications. Let us hope for the benefit of mankind that the care for and interest in ideas which has characterised French psychiatry since the nineteenth century is preserved in the midst of these changes.

6.2 Great Britain

Twentieth-century British psychiatry has retained its earlier interest in the management of the insane (for a full history, see Berrios and Freeman 1991b; Freeman and Berrios 1996). Of course, there is also a conceptual story to tell which invariable reflects changes in German and French frameworks. For example, since the 1850s, the English term "psychopathology" is known to have changed its meaning at least three times. It started life in 1847 as a transliteration of the German *Psychopathologie*, as used by von Feuchtersleben (1845). At this stage, German usage was explicative rather than descriptive: According to von Feuchtersleben (1847, p. 70), "psychopathology has not yet acquired sufficient light respecting these critical processes", but this perspective was not influential in England; Winslow (1848), for example, disregarded it altogether. The term re-emerged at the end of the nineteenth century with a forensic meaning: "the science which treats of the legal aspects of insanity; i.e., of the rights and responsibilities of lunatics" (Tuke 1892b, p. 1014).

An additional explanation for the lack of success of "psychopathology" in Great Britain is likely to be that there were established competitors such as "psychological medicine", "mental science", "mental pathology" and "mental physiology". At the beginning of the twentieth century, the forensic meaning petered out, and the term started to gain the meanings debated in continental Europe. In 1903, Pierre Janet and Georges Dumas founded the *Journal de Psychologie Normale et Pathologique*, and 5 years later Dumas published his "Qu'est-ce que la psychologie pathologique" (1908), in which he made a case for psychological pathology becoming a branch of normal psychology. In 1889, Binet described the situation accurately: "with few exceptions, the psychologists of my country have left psychophysical research to the Germans and comparative psychology to the English, to dedicate themselves completely to pathological psychology" (Beauchesne 1986, p. 67).

At the beginning of the twentieth century, the French debate on whether there was continuity or discontinuity between normal and abnormal behaviour was rehearsed in Great Britain. Likewise, the need for a term was felt to refer to that branch of psychology dedicated to the psychological (rather than somatic) explanations of mental dysfunction. By then, "mental science", "psychiatry" and "psychological medicine" were all tainted by their use by psychiatrists and were not acceptable to psychologists (or psychiatrists of a psychodynamic persuasion). Thus, with Jastrow (1902), psychopathology acquired its widest definition:

"the general study of diseased mental conditions; a synonym of psychiatry and abnormal psychology but rather more comprehensive that either, because it emphasizes the general scientific study of all forms of mental aberration" (p. 391). This definition was to remain popular, particularly in the United States, where the term psychopathology was for a time used as a synonym of psychiatry (Berrios 1991). Soon enough, however, British psychiatry was affected by the arrival of psychoanalysis.

6.2.1 The Psychodynamic Period

The early impact of the psychodynamic ideas (those of Janet and Freud) on psychiatric matters in Great Britain has not yet been fully studied (Hinshelwood 1991). It is assumed that, in general, this impact was not marked, as British psychiatry remained protected by its "empirical" mentality and by the fact that it main clinical concern was the management of the insane. This was to change during and after the Great War, when the understanding of "shell-shock" (Merskey 1991) led to a gradual acceptance of psychodynamic views. Analysis of works published around that period suggests that these views may have been more influential than it has been hitherto considered (e.g., Hart 1927).

By the onset of the First World War, the term "psychopathology" had also reached its widest meaning in Great Britain. In the 1926 Goulstonian Lectures, Bernard Hart, an important psychiatrist of the period stated: "Psychopathology connotes not a mere description of mental symptoms, but an endeavour to explain disorder or certain disorders in terms of psychological processes" (Hart 1927, p. 2). He also differentiated this type of explanation from "somatic ones" used in "psychiatry". Hart's work illustrates well the manner in which psychoanalysis became assimilated into the epistemology of British psychiatry (Pines 1991). While using a number of classical Freudian categories, Hart constantly felt the need to appear to be practising a "scientific" and "empiricist" approach (in the British sense of these terms). This need is almost painfully expressed in the way in which he clung on to the definition of science developed by K. Pearson (1892), which was incompatible with Freud's.

Hart was not alone in this endeavour. Myers and Rivers, two Cambridge academics, were also influenced by psychoanalytic concepts and planned to develop a "science of psychopathology" (Crampton 1978). To this end, they were successful in the creation of a lectureship in psychopathology at the University of Cambridge. The first incumbent, McCurdy, a psychiatrist and psychologist from Canada, was at one with Rivers and Myers in regarding psychopathology as both the description and psychological (not somatic) explanation of mental symptoms (Banister and Zangwill 1949). This academic view reflected the zeitgeist. It was under McCurdy's tutelage that Gillespie (1929) developed the concept of "reactive" and "neurotic" depression.

Another interesting figure in this context is McDougall, who also carried out psychopathological research in Cambridge. His early conception of "clinical psychology" included the study of mental disorders, and he spoke persuasively on this subject in his presidential address before the Medico-Psychological Association of Great Britain and Ireland (McDougall 1919). In his *Outline of Abnormal Psychology*, McDougall (1926) used "abnormal psychology" for psychopathology, approvingly quoting Eugen Bleuler's statement that "one of the most important, if not the most important, of all paths to a knowledge of the human soul is by way of psychopathology" (p. vii). Later on, McDougall proposed an idiosyncratic model where the central role was played by an energising principle or "horme", which owed much to Freud's *libido*, Bergson's *élan vital* and Oskar Vogt's *neurokyme*. In his autobiography, McDougall wrote: "I built up my abnormal psychology incorporating what seemed most sound in the teachings of Freud, and Jung and Morton Prince, especially the principles of conflict, repression and dissociation, and the subconsciously working complex" (McDougall 1930, pp. 215–216). The influence of McDougall on British psychiatry has not yet been fully evaluated.

6.2.2 The Descriptive and "Phenomenological" Period

In the 1930s, Great Britain had the great fortune of becoming the country of exile for a group of major continental figures, including Willy Mayer-Gross, Eric Guttmann, Alfred Meyer, Stephen Krauss and the Austrian Erwin Stengel (for a history, see Peters 1996). Mayer-Gross arrived in Britain in 1933, and on his death he left a great intellectual heritage. As Slater and Roth (1969) were to write years later: "Mayer-Gross was educated in the German school of psychiatry, and was a pioneer contributor, with such other great men as Beringer, Gruhle and Jaspers, to the remarkable flowering of clinical psychiatry in the development of 'phenomenology', i.e., the exact study and precise description of psychic events, which are a primary requisite for their understanding". The same authors also suggested that Mayer-Gross helped to rejuvenate British psychiatry which was "until then bogged down in the unproductive generalities of Meyerian psychobiology". This statement is historically correct. Up to the 1930s, British psychiatry was drifting into a Meyerian and psychodynamic limbo and was closed to everything else.

For example, the four lectures on the "Phenomenological Method" delivered by Husserl at the University of London in 1922 had no impact whatsoever (Spiegelberg 1982). At an early stage of his career, Ryle is said to have been asked by his professor to learn German in order to read Husserl's *Logische Untersuchungen.* Thus armed, he went on to write a review of Heidegger's *Sein und Zeit* (Ryle 1929). In his autobiography, Ryle (1970) recalled the lack of interest for phenomenology in the Oxford of the 1920s and stated: "it is sometimes suggested that in my well or ill spent youth I had been for a while a disciple of Husserl's phenomenology. There is no much truth in this" (p. 9).

Mayer-Gross encouraged a precise description of psychic events and took pains to construct structured instruments to organise clinical information (Mayer-Gross and Guttmann 1937; Lewis 1970b). He also inculcated in his disciples a sensible combination of descriptivism and neurobiology. Moreover, he is said to have introduced many to the work of Jaspers. In spite of this teaching, however, it is of some historical interest that the historian cannot find any tangible proof of Jaspersian influence until the 1950s, when a group of psychiatrists in Manchester, under the leadership of Anderson, began to translate some German classics, including the 1947 edition of Jaspers' *Allgemeine Psychopathologie*, which rather surprisingly appeared for the first time in English in 1963.

Evidence for this lack of influence can be found in Ernest Nicole's book. The honorary secretary of the Psychopathology Subcommittee of the Royal Medico-Psychological Association, Nicole wrote in 1930 the first edition of what was to become a very popular book (it went through six editions). In its fourth edition, published in 1946, this work included about 1400 references, by far the longest list ever included in a book on psychopathology. Surprisingly, therefore, no reference can be found to either Jaspers or Mayer-Gross, and this because as late as 1946, Nicole's more or less official conception of psychopathology (Nicole 1946) remained pre-1930 in spirit and amounts to no more than a list of explanations for all mental disorders, including some somatic theories.

Yet another piece of evidence that Jaspers' influence on British psychiatry was much more tardy than is retrospectively allowed is the perceptive comment by Anderson (1963) "that this school [phenomenology] should have been insufficiently recognized in England is perhaps hardly surprising. Many reasons can be adduced for this, first and foremost of which is the linguistic barrier. Secondly, even for those familiar with the German language, Jaspers' thought and style are difficult, due no doubt to his training as a philosopher, and his consequent use at times of terms specially devised to express some nuance of meaning not easily, if at all, to be translated intelligibly. To the English with their ingrained empiricism such an approach might well repel . . . yet this is a superficial evaluation". And referring to the type of psychopathology which was popular in the United Kingdom, Anderson added: "By a remarkable paradox, however, our supposedly empirical countrymen have accepted readily enough, and on the whole with an astonishing lack of criticism, the unproven and unprovable assertions of the so-called psychodynamic schools" (Anderson 1963).

Other important accounts of psychopathology published during that period also remain unenthusiastic about Jaspers. For example, in the first edition of his *Psychopathology*, Taylor (1966) stated that "psychopathology is a word of many meanings. It has a chequered career in the history of modern psychiatric thought, acquiring new meanings and connotations with the advent of every new theory of mental illness", and in spite of his overt support for phenomenology, Taylor still conceived of psychopathology as a form of "physiology of the mind" (p. 10). The same year, Frank Fish, a psychiatrist of German descent, was invited to give a seminar on symptoms at the University of Manitoba, which he duly published as *Clinical Psychopathology* (Fish 1967). Although his emphasis was then descriptive, Fish (1967) felt forced to state that "this book has been written from the descriptive stand-point and hence over-emphasizes certain aspects of psychiatry. This does not imply that the author believes that interpretative psychology, such as Freudian psychopathology, and experimental psychology have nothing to contribute to our understanding of psychiatric signs and symptoms". After Fish's untimely death, the book was updated by Hamilton (1974) (also German born), who wrote that "anyone who is acquainted with Anglo-American psychiatric literature will know that the careful description of psychiatric symptoms in English is conspicuous for its absence" (p. 1) and, referring to Jaspers' *General Psychopathology*; "it is unfortunate that this should be the only account of German views on symptomatology in English, because the book is overloaded with philosophy, is somewhat out of date, and does not do justice to the views which Jaspers does not accept" (p. 1). A fairer and more succinct judgement would be difficult to make.

In 1960, another German writer, Hans Eysenck, edited his epoch-making *Abnormal Psychology*, which he dedicated to Kraepelin. This book was to be influential among psychologists and in fact included chapters by the best research psychologists of the late 1950s. Eysenck (1960) attempted an "integration through theory, an attempt to see abnormal psychology as a part of general, experimental psychology". His view was that symptoms were dimensional and

continuous with normal function and hence that psychopathology had as its task the working out of "the details of abnormal behaviour from general laws".

To summarise, the first psychopathological tradition to become established in Great Britain was a watered down variant of the Freudian approach. Its exponents strove for the development of a science of the abnormal mind that contained a set of descriptions and psychological explanations of mental symptoms. This tradition was predominant up to the 1930s, when the German-speaking psychiatrists brought with them views that combined Jaspersian descriptive psychopathology and a welcome search for organic causation.

6.2.3 The Present

How does the average British psychiatrist understand "psychiatry" and "psychopathology" and practice them in everyday clinical work? It is still customary to teach that "descriptive" psychopathology (divided itself into "phenomenological" or verbal or descriptive and experimental) studies the "form", whereas "psychodynamic" psychopathology studies the "content" of the symptoms. No set of rules exists, however, to cross-refer from one to the other; hence this classification is little more than a compromise between parallel psychopathological approaches.

The historian of the present can identify in current British psychiatry a combined or eclectic approach that has become its hallmark. However, while paying lip service to the objectivity of "atheoretical" descriptions, researchers in Britain make assumptions as to the nature of the mind and the brain causation of mental disease. They have also succumbed to an absolute need for "reliability" and increasingly rely on closed psychopathological glossaries (e.g., DSM IV). Most feel that a science of psychopathology is no longer needed, as the description of symptoms is now complete. In this regard, it is revealing that the Psychopathology Subcommittee of the old Royal Medico-Psychological Association has no counterpart in the present Royal College of Psychiatrists.

7 Conclusions

This chapter has given just few examples of how some of the concepts reigning in modern psychiatry (and constituting its hidden scaffolding) came into being. It has suggested that psychiatry is a complex medico-social activity whose historical success depends upon the progressive goodness of fit between man-made descriptions and mental symptoms, i.e., behaviours (mostly) resulting from neurobiological lesions. In all cases, however, social factors are crucial to their modulation. In practice, this means that more or less featureless experiences are always formatted by personal, social and cultural codes.

Behavioural signals and their biological mechanisms are in general opaque to understanding, for they are proto-behaviours that, in order to be intelligible at all, require severe cultural formatting. This means that the description and mapping of mental disorders are language dependent. This has led some to conclude (wrongly) that mental disorders can therefore be reduced (without residuum) to linguistic events. Although it cannot be denied that mental disorders are semantic and social artefacts, it is also undeniable that they are a manifestation of disordered neurobiology. From the point of view of explaining and understanding mental illness, both components are essential.

It also follows that the language of descriptive psychopathology is an intellectual, technical and social event. Like any other descriptive system, it includes terms, assumptions and application rules and can be studied by means of a cross-sectional (synchronic) or a longitudinal (diachronic or historical) methodology. While the former pertains to the philosophy of psychiatry, the latter gives rise to psychiatric history proper. If it is to be of any use, psychiatric history must do more than chronicle surface events. It must also unveil the particular structures of successive psychiatric discourses (i.e., the hidden scaffoldings often manipulated by the psychiatric elite). Current descriptive psychopathology, for example, emerged from a compromise between the need to capture the invariant features of mental disorder (i.e., the biological signal) and the need to achieve public order and contribute to the good functioning of society; the former was the expression of nineteenth-century science, the latter of socio-political demands. This encouraged psychiatrists to create a descriptive system that was both based on reality and socially committed.

However, the descriptive system also needed to be stable enough to withstand signal variation and redundant enough to cope with unskilful usage and perfunctory teaching. The conceptual stabilisers built into the language of description by the great alienists of the nineteenth century proved adequate for these tasks.

Unfortunately, the ensuing stability has been interpreted by some incautious twentieth-century psychiatrists as meaning that DP is correct in every detail and exhaustive and hence can be considered as transparent. This has encouraged a premature closure of the psychopathological glossaries. Closure has contributed to reliability, stability and communication, but might have curtailed clinical validity and the rights of psychiatrists to describe new mental symptoms.

Analysis of this historical process is only possible because records exist that document the language of

description and show that it has changed throughout time. Since the clinical historian cannot study first-hand the behavioural signals emitted by the countless patients that once came within the observational purview of his or her predecessors (and on which the language of description was calibrated in the first place), he or she must rely on documents enshrining partial accounts of what those signals were like. Confronted with a multiplicity of descriptions, the historian must map and "triangulate" these systems of words and assess the relative contribution of biological signal and social noise.

This sort of analysis shows that, from time to time, changes take place in the way in which mental illness is described. In principle, these might have resulted from either shifts in the language of description itself or mutations in the gene system that controls the biological signal. In this respect, clinical historians have the advantage of being able to resort to an internal catalogue of bedside experiences, i.e., to "knowledge by acquaintance".

When explaining the development of the language of psychiatry during the nineteenth century, the following factors have been considered:

1. The availability of new semiological and psychological theories that allowed for a fractionation and (eventually) a quantification of behaviour
2. Changes in the conception of physical illness itself, which demanded correspondences between lesion and signal and placed the disease firmly in the internal space of the body
3. The availability of patient cohorts (as a result of the creation of asylums), which allowed for longitudinal observation, thereby introducing a time dimension in the interpretation of symptoms
4. The acceptance of subjective experiences as legitimate sources of mental symptoms

Other factors will no doubt be uncovered by further research.

Once descriptive languages have reached a steady state, they show surprising stability and become almost transparent to their users. Mentally ill patients are perceived in a similar manner by most psychiatrists the world over. Stability and uniformity, however, should not be considered as features intrinsic to the descriptive system (or as necessarily advantageous). Part of the task of evaluating their usefulness is to explain their origin. In spite of the fact that an adequate theory of language stability (and change) would be of great use to current clinical practice, clinical historians have not yet started delving into these processes. Explanations for language stability can be based on assumptions of neurobiological invariance or on the effective and calming operation of social macro-concepts.

The historical and conceptual hypotheses outlined in this chapter have already opened a major research space for younger colleagues. There is urgent need to see whether the models in question apply to the history of each individual symptom and disease. The history of DP will achieve full usefulness only after such a painstaking analysis has been completed. Furthermore, the detailed historical study of concepts, documents, biographies, socio-political contexts and patient cohorts will in due course throw light on the mechanisms and rules whereby small cliques take upon themselves the honourable and awesome task of periodically reconstructing psychiatric nosology for all patients in the world and for the ordinary psychiatrists who look after them.

8 References

Abbagnano N (1961) Dizionario di filosofia. Unione Tipografico Torinese, Turin

Ackerknecht E (1957) Kurze Geschichte der Psychiatrie. Enke, Stuttgart

Ackerknecht E (1967) Medicine at the Paris Hospital 1794–1848. Johns Hopkins Press, Baltimore

Ackerknecht E (1985) Kurze Geschichte der Psychiatrie, 3rd edn. Enke, Stuttgart

Albarracín Teulón A (1983) La teoria celular. Historia de un paradigma. Alianza, Madrid

Albrecht FM (1970) A reappraisal of faculty psychology. J Hist Behav Sci 6: 36–40

Anderson EW (1963) Foreword. In: Jaspers K (1963) General psychopathology (translated by Hoenig J, Hamilton M). Manchester University Press, Manchester

Apert J (1919) L'hérédité morbide. Flammarion, Paris

Arnold T (1806) Observations on the nature, kinds, causes, and prevention of insanity, 2 vols, 2nd edn. Phillips, London

Babb L (1951) The Elizabethan malady. Michigan State College Press, East Lansing

Bacon F (1858, [1]1620) The novum organum. In: Devey J (ed) The physical and metaphysical works of Lord Bacon. Bohn, London

Baillarger JGF (1853) Essai sur une classification des différents genres de folie. Ann Méd Psychol 5: 545–566

Bain A (1859) The emotions and the will. Parker, London, pp 599–646

Banister H, Zangwill OL (1949) John Thomson McCurdy, 1886–1947. Br J Psychol 40: 1–4

Bannour W (1992) Jean Martin Charcot et l'hysterie. Métailié, Paris

Barrucand D (1967) Histoire de l'hypnose en France. Presses Universitaires de France, Paris

Barthes R (1972) Sémiologie et médecine. In: Bastide R (ed) Les sciences de la folie. Mouton, Paris, pp 34–46

Baruk H (1926) Les troubles mentaux dans les tumeurs cérébrales. Etude clinique, pathogénie, traitement. Doin, Paris

Baruk H (1967) La psychiatrie française de Pinel à nos jours. Presses Universitaires de France, Paris

Bastian HC (1870) Consciousness. J Ment Sci 15: 501–523

Battie W (1758) A treatise on madness. Whiston and White, London

Bayle ALJ (1826) Traité des maladies du cerveau. Gabon, Paris

Beauchesne H (1986) Histoire de la psychopathologie. Presses Universitaires de France, Paris
Beaugrand E (1865) Alienation. In: Dechambre A (ed) Dictionnaire encyclopédique des sciences médicales, vol 3. Asselin, Paris, pp 11–50
Beer D (1996a) The endogenous psychoses: a conceptual history. Hist Psychiatry 7: 1–29
Beer D (1996b) The dichotomies: psychoses/neuroses and functional/organic: a historical perspective. Hist Psychiatry 7: 231–255
Bentley M (1916) The psychological antecedents of phrenology. Psychol Rev Monogr 21: 102–115
Bercherie P (1980) Les fondements de la clinique. La Bibliothèque d'Ornicar, Paris
Bercherie P (1983) Genèse des concepts freudians. Navarin, Paris
Bercherie P (1988) Géographie du champ psychanalytique. Navarin, Paris
Berrios GE (1977) Henri Ey, Jackson et les idées obsédantes. Evol Psychiatr 62: 685–699
Berrios GE (1981a) Delirium and confusion in the 19thC: a conceptual history. Br J Psychiatry 139: 439–449
Berrios GE (1981b) Stupor: a conceptual history. Psychol Med 11: 677–688
Berrios GE (1981c) The two manias. Br J Psychiatry 139: 258–259
Berrios GE (1982a) Disorientation states and psychiatry. Comprehensive Psychiatry 23: 479–490
Berrios GE (1982b) Tactile hallucinations: conceptual and historical aspects. J Neurol Neurosurg Psychiatry 45: 85–293
Berrios GE (1983) Investigación biológica y psicopatología descriptiva. Rev Psicol (Lima) 1: 39–52
Berrios GE (1984a) Descriptive psychopathology: conceptual and historical aspects. Psychol Med 14: 303–313
Berrios GE (1984b) Epilepsy and insanity during the early 19th century. Arch Neurol 41: 978–981
Berrios GE (1985a) 'Depressive pseudodementia' or 'melancholic dementia'. A nineteenth century view. J Neurol Neurosurg Psychiatry 48: 393–400
Berrios GE (1985b) The psychopathology of affectivity: conceptual and historical aspects. Psychol Med 15: 745–758
Berrios GE (1987a) Dementia during the 17th and 18th Centuries. Psychol Med 17: 829–837
Berrios GE (1987b) Historical aspects of the psychoses: 19th century issues. Br Med Bull 43: 484–498
Berrios GE (1987c) The historical development of abnormal psychology. In: Miller E, Cooper P (eds) Historical background to abnormal psychology. Churchill Livingstone, Edinburgh, pp 26–51
Berrios GE (1988a) Depressive and manic states during the 19th century. In: Gorgotas A, Cancro R (eds) Depression and mania. Elsevier, New York, pp 13–25
Berrios GE (1988b) Melancholia and depression during the 19th century. Br J Psychiatry 153: 298–394
Berrios GE (1991) Delusions as wrong beliefs: a conceptual history. Br J Psychiatry 159[Suppl 14]: 6–13
*Berrios GE (1992a) Phenomenology, psychopathology and Jaspers: a conceptual history. Hist Psychiatry 3: 303–327
Berrios GE (1992b) Ajuriaguerra, Francia e Inglaterra. In: Aguirre JM, Guimón J (eds) Vida y obra de Julián de Ajuriaguerra. Aran, Madrid, pp 83–89
Berrios GE (1993) European views on personality disorders: a conceptual history. Comprehensive Psychiatry 34: 14–30
Berrios GE (1994a) Historiography of mental symptoms and diseases. Hist Psychiatry 5: 175–190
Berrios GE (1994b) Hallucinations: selected historical and clinical aspects. In: Critchley EMR (ed) The neurological boundaries of reality. Farrand, London, pp 229–250
Berrios GE (1994c) Delusions: selected historical and clinical aspects. In: Critchley EMR (ed) The neurological boundaries of reality. Farrand, London, pp 251–267
Berrios GE (1995) Conceptual problems in diagnosing schizophrenic disorders. In: Den Boer JA, Westenberg HGM, Van Praag HM (eds) Advances in the neurobiology of schizophrenia. Wiley, Chichester, pp 7–25
*Berrios GE (1996) The history of mental symptoms. Descriptive psychopathology since the 19th century. Cambridge University Press, Cambridge
Berrios GE, Beer D (1994) The notion of unitary psychosis: a conceptual history. Hist Psychiatry 5: 13–36
Berrios GE, Chen E (1993) Symptom-recognition and neural-networks. Br J Psychiatry 163: 308–314
Berrios GE, Dening T (1990) Biological and quantitative issues in neuropsychiatry. Behav Neurol 3: 247–259
Berrios GE, Freeman H (eds) (1991a) Alzheimer and the dementias. Royal Society of Medicine, London
Berrios GE, Freeman H (eds) (1991b) 150 years of British psychiatry 1841–1991. Gaskell, London
Berrios GE, Gili M (1995) Will and its disorders. A conceptual history. Hist Psychiatry 6: 87–104
Berrios GE, Hauser R (1988) The early development of Kraepelin's ideas on classification. A conceptual history. Psychol Med 18: 813–821
**Berrios GE, Porter R (eds) (1995) The history of clinical psychiatry. Athlone, London
Berrios GE, Marková IS, Olivares JM (1995) Hacia una teoría de la formación del síntoma. Psiquiatr Biol 2: 13–24
Billod T (1861) De la lésion de l'association des idées. Ann Méd Psychol 18: 540–552
Blakey R (1850) History of the philosophy of the human mind. Longman, Brown, Green and Longmans, London
Blasius D (1980) Der verwaltete Wahnsinn. Fischer, Frankfurt
Bollote G (1973) Moreau de Tours 1804–1884. Confront Psychiatr 11: 9–26
Bondy M (1974) Psychiatric antecedents of psychological testing (before Binet). J Hist Behav Sci 10: 180–194
Boring EG (1950) A history of experimental psychology. Appleton-Century-Crofts, New York, pp 250–261
Boring EG (1953) A history of introspection. Psychol Bull 50: 169–189
Boring EG (1961) The beginning and growth of measurement in psychology. Isis 52: 238–257
Boutroux E (1908) Etudes d'histoire de la philosophie. Alcan, Paris, pp 413–443
Bowman IA (1975) William Cullen (1710–1790) and the primacy of the nervous system. PhD Thesis, Indiana University
Boyer L (1873) Histoire de la médicine. In: Dechambre A (ed) Dictionnaire encyclopédique des sciences médicales, vol 6. Asselin and Masson, Paris, pp 1–209
Brett GS (1953) History of psychology. Allen and Unwin, London
Brooks GP (1976) The faculty psychology of Thomas Reid. J Hist Behav Sci 12: 65–77
Brown T (1828, [1]1820) Lectures on the philosophy of the human mind. Tait, Edinburgh
Browne J (1985) Darwin and the face of madness. In: Bynum WF, Porter R, Shepherd M (eds) The anatomy of madness, vol 1. Tavistock, London, pp 151–165

Buchner EF (1897) A study of Kant psychology. Psychol Rev (Monogr Suppl) 4: 1–87
Bucknill JC, Tuke DH (1858) A manual of psychological medicine. Churchill, London
Bühler K (1968) Ausdruckstheorie. Das System an der Geschichte aufgezeigt. Fischer, Sttutgart
Bulbena A, Berrios GE (1986) Pseudodementia: facts and figures. Br J Psychiatry 148: 87–94
Burt C (1962) The concept of consciousness. Br J Psychol 53: 229–242
Burton R (1883, [1]1621) The anatomy of melancholy. Chatto and Windus, London
Calmeil LF (1839) Manie. In: Adelon, Béclard, Bérard et al. (eds) Dictionnaire de médecine, vol 19. Béchet, Paris
Calmette E (1874) Considérations sur la valeur des symptoms en pathologie mentale. Thesis. Parent, Paris
Canguilhem G (1966) Le normal et le pathologique. Presses Universitaires de France, Paris
Cantor GN (1975) The Edinburgh phrenology debate: 1803–1828. Ann Sci 32: 195–218
Caro Baroja J (1988) Historia de la fisionomica. ISTMO, Madrid
Chambeyron AM (1827) Translator's introduction. In: Hoffbauer JC (1827) Médecine légale. Baillière, Paris
Charcot JM (1881) Clinical lectures on senile and chronic diseases (translated by Tuke WS). The New Sydenham Society, London, p 4
Charcot J M (1971, [1]1887) L'hystérie. Textes choisis et présentés par E Trillat. Privat, Paris
Chaslin P (1912) Eléments de sémiologie et clinique mentales. Asselin and Houzeau, Paris
Chaslin P (1914) La 'psychiatrie' est-elle une langue bien faite? Rev Neurol 26: 16–23
Chiarugi V (1793) Della pazzia in genere, e in specie. Tratatto medico-analitico con una centuria di osservazioni. Carlieri, Florence (recently translated by Mora G (1987) On insanity and its classification. Science History Publications, Canton)
Clark MJ (1988) 'Morbid introspection' unsoundness of mind, and British psychological medicine, c.1830–c.1900. In: Bynum WF, Porter R, Shepherd M (eds) The anatomy of madness, vol 3. Tavistock, London, pp 71–101
*Clarke E, Jacyna L S (1987) Nineteenth century origins of neuroscientific concepts. University of California Press, Berkeley, pp 220–241
Cooter RJ (1976) Phrenology and British alienists, c.1825–1845. Med Hist 20: 1–21, 135–151
Coser LA (1992) Introduction. In: Halbwachs M (1992) On collective memory. University of Chicago Press, Chicago
Couchoud PL (1913) Histoire de la manie jusqu'à Kraepelin. Rev Sci Psychol 1: 149–173
Crampton C (1978) The Cambridge School: the life, work and influence of J Ward, WHR Rivers, CS Myers and Sir F Bartlett. PhD dissertation, University of Edinburgh
Dagonet H (1881a) Conscience et aliénation mentale. Ann Méd Psychol 5: 368–397
Dagonet H (1881b) Conscience et aliénation mentale. Ann Méd Psychol 6: 19–32
Damiron P (1828) Essai sur l'histoire de la philosophie en France. Schubart and Heideloff, Paris
Danion JM, Keppi J, Singer L (1985) Une approche historique de la doctrine des dégénérescences et des constitutions psychopathiques. Ann Méd Psychol 146: 271–280
*Danziger K (1980) The history of introspection reconsidered. J Hist Behav Sci 16: 241–262
Darwin C (1904, [1]1872) The expression of the emotions in man and animals. Murray, London
Darwin E (1796) Zoonomia; or the laws of organic life. Johnson, London
Daston LJ (1982) The theories of the will versus the science of mind. In: Woodward WR, Ash MG (eds) The problematic science. Psychology in 19thC thought. Praeger, New York, pp 88–115
Daumezon G (1957) Reflexions sur la sémiologie psychiatrique. Evol Psychiatr 22: 207–237
*de Boor W (1954) Psychiatrische Systematik. Springer, Berlin Heidelberg New York
Del Pistoia L (1971) Le probléme de la temporalité dans la psychiatrie française classique. Evol Psychiatr 36: 445–474
Delasiauve M (1861) Des diverses formes mentales. J Méd Ment 1: 4–14
Deshaies G (1967) Psychopathologie générale. Presses Universitaires de France, Paris
Despine P (1876) Du role de la psychologie dans la question de la folie. Ann Méd Psychol 34: 161–175
Desruelles M, Léculier P, Gradien MP (1934) Contribution à l'histoire des classifications psychiatriques. Ann Méd Psychol 92: 637–675
Destutt de Tracy ALC Comte (1818, [1]1801) Elements d'idéologie. Courcier, Paris
Devereux G (1980) Basic problems of ethnopsychiatry. The University of Chicago Press, Chicago, pp 3–71
Diethelm O, Heffernan TF (1965) Felix Platter and psychiatry. J Hist Behav Sci 1: 10–23
Dijksterhuis EJ (1961) The mechanization of the world picture. Oxford Universitry Press, Oxford
Donalies C (1971) Zur Systematik in der Psychiatrie vor Wernicke, Kraepelin und Bonhoeffer. Psychiatrie Neurol Psychol 23: 411–419
Donelly M (1983) Managing the mind. A study of medical psychology in early 19thC Britain. Tavistock, London
Dörner K (1969) Bürger und Irre. Europäische Verlagsanstalt, Frankfurt am Main
Dowbiggin I (1985) Degeneration and hereditarianism in French mental medicine 1840–1990: psychiatric theory as ideological adaptation. In: Bynum WF, Porter R, Shepherd M (eds) The anatomy of madness, vol 1. People and ideas. Tavistock, London, pp 189–232
Drabkin IE (1955) Remarks on ancient psychopathology. Isis 46: 223–234
Drevet A (1968) Maine de Biran. Presses Universitaires de France, Paris
*Drinka GF (1984) The birth of neurosis. Myth, malady, and the Victorians. Simon and Schuster, New York
Dumas G (1908) Qu'est-ce que la psychologie pathologique? J Psychol Norm Pathol 5: 10–22
Dwelshauvers G (1920) La psychologie française contemporaine. Alcan, Paris
Ebbinghaus H (1964, [1]1885) Memory. A contribution to experimental psychology (translated by Ruger HA, Bussenius CE, Hilgard ER). Dover, New York, pp 6–21
Eigen JP (1995) Witnessing insanity. Yale University Press, New Haven
**Ellenberger HF (1970) The discovery of the unconscious. The history and evolution of dynamic psychiatry. Lane, London
Emerton NE (1984) The scientific reinterpretation of form. Cornell University Press, Ithaca
Esquirol E (1838) Des maladies mentales. Baillière, Paris

Ey H (1952) Le développement 'mecaniciste' de la psychiatrie. In: Ey H (ed) Études psychiatriques. Desclée de Brouwer, Paris
Ey H (1966) La conscience. Presses Universitaires de France, Paris
Ey H (1978) La notion de 'maladie morale' et le 'traitement moral' dans la psychiatrie française et allemande du début du XIXe siècle. Perspect Psychiatr 65: 12–35
Ey H, Mignot H (1947) La psychologie de J Moreau de Tours. Ann Méd Psychol 2: 225–241
Eysenck HJ (ed) (1960) Handbook of abnormal psychology. Pitman, London
Faber K (1923) Nosography in modern internal medicine. Oxford University Press, London
Falret JP (1864) Des maladies mentales et des asiles d'aliénés. Ballière, Paris
Fancher RE (1977) Brentano's psychology from an empirical standpoint and Freud's early metapsychology. J Hist Behav Sci 13: 207–227
Ferrater-Mora J (1958) Diccionario de filosofía. Editorial Sudamericana, Buenos Aires
Finzen A (1996) Massenmord ohne Schuldgefühl. Die Tötung psychisch Kranker und geistig Behinderter auf dem Dienstweg. Psychiatrie-Verlag, Bonn
Fish F (1967) Clinical psychopathology. Wright, Bristol
Flashar H (1966) Melancholie und Melancholiker in den medizinischen Theorien der Antike. de Gruyter, Berlin
Fodor J (1983) The modularity of mind. MIT, Massachusetts
Foucault M (1972a, [1]1961) Histoire de la folie à l'âge classique. Gallimard, Paris
Foucault M (1972b, [1]1963) Naissance de la clinique, 2nd edn. Presses Universitaires de France, Paris
Fourquet F, Murard L (1980) Histoire de la psychiatrie de secteur, 2nd edn. Recherches, Paris
Foville A Jr (1872) Nomenclature et classification des maladies mentales. Ann Méd Psychol 30: 5–35
Freeman H, Berrios GE (eds) (1996) 150 years of British psychiatry. II. The aftermath. Athlone, London
Friedlander R (1973) Benedict Augustin Morel and the development of the theory of degenerescence. PhD dissertation, University of California
Fritzsch T (1932) Juan Federico Herbart. Labor, Barcelona
Fuentenebro F, Berrios GE (1997) Charles Blondel and la conscience morbide. Hist Psychiatry 8: 277–295
Gardiner HM, Metcalf RC, Beebe-Center JG (1937) Feeling and emotion. A history of theories. American Book, New York
Gaston A, Tatarelli M (1984) Analyse critique de l'évolution du concept d'endogène. Evol Psychiatr 2: 569–575
Gauchet M, Swain G (1980) Du traitement de la manie aux passions: la folie et l'union de l'âme et du corps. In: Esquirol E (ed) Des passions. Deux Mondes, Paris
*Genil-Perrin GPH (1913) Histoire des origines et de l'évolution de l'idée de dégénérescence en médecine mentale. Leclerc, Paris
Georgin B (1980) Remarques sur le discours nosologique en psychiatrie. Evol Psychiatr 45: 5–17
Gillespie RD (1929) The clinical differentiation of types of depression. Guy's Hosp Rep 79: 306–344
Gilman SL (1982) Seeing the insane. Wiley, New York
Goldstein J (1988) Console and classify. Cambridge University Press, Cambridge
Grave SA (1960) The Scottish philosophy of common sense. Clarendon, Oxford
Greenway AP (1973) The incorporation of action into associationism. The psychology of Alexander Bain. J Hist Behav Sci 9: 42–52
Griesinger W (1865) La pathologie mentale au point de vue de l'école somatique allemande. Ann Méd Psychol 23: 1–31
Griesinger W (1867) Mental pathology and therapeutics (translated by Robertson CL, Rutherford J). The New Syndenham Society, London
Guiraud P (1943) Evolution de l'idée de démence. Ann Méd Psychol 101: 186–199
Gurney E (1885) Hallucinations. Mind 10: 161–199
Haas FJ (1864) Essai sur les avantages cliniques de la doctrine de Montpellier. Baillière, Paris, pp 115–154
Hamilton M (1974) Fish's clinical psychopathology. Wright, Bristol
Hart B (1927) Psychopathology. Its development and its place in medicine. University Press, Cambridge
Haslam J (1809) Observations on madness. Callow, London
Heiberg JL (1927) Geisteskrankheiten in klassischen Altertum. Z Psychiatrie 86: 1–44
Heinroth JC (1818) Lehrbuch der Störungen des Seelenlebens [translated by Schmorak J (1975) Textbook of disturbances of mental life]. Johns Hopkins, Baltimore
Helmchen H (1985) Verbal and non-verbal psychopathology as a necessary element of classification. In: Mental disorders, alcohol- and drug-related problems. International perspectives on their diagnosis and classification. Excerpta Medica, Amsterdam, pp 177–181
Herbart JF (1884, [1]1806) Johann Friedrich Herbart's Kurze Encyklopädie der Philosophie. Voss, Hamburg
Hermle L (1986) Die Degenerationslehre in der Psychiatrie. Fortschr Neurol Psychiatrie 54: 69–79
Heron MJ (1965) A note on the concept endogenous–exogenous. Br J Med Psychol 38: 241–245
Herzen A (1880) Fisiología de la voluntad (translated by Ocina A). Iravedra, Madrid
Hesnard A (1971) De Freud à Lacan. Editions ESF, Paris
Hilgard ER (1980) The trilogy of mind: cognition, affection and connotation. J Hist Behav Sci 16: 107–117
Hilts VL (1981) Statist and statistician. Arno, New York
Hinshelwood RD (1991) Psychodynamic psychiatry before World War I. In: Berrios GE, Freeman H (eds) 150 years of British psychiatry 1841–1991. Gaskell, London, pp 197–205
*History of Psychiatry (1996) The psychoses. Hist Psychiatry 7: 1–192
Hoeldtke R (1967) The history of associationism and British medical psychology. Med Hist 11: 46–64
Hoff P (1985) Zum Krankheitsbegriff bei Emil Kraepelin. Nervenarzt 56: 510–513
Hoff P (1994) Emil Kraepelin und die Psychiatrie als klinische Wissenschaft. Springer, Berlin Heidelberg New York
Huertas R (1985) Valentín Magnan y la teoría de la degeneración. Rev Asoc Esp Neuropsiquiatr 5: 361–367
Huertas R (1987) Locura y degeneración. CSIC, Madrid
Hunter R, Macalpine I (1963) Three hundred years of psychiatry 1535–1860. Oxford University Press, New York
*Jackson SW (1986) Melancholia and depression. From Hippocratic to modern times. Yale University Press, New Haven
Jacyna LS (1982) Somatic theories of mind and the interests of medicine in Britain. Med Hist 26: 233–258
Jalley M, Lefebvre JP, Feline A, Kaufmann E et al (1977) Essai sur les maladies de la tête par E Kant. Evol Psychiatr 42: 203–230

Janzarik W (1959) Dynamische Grundkonstellationen in endogenen Psychosen. Springer, Berlin
*Janzarik W (1969) Nosographie und Einheitpsychose. In: Huber G (ed) Schizophrenie und Zyklothymie. Ergebnisse und Probleme. Thieme, Sttutgart
Jaspers K (1913) Allgemeine Psychopathologie, 1st edn. Springer, Berlin
Jaspers K (1957) Philosophical autobiography. In: Schilpp A (ed) The philosophy of Karl Jaspers. Open Court, Illinois, pp 5–94
Jaspers K (1963) General psychopathology (translated by Hoenig J, and Hamilton MW). Manchester University Press, Manchester
Jastrow J (1902) Psychopathology. In: Baldwin JM (ed) Dictionary of philosophy and psychology. McMillan, London
Jastrow J, Baldwin JM (1901) Psychosis. In: Baldwin JM (ed) Dictionary of philosophy and psychology. McMillan, London, p 392
Jones K (1972) A history of the mental health services. Routledge and Kegan Paul, London
Juliard P (1970) Philosophies of language in eighteenth-century France. Mouton, The Hague
Kageyama J (1984) Sur l'histoire de la monomanie. Evol Psychiatr 49: 155–162
Kahlbaum K (1863) Die Gruppierung der psychischen Krankheiten. Kafemann, Danzig [translation by Berrios GE (1996) The classification of mental disorders, part III. Hist Psychiatry 7: 167–181]
Kant E (1914, [1]1790) Critique of judgement (translated by Bernard JH). Longmans, London
Kant E (1974, [1]1798) Anthropology from a pragmatic point of view (translated with an introduction and notes by Gregor MJ). Nijhoff, The Hague, pp 73–89
Kauffmann F (1957) Karl Jaspers and a philosophy of communication. In: Schilpp A (ed) (1957) The philosophy of Karl Jaspers. Open Court, Illinois, pp 210–295
Keller W (1954) Psychologie und Philosophie des Wollens. Reinhardt, Munich
King LS (1968) Signs and symptoms. J Am Med Assoc 206: 1063–1065
King LS (1982) Medical thinking. Princeton University Press, Princeton, pp 131–183
Kolle K (1956) Große Nervenärzte, vol 1. Thieme, Stuttgart
Kolle K (1959) Große Nervenärzte, vol 2. Thieme, Stuttgart
Kolle K (1963) Große Nervenärzte, vol 3. Thieme, Stuttgart
Kraepelin E (1896) Der psychologische Versuch in der Psychiatrie. Psychol Arbeiten 1: 1–91
Kraepelin E (1920) Die Erscheinungsformen des Irreseins. Z Ges Neurol Psychiatrie 22: 1–29
Kraepelin E (1921) Manic-depressive insanity and paranoia (translated from the by Barclay RM). Livingstone, Edinburgh
Kraepelin E (1924) Paul Julius Mòbius (1853–1907). In: Kirchhoff T (ed) Deutsche Irrenärzte, vol 2. Springer, Berlin, pp 274–279
Kraepelin E (1983) Lebenserinnerungen. Springer, Berlin Heidelberg New York
Krafft-Ebing R (1893) Lehrbuch der Psychiatrie. Enke, Sttutgart
Kreuter A (1996) Deutschsprachige Neurologen und Psychiater. Ein biographisch-bibliographisches Lexikon von den Vorläufern bis zur Mitte des 20. Jahrhunderts. Saur, Munich
Lacan J (1977, 1932) De la psychose paranoïaque dans ses rapports avec la personnalité suivie de premiers écrits sur la paranoïa. Seuil, Paris
Laín Entralgo P (1961) La historia clínica. Salvat, Barcelona
Laín Entralgo P (1978) Historia de la medicina. Salvat, Barcelona
Laín Entralgo P (1982) El diagnóstico médico. Salvat, Barcelona
Land SK (1974) From signs to propositions. The concept of form in 18th century semantic theory. Longman, London
Landre-Beauvais AJ (1813) Séméiotique ou traité des signes des maladies, 2nd edn. Brosson, Paris
Lanteri Laura G (1966) Les apports de la linguistique à la psychiatrie contemporaine. Masson, Paris
Lanteri Laura G (1968) Psychologie pathologique. In: Encyclopédia medico-chirurgicale, vol 1. Paris
Lanteri Laura G (1970) Histoire de la phrénologie. Presses Universitaires de France, Paris
Lanteri Laura G (1972) La chronicité dans la psychiatrie moderne française. Annales 3: 548–568
Lanteri Laura G (1982) La connaissance clinique: histoire et structure en médecine et en psychiatrie. Evol Psychiatr 47: 423–469
Lanteri Laura G (1983) La sémiologie psychiatrique: son évolution et son état en 1982. Evol Psychiatr 48: 327–363
*Lanteri Laura G (1984) La sémiologie de J P Falret. Perspect Psychiatr 22: 104–110
Lanteri Laura G (1986) Acuité et pathologie mentale. Evol Psychiatr 51: 403–418
Lapie P (1902) Logique de la volonté. Alcan, Paris
Laromiguière P (1820) Leçons de philosophie ou essai sur les facultés de l'âme. Brunot-Labbe, Paris
Larson JL (1971) Reason and experience. The representation of natural order in the work of Carl Von Linné. University of California Press, Berkeley
Lavater JC (1891, [1]1772) Essays on physiognomy. Ward and Lock, London
Leary DE (1982) Immanuel Kant and the development of modern psychology. In: Woodward WR, Ash M (eds) The problematic science: psychology in 19thC thought. Praeger, New York, pp 17–42
*Leibbrand W, Wettley A (1961) Der Wahnsinn. Geschichte der abendländischen Psychopathologie. Alber, Freiburg
Lesky E (1970) Structure and function in Gall. Bull Hist Med 44: 297–314
Lewis A (1970a) Paranoia and paranoid: a historical perspective. Psychol Med 1: 2–12
Lewis A (1970b) William Mayer-Gross: an appreciation. Psychol Med 7: 11–18
Lewis A (1971) 'Endogenous' and 'exogenous': a useful dichotomy. Psychol Med 1: 191–196
Linas A (1871) Manie. In: Dechambre T (ed) Dictionnaire encyclopédique de sciences védicales, vol 4 (2nd series). Asselin, Paris, pp 507–560
Llópis B (1954) La psicósis única. Arch Neurobiol 17: 3–39
Locke J (1959, [1]1690) An essay concerning human understanding. Dover, New York
López-Piñero JM (1983) Historical origins of the concept of neuroses (translated by Berrios D). Cambridge University Press, Cambridge
López-Piñero JM, Morales Meseguer JM (1970) Neurosis y psicoterapia. Espasa Calpe, Madrid
Losserand J (1967) Les rapports du physique et du moral de l'homme de Cabanis à Auguste Comte. Evol Psychiatr 32: 573–601
MacDonald M (1981) Mystical bedlam. Cambridge University Press, Cambridge
Mackenzie TB, Rosemberg SD, Bergen BJ, Tucker GJ (1978) The manipulative patient: an interactional approach. Psychiatry 41: 264–271

Macmillan NA, Creelman CD (1991) Detection theory. Cambridge University Press, Cambridge
Magnan V, Sérieux P (1911) Délire chronique. In: Marie A (ed) Traité international de psychologie pathologique, vol 2. Alcan, Paris, pp 605–639
Mairet A, Ardin-Delteil P (1907) Hérédité et prédisposition. Coulet, Montpellier
Malberg B (1977) Teoría de los signos. Siglo Veintiuno, Mexico
Manetti G (1993) Theories of the sign in classical antiquity (translation by Richardson C). Indiana University Press, Bloomington
Mangin-Lazarus C (1994) Maurice Dide: Paris 1873 Buchenwald 1944. Érès, Paris
Mantegazza P (1878) Physiognomy and expression. Scott, London
Marchais P (1995) 'L'automatisme mental' de Clérambault et ses liens avec la pensée psychiatrique française. In: Un siglo de psiquiatria en España. Coleccion salud mental: pensamiento y practica. Extra Editorial, Madrid, pp 285–301
Markman E (1994) Natural kinds. In: Kornblith H (ed) Naturalizing epistelomology, 2nd edn. MIT Press, Cambridge, pp 77–104
Marshall ME (1982) Physics, metaphysics and Fechner's psychophysics. In: Woodward WR, Ash M (eds) The problematic science: psychology in 19thC thought. Praeger, New York, pp 65–87
Martinet J (1973) Clefs pour la sémiologie. Seghers, Paris
*Marx OM (1970) What is the history of psychiatry. J Orthopsychiatry 40: 593–605
Marx OM (1980) The case of the chronic patient seen in a historical perspective. In: Wallace ER, Pressley LC (eds) Essays in the history of psychiatry. Hall Psychiatric Institute, Columbia, pp 22–27
Matthews JR (1995) Quantification and the quest for medical certainty. Princeton University Press, Princeton
Maudsley H (1885) Responsibility in mental illness. Paul and Trench, London
Mayer-Gross W, Guttmann E (1937) Schema for the examination of organic states. J Ment Sci 83: 440–448
McDougall W (1919) The present position in clinical psychology. J Ment Sci 45: 141–152
McDougall W (1926) An outline of abnormal psychology. Methuen, London
McDougall W (1930) William McDougall. In: Murchison C (ed) A history of psychology in autobiography, vol. 1. Clark University Press, New York, pp 191–223
Mechler A (1963) Degeneration und Endogenität. Nervenarzt 5: 219–226
Meeus F (1908) Epilepsie et délire chronique. Contribution à l'étude des psychoses combinées. Ann Méd Psychol 7: 353–382
Mendel E (1907, [1]1907) Textbook of psychiatry. A psychological study of Insanity (translated by Krauss WC). Davis, Philadelphia
Menninger K (1964) The vital balance. Viking, New York, pp 419–509
Menninger K, Ellenberger H, Pruyser P, Mayman M (1958) The unitary concept of mental illness. Bull Menninger Clin 22: 4–12
Merani AL (1976) Historia crítica de la psicología. Grijalbo, Barcelona
Merskey H (1991) Shell-shock. In: Berrios GE, Freeman H (eds) 150 years of British psychiatry 1841–1991. Gaskell, London, pp 245–267
Meyer A (1901) Psychosis. In: Baldwin JM (ed) Dictionary of philosophy and psychology, vol 2. McMillan, London, pp 392–394
Middleton E, Turnbull W, Ellis T, Davison J (1780) The new complete dictionary of arts and sciences. Hogg, London
Mill J (1829) Analysis of the phenomena of the human mind. Longmans and Dyer, London
Mill JS (1898, [1]1843) A system of logic. Longmans and Green, London, pp 76–86
Monakow C, Mourgue R (1928) Introduction biologique à l'étude de la neurologie et de la psychopathologie. Alcan, Paris
Moore FCT (1970) The psychology of Maine de Biran. Clarendon, Oxford
Mora G (1975) Heinroth's contribution to psychiatry. In: Heinroth JC (ed) Textbook of disturbances of mental life (translated by Schmorak J). Johns Hopkins University Press, Baltimore, pp ix–lxxv
Mora G (1981) Cabanis, neurology and psychiatry. In: Mora G (ed) On the relations between the physical and moral aspects of man by PJG Cabanis, vol 1. Johns Hopkins Press, Baltimore, pp 45–90
Moravia S (1983) The capture of the invisible for a (pre)history of psychology in eighteenth century France. J Hist Behav Sci 19: 370–378
Mordier JP (1981) Les débuts de la psychanalyse en France 1895–1926. Maspero, Paris
Moreau (de Tours) J (1859) La psychologie morbide dans ses rapports avec la philosophie de l'histoire ou de l'influence des névropathies sur le dynamisme intellectuel. Masson, Paris, pp 193–243
Morel BA (1857) Traité des dégénérescences physiques, intellectuelles et morales de l'espèce humaine et des causes qui produisent ces variétés maladives. Baillière, Paris
Morel BA (1860) Traité de maladies mentales. Baillière, Paris
Murphy TD (1981) Medical knowledge and statistical methods in early nineteenth-century France. Med Hist 25: 301–319
Nicole JE (1946) Psychopathology, 4th edn. Baillière, London
O'Shaughnessy B (1980) The will. Cambridge University Press, Cambridge
Owen ARG (1971) Hysteria, hypnosis and healing. The work of JM Charcot. Dobson, London
Parchappe MJB (1856) Rapport sur la statistique de l'aliénation mentale. Ann Méd Psychol 2: 1–6
Parish E (1897) Hallucinations and illusions. Scott, London
Parsons J (1747) Human physiognomy explain'd. Cronian Lectures on muscular motion for the year 1746. Transactions Royal Society, London, pp 60–62
Paulhan F (1903) La volonté. Doin, Paris
Pearson ES (ed) (1978) The history of statistics in the 17th and 18th centuries. Griffin, London
Pearson K (1892) The grammar of science. Scott, London
Perrot J-C, Woolf S J (1984) State and statistics in France 1789–1815. Harwood, London
Peset JL (1983) Ciencia y marginación. Sobre negros, locos y criminales. Crítica, Barcelona
*Peters UH (1996) The emigration of German psychiatrists to Britain. In: Freeman H, Berrios GE (eds) 150 years of British psychiatry, vol II. The aftermath. Athlone, London, pp 565–580

*Pichot P (1995) The birth of the bipolar disorder. Eur Psychiatry 10: 1–10
Pick D (1989) Faces of degeneration. Cambridge University Press, Cambridge
Pierce CS (1931–1935) Collected papers, vol II. Harvard University Press, Cambridge
Pigeaud J (1986) La génie et la folie: étude sur la psychologie morbide de J Moreau de Tours. Evol Psychiatr 51: 587–608, 193–255
Pinel P (1809) Traité Médico-Philosophique sur l'aliénation mentale, 2nd edn. Brosson, Paris
Pinel P (1818) Nosographie philosophique ou la méthode de l'analyse appliquée à la médecine, 6th edn. Brosson, Paris
Pines M (1991) The development of the psychodynamic movement. In: Berrios GE, Freeman H (eds) 150 years of British psychiatry 1841–1991. Gaskell, London, pp 206–231
Platt AM, Diamond BL (1965) The origins and development of the 'wild beast' concept of mental illness and its relation to theories of criminal responsability. J Hist Behav Sci 1: 355–367
*Porter R (1987) Mind-forg'd manacles. A history of madness in England from the Restoration to the Regency. Athlone, London
Porter TM (1986) The rise of statistical thinking. Princeton University Press, Princeton
*Postel J (1984) Images de la folie au XVIIIe siècle: quelques differénces de sa représentation dans les littératures française et britannique au Siècle des Lumières. Evol Psychiatr 49: 707–718
Prichard JC (1835) A treatise on insanity and other disorders affecting the mind. Sherwood, Gilbert and Piper, London
Quetel C, Morel P (1979) Les fous et leurs médecines. De la Renaissance au XXe siècle. Hachette, Paris
Radden J (1996) Lumps and bumps: Kantian faculty psychology. Philosophy Psychiatry Psychol 3: 1–14
Radicke (1861) On the application of statistics to medical enquiries (translated by Bond FT). The New Sydenham Society, London (first appeared in 1958 in Wunderlichs Arch Physiol Heilkd 2)
Ramul K (1960) The problem of measurement in the psychology of the eighteenth century. Am Psychol 15: 256–265
Ravaisson F (1885) La philosophie en France au XIXe siècle, 2nd edn. Hachette, Paris
*Raynier J, Beaudouin H (1961) L'assistance psychiatrique française, 3rd edn. Librarie Le François, Paris
Régis E (1906) Précis de psychiatrie. Doin, Paris, pp 116–118
Remond MM, Lagriffe L (1902) Essai sur la classification en psychiatrie. Gazette Hop Civ Milit 75: 973–976, 983–987
Renaudin E (1856) Observations sur les recherches statistiques relatives à l'aliénation mentale. Ann Méd Psychol 2: 339–360
Rennert H (1968) Wilhelm Griesinger und die Einheitpsychose. Wissenschaftl Z Humboldt Uni 17: 15–16
Reports of the Sessions of the Société Médico-Psychologique (1854) Discussion sur la monomanie. Ann Méd Psychol 6: 99–118, 273–298, 464–474, 629–644
Ribot T (1871) L'hérédité psychologique. Alcan, Paris
Ribot T (1885) La psychologie allemande contemporaine, 2nd edn. Alcan, Paris
Ribot T (1904) Les maladies de la volonté. Alcan, Paris
Roccatagliata G (1973) Storia della psichiatria antica. Hoepli, Milan
Rogers D (1988) Psychiatry and the Necker cube. Neurological and psychological conceptions of psychiatric disorder. Behav Neurol 1: 3–10
Rothschuh KE (1973) History of physiology (translated by Risse GB). Krieger, New York
Roubinovitch J. (1896) Des variétés cliniques de la folie en France et en Allemagne. Doin, Paris
Royer-Collard AA (1843) Examen de la doctrine de Maine de Biran. Ann Méd Psychol 2: 1–45
Ryle G (1929) Critical notice of Martin Heidegger's 'Sein und Zeit'. Mind 38: 355–370
Ryle G (1970) Autobiographical. In: Wood O, Pitcher G (eds) Ryle. McMillan, London, pp 1–15
Salas J (1920) Los degenerados en sociedad. Moya, Madrid
Sauri JJ (1969) Historia de las ideas psiquiátricas. Lohla, Buenos Aires
Sauri JJ (1972) Las significaciones del vocablo psicósis. Acta Psiquiátr Psicol Am Latin 18: 219–226
Saury H (1886) Étude clinique sur la folie héréditaire. Delahaye and Lecrosnier, Paris
Savage G H (1898) Insanity and allied neuroses. Practical and clinical. Cassell, London
Schiller F (1982) A Möbius strip. Fin-de-siècle neuropsychiatry and Paul Möbius. University of California Press, Berkeley
Schliack H, Hippius H (eds) (1998) Nervenärzte. Biographien. Thieme, Stuttgart
Schofield RE (1970) Mechanism and materialism. British natural philosophy in an age of reason. Princeton University Press, Princeton
Scull A (1979) Museums of madness, Penguin, London
Sedler MJ, Dessain EC (1983) Falret's discovery: the origins of the concept of bipolar affective illness. Am J Psychiatry 140: 1127–1133
Séglas J (1895) Leçons cliniques sur les maladies mentales. Asselin, Paris
Semelaigne R (1932) Les pionniers de la psychiatrie française (après Pinel), vol 2. Baillière, Paris
Seth A (1890) Scottish philosophy. Blackwood, Edinburgh
*Shallice T (1988) From neuropsychology to mental structure. Cambridge University Press, Cambridge, pp 18–34, 269–404
Shrout PE, Spitxer RL, Fleiss JL (1987) Quantification of agreement in psychiatric diagnosis revisited. Arch Gen Psychiatry 44: 172–177
Shryock RH (1961) The history of quantification in medical science. Isis 52: 215–237
Siegel RE (1973) Galen on psychology, psychopathology, and function and diseases of the nervous system. Karger, Basel, pp 272–274
Simon B (1978) Mind and madness in ancient Greece. Cornell University Press, Ithaca
Slater E, Roth M (1969) Clinical psychiatry, 3rd edn. Baillière, London
Slaughter MM (1982) Universal languages and scientific taxonomy in the seventeenth century. Cambridge University Press, Cambridge
Smith R (1979) Mental disorder, criminal responsibility, and the social history of theories of volition. Psychol Med 9: 13–19
Smith R (1992) Inhibition. History and meaning in the sciences of mind and brain. Free Association, London
Sokal MM, Davis AB, Merzbach UC (1976) Laboratory instruments in the history of psychology. J Hist Behav Sci 12: 59–64
Spiegelberg H (1982) The phenomenological movement, 3rd edn. Nijhoff, The Hague
Spoerl HD (1936) Faculties versus traits: Gall's solution. Character Personality 4: 216–231

Stierlin H (1974) Karl Jaspers' psychiatry in the light of his basic philosophical position. J Hist Behav Sci 10: 213–226
Swain G (1978) L'aliéné entre le médecin et le philosophe. Perspect Psychiatr 65: 90–99
Talbot ES (1898) Degeneracy. Scott, London
Tamburini N (1881) La théorie des hallucinations. Rev Sci Fr Etrang 27: 138–142
Taylor FK (1966) Psychopathology. Butterworths, London
Temkin O (1965) The history of classification in the medical sciences. In: Katz MM, Cole JO, Barton WE (eds) The role and methodology of classification in psychiatry and psychopathology. United States Department of Health, Washington, pp 11–20
Tuke DH (1892a) Dictionary of psychological medicine. Churchill, London
Tuke DH (1892b) Psychosis. In: Tuke DH (ed) A dictionary of psychological medicine, vol 2. Churchill, London, p 1025
Underwood EA (1951) The history of the quantitative approach in medicine. Br Med Bull 7: 265–274
Verwey G (1985) Psychiatry in an anthropological and biomedical context. Philosophical presuppositions and implications of German psychiatry, 1820–1870. Reidel, Dordrecht
Vié MJ (1940) Sur l'existence d'entités morbides en psychiatrie, l'utilité et l'orientation de l'effort nosologique. Ann Méd Psychol 98: 347–358
Viziolo R, Bietti C (1966) Il problema della coscienza in neuropsichiatria. Omnia Medica, Pisa
*Vliegen J (1980) Die Einheitpsychose. Enke, Stuttgart
von Feuchtersleben E (1845) Lehrbuch der ärztlichen Seelenkunde. Gerold, Vienna
von Feuchtersleben E (1847) The principles of medical psychology (translated by Lloyd HE, Babington BG). Sydenham Society, London
Wahrig-Smith B (1985) Der junge Wilhelm Griesinger. Narr, Tubingen
Walk A (1964) Mental Hospitals. In: Pointer FNL (ed) The evolution of hospitals in Britain. Pitman, London, pp 123–146
Walker C (1988) Philosophical concepts and practice: the legacy of Karl Jaspers' psychopathology. Curr Opin Psychiatry 1: 624–629
Walker C (1993) Karl Jaspers as a Kantian psychopathologist. I. The philosophical origins of the concept of form and content. Hist Psychiatry 4: 209–238
Walker N (1968) Crime and insanity in England. 1. The historical perspective. Edinburgh University Press, Edinburgh
Walter RD (1956) What became of the degenerate. J Hist Med 11: 422–429
Warner F (1892) Psychosis. In: Tuke DH (ed) A dictionary of psychological medicine, vol 2. Churchill, London, pp 1025–1034
Warren HC (1921) History of the Association Psychology. Scribners, New York
Watson RI (1963) The great psychologists. Lippincott, New York, pp 233–237
Werlinder H (1978) Psychopathy: a history of the concepts. Borgstrom Tryckeri, Motala
West DJ, Walk A (eds) (1977) Daniel McNaughton. Gaskell, London
Wettley A (1959) Zur Problemgeschichte der 'dégénérescences'. Sudhoffs Arch 43: 193–212
Whewell W (1857) History of the inductive sciences. Parker, London
Windelband W (1948) Historia de la filosofía moderna. Nova, Buenos Aires
Winslow F (1848) Review of the principles of medical psychology. J Psychol Med Ment Pathol 1: 247–263, 499–512
Wittchen HU (ed) (1996) Comorbidity of mood disorders. Br J Psychiatry 168[Suppl 30]: 7–134
Wulff HR, Pedersen SA, Rosenberg R (1986) Philosophy of medicine. Blackwell, Oxford
Young RM (1970) Mind, brain and adaptation in the nineteenth century. Clarendon, Oxford
Zupan ML (1976) The conceptual development of quantification in experimental psychology. J Hist Behav Sci 12: 145–158

CHAPTER

2

H. Dilling

Psychiatric Classification

Nomina si nescis, pereat et cognitio rerum.
CARL VON LINNÉ

Translator: E. Taub

1 Diagnosis and Classification

Medical diagnosis is the differential recognition, assessment, and naming of disease patterns. The term "diagnosis" may refer either to the description of the disease pattern itself or to the process of formulating this description, i.e., the diagnostic process. "Classification" refers to the organization of a large number of different cases, all of which differ from each other in certain defined ways, into a system of classes or to the process of assigning individual cases or diagnoses to the classes within a system (Mombour 1975). Diagnosis and classification are closely related to each other: diagnoses require organization, just as classification requires content.

Diagnosis serves multiple purposes: differential description, the identification of cases, the identification of indications for therapy, the determination of a prognosis, scientific classification, and, finally, documentation in the health sector for administrative, accounting, and planning purposes (Dilling and Dittmann 1990). The multiplicity of demands placed upon diagnostic classification systems can perhaps be satisfied only by the use of several systems or system versions. In light of the remarkable progress made recently by scientific medicine, including in biological psychiatry, the field of clinical diagnosis and classification has considerable catching up to do (Jablensky 1988).

2 History of Psychiatric Classification

The systematic classification of medical diagnoses has its antecedents in the development of descriptive natural science in the eighteenth century. Carl von Linné (Linnaeus), after publishing his landmark work on the classification of plants, went on to produce the *Genera morborum*, a systematic classification of diseases, the fifth chapter of which deals with mental disorders (Linné 1742). William Cullen's *Synopsis nosologiae methodicae*, which also contained a classification of mental illnesses, appeared shortly thereafter (1772). These two broad surveys marked the beginning of the steady development of psychiatric diagnosis and diagnostic classification continuing over 200 years until the present.

In Germany, in 1768, Immanuel Kant published his *Kleine Onomastik der Gebrechen des Kopfes* ("Brief Nomenclature of Afflictions of the Head"), ranging from "Dummköpfigkeit" ("stupidity") to "Narrheit" ("folly") and from "Blödsinnigkeit" ("feeble-mindedness") to "Tollheit" ("madness") and including a classical description of delusional thinking (Kant 1983). In 1818, Heinroth, then serving as the first professor of psychiatry in Germany, published a quite extensive classification of mental illnesses, subdivided like Linnaeus' system into classes, orders, genera, species, and varieties (Heinroth 1818). Heinroth failed to base his system on unambiguous characteristics of illnesses, as Linnaeus had classified plants, for example, on the basis of distinct features such as anthers and calyces. What resulted was a set of disease designations rather than adequate descriptions that could be used to identify diseases reproducibly.

Pinel (1809) made a more practical, and thus superior, contribution by limiting the number of diagnoses to a few major ones such as mania, melancholy, dementia, and idiocy. He and his pupil Esquirol (1838) had wide-reaching influence in the mid-nineteenth century; psychiatric diagnosis in many countries, including Germany (Eschenburg 1855), followed the lines they laid down. In 1845, Griesinger published the first edition of his textbook, in which he propounded a unified concept of psychosis, like his teacher Zeller before him. He considered the various different phenomena such as depression, "exaltation" (i.e., mania), and dementia to be different symptomatic manifestations of a single underlying process. Griesinger seems to have switched to a nosological approach in his later years by adopting Snell's concept (1865) of separate disease entities.

2.1 Development of a Nosological Approach

The descriptions of "démence précoce" by Morel (1853), of "dementia praecox" by Kahlbaum (1863), and of "hebephrenia" by Kahlbaum's pupil Hecker (1871) may be considered as descriptions of groups of symptoms rather than diseases; Emil Kraepelin's approach, on the other hand, was clearly nosological. He created two inclusive disease entities, dementia praecox and manic-depressive insanity (Kraepelin 1896), and stated the characteristic symptoms and course of each, thereby becoming the major figure in psychiatry around the turn of the century. Kraepelin's theory found acceptance around the world, although not without opposition. Hoche (1912), for example, who propounded his own system of syndromes, compared the nosological classification of mental disorders to an attempt to cleanse muddy waters by "pouring them assiduously from one vessel into another." Kraepelin also created the triadic classification of mental disorders that is still used today, dividing them into mental disorders with physical causes, endogenous psychoses, and psychogenic disorders.

Eugen Bleuler took a position between those of Kraepelin and Hoche. He made use of Kraepelin's description of dementia praecox, while rejecting his pessimistic view of its prognosis, and instead used the new category of "schizophrenia," whose course he considered variable. The title of his most famous work (Bleuler 1911), which speaks of the "group of schizophrenias," indicates that he did not consider schizophrenia to be a single disease entity.

The scientific contributions of the nosologically inclined authors were incorporated into the Würzburg Diagnostic Scheme (*Würzburger Diagnosenschema*; Wilmanns 1930), which remained in general use in German-speaking countries until the 1960s. This classification scheme for both neurological and psychiatric disorders, however, failed to make adequate diagnostic distinctions in several areas, particularly among the psychogenic disorders.

2.2 National Diagnostic Classification Systems

While the classification systems of the nineteenth century generally arose in individual schools of psychiatry, many national classification systems were developed in the twentieth century. These largely made use of the diagnoses that had been propounded earlier in Central Europe. Numerous diagnostic classifications current in the early 1960s were described in 1961 by J.E. Meyer in an earlier edition of this textbook (Meyer 1961). By the mid-twentieth century, approximately 100 systems of classification of mental disorders were in use around the world (Wittchen 1994). Repeated suggestions were made for the improvement of these classifications, which were still considered unsatisfactory (Helmchen et al. 1966).

As it is impossible to give a full account of all of the historically developed and currently used classification systems, this discussion will be limited to a few examples illustrating their wide variety.

In French psychiatry, Magnan and Serieux (1893) developed a classification parallel to that of Kraepelin, which, unlike the latter, failed to achieve general acceptance (Pull et al. 1988) and was used only in French-speaking countries. This classification introduced a number of new categories of mental disorder, such as the various subforms of acute delusion ("bouffées délirantes") and of chronic delusional disorder ("délires chroniques"). The meaning of these terms, however, was only understood in France. The concept of schizophrenia was acknowledged despite some resistance, but was assigned a relatively limited scope. The concept of manic-depressive psychosis, however, found acceptance in France more easily, probably because of similar earlier descriptions by Falret (1854) and Baillarger (1854).

While no special diagnostic system was developed in England (Kendell 1990), it was Scandinavian psychiatry in particular that contributed not only the concept of psychogenic psychoses (Wimmer 1916), but also a number of early attempts toward multidimensional diagnostic assessment (Sjöbring 1919). Strömgren's repeated efforts (1988, 1989) led to international use of the concept of psychogenic psychoses.

Kleist and his pupil Leonhard (1986) proposed a classification of psychoses that provided definite prognoses on the basis of clinical phenomenology. This special, separate system was used in several places in Germany.

The Chinese Classification of Mental Disorders (CCMD) (Yu-cun and Changhui 1988) is an example of a system adapted to national needs. It is the product of an attempt to introduce modern diagnostic concepts while retaining certain traditional entities. The system is nosologically oriented and lists etiologic factors where possible. Neurasthenia is one of the subclassifications used in furtherance of tradition; this category is not present in DSM-IV, but appears again in ICD-10, certainly at least in part because of its use in China.

In the Meiji period, beginning in 1868, Japanese psychiatry turned away from traditional diagnostic procedures and took note of the latest developments in Europe. Kraepelin's system was adopted by the nation's leading university (Tokyo) and was largely adhered to until after the Second World War. In the postwar years, ICD-8 and ICD-9 were influential, followed by operationalized diagnosis, exemplified by the DSM-III. At present, the translated version of ICD-10 is in widespread use. Japanese centers participated in the international studies carried out by the World Health Organization (WHO) (Fujinawa 1994).

In 1989, in connection with the Japanese adoption of ICD-10, a special Japanese clinical modification was suggested (ICD-10 JCM) (Hanada 1994), which was based on the findings of a large national study.

In consultations on the construction of international classification systems, representatives of the so-called Third World have pointed out that many common disorders in Europe and North America, such as anorexia, anxiety disorders, sexual deviation, and borderline personality disorders, are rarely seen in their countries. The more common disorders there generally include acute psychotic episodes, hysterical manifestations, obsessions, and various somatization disorders (Wig 1990). Special transcultural syndromes, such as *koroh*, *latah*, and *dhat*, play a subordinate role, but should nevertheless be included in international classifications in appropriate places. The dualism concerning body and soul, characteristic of western psychiatry

until the present, is absent from many traditional ways of thinking. Thus the transition from ayurvedic medicine or from the medicine of traditional healers in Africa, for example, to modern, western medicine occurs across a conceptual gap that may be difficult to bridge. Nonetheless, ICD-10 fulfills many of the diagnostic needs of Third World psychiatry (Wig 1990).

3 Principles of Current Diagnostic Classification

The 1960s and 1970s were marked both by attacks on psychiatric diagnostic concepts, emanating from psychoanalysis and, above all, from antipsychiatry, and by a growing awareness of the need for unambiguous diagnostic criteria, particularly because of the increasing importance of psychopharmacologic research (Kendell 1975; Saß 1987).

In the United States, a reaction to the dominant psychoanalytic current of the 1950s and 1960s led to a reconsideration of Kraepelin's diagnostic system, termed neo-kraepelinianism (Klerman 1990). The Feighner Criteria (Feighner et al. 1972) and, later, the Research Diagnostic Criteria (RDC; Spitzer et al. 1975) were both developed for scientific purposes and served as forerunners of the more comprehensive modern systems of classification, starting with DSM-III.

The development of psychiatric classification from the diagnostic schemes of individual, prominent psychiatrists toward a consensus of many experts on the national and international level (Kendler 1990) culminated in the two major current systems: that of the World Health Organization (WHO), the *International Classification of Diseases* (ICD-10), and the American national psychiatric classification, the *Diagnostic and Statistical Manual* of the American Psychiatric Association (DSM-IV). The development has thus also been toward a reduction in the number of major systems in use. Nonetheless, geographic and ethnic variations remain, and will likely persist, in the frequency of certain diseases and disorders as well as in the tendency to make use of certain diagnostic categories (Sartorius et al. 1990).

3.1 International Cause-of-Death and Disease Statistics

A uniform international system of diagnosis was sought for general medicine long before the WHO tried to provide a classification of diagnoses. At the First International Statistical Congress in Brussels, in 1863, William Farr and Marc D'Espine called for the adoption of a common international nomenclature of causes of death (Kramer 1988). Their system, based on five major groups of categories, had originated in 1855 and underwent three revisions by the end of the century. A fundamentally new version was introduced in 1893 by Jacques Bertillon, then head of the Statistics Bureau of the city of Paris (Kendell 1975). This classification distinguished between generalized and localized diseases. It was revised every 10 years thereafter by a commission comprising representatives of the International Institute of Statistics and of the Health Organization of the League of Nations. A listing of diseases, consisting of a combination of previous national and international listings, was added to the cause-of-death classification in 1938. These existing international listings were adopted after the Second World War by the WHO and were incorporated into the sixth edition of the *International Statistical Classification of Diseases, Injuries and Causes of Death* (WHO 1948). The psychiatric part of this classification was accepted in only a few countries, even though the increasingly frequent contact among psychiatrists from different countries made the need for international agreement increasingly apparent.

3.2 ICD-8 and ICD-9 and the Development of DSM-III and DSM-IV

The eighth edition of ICD-10, used from 1969 onward in numerous countries, including West Germany, was the first to contain a glossary (i.e., a collection of brief definitions) and made progress toward more exact descriptions of disease (WHO 1967; Degkwitz et al. 1971). The classification systems of this edition and the following ninth edition, which is currently still valid (WHO 1978; Degkwitz et al. 1980), were recognized as official by the Federal Government of West Germany, thus enabling the collection of nationwide statistics, and were introduced into clinical practice by the Bundesarbeitsgemeinschaft der Träger psychiatrischer Krankenhäuser (Federal Psychiatric Hospital Association). Many comparative studies were thus made possible on both the national and the international levels.

The official American diagnostic classification, DSM-II (APA 1968), was still largely in accord with ICD-8. However, both ICD-8 and ICD-9 had major weaknesses, including lack of multiaxiality, nonuniform classification of depression, and problems relating to the diagnosis of sexual disorders. The American psychiatrists therefore chose to pursue a different route than their WHO colleagues with the third edition (and, later, the revised third edition) of the *Diagnostic and Statistical Manual of Mental Disorders* (APA 1980,

1987). The most important difference was the so-called operationalization of diagnosis, with the aid of clearly defined criteria; this stood in contrast to ICD-9, which had only a glossary containing brief descriptions. DSM-III was widely accepted in practical use, and even more widely in research, within a few years. Many scientific publications now require the description of patient populations by the criteria of DSM-III, DSM-III-R, and the most current version at the time of writing, DSM-IV (APA 1994). This development put the WHO under considerable pressure to conform. While the WHO was still obliged by multiple international agreements to construct and propagate its own classification system (Strömgren 1994), it could not deviate much from the model of DSM-III, which had already been adopted into common use.

3.3 Operational Diagnosis

Operational diagnosis is based on the phenomenological characterization of disorders without regard to their etiology. The typical diagnostic features of each disorder are listed, including the intensity and exact duration of the symptoms. A diagnostic algorithm including required and optional criteria, such as psychopathological abnormalities, is used to make a diagnosis. In addition, inclusion and exclusion criteria must be fulfilled.

Operational diagnosis can be contrasted with typological diagnosis (Dilling 1994), which relies mainly on typical case descriptions, making only secondary use of characteristic cross-sectional symptoms and symptom course. A diagnosis is made based on maximal similarity to a typical case, the so-called prototype. The diagnostician thus has considerable freedom in determining which cases are sufficiently comparable to his or her internalized prototype case to merit being given the corresponding diagnosis.

Typological and operational diagnosis differ primarily with respect to the diagnostic process. Even operational diagnosis has prototypes; these are the ideal cases that fulfill all of the diagnostic criteria for a given disorder. The number of criteria fulfilled can be considered an index of the nearness of a given constellation of symptoms to the prototype (Maier and Philipp 1988). Operational diagnosis is not concerned with the etiology of disorders and thus does not obey the classical principle of nosology, according to which an etiology and a constellation of symptoms together constitute a disease entity. Because of the relative lack of attention to etiology in operational diagnosis, the diagnostic categories of ICD-10 are deliberately referred to as disorders, not diseases.

3.4 Preparatory Steps Taken by the World Health Organization Toward the Introduction of ICD-10

In 1982, the WHO and the United States Alcohol and Drug Abuse and Mental Health Administration (ADAMHA) held a conference in Copenhagen to address the diagnosis and classification of mental disorders and alcohol and drug problems. This conference was based on extensive preparatory work (WHO 1981; Research Report 1983; Table 1). In 1984, the WHO organized expert sessions in Jakarta and in Geneva. These meetings led to a preliminary classification scheme for the psychiatric chapter of ICD-10. The individual sections were produced by various groups of mostly English-speaking experts and were provided with clinical descriptions and diagnostic guidelines. The first English version was submitted to the World Psychiatric Association (WPA) (Mezzich and von Cranach 1988), the national psychatric societies, and individual experts for comment and critical review (Dilling and Dittmann 1990).

The WHO received numerous suggestions for improvement, some of which were incorporated into further drafts. The so-called *Blue Book* underwent a series of revisions and was finally published in 1992 as *Clinical Descriptions and Diagnostic Guidelines*. In parallel with the work of the WHO, translations were prepared in Spanish, French, Arabic, German, Chinese, and many other languages; in 1997, the text was available in 22 languages and in preparation in a further nine. The German edition appeared in 1991 (Dilling et al. 1991). Ten so-called WHO Reference and Training Centres in Classification, Diagnosis and Assessment of Mental and Behavioural Disorders were established worldwide to develop the psychiatric section of ICD-10 by means of translations, field studies initiated by the WHO, and their own investigations (Sartorius 1995). The center responsible for the German-speaking countries is located in the Department of Psychiatry of the Medical University of Lübeck, Germany (Dilling et al. 1994b).

4 International Classification of Diseases (ICD-10)

4.1 Structure

Psychiatric classification is part the general medical classification of ICD-10 (WHO 1992a; DIMDI 1995/1996). Many more diagnoses are included in ICD-10

Table 1. Timetable of the development of psychiatric diagnosis from ICD-8 to ICD-10

Year	Classification system	Comments
1967	ICD-8	
1978	ICD-9	
1980	DSM-III	
1982		WHO/ADAMHA conference, Copenhagen
1984	ICD-10 Chap. V (F): first preliminary draft of classification	WHO conferences in Jakarta and Geneva
1986	ICD-10 Chap. V (F): Draft (0)	WPA consultation
1987	ICD-10 Chap. V (F): Draft I	Field study of the Clinical Descriptions and Diagnostic Guidelines
	DSM-III-R	
1989	ICD-10 Chap. V (F): Draft III	Beginning of the study on the Research Criteria
		Revision conference, Geneva
1990	ICD-10 Chap. V (F): Draft IV	
1991/92	Publication of the ICD-10 Guidelines	
1993	Publication of the ICD-10 Research Criteria	ICD-10 becomes the official WHO classification
1994	DSM-IV	

ADAMHA, Alcohol and Drug Abuse and Mental Health Administration; WHO, World Health Organization; WPA, World Psychiatric Association.

than in ICD-9 because of the need for refined diagnosis in many specialized areas. An alphanumeric system is used in which a capital letter designates one of a total of 21 different chapters. Psychiatric diseases are designated hierarchically by the letter F (Chap. V), followed by a number for the major diagnostic group, a decimal point, and a further number for the subordinate diagnostic group. A designation may include three or four digits. There are thus 1000 possible diagnoses for psychiatric disorders (F00.0–F99.9), of which approximately one third are now in use. This system allows the addition of new diagnoses in the future without the need for major changes in the classification. In contrast, both ICD-8 and ICD-9 had room for only 300 possible psychiatric diagnoses.

A fifth or sixth digit may be added to provide more precise documentation of the course of illness or other characteristics. Certain nonspecific symptoms or special circumstances, such as suicide, self-injury, or certain psychosocial conditions, are classified in other chapters (Chaps. X, Y, and Z). Somatic comorbidities are classified in the corresponding chapters, e.g., respiratory diseases in Chap. J (X). Psychiatric disorders occupy a special position in ICD-10, because they are the only disorders for which definitions and inclusion and exclusion criteria are provided; in all other areas, the diagnoses are merely named, without further explanation.

4.2 Organization of Chapter V (F)

The WHO intended Chap. F of ICD-10 to serve not just as a classification scheme for statistical purposes, but also, depending on its users, as a clinical manual, a textbook of diagnosis, and a research tool. A "family of documents" for different uses thus came into being:

- Various versions of the classification
  - Brief glossary
  - Clinical diagnostic guidelines
  - Research criteria
  - Multiaxial system (MAS)
  - Classification of mental disorders in primary health care (PHC) in child and adolescent psychiatry
  - Classification of mental disorders in PHC in adult psychiatry
  - Fascicles for special areas (e.g., mental retardation, neuroses, personality disorders)
- Cross-reference tables
- Teaching and learning materials (lexica, multilingual dictionaries of ICD-10 terminology, specialized texts, teaching overhead transparencies, adaptations, computer versions)
- Instruments for diagnostic assessment (Schedule for Assessment in Neuropsychiatry, SCAN; Composite

International Diagnostic Interview, CIDI; International Personality Disorder Examination, IPDE; Disability Assessment Schedule, DAS; International Symptom Checklist, ISCL; International Diagnostic Checklists, IDCL)

The concise version of ICD-10, Chap. V (F) is part of the system of ICD-10 diagnoses (WHO 1992a; DIMDI 1995/1996). The appended glossary provides brief explanatory descriptions of each disorder, which make psychiatric terminology accessible to many people in the health sector who may lack familiarity with it, such as statisticians, administrators, and insurance personnel.

The central portion of this work consists of the *Clinical Descriptions and Diagnostic Guidelines*, or *Blue Book*, whose approximately 350 pages contain extensive descriptions of all diagnoses and their clinical operationalization (WHO 1992b; Dilling et al. 1991, 1993).

Although the clinical descriptions are meant to be definitions of diagnostic entities with a discussion of their most important characteristics, and the diagnostic guidelines are meant to be listings of features in connection with temporal criteria and rules for inclusion and exclusion, these two areas inevitably overlap. This version of Chap. V (F) is meant primarily for psychiatrists providing care in hospital wards and in the outpatient setting.

The *Diagnostic Criteria for Research* (DCR; the *Green Book*) are more rigorous than the guidelines, in accordance with the needs of research studies (WHO 1993a; Dilling et al. 1994a). The inclusion and exclusion criteria are more stringent, and the temporal limits are more precisely defined. This greater stringency is achieved at the cost of the loss of a fairly large group of cases to which definite diagnoses can no longer be assigned. This perspicuous book of some 150 pages, containing no descriptive definitions of the disorders, is not only of use to researchers, but also practical for psychiatrists in both hospital and outpatient settings.

A special version of ICD-10, the Primary Health Care (PHC) version, was prepared for primary care providers in view of the high incidence of mental disorders among the general population, e.g., the highly prevalent depressive and dependency disorders (WHO 1996; Müßigbrodt et al. 1996). A total of 24 syndromes were chosen for inclusion in the English version. Their most important symptoms are described, as are diagnostic criteria, differential diagnoses, treatment, methods of dealing with patients and their families, and criteria for referral to specialists.

There is a German version is particularly aimed at the general practitioner (Müßigbrodt et al. 1996). It goes significantly beyond the requirements of the WHO concerning primary health care (e.g., in the Third World). Help is provided for the encoding of three- and four-digit diagnoses, as this will soon be required by the German health insurance system.

4.3 Multiaxial System

An overview of the WHO multiaxial system (MAS) is shown below:

I Clinical diagnoses
  Ia Mental disorders according to Chap. V (F)
  Ib Underlying or concurrent physical illnesses, as covered in the somatic chapters of ICD-10
II Assessment of social functional disability attributable to either mental or physical disorders (coding on a five-point scale)

  Specific functional areas:

  IIa Self-care and coping with everyday life (bodily hygiene, clothing, nutrition, etc.)
  IIb Ability to function in an occupation (gainful employment, study, housework, etc.)
  IIc Family and household (interaction with spouse or other partner, parents, children, and other relatives)
  IId Ability to function in other social roles and activities (relationship to members of the community, participation in leisure and social activities)
  IIe Global functional disability (overall disability)
III Environmental factors (selected Z codes)
  1. Negative life experiences in childhood
  2. Upbringing and education
  3. Primary relation group, including family
  4. Social environment
  5. Living accommodations and financial situation
  6. Occupational activity and unemployment
  7. Environmental stresses
  8. Psychosocial or legal problems
  9. Diseases or disabilities in the family
  10. Conduct of life, coping

As far back as the first half of the twentieth century, attempts were made to enlarge the excessively narrow scope of the diagnostic process by introducing a multiplicity of variables. The multidimensional diagnostic system of Kretschmer (1919) and, later, the multiaxial system of Essen-Möller (1973) were of this type. The multiaxial diagnostic system of Rutter et al. (1975) has been used in child and adolescent psychiatry for many years; it provides classification according to clinical syndrome, possible developmental delay, intelligence level, somatic

illnesses, and psychosocial stresses (Remschmidt and Schmidt 1996). Numerous proposals have been made for the introduction of multidimensional criteria into adolescent psychiatry (Mezzich 1988). A number of these were selected for inclusion in the multiaxial system of ICD-10. Unfortunately, there was a large interval between the completion of Chap. V (F) and the development of the MAS, with the result that the additional diagnostic axes have only seldom been used so far. DSM-III and DSM-IV share this problem with ICD-10. DSM makes frequent use of axis V, or global assessment of functional level (GAF), but relatively little of axis IV, psychosocial and environmental problems, which corresponds to axis III of ICD-10. Axes II and III of DSM-IV are included under axis I of ICD-10 under comorbidities. Thus personality disorders and somatic illnesses are not assigned to separate axes.

To simplify the assessment of axis II (disability), the WHO has developed a brief instrument, the Short Disability Assessment Schedule (WHO DAS-S). This may help in the assessment of the sequelae of axis I disorders (Janca et al. 1996).

In connection with the axis of social functional disability, the International Classification of Impairments, Disabilities and Handicaps (ICIDH) deserves to be mentioned (WHO 1980; Matthesius 1995). This classification was created by the WHO almost 20 years ago for the entire area of rehabilitation medicine, far beyond the borders of psychiatry. It is currently being revised and restructured by the WHO, with the participation of many centers (Dilling and Siebel 1995).

4.4 Further Texts on Chapter V (F)

Alongside these versions of and additions to ICD-10, a number of ancillary diagnostic works and specially covered areas deserve to be mentioned.

The successive replacement of ICD-8 by ICD-9 and ICD-10 created a partial lack of continuity of diagnostic categories, a problem which had to be overcome to enable various kinds of clinical and research work to be continued. Although many disorders are still designated by the same or similar names, the introduction of operational diagnosis meant a fundamental change in the system. A basis for statistical comparisons is fairly easy to create for disorders such as catatonic schizophrenia or obsessive-compulsive neurosis, but can be very difficult in other cases, e.g., the reconsideration of "neurotic depression" under "dysthymia" or the comparison of older diagnoses with newly proposed entities such as panic disorder and somatization disorder. Thus the cross-reference tables of the WHO (Freyberger et al. 1993a,b) should not be taken as a mechanical system of translation from one system to the other, because the recasting of diagnoses can only be handled on a case-by-case basis. These tables are to be used as an aid in this process (see also Chaps. 3, 4, this volume).

The WHO has also produced lexica of the psychopathological and other terminology of ICD-9 (WHO 1985), the terminology of ICD-10 (WHO 1994a), and terms relating to alcoholism and drug dependency (WHO 1994b). A further lexicon of transcultural terms is in preparation (Sartorius 1995).

Casebooks have been published in English and in German to give the users of ICD-10 a basis for clinical comparisons and to provide case material for introductory courses. The 100 cases contained in the English casebook are systematically organized to cover the entire field of Chap. V (F) (Üstün et al. 1996). The approximately 50 cases contained in the German text (Freyberger and Dilling 1993) follow no particular organizational principle and reflect the individual contributions of their authors, whose names are signed to the cases.

The WHO has issued numerous overhead transparencies for use in introductory courses, training seminars, and continuing education. These materials highlight various aspects of ICD-10, and some of them have been translated into other languages. A comparable collection of German-language transparencies on the research criteria has been prepared (Dilling et al. 1999).

The psychiatric and neurological adaptations of ICD-10 may be used as aids to coding. The neurological adaptation (Kessler and Freyberger 1996) contains all diagnostic codes of importance to neurologists, i.e., not only neurologic diagnoses, but also a number of psychiatric diagnoses and diagnoses in a series of special areas. The projected psychiatric adaptation (ICD-10 PA) will include diagnoses from the fields of psychiatry, neurology, and internal medicine and from Chaps. X, Y, and Z.

The parallel development of operational diagnostic systems and of computer technology has enabled a collaboration between the two areas (Andrews et al. 1994). A German-language ICD-10 tutorial is available in hypertext form and may be used for many purposes, particularly didactic ones (Malchow et al. 1995). Further texts and links to various texts will appear in the near future.

4.5 Operationalized Psychodynamic Diagnosis

In the Federal Republic of Germany, a working group of specialists in psychosomatic medicine, with the collaboration of several psychiatrists, developed a

multiaxial operational system for psychodynamic diagnosis (OPD) (Arbeitskreis OPD 1996). In addition to the ICD-10 diagnoses, the five axes of "illness experience," "treatment expectation," "relationship," "conflict," and "structure" are considered. This system is potentially very useful in the determination of indications for psychotherapy and in the assessment of the clinical course. An English translation is being prepared.

4.6 Diagnostic Instruments

Diagnostic instruments with an operational orientation have been developed to aid in the diagnosis of mental disorders according to either the ICD-10 or the DSM system.

A group of experts from 14 countries, led by J.K. Wing, developed the Schedule for Assessment in Neuropsychiatry (SCAN), a structured interview for clinical use (WHO 1994d). SCAN is comparable to the tenth edition of the Present State Examination (PSE) (Wing et al. 1974). It consists of a semistandardized interview, a glossary with symptom definitions, an Item Group Checklist, and a history sheet and takes 60–90 min to perform. The center for SCAN-based diagnosis in Germany is the Zentralinstitut für Seelische Gesundheit (Central Institute of Mental Health) in Mannheim.

The Composite International Diagnostic Interview (CIDI) was developed on the basis of an earlier interview, the Diagnostic Interview Schedule (DIS), and is now available in several languages (WHO 1993b; Wittchen and Semler 1991). CIDI can be performed by lay interviewers after a relatively short training period of a few days. Unlike SCAN, which relies on the judgment of the interviewer, the CIDI is a fully standardized interview in which the responses of the interviewee constitute the entirety of the data. The interview takes approximately 1 h. The center for CIDI-based diagnosis in Germany is the Max Planck Institute for Psychiatry in Munich.

The International Personality Disorder Examination (IPDE) was developed by the WHO for the diagnosis of personality disorders (Loranger 1996; Loranger et al. 1994). It is intended principally for use by experienced clinicians.

An International Symptom Checklist (ISCL) was developed and tested by the WHO (WHO 1994c; Janca et al. 1993) for the assessment of the severity, course, and duration of symptoms. This instrument has also been published in German under the title *ICD-10-Checklisten* (Hiller et al. 1995). The Symptom Checklist (SCL) is a screening procedure; after it is used, more precise ICD-10 diagnoses can be given with the aid of 32 International Diagnostic Checklists (IDCL). A further checklist has been developed especially for personality disorders (Bronisch et al. 1995).

Meanwhile, a separate German-language working group developed an ICD-10 checklist corresponding to the Diagnostic Guidelines; this checklist contains 750 criteria divided among 14 feature groups and was validated in eight centers by use in 858 patients. The validating study investigated the frequency, sensitivity, specificity, and predictive value of various features in ICD-10 disorders as well as the suitability of the diagnoses (Dittmann et al. 1992).

A further diagnostic instrument is the Structured Interview for the Diagnosis of Dementia in Alzheimer's Disease, Vascular Dementia, and Dementias in Other Diseases According to DSM-III-R, DSM-IV and ICD-10 (SIDAM) (Zaudig and Hiller 1996).

4.7 Structure and Subdivisions of Chapter V (F)

At first glance, the structure of Chap. V appears to follow the familiar sequence of ICD-8 and ICD-9. The classification begins with mental disorders of organic origin, proceeds through disorders due to the use of psychoactive substances, schizophrenic psychotic disorders, affective disorders, psychogenic disorders, and personality disorders, and concludes with mental retardation and disorders from the realm of child and adolescent psychiatry. However, a closer look reveals that the traditional division of diagnoses into the two major areas of psychoses (ICD-9: 290–299) and neuroses (ICD-9: 300–310) has been abandoned. Diagnostic terms are now used predominantly in a phenomenological, descriptive sense. According to the authors of ICD-10, a single mental disorder may have both psychotic and nonpsychotic manifestations, if we take "psychosis" to mean the appearance of productive symptoms. It was originally planned to leave the term "neurosis" out of ICD-10 entirely, as it had been used in too many different contexts and is based on theories of intrapsychic causation that are no longer generally accepted. The later decision to retain the term "neurotic disorders" in ICD-10 was a concession to traditional usage and does not imply any assumption concerning etiology.

The individual sections of the chapter are subdivided hierarchically, i.e., the descriptions of major diagnostic terms apply also to the three- and four-digit diagnostic categories that follow. As in the DSM, mental illnesses or diseases are no longer spoken of, and the term "disorder" is used throughout, with the exception of a few diseases in F0.

4.8 Particular Aspects of Individual Sections

4.8.1 F0 – Organic, Including Symptomatic, Mental Disorders

F0 includes all disorders of demonstrably organic etiology, regardless of whether they have psychotic or nonpsychotic symptoms. The scope of the term "dementia" has been considerably widened in accordance with the American usage of DSM, although a minimal duration of symptoms of 6 months is required. The diagnosis "delirium" is also less specific than before and no longer requires the presence of perceptual abnormalities.

4.8.2 F1 – Disorders Due to Psychoactive Substance Use

The consideration of mental and behavioral disorders due to the use of psychoactive substances in F1 is an improvement over their previous treatment in two separate places in the section in ICD-9. In F1, the third digit of the diagnosis code designates the causative substance or substance class, while the fourth and (sometimes) fifth digits designate the type of disorder. Distinctions are drawn between acute intoxication, harmful use, dependency syndrome, withdrawal syndrome, psychotic conditions, alcohol- and drug-induced amnesic syndromes, and alcohol- and drug-induced residual conditions. Thus each code designates not only the psychopathological syndrome, but also the responsible substance class. On the other hand, the simultaneous diagnosis of dependency and acute intoxication requires the assignment of two diagnostic codes.

4.8.3 F2 – Schizophrenia, Schizotypal and Delusional Disorders

The schizophrenias, acute psychotic disorders, and schizoaffective, schizotypal, and delusional disorders appear in F2. New entities include undifferentiated schizophrenia, post-schizophrenic depression, and schizotypal disorder. The presence of typical symptoms for at least 1 month is a prerequisite for the diagnosis of schizophrenia. This is in contrast to DSM-IV, which requires the presence of symptoms in the 6 months preceding the time of diagnosis, with at least 1 month of florid psychotic manifestations. Schizoaffective disorders and acute psychoses of brief duration arising within a period of 2 weeks, or even abruptly within 2 days, are presented in particular detail. This emphasis on the diagnosis of acute psychoses is a reflection of the concerns of psychiatrists from the Third World, where acute psychoses of brief duration are quite common and usually carry a good prognosis.

4.8.4 F3 – Mood (Affective) Disorders

The presentation of the affective disorders in F3 has undergone particularly extensive changes. The distinction between endogenous and neurotic depressions has been abandoned. The previously quite common category of "neurotic depression" (ICD-9: 300.4) no longer appears in the system, and its role has been taken over primarily by dysthymia, which, however, is said to be present only if the symptoms have persisted for at least 2 years. Regretfully, manic disorder with recurrent episodes is now absent, while single manic episodes still receive a diagnostic code. Particular emphasis is laid on the determination of the severity of depressive disorders; these are classified as mild, moderate, or severe. The overall subdivision and the introduction of a grading of severity are both analogous to features of DSM-IV, although some nuances of the latter are missing, such as the concept of major depression, which corresponds to the moderate and severe depressive episodes of ICD-10.

4.8.5 F4 – Neurotic, Stress-Related, and Somatoform Disorders

As mentioned above, the occurrence of the term "neurosis" in this section is merely a conventional usage, rather than an indication of acceptance of earlier etiologic theories of neurosis. The individual disorders are highly subdivided. For example, there are seven diagnostic subcategories of dissociative disorders, some of which are quite rare. On the other hand, the term "hysteria" has been definitively abandoned. Stress-induced disorders are classified according to their duration and severity as acute stress reaction, post-traumatic stress disorder, and adjustment disorder. The somatoform disorders appear for the first time in this classification, including somatization disorder, hypochondriacal and somatoform autonomic disorder, and, finally, the somatoform pain disorders. The diagnosis of neurasthenia, closely related to the somatoform disorders, is used with variable frequency.

4.8.6 F5 – Behavioral Syndromes Associated with Physiological Disturbances and Physical Factors

These elaborate enumerations occupy considerable space, with extensive and detailed descriptions. The sexual disorders, which were considered together in ICD-9, are found in two separate sections in ICD-10;

some are listed under physiological disturbances (F52), while others are listed in F6 under the heading "disorders of personality and behaviour," including disorders of gender identity (F64) and sexual preference (F65) as well as disorders of sexual development and orientation (F66). This revealing separation results in a better integration of the various sexual disorders into the context of the larger categories of disorders to which they belong.

4.8.7 F6 – Disorders of Adult Personality and Behavior

Cyclothymic personality disorder is conspicuously absent from the list of personality disorders in F6, appearing instead as "cyclothymia" in the chapter on affective disorders. Schizotypal disorder, too, might have been chosen for inclusion in F6, but was not. As in DSM-IV, the listing of personality disorders includes emotional instability of the borderline type, a diagnosis given very often (too often) in clinical practice. Artificial disorders, i.e., disorders generated or simulated by patients without any discernible purpose, appear here as a new entity. Accentuated personality traits are not included in the classification, although this was originally planned. The absence of narcissistic personality disorder is frequently decried, although this, like passive-aggressive personality disorder, can be found in the appendix to the research criteria.

The inclusion in this chapter of lasting personality changes after extreme stress, such as torture or incarceration in a concentration camp, is of great importance. German medicolegal practitioners often doubted or denied the existence of this category in the years, and even decades, after the end of the Second World War.

4.8.8 F7 – Mental Retardation

As in ICD-9, the section on mental retardation (F7) distinguishes between mild, moderate, severe, and extremely severe forms.

4.8.9 F8 – Disorders of Psychological Development/F9 – Behavioral and Emotional Disorders with Onset Usually Occurring in Childhood and Adolescence

The developmental disorders and the behavioral and emotional disorders with onset usually occurring in childhood and adolescence (F9) are treated much more extensively than in ICD-9. If possible, all disorders beginning in childhood and adolescence are to be given the diagnostic codes of the preceding sections (F0–F6); F8 and F9 contain only those special disorders that either do not occur in adults or typically begin in childhood.

5 Empirical Research Relating to ICD-10

The initial version of ICD-10, Chap. V, was based predominantly on diagnostic concepts emerging out of clinical practice, which then needed to be tested for suitability, reliability, and above all validity. Studies for these purposes were initiated primarily by the WHO and described in numerous publications. In 1987, the WHO initiated an international field study concerning the diagnostic guidelines, in which 112 institutions from 39 countries participated (Sartorius et al. 1993). A total of 711 clinicians performed 15,302 evaluations. In the German-speaking countries (the Federal Republic of Germany, Austria, and Switzerland), 134 diagnosticians from ten institutions for adult psychiatry and seven diagnosticians from two institutions for child and adolescent psychiatry participated (Dilling et al. 1990), and 2058 evaluations of joint case conferences and written cases were performed. The interrater reliability was highly variable for different diagnoses, reaching an overall value of $\kappa = 0.51$ for all cases. The Department of Psychiatry of the Medical University of Lübeck was responsible for coordinating the study in the German-speaking countries. In the wake of this study, a number of changes in ICD-10, Chap. V, were proposed to the WHO, some of which were incorporated into the final version.

The research criteria were the subject of a further study initiated by the WHO. A total of 151 centers in 32 countries participated (Sartorius et al. 1995), and 942 interviewers performed 11,491 evaluations. In most sections, the interrater reliability of the research criteria was found to be significantly higher than that of the clinical guidelines; in all sections other than F6 (personality disorders), the interrater reliability lay between $\kappa = 0.8$ and 1.0.

In the German-speaking countries, 451 raters from 34 centers participated in this study (Freyberger et al. 1996). A total of 2228 evaluations of 39 videotaped cases were performed. Here, too, there was a high interrater reliability ($\kappa = 0.86$) for three-digit diagnoses, although the value was lower than in the international study. There may be several reasons for this, e.g., the variable independence of raters who may evaluate patients, or videotapes, together as a working group. The suitability of the diagnoses and the degree of difficulty of assigning them were judged by the raters as favorable.

A multicenter study was carried out in the German-speaking countries to assess the symptoms and criteria included in the checklists underlying ICD-10 diagnosis (Dittmann et al. 1992). The practicality, quality, and validity of the clinical diagnostic guidelines were investigated.

The WHO further initiated a study of multiaxiality, in which ten cases were presented to numerous centers around the world for evaluation. Various studies of interrater reliability and validity were also carried out, including a study in the German-speaking countries involving the evaluation of typical cases on videotape (Michels et al. 1996).

A study was performed at the Zentralinstitut für seelische Gesundheit in Mannheim, in collaboration with the Departments of Psychiatry of the Universities of Munich and Mainz, to assess the reliability of ICD-10 diagnoses by means of video recordings. At the Department of Psychiatry of the University of Mainz, the DSM-III and ICD-10 classifications of psychotic disorders were compared in a study of clinical course (Maier and Philipp 1990).

The PHC version was the subject of a comprehensive international study. In the German-speaking portion of this study, the diagnostic categories were tested among groups of general medical practitioners, social workers, nurses, medical students, etc., for practicability and reliability (Kleinschmidt et al. 1995).

6 Diagnostic and Statistical Manual of Mental Disorders (DSM-III and DSM-IV)

The *Diagnostic and Statistical Manual of Mental Disorders* was developed in North America because of the perceived need for operational diagnosis, which was not yet a feature of the WHO system. DSM-III appeared in 1980 (APA 1980); its revision, DSM-III-R, in 1987 (APA 1987,Wittchen et al. 1989); and DSM-IV in 1994 (APA 1994; Saß et al. 1996).

With the introduction of DSM-III came the application of explicit, operationally defined diagnostic criteria, a multiaxial descriptive system and a descriptive approach intended to preserve maximum neutrality with regard to etiologic assumptions. In the 1980s, the two versions of DSM were the subject of intense scientific study, many corrections were suggested, and, finally, a systematic, consensus-based revision was carried out, taking account of the world literature. The WHO was also involved in this process; representatives of the APA and the WHO held joint meetings before the completion of DSM-IV. Agreement was reached in some areas, while differences remained in others.

7 Comparison of ICD-10 and DSM-IV

The overall difference between ICD-10 and DSM-IV is considerably less than that between ICD-9 and ICD-10. American diagnosticians were involved in the development of ICD-10 from its inception and exerted a strong influence on its content. Nevertheless, despite the convergence of these two major systems, a thorough congruence has not yet been achieved. The development of DSM-IV is discussed in great detail in the introduction, as well as the appendix, of its German edition. The German DSM-IV editors consider ICD-10 and DSM-IV to represent different dialects of the same language (Saß et al. 1996). Although they do not minimize the areas of divergence, these editors approvingly cite Thangavelu and Martin (1995) as follows: "The course of development of these two systems, and their emphasis on the operationalization of diagnoses in research and in inpatient and ambulatory care, has brought the world's psychiatrists closer together than ever before" (cited from Saß et al. 1996).

The structure of DSM-IV corresponds in broad outline to that of ICD-10, except that disorders occurring in infancy, childhood, and adolescence are listed before, rather than after, the organic mental disorders; these are listed under the section heading "delirium, dementia, amnestic and other cognitive disorders." This renaming was a consequence of the consideration that many other mental disorders also have bodily causes, so that restricting the use of the term "organic" to these disorders alone appears unjustified.

There are further differences in the diagnosis of individual disorders. In DSM-IV, as already mentioned, symptoms of schizophrenia are required to have been observable for at least 6 months before this disorder can be diagnosed, while ICD-10 requires only 1 month. ICD-10 includes conversion disorders among the dissociative disorders, while DSM-IV considers them separately. The eating disorders are defined differently in the two classifications. Finally, unlike ICD-10, DSM-IV exists in only one version, which makes its application less ambiguous, because no conflict can arise between the criteria of two different versions, such as ICD-10 Guidelines and Research Criteria. In many cases, however, and for general practitioners in particular, the absence of graded sets of criteria makes DSM-IV more difficult to apply.

The difference with respect to multiaxiality has already been mentioned; ICD-10 puts all medical diagnoses together on a single axis. In contrast, the multiaxial classification of DSM-IV contains five axes:

axis I, mental disorders; axis II, personality disorders and mental retardation; axis III, medical disease factors, i.e., somatic diagnoses; axis IV, psychosocial and environmental problems; and axis V, global assessment of functional level (GAF). Combining axis I and axis II results in a fairly close duplication of the multiaxial system of ICD-10.

The international and German-language editions of DSM-IV are oriented to the ICD-10 codes. In the United States, the existing coding system cannot be replaced until approximately the year 2000 because of technical and logistical considerations; at that time, ICD-10 can be introduced in its entirety, i.e., for all medical disciplines, including psychiatry. The differences between the American and other coding systems are thus limited at present to Chap. V (F). DSM-IV is much more comprehensive than Chap. V (F), in that it includes discussions of related findings such as laboratory tests as well as data on prevalence, clinical course, familial incidence, and differential diagnosis. DSM-IV thus takes on the character of a textbook rather than that of a diagnostic manual (although it contains no discussion of treatment).

8 Final Observations

8.1 Advantages of International Classification

What, then, are the advantages of international psychiatric classification? Uniform diagnosis allows the international standardization of treatment, as has been shown in recent years by a number of studies performed around the world. Not only are multicenter psychopharmacologic studies required, but also evaluative studies of psychotherapy and studies of the effectiveness of interventions involving social therapy. The use of standardized and structured interviews that enable diagnosis according to ICD-10 or DSM-IV has become well established; scientific studies will have to be based henceforth on this kind of diagnostic technique.

Cross-cultural comparative psychiatry is aided by a uniform classification that enables epidemiologic studies to be performed with identical instruments in different countries and different cultures, which can then provide data for valid clinical and epidemiological comparisons. One advantage of ICD-10 over its regional American counterpart, DSM-IV, is its more thorough consideration of European needs and, above all, the needs of the Third World. International and interregional comparisons of diagnosis-based administrative data would be of the utmost interest.

8.2 Risks and Disadvantages

The international operational diagnostic classification has risks and disadvantages in addition to its obvious advantages (Jablensky 1991). Diagnosticians must always remain aware that the psychiatric disorders listed in ICD-10 are conventional and not scientifically well founded disease entities. Further changes will undoubtedly be made in the coming decades, and a certain flexibility is required in the definitions of these disorders. Despite a high interrater reliability, most of the disorders in ICD-10 have been insufficiently tested for validity, and many of them may be redefined in the future on the basis of different characteristics than at present.

In the field of operational diagnosis, there is a strong compulsion to define diagnostic categories and to furnish the criteria that distinguish them, so that the individual disorder entities may be separated into groups according to their common syndrome characteristics. Both of the current diagnostic classifications are categorical in this sense and have little hint of dimensionality, except in a few peripheral areas, such as the personality disorders. The somatic medical categorical systems are based on etiology, but this is not possible for the majority of mental disorders, so that we are forced to work with conventional definitions for many disorders, i.e., with constructs whose long-term stability is uncertain. Clearly, while these constructs are acceptable in terms of their interrater reliability, stronger evidence is needed for their validity.

The traditional diversity of clinical schools and local diagnostic customs cannot be maintained into the future. Even if the WHO encourages the preservation of distinct psychiatric traditions, the worldwide standardization of diagnostic classification will make this impossible in the long run, much as local ethnic traditions and dialects inexorably die out as the world's cultures come closer together. Nonetheless, one may hope that sensitivity and openness toward local traditions in psychiatry will not be entirely lost, so that these can continue to serve as points of departure for new creative impulses.

The "language-conditioning effect" of any diagnostic system must also be borne in mind. In the current diagnostic systems, formulations have been sought that will no longer offend or insult the affected persons; thus the "asocial psychopath" of yesteryear now suffers from a "dissocial personality disorder," and "sexual perversions" are now referred to as "disorders of sexual preference." A danger exists that the presence of certain conditions, such as pyromania and kleptomania, in ICD-10 will automatically convert

all individuals with these characteristics into psychiatric patients. This would amount to a dangerous reification of what was intended to be merely a construct in ICD-10, applying only to a restricted group of people. Other conditions have been removed from the classification, such as homosexuality, on the ground that homosexuality ought to have the same social and juridical status as heterosexuality; a certain "political correctness effect" can be discerned here. Nevertheless, if psychiatric symptoms are connected to homosexuality as well as to heterosexuality, this can be coded.

A further danger of diagnostic classifications laid down along operational lines is that newer and perhaps better systems will not be allowed to develop, because scientific publications will require the use of the officially recognized systems. The same applies to research grants, scholarships, and so on.

The greatest danger is one inherent in any codified classification – that of unrestricted application by the social security and health insurance systems. Control mechanisms may be devised that operate to the patient's disadvantage; in this situation, the treating psychiatrist has a duty to prevent negative consequences to the patient. For example, the assigned diagnostic code may be used by health insurance companies to preset the allowed duration of the patient's stay in the hospital. Treating clinicians must firmly oppose this procedure, which fails to take account of the patient's individual circumstances, including the course of the disease in the individual case. In both the economic and the forensic spheres of the health care system, the danger exists that our diagnostic conventions, which may well represent erroneous notions, will be mistaken for established scientific facts and will thereby be granted immunity to further criticism.

Yet a further danger of these diagnostic schemes lies in their possible oversimplified use; diagnosticians may be content to assign a typological label to the patient's illness, as in the past, and the use of the stated criteria may turn out to be more of a pretense than a reality (Arolt and Dilling 1994). Most of the diagnoses listed in ICD-10 – and there are 50 codes just for various types of depression, for example – have the inherent potential of being turned into globalized "disorders" that will play a role equivalent to that of the previous global "disease patterns" and fail to be differentiated on the basis of the criteria required by ICD-10. This would amount to a reductionistic distortion of the diagnostic system. Indeed, the preliminary findings of a questionnaire we distributed, with the support of the WHO, regarding the frequency of various diagnoses in psychiatric inpatient wards and ambulatory services reveal a lack of appreciation and application of the differentiating power of the new classification. A small number of diagnoses were found to be used especially frequently; in the Federal Republic of Germany, 40% of initial diagnoses are in one of only three categories – alcoholism, depression, and adjustment disorder (Müßigbrodt et al. 1999).

While scientific papers that do not apply operational diagnostic systems run the risk of being rejected by prestigious journals, and the awarding of support for research is increasingly linked to the use of such systems, the situation in clinical practice appears markedly different. The new diagnostic concepts are indeed used by many clinicians to describe the patient's abnormalities, but the planning and delivery of treatment frequently appear to be based on other concepts. In some special areas, e.g., the psychogenic disorders, and for certain special applications, there is a need for increased differentiation and the construction of new diagnostic modules.

Thus, despite the introduction of ICD-10 around the world, operational diagnosis remains far from being generally accepted, particularly in ambulatory care. There are many practical difficulties, such as that of introducing new software, but there are also deeper sources of resistance. Many neurologists and psychiatrists are uneasy with the dissolution of deeply rooted traditional modes of thought, such as the triadic system, with their own lack of detailed knowledge of current diagnosis, and with being required to use a new code, possibly under outside control (Arolt 1994). Consequently, there is a need not merely for occasional introductory lectures on the new diagnostic system, but for systematic training, including introductory and advanced seminars and group discussions of case material, if a uniform, high standard of diagnosis is to be achieved. All of the modern media techniques and methods of information transfer should be applied. The acceptance or nonacceptance of operational diagnosis will ultimately depend on whether it improves the indications for treatment and therefore its results. In this respect, many sections of ICD-10, such as that concerning affective disorders, have a long road ahead.

8.3 Future Development

Both the international classification of mental disorders and the North American classification, beyond their global aspects, have led to a large number of developments in special areas, new diagnostic instruments, and various national and international versions and subclassifications. A far-reaching system of descriptive diagnosis has become established that no longer takes account of certain traditional models in psychiatry and psychotherapy, such as endogenous disorders or the concept of neurosis.

The postulation of the axes of "illness experience," "treatment expectation," "relationship," "conflict," and "structure" by the Task Force for Arbeitskreis OPD (Task Force on OPD) indicates that the operational method of the two major classifications has extended its influence to fields that initially resisted exact diagnosis, even if disagreement remains regarding content. As Hoffmann stated, "Our intention was to reduce the psychoanalytic arbitrariness of the diagnostic process, and to set down accepted and communicable standards" (Arbeitskreis OPD 1996).

The foregoing example shows one way in which future perspectives may open up for the emulation and furtherance of the WHO classification. Other possibilities include a behavior-theoretic classification of neurotic disorders and the current reworking of the ICIDH.

A series of detailed special classifications will thus be produced in the context of Chap. V of ICD-10, in which special disorder areas will benefit from a closer look, and the multiaxiality of the present system will be further elaborated. For example, the unsatisfactory axis III of ICD-10 might be replaced by a more comprehensive psychosocial diagnostic system.

Of the two major classification systems now in use for scientific diagnosis in psychiatry, ICD-10 will no doubt be the more used in future for routine statistical purposes. An eventual merging of the two systems into a single classification, with a number of different versions, would be desirable. A dichotomous use of ICD-10 in clinical practice and DSM-IV in research would be disadvantageous to both sides.

An especially important future version will be concerned with psychiatric disorders in general practice, where the diagnostic process necessarily differs from that of the specialist. Although severe psychiatric disturbances requiring specialized care have occupied the center of attention hitherto, the need for adequate treatment of mental disorders is now strongly felt in nonpsychiatric practice as well. The provision of adequate diagnostic aids should be a common concern of the two systems of classification.

Finally, even when all of the diagnostic rules are strictly followed, some cases will prove resistant to operational diagnosis. The diagnostician must then ignore the official criteria and make an individual judgment, as in the days before these systems were introduced.

Even though the diagnostic process has moved a good distance away from the intuitive typological view of diseases and the patient, the physician still must not forget to consider the totality of the patient's illness experience and his or her individuality. Etiology, diagnosis, and therapy ought to stand in a logical relation to each other, and, despite the current tendency toward description, nosology may someday come back into the foreground. The next few years may bring progress in the clarification of etiologic aspects of mental disorders, particularly when attention is focused on smaller disease entities that may come closest to the underlying substrate. These might include bipolar affective disorders, seasonal depression, and individual types of schizophrenia with a strong tendency to occur in families, such as periodic catatonia of the Kleist-Leonhard type (Leonhard 1986). Thus, for both psychotic and nonpsychotic disorders, the further development of the concept of disease and the overcoming of the currently espoused atheoretic approach to diagnosis may be regarded as important challenges for the future.

According to Sartorius (1988), classification is a way of seeing the world. We live in a world of increasingly rapid change, and we may thus confidently expect that diagnosis and classification will change more rapidly in the coming century than in the one now reaching its end. The shifting view of the world, and of psychiatry, through the eyes of the diagnostician is bound to yield many surprises.

9 References

Andrews G, Dilling H, Üstün T, Briscoe M (1994) Computers in mental health, vol 1. WHO, Geneva/Churchill Livingstone, Edinburgh

APA (1968) Diagnostic and statistical manual of mental disorders, 2nd edn (DSM-II). American Psychiatric Association, Washington DC

APA (1980) Diagnostic and statistical manual of mental disorders, 3rd edn (DSM-III). American Psychiatric Association, Washington DC

APA (1987) Diagnostic and statistical manual of mental disorders, 3rd edn, revised (DSM-III-R). American Psychiatric Association, Washington DC

APA (1994) Diagnostic and statistical manual of mental disorders, 4th edn (DSM-IV). American Psychiatric Association, Washington DC

Arbeitskreis OPD (ed) (1996) Operationalisierte psychodynamische Diagnostik - OPD. Huber, Bern

Arolt V (1994) Die Einführung kriterienorientierter Klassifikationen: Probleme und Risiken in der Praxis. Krankenhauspsychiatrie 5: 21–24

Arolt V, Dilling H (1994) Confounding diagnostic systems: a major risk in the use of criteria-based manuals. Psychopathology 27: 58–63

Baillarger J (1854) De la folied à double forme. Ann Med Psychol 6: 369–384

Bleuler E (1911) Dementia praecox oder Gruppe der Schizophrenien. Deuticke, Leipzig

Bronisch T, Hiller W, Mombour W, Zaudig M (1995) IDCL-P Internationale Diagnosen Checkliste für Persönlichkeitsstörungen nach ICD-10. Huber, Bern

Cullen W (1772) Synopsis nosologiae methodicae. Creech, Edinburgh

Degkwitz R, Helmchen H, Meyer JE, Mombour W (1971) Diagnosenschlüssel und Glossar psychiatrischer Krankheiten. (German translation of ICD, 8th revision). Springer, Berlin Heidelberg New York

Degkwitz R, Helmchen H, Kockott G, Mombour W (1980) Diagnosenschlüssel und Glossar psychiatrischer Krankheiten. (German translation of ICD, 9th revision). Springer, Berlin Heidelberg New York

Dilling H (1994) Diagnostische Modelle in der Psychiatrie. In: Janssen PL, Schneider W (eds) Diagnostik in Psychotherapie und Psychosomatik. Fischer, Stuttgart, pp 7–15

Dilling H, Dittmann V (1990) Die psychiatrische Diagnostik nach der 10. Revision der Internationalen Klassifikation der Krankheiten (ICD-10). Nervenarzt 61: 259–270

Dilling H, Siebel U (1995) Kommentierung der ICIDH aus psychiatrisch-rehabilitativer Sicht. In: Matthesius RG, Jochheim KA, Barolin GS, Heinz C (eds) International classification of impairments, disabilities and handicaps (ICIDH). Ullstein, Berlin, pp 143–152

Dilling H, Dittmann V, Freyberger HJ (eds) (1990) ICD-10 field trial in German-speaking countries. Pharmacopsychiatry 23[Suppl IV]: 135–216

Dilling H, Mombour W, Schmidt MH (eds) (1991) Weltgesundheitsorganisation. Internationale Klassifikation psychischer Störungen. ICD-10, Kapitel V (F): Klinisch-diagnostische Leitlinien. Huber, Bern

Dilling H, Mombour W, Schmidt MH (eds) (1993) Weltgesundheitsorganisation. Internationale Klassifikation psychischer Störungen. ICD-10, Kapitel V (F): Klinisch-diagnostische Leitlinien, 2nd edn. Huber, Bern

Dilling H, Mombour W, Schmidt MH (eds) (1994a) Weltgesundheitsorganisation. Internationale Klassifikation psychischer Störungen. ICD-10, Kapitel V (F): Forschungskriterien. Huber, Bern

Dilling H, Schulte-Markwort E, Freyberger HJ (eds) (1994b) Von der ICD-9 zur ICD-10. Huber, Bern

Dilling H, Freyberger HJ, Kanitz RD, Müßigbrodt H (eds) (1998) Weltgesundheitsorganisation. Internationale Klassifikation psychischer Störungen. ICD-10, Kapitel V (F): Didaktische Folien (CD-Rom). Huber, Bern

DIMDI (Deutsches Institut für medizinische Dokumentation und Information) (ed) (1995/1996) ICD-10 Internationale Klassifikation der Krankheiten und verwandter Gesundheitsprobleme, 10. Revision. Deutscher Ärzteverlag, Cologne

Dittmann V, Dilling H, Freyberger HJ (eds) (1992) Psychiatrische Diagnostik nach ICD-10 – Klinische Erfahrungen bei der Anwendung. Ergebnisse der ICD-10-Merkmalslistenstudie. Huber, Bern

Esquirol E (1838) Des maladies mentales considérées sous les rapports médical, hygiénique et médico-légal. Baillière, Paris

Eschenburg BG (1855) Die Irrenstatistik des lübeckischen Staates. Neue Lübeckische Blätter 21: 329–331

Essen-Möller E (1973) Standard list for threefold classification of mental disorders. Acta Scand Psychiatr 49: 198–212

Falret JP (1854) De la folie circulaire. Bull Acad Med 19: 382–395

Feighner J, Robins E, Guze S et al (1972) Diagnostic criteria for use in psychiatric research. Arch Gen Psychiatry 26: 57–63

Freyberger HJ, Dilling H (eds) (1993) Fallbuch Psychiatrie. Kasuistiken zum Kapitel V (F) der ICD-10. Huber, Bern

Freyberger HJ, Schulte-Markwort E, Dilling H (1993a) Referenztabellen der WHO zum Kapitel V (F) der 10. Revision der Internationalen Klassifikation der Krankheiten (ICD-10): ICD-9 vs. ICD-10. Fortschr Neurol Psychiatr 61: 109–127

Freyberger HJ, Schulte-Markwort E, Dilling H (1993b) Referenztabellen der WHO zum Kapitel V (F) der 10. Revision der Internationalen Klassifikation der Krankheiten (ICD-10): ICD-10 vs. ICD-9. Fortschr Neurol Psychiatr 61: 128–143

Freyberger HJ, Dilling H, Stieglitz RD (1996) ICD-10 Field trial of the diagnostic criteria for research in German-speaking countries. Psychopathology 5: 258–314

Fujinawa A (1994) Japanese experiences in psychiatric diagnosis. Overview of Japanese experience in diagnostic classification: past and present classification of mental disorders in Japan. In: Mezzich JE, Honda Y, Kastrup MC (eds) Psychiatric diagnosis. Springer, Berlin Heidelberg New York, pp 81–83

Griesinger W (1845) Die Pathologie und Therapie der psychischen Krankheiten für Aerzte und Studirende. Verlag von Adolph Krabbe, Stuttgart

Hanada K (1994) Japanese experiences in psychiatric diagnosis. The new classification of mood disorders in Japan. In: Mezzich JE, Honda Y, Kastrup MC (eds) Psychiatric diagnosis. Springer, Berlin Heidelberg New York, pp 93–95

Hecker E (1871) Die Hebephrenie. Virchows Arch 52: 394–429

Heinroth J (1818) Lehrbuch der Störungen des Seelenlebens. Vogel, Leipzig

Helmchen H, Hippius H, Meyer JE (1966) Ein neues psychiatrisches Diagnosenschema. Nervenarzt 37: 115–118

Hiller W, Zaudig M, Mombour W, WHO (1995) ICD-10 Checklisten. Huber, Bern

Hoche AE (1912) Die Bedeutung der Symptomkomplexe in der Psychiatrie. Z Ges Neurol Psychiatr 12: 540–551

Jablensky A (1988) Methodological issues in psychiatric classification. Br J Psychiatry 152: 15–20

Jablensky A (1991) Diagnostic criteria in psychiatry: a straitjacket or a prop? Eur Psychiatry 6: 323–329

Janca A, Üstün T, Early T, Sartorius N (1993) The ICD-10 symptom checklist: a companion to the ICD-10 classification of mental and behavioural disorders. Soc Psychiatry 28: 239–242

Janca A, Kastrup M, Katschnig H, López-Ibor JJ, Mezzich JE, Sartorius (1996) The World Health Organization Short Disability Assessment Schedule (WHO DAS-S): a tool for the assessment of difficulties in selected areas of functioning of patients with mental disorders. Soc Psychiatry 31: 349–354

Kahlbaum K (1863) Die Gruppierung der psychischen Krankheiten. Kafemann, Danzig

Kant I (1983) Versuch über die Krankheiten des Kopfes. In: Weischedel W (ed) Vorkritische Schriften bis 1768, part 2. Wissenschaftliche Buchgesellschaft, Darmstadt

Kendell RE (1975) The role of diagnosis in psychiatry. Blackwell, Oxford

Kendell RE (1990) A brief history of psychiatric classification in Britain. In: Sartorius N, Jablensky A, Regier DA, Burke JD, Hirschfeld RMA (eds) Sources and traditions of classification in psychiatry. Huber, Bern, pp 129–151

Kendler KS (1990) Toward a scientific psychiatric nosology. Arch Gen Psychiatry 47: 969–973

Kessler C, Freyberger HJ (eds) (1996) Weltgesundheitsorganisation. Internationale Klassifikation neurologischer Erkrankungen. Huber, Bern

Kleinschmidt S, Müßigbrodt H, Schürmann A, Freyberger HJ, Dilling H (1995) Psychiatrische Diagnostik in der Allgemeinpraxis. Fortschr Neurol Psychiatr 63: 227–231

Klerman GL (1990) The contemporary American scene: diagnosis and classification of mental disorders, alcoholism and drug abuse. In: Sartorius N, Jablensky A, Regier DA, Burke JD,

Hirschfeld RMA (eds) Sources and traditions of classification in psychiatry. Huber, Bern, pp 93–138
Kraepelin E (1896) Lehrbuch der Psychiatrie, 5th edn. Barth, Leipzig
Kramer M (1988) Historical roots and structural basis of the International Classification of Diseases. In: Mezzich JE, von Cranach M (eds) International Classification in Psychiatry. Cambridge University Press, Cambridge, pp 3–29
Kretschmer E (1919) Gedanken über die Fortentwicklung der psychiatrischen Systematik. Z Neurol 48: 370–377
Leonhard K (1986) Aufteilung der endogenen Psychosen und ihre differenzierte Ätiologie. Akademie-Verlag, Berlin
Loranger AW (1996) Weltgesundheitsorganisation. International personality disorder examination. IPDE. ICD-10 Modul. Huber, Bern
Loranger AW, Sartorius N, Andreoli A et al (1994) The international personality disorder examination. Arch Gen Psychiatry 51: 215–224
Magnan V, Serieux P (1893) Le délire chronique à évolution systématique. Villars/Masson, Paris
Maier W, Philipp M (1988) Die empirische Erforschung der Klassifikation psychischer Störungen. Nervenarzt 59: 449–445
Maier W, Philipp M (1990) Die Mainzer Verlaufsstudie zur Validierung von ICD-10. In: Lungershausen E, Kaschka WP, Witkowski, RJ (eds) Affektive Psychosen. Schattauer, Stuttgart, pp 497–501
Malchow CP, Kanitz RD, Dilling H (eds) (1995) ICD-10-Computer-Tutorial: Psychische Störungen. Huber, Bern
Matthesius RG (1995) Internationale Klassifikation der Schädigungen, Fähigkeitsstörungen und Beeinträchtigungen (ICIDH). Ullstein, Berlin
Meyer JE (1961) Diagnostische Einteilungen und Diagnosenschemata in der Psychiatrie. In: Gruhle WH, Jung R, Mayer-Gross W, Müller M (eds) Psychiatrie der Gegenwart, vol III. Springer, Berlin Heidelberg, pp 131–180
Mezzich JE (1988) On developing a psychiatric multiaxial schema for ICD-10. Br J Psychiatry 152: 38–43
Mezzich JE, von Cranach M (eds) (1988) International classification in psychiatry. Cambridge University Press, Cambridge
Michels R, Siebel U, Freyberger HJ, Stieglitz RD, Schaub RT, Dilling H (1996) The multiaxial system of ICD-10: evaluation of a preliminary draft in a multicentric field trial. Psychopathology 29: 347–356
Mombour W (1975) Klassifikation, Patientenstatistik, Register. Psychiatrie der Gegenwart, vol III. Springer, Berlin Heidelberg New York
Morel BA (1853) Traité des maladies mentales. Masson, Paris
Müßigbrodt H, Kleinschmidt S, Schürmann A, Freyberger HJ, Dilling H (1996) Psychische Störungen in der Praxis. Huber, Bern
Müßigbrodt H, Michels R, Malchow CP, Dilling H, Munk-Jørgensen P, Bertelsen A (1999) Use of the ICD-10-classification in psychiatry – an international survey. Psychopathology
Pinel P (1809) Traité médico-philosophique sur l'aliénation mentale. Brosson, Paris
Pull CB, Pull MC, Pichot P (1988) The French approach to psychiatric classification. In: Mezzich JE, von Cranach M (eds) International classification in psychiatry. Cambridge University Press, Cambridge, pp 37–47
Remschmidt H, Schmidt MH (eds) (1996) Multiaxiales Klassifikationsschema für psychische Störungen des Kindes- und Jugendalters nach ICD-10 der WHO. Huber, Bern
Research Report (1983) Diagnosis and classification of mental disorders and alcohol- and drug-related problems: a research agenda for the 1980s. Psychol Med 13: 907–921
Rutter M, Shaffer D, Sturge C (1975) A multi-axial classification of child psychiatric disorders. WHO, Geneva
Sartorius N (1988) International perspectives of psychiatric classification. Br J Psychiatry 152: 9–14
Sartorius N (1995) Understanding the ICD-10 classification of mental disorders. Science, London
Sartorius N, Jablensky A, Regier DA, Burke JE, Hirschfeld RMA (eds) (1990) Sources and traditions of classification in psychiatry. Huber, Bern
Sartorius N, Kaelber CT, Cooper JE et al (1993) Progress toward achieving a common language in psychiatry. Results from the field trial of the clinical guidelines accompanying the WHO classification of mental and behavioural disorders in ICD-10. Arch Gen Psychiatry 50: 115–124
Sartorius N, Üstün TB, Korten A, Cooper JE, van Drimmelen J (1995) Progress toward achieving a common language in psychiatry. II. Results from the international field trials of the ICD-10 diagnostic criteria for research for mental and behavioural disorders. Am J Psychiatry 152: 1427–1437
Saß H (1987) Die Krise der psychiatrischen Diagnostik. Fortschr Neurol Psychiatr 55: 255–258
Saß H, Wittchen HU, Zaudig M (1996) Diagnostisches und statistisches Manual psychischer Störungen DSM-IV. Hogrefe, Bern
Sjöbring H (1919) Mental constitution and mental illness. Svenska Läkare Sällskap 45: 462–493
Snell L (1865) Über Monomanie als primäre Form der Seelenstörung. Allg Z Psychiatr 22: 368–381
Spitzer RL, Endicott J, Robins E (1975) Research diagnostic criteria (RDC) for a selected group of functional disorders. New York State Department of Mental Hygiene, Biometrics Branch, New York
Strömgren E (1988) Scandinavian approaches to psychiatric diagnosis. In: Mezzich JE, von Cranach M (eds) International classification in psychiatry. Cambridge University Press, Cambridge
Strömgren E (1989) Aktuelle Probleme der psychiatrischen Klassifikation. In: Kisker KP, Lauter H, Meyer JE, Müller C, Strömgren (eds) Psychiatrie der Gegenwart, vol 9. Springer, Berlin Heidelberg New York, pp 47–84
Strömgren E (1994) Scandinavian contributions to psychiatric nosology. In: Mezzich JE, Honda Y, Kastrup MC (eds) Psychiatric diagnosis. Springer, Berlin Heidelberg New York, pp 48–54
Thangavelu R, Martin RL (1995) ICD-10 and DSM IV: depiction of the diagnostic elephant. World Psychiatry 3: 3–11
Üstün TB, Bertelsen A, Dilling H, van Drimmelen J, Pull C, Okasha A, Sartorius N (1996) ICD-10 casebook. The many faces of mental disorders. American Psychiatric Press, Washington DC
von Linné C (1742) Genera morborum in auditorium usum. Buchenroeder and Ritter, Hamburg
WHO (1948) Manual of the international statistical classification of diseases, injuries and causes of death (ICD-6). Bulletin of the World Health Organisation. World Health Organization, Geneva
WHO (1967) Glossary of mental disorders and guide to their classification, for use in conjunction with the International Classification of Diseases, 8th revision. World Health Organization, Geneva

WHO (1978) Mental disorders: glossary and guide to their classification, for use in conjunction with the Ninth Revision of the International Classification of Diseases. World Health Organization, Geneva

WHO (1980) International classification of impairments, disabilities and handicaps. World Health Organization, Geneva

WHO (1981) Current state of diagnosis and classification in the mental health field. World Health Organization, Geneva

WHO (1985) Lexicon of psychiatric and mental health terms, vol 1. World Health Organization, Geneva

WHO (1992a) International statistical classification of diseases and related health problems, 10th revision. World Health Organization, Geneva

WHO (1992b) The ICD-10 classification of mental and behavioural disorders – clinical descriptions and diagnostic guidelines. World Health Organization, Geneva

WHO (1993a) The ICD-10 classification of mental and behavioural disorders – diagnostic criteria for research. World Health Organization, Geneva

WHO (1993b) Composite international diagnostic interview (CIDI). American Psychiatric Press, Washington, DC

WHO (1994a) Lexicon of psychiatric and mental health terms, 2nd edn. World Health Organization, Geneva

WHO (1994b) Lexicon of alcohol and drug terms. World Health Organization, Geneva

WHO (1994c) Division of mental health. The ICD-10 symptom checklist. Version 2.0. World Health Organization, Geneva (prepared by Janca A, Üstün T, van Drimmelen-Krabbe J, Dittmann V, Isaac M)

WHO (1994d) Schedules for clinical assessment in neuropsychiatry (SCAN). Version 2.0. American Psychiatric Press, Washington, DC

WHO (1996) Diagnostic and management guidelines for mental disorders in primary care. ICD-10, Chap. V, primary care version. Huber, Bern

Wig NN (1990) The Third-World perspective on psychiatric diagnosis and classification. In: Sartorius N, Jablensky A, Regier DA, Burke JD, Hirschfeld RMA (eds) Sources and traditions of classification in psychiatry. Huber, Bern

Wilmanns K (1930) Entwurf einer für die Reichsstatistik bestimmten Diagnosentabelle für die Geisteskrankheiten. Allg Z Psychiatr 93: 223–234

Wimmer A (1916) Psykogene sindssygdomsformer [Psychogenic mental disorders]. In: Sct. Hans Mental Hospital 1816–1916, jubilee publication. Gad, Copenhagen, pp 85–216

Wing JK, Cooper JE, Sartorius N (1974) Present state examination. Cambridge University Press, Cambridge

Wittchen HU (1994) Klassifikation. In: Stieglitz R, Baumann U (eds) Psychodiagnostik psychischer Störungen. Enke, Stuttgart, pp 47–60

Wittchen HU, Semler G (eds) (1991) Composite international diagnostic interview. Beltz, Weinheim

Wittchen HU, Saß H, Zaudig M, Koehler K (1989) Diagnostisches und statistisches Manual psychischer Störungen DSM-III-R. Beltz, Weinheim

Yu-cun S, Changhui C (1988) Principles of the Chinese classification of mental disorders (CCMD). In: Mezzich JE, von Cranach M (eds) International classification in psychiatry. Cambridge University Press, Cambridge

Zaudig M, Hiller W (1996) SIDAM. Strukturiertes Interview für die Diagnose einer Demenz vom Alzheimer-Typ, der Multiinfarkt- (oder vaskulären) Demenz und Demenzen anderer Ätiologie nach DSM-III-R, DSM-IV und ICD-10. Huber, Bern

CHAPTER

3

J. van Drimmelen-Krabbe, A. Bertelsen, C. Pull

International Psychiatric Classification: ICD-10 and DSM-IV

... ein Diagnosenschema ... hat daher nur einen stets vorläufigen Ordnungswert. Sie [eine solche Einteilung] ist eine Fiktion, die ihre Aufgabe erfüllt, wenn sie die zur Zeit relativ richtigste ist.
JASPERS (1973)

1 Background and Development

Classification is a systematic arrangement of the world as a result of a human ability to create systems in a multitude of otherwise chaotic entities in order to comprehend and master the surrounding world, an ability that probably corresponds to the structure of the human neural network. In psychiatry, classification is as old as history itself, from the Hippocratic school onwards. In the course of the last century, psychiatric classification has become increasingly refined but also subdivided among various schools and traditions, ending up with a number of different major national classifications, such as the French, German, Russian and Anglo-Saxon classifications and, outside Europe, the American, Chinese, Indonesian and Japanese ones. Moreover, within psychiatry, separate groups used to produce their own classification, e.g., for childhood mental disorders. In the years after the Second World War, with the re-establishment of international communication on psychiatric research, the need for a common classification became apparent (Stengel 1959). One of the goals of the Mental Health Division of the World Health Organization was to develop such a classification, and a committee of expert consultants prepared the psychiatric chapter of the eighth revision of the *International Classification of Diseases*, ICD-8 (World Health Organization 1965), which was followed by a *Glossary of Mental Disorders and Guide to Their Classification* a few years later (World Health Organization 1974). The glossary provided prototypical descriptions of the major diagnostic categories and indications of inclusion terms. The categories were arranged in a continuous code number system from 0 to 999, each with a fourth-character decimal subdivision into ten subcategories. Main psychiatric categories were distributed among the numbers 290–315. Several countries changed from their national classification to ICD-8, but overall use worldwide was still rather limited. The next revision, ICD-9, appeared in 1975 (World Health Organization 1975), followed by a glossary in 1978 (World Health Organization 1978). The psychiatric chapter was enlarged within the same code numbers 290–315 by a number of subcategories, particularly for child psychiatry disorders, but otherwise the changes in categorization and glossary descriptions were minor (Kramer et al. 1979). ICD-9 became more widely used, but a number of countries continued using ICD-8 because they did not find the changes sufficiently significant to justify the costs and inconvenience of replacing ICD-8 with ICD-9.

In the late 1960s and early 1970s, the U.S–U.K. diagnostic study demonstrated the lack of comparability between the psychiatric classifications in the two countries, using the Present State Examination for comparable registration of psychopathology (Cooper et al. 1972). As a consequence, more accurate and reliable diagnostic criteria were provided for research (Feighner et al. 1972; Spitzer et al. 1977). They were well accepted internationally and inspired a major revision of the U.S. national psychiatric classification, the *Diagnostic and Statistical Manual*, 3rd revision (DSM-III; American Psychiatric Association 1980). The revision was radical, implementing a change from more or less aetiologically defined, broadly described categories (American Psychiatric Association 1968) to purely descriptive, non-aetiological, criteria-based diagnostic categories based on empirical evidence for use in clinical practice and research. In spite of some shortcomings, revised in 1987 (DSM-III-R; American Psychiatric Association 1987), the DSM-III classifications (DSM-III and DSM-III-R) became a major success and were soon taken up outside the United States as well, primarily in research but in some countries also to define the content of the ICD-9 categories to which the DSM-III categories were linked. The DSM-III system seemed to meet a worldwide need for comparable diagnostics and common classification. In this way, psychiatric research and clinical practice was improved, increasing the identity and status of psychiatry as a medical discipline.

2 WHO ICD-10 Revision Principles

DSM-III also had a major impact on the ICD-10 revision of the psychiatric classification, which began in the early 1980s. The WHO-coordinated revision was based on a major and complex collaborative effort including groups of expert consultants, national psychiatric societies and the World Psychiatric Association (Sartorius 1991, 1993). The main obligation was to provide a major revision promoting a worldwide common language in psychiatric classification. The revision therefore had to follow certain principles (Sartorius 1988, 1995; Jablensky 1988; Cooper 1988).

It had to be comprehensive, including the majority of categories used in the various national classifications, particularly taking into consideration special categories seen in the developing countries which did not easily fit in with the previous classifications. The heading was changed from mental disorders to also include behavioural disorders. In this way, the total number of categories and subcategories was increased from about 300 to more than 1000, facilitated by an overall change from the numeric coding to an alpha-

numeric system based on codes with a single letter followed by two numbers at a three-character level (A00–Z99). Further detail is then provided by means of a decimal numeric subdivision at a four- or five-character level (World Health Organization 1992a). The fifth chapter of ICD-10, containing mental and behavioural disorders, received the letter F, followed by two numbers: F00–F99 for the main categories at the three-character level with decimals for subcategories (Fxx.xx), allowing a total of $10^4 = 10{,}000$ subdivisions at the five-character level. The system provides a possibility for future revisions to enlarge or change within wide frames with only minor changes in computer registration. The neutral term "disorder" was used throughout the classification to avoid problems inherent in the use of terms such as "disease", "illness" or "abnormality". To be included in the psychiatric classification, a "disorder" had to be clinically significant and defined by a recognizable set of symptoms or pattern of behaviour associated with distress or interference with daily-life personal functioning, avoiding inclusion of social deviance or conflict as such.

To provide a meaningful classification, diagnostic categories have to be mutually exclusive without substantial overlapping or ambiguity. Consequently, the categories had to be defined by clear and precise boundaries by operational criteria. They were presented by the WHO Mental Health Division in two major publications: *Clinical Descriptions and Diagnostic Guidelines*, with the criteria presented as diagnostic guidelines for flexible clinical use (World Health Organization 1992b), and *Diagnostic Criteria for Research*, with criteria presented in the same way as in the DSM-III system (World Health Organization 1993a). The guidelines and the research criteria were almost completely identical, with the exception of a few categories in which the research criteria had to be made more precise to ensure mutually exclusive categorization. Thus ICD-10 has no border areas or borderline states. The borders are sharp lines, and a disorder will be either on the one or the other side of the line. The criteria were, as far as possible, operationally defined, relying heavily on the glossary definitions of the Present State Examination symptom items.

Worldwide acceptability was a third principle the revision of the classification had to follow. This implied an atheoretical descriptive approach and the avoidance of social criteria. The national classifications were based on aetiological assumptions according to various traditions or schools, these in turn being based on different aetiological theories which were mutually incompatible. An atheoretical approach therefore requires a non-aetiological classification for most of the diagnostic categories, with the exception of those with a proven aetiology, e.g., the so-called organic disorders, which can be attributed to brain damage and dysfunction caused by an independently diagnosable cerebral or systemic physical disease or disorder. This implies that disorders caused by alcohol or other psychoactive substances also are "organic". Use of the term "organic" does not imply that conditions elsewhere in the classification are "non-organic" in the sense of having no cerebral substrate. The other group of disorders for which an aetiology has been universally accepted are the stress-related disorders with evidence of psychological causation, either in the form of exceptional mental stress or of more ordinary psychosocial stressors or life events. Otherwise, the non-aetiological approach has been strictly followed throughout, also in the choice of nomenclature or terms. The previous concepts of "psychosis" and "neurosis" have been abandoned and the terms deleted because of the aetiological implications. Only the adjective forms "psychotic" and "neurotic" have been retained as convenient descriptive terms. "Psychotic" is now purely descriptive, defined by the presence of "psychotic" symptoms such as hallucinations or delusions or a limited number of other severe abnormalities such as disorganized speech, catatonic behaviour and extreme psychomotor excitement or stupor, but with no implications of loss of sense of reality or lack of insight. "Neurotic" is used in the meaning "non-psychotic" and mainly as a descriptive term for anxiety, obsessive–compulsive and dissociative symptoms or disorders. These descriptive approaches are fully in line with the DSM-III classification.

Social criteria have been avoided almost completely in order to make the ICD-10 classification applicable and thus acceptable in the various cultures all over the world. Major areas of social functioning such as work or interpersonal relations differ substantially between the various cultures to a degree that makes it impossible to obtain universally applicable criteria in these areas. Only basic personal daily-life functioning such as basic self-care, which is universal to all cultures, may be applied, and this is already included in the definition of a "disorder". The avoidance of social criteria is one of the major principal differences between ICD-10 and the DSM-III system, which applies social criteria including occupation and personal relationships throughout. In ICD-10, social criteria only appear as an exception in a few categories such as "Simple Schizophrenia" and unavoidably in "Dyssocial Personality Disorder and Childhood Conduct Disorder".

In ICD-10, the ordering of the diagnostic categories follows the traditional division between (a) disorders with onset specific to childhood or adolescence but which may, however, persist into adulthood and (b) "adult" mental and behavioural disorders without

onset specific to childhood. The latter may occur in childhood as well, but most often appear in adult life. For adult disorders, the order of categories follows a hierarchical principle. The "organic" disorders, including the substance use disorders, are placed at the top of the hierarchy because they may imitate all other kinds of disorders; exclusion criteria for organic aetiology are thus required. In the hierarchical order, they are followed by schizophrenia and disorders descriptively related to schizophrenia, which again may imitate the disorders at a lower level in the hierarchy and therefore also have to be ruled out in the lower-level diagnoses. These are followed by mood or affective disorders, which, however, in some cases compete with the schizophrenia level as to hierarchical preference. The next disorders are neurotic, stress-related and somatoform disorders, behavioural syndromes associated with physiological disturbances and physical factors and, finally, disorders of adult personality and behaviour. This means that, if there is a simultaneous course of a depressive episode and an anxiety disorder, for example, the diagnosis of the depressive episode will take precedence and incorporate the anxiety disorder as part of the depressive episode. To some extent, this principle precludes the registration of co-morbidity. This is another major principal difference from the DSM system, which encourages the diagnosis of co-morbidity.

Last but not least, the ICD-10 revision needed to be based on demonstrated reliability and was designed be fairly easy to use in the daily clinic. This was tested by WHO-coordinated worldwide field trials of the pre-final drafts of the *Clinical Description and Diagnostic Guidelines* and the *Diagnostic Criteria for Research* (Sartorius et al. 1993, 1995). Categories with low reliability were deleted (e.g., "hazardous alcohol use" and "accentuated personality" traits) or re-defined (e.g., "specific personality disorders"). The final draft prepared for the complete ICD-10 was presented for and approved by the World Health General Assembly in 1989 to come into effect as of 1 January 1993.

The centres that coordinated the field trials continued their participation as WHO ICD-10 reference centres for the introduction and training in the use of ICD-10; these centres also serve as members of the WHO ICD-10 Advisory Committee.

3 DSM-IV Revision

Following DSM-III-R, the American Psychiatric Association began the preparation of another revision for the DSM-IV. A task force and its work groups conducted a three-stage empirical process that included the following: (a) comprehensive and systematic reviews of the published literature, (b) re-analysis of the previously collected data and (c) extensive field trials focused on specific issues. The revision followed the tradition embodied in DSM-III and DSM-III-R, and changes were introduced only if based on substantial evidence from literature reviews, data analyses and field trials results. A special effort to ensure compatibility with ICD-10 included mutual meetings between the task force and the WHO ICD-10 Expert Committee and between work group members from both systems. This led to an approximation of definitions and diagnostic criteria in a number of categories, but some of the major principal differences remained, such as the avoidance of social criteria and the hierarchical order of categories with restriction of co-morbidity in ICD-10. The DSM-IV revision was to serve as a national classification system in the United States, which allowed a more evidence-directed approach and the application of social criteria; in contrast, the WHO ICD-10 classification had to be the result of a universal worldwide consensus that not only required an evidence-based approach but also had to take diplomatic aspects into consideration. DSM-IV was finalized and introduced in 1994 (American Psychiatric Association 1994). Because of difficulties with the registration and insurance systems, the code numbers still follow the ICD-9 clinical modification system (ICD-9 CM), but will eventually change to the alphanumeric system of ICD-10 (American Psychiatric Association 1995). The similarities between the two systems are substantial, and for moderate and severe cases within the various diagnostic categories the correspondence is almost complete. The differences appear mainly in the mild to moderate cases because of differences in the definitions of various diagnostic categories affecting the diagnostic thresholds. The differences due to the ICD-10 lack of social criteria and restricted use of co-morbidity are also most prominent in these areas. This is particularly relevant for psychiatric research, because variation in diagnostic thresholds may create quite different diagnostic samples of corresponding categories for which the resulting evidence will not be comparable.

4 Differences Between the ICD-10 and DSM-IV Classifications

The presentation of the diagnostic categories differs substantially between the two systems. Disorders usually first diagnosed in infancy, childhood or adolescence are presented first in DSM-IV, whereas ICD-10 has placed them as the last part of the classification

with the main group code numbers F7–F9. Organic disorders and substance-related disorders are presented in DSM-IV in a new and unconventional way contrary to the hierarchical principle of ICD-10. Delirium, dementia and amnestic disorders are presented as main groups, as are substance dependence, abuse, intoxication and withdrawal. Other organic mental disorders are presented in the various syndrome sections under the headings "syndrome disorder due to a general medical condition" and "substance induced syndrome disorder". Otherwise, DSM-IV follows almost the same order of presentation as ICD-10. DSM-IV has an additional group of "other conditions that may be a focus of clinical attention", including a number of disorders or factors covered by categories outside Chap. V (F) of ICD-10.

4.1 Organic Disorders

Dementia (F00–F03), amnesic syndrome (F04) and delirium (F05) tend to be more narrowly defined in ICD-10 than in DSM-IV. For delirium, ICD-10 requires specific impairment of immediate recall and recent memory with relatively intact remote memory, psychomotor disturbances and disturbed sleep. ICD-10 differentiates between delirium not superimposed or superimposed on dementia and other delirium, including cases with mixed origin. Delirium may further be associated with substance use intoxication or withdrawal. DSM-IV differentiates between delirium due to a general medical condition, substance-induced delirium and delirium due to multiple aetiologies.

For dementia, the ICD-10 criteria requires a minimum duration of 6 months and an additional decline in emotional control, motivation or social behaviour. Both systems require a decline in memory and other cognitive abilities, which are more extensively presented in DSM-IV, also including aphasia, apraxia and agnosia. They both differentiate between dementia in Alzheimer's disease, vascular dementia and dementia in other neurological diseases affecting the brain, to which DSM-IV also adds dementia due to other general medical conditions, substance-induced persisting dementia and dementia due to multiple aetiologies.

In ICD-10, organic amnesic syndrome (referred to as amnestic disorder in DSM-IV) requires an additional defect of recent memory together with a reduced ability to recall past experiences; moreover, it requires that immediate recall is preserved in order to make a clear distinction from delirium. ICD-10 differentiates between amnesic syndrome not induced by alcohol and other psychoactive substances (F04) and amnesic syndrome due to psychoactive substance use (F1x.6). DSM-IV differentiates between amnestic disorder due to a general medical condition and substance-induced persistent amnestic disorder.

The following two ICD-10 subsections – F06 (other mental disorders due to brain damage and dysfunction and to physical disease) and F07 (personality and behavioural disorders due to brain disease, damage and dysfunction) – are also found in DSM-IV under the specific syndrome due to "a general medical disorder". ICD-10 includes a new category, F06.7 (mild cognitive disorder), which has no equivalent in DSM-IV except that it has been appended in the subsection of "criteria sets and axes provided for further study" in appendix B, under the heading "mild neurocognitive disorder". The same applies to post-concussional syndrome (F07.2), whereas post-encephalitic syndrome may be covered by DSM-IV in "cognitive disorder not otherwise specified".

For substance dependence, the diagnostic criteria are quite similar in the two systems. ICD-10 requires a strong desire or sense of compulsion to take a substance (craving), whereas DSM-IV considers this implicit and does not have a specific criterion for "craving". ICD-10 requires three symptoms simultaneously for at least 1 month or repeatedly within a 12-month period, whereas DSM-IV requires clinically significant impairment or distress manifested by three or more symptoms occurring at any time in the same 12-month period.

For harmful use (referred to in DSM-IV as substance abuse), the differences appear to be substantial. ICD-10 requires clear evidence of physical or psychological harm due to a pattern of use that has persisted for at least 1 month or has occurred repeatedly within a 12-month period. DSM-IV requires clinical and significant impairment or distress manifested by one or more of the following symptoms occurring within a 12-month period: (a) failure to fulfil major role obligations at work, school or home, (b) recurrent substance use in physically hazardous situations, (c) recurrent substance-related legal problems and/or (d) continuous substance use despite social or interpersonal problems. Here, the principal avoidance of social criteria in ICD-10 becomes apparent and seems to make the criteria for the two systems completely incompatible. The ICD-10 formulation, however, has been made flexible so that psychological harm can include "impaired judgement or dysfunctional behaviour, which may lead to disability or have adverse consequences for interpersonal relationships". Thus the DSM-IV social criteria may be interpreted as expressions of impaired judgement or dysfunctional behaviour, i.e., psychological harm, thus also meeting the ICD-10 criteria.

Substance intoxication and withdrawal broadly follow the same criteria, except for the DSM-IV requirement of "clinically significant distress or impairment",

which is implicit in the ICD-10 definition of a disorder. Both systems provide detailed substance-specific symptoms for intoxication and withdrawal, which are broadly equivalent, with the exception that ICD-10 includes phencyclidine use under F19 ("multiple drug use and use of other psychoactive substances").

The ICD-10 categories "psychoactive substance use psychotic disorders" (F1x.5) and "residual disorders and late onset psychotic disorder" (F1x.7) are dispersed among the syndrome sections in DSM-IV under the heading of "substance-induced [syndrome]" disorders. ICD-10 substance-induced psychotic disorders include disorders with predominantly depressive or manic symptoms. As they are defined, they have to be psychotic, i.e., include hallucinations or delusions. Substance-induced non-psychotic affective disorder has to be classified as residual affective disorder (F1x.72).

4.2 Schizophrenia and Related Psychotic Disorders

For schizophrenia, the ICD-10 and DSM-IV criteria sets are different as to duration, but appear more similar as to characteristic symptoms. For duration, ICD-10 requires 1 month of symptoms be present most of the time irrespective of treatment, whereas DSM-IV requires 6 months of symptoms including a 1 month active phase with characteristic symptoms or less than 1 month if successfully treated. Thus schizophrenia is of a chronic nature in DSM-IV, whereas in ICD-10 it may correspond to DSM-IV's schizophreniform disorder if it lasts for less than 6 months. The presentation of the characteristic symptoms also shows subtle but important differences. ICD-10 requires either at least one first-rank symptom (Schneider 1967) or "bizarre" delusions or, instead, a combination of at least two of the following: (a) hallucinations with non-affective delusions, (b) disorganized speech, (c) catatonic behaviour and/or (d) negative symptoms. The DSM-IV presentation appears more user-friendly, requiring two or more of the following: (a) delusions, (b) hallucinations, (c) disorganized speech, (d) grossly disorganized or catatonic behaviour and/or (e) negative symptoms. In a footnote, it is added that only one symptom is required if the delusions are bizarre or if the hallucinations consist of commenting or discussing voices. Thus the combination of hallucinations and delusions, whether affective or non-affective, is sufficient for the diagnosis, whereas the combination of hallucinations with non-affective delusions is only counted as one out of two necessary symptoms in ICD-10. As far as bizarre delusions are concerned, the quality of bizarreness is defined differently in the two systems. In ICD-10, they are not called bizarre, but are defined as "persistent delusions that are culturally inappropriate *and* completely impossible", whereas DSM-IV states that "delusions are deemed bizarre if they are clearly implausible and not understandable, and do not derive from ordinary life experiences" with the caution that "bizarreness" may be difficult to judge, especially across different cultures. It is obvious that delusions that are clearly implausible may all the same not be completely impossible. Furthermore, the bizarre delusions in DSM-IV are meant to cover the Schneiderian first-rank symptoms of experiences of subjective thought disorder and replacement of will, which, however, are combinations of subjective experiences and delusional explanations, albeit of a highly implausible character. In ICD-10, the Schneiderian first-rank auditory hallucinations are meant to be in the third person, which, however, is not clearly stated in either ICD-10 or DSM-IV. In ICD-10, audible thoughts and hallucinatory voices coming from some part of the body also qualify as a single symptom characteristic of schizophrenia. Furthermore, ICD-10 has no social criteria including occupation and interpersonal relationships, but as usual distress or interference with personal daily-life functioning is implicit in the definition of an ICD-10 disorder. For severe or chronic cases, these differences are of no practical significance, but for new cases with a recent onset and moderate or subtle symptoms, the differences are of substantial importance. As to the exclusion criteria, ICD-10 excludes primary or simultaneous presence of a manic or depressive syndrome, which will be classified under schizo-affective disorder, requiring for schizophrenia that the full schizophrenia syndrome must have been met before the mood disturbance developed. In DSM-IV, a concurrent mood syndrome is either ruled out or, if it occurs with the active-phase symptom, the duration must have been brief relative to the total duration of the active-phase and residual symptoms. The criteria to rule out substance use or a general medical condition are identical. Finally, DSM-IV requires prominent delusions or hallucinations for at least 1 month if the disorder appears in addition to autistic or other pervasive developmental disorders.

The schizophrenia subtypes are not fully equivalent and are defined in a slightly different way. Paranoid schizophrenia, however, has the same criteria in the two classifications. Hebephrenic schizophrenia in ICD-10 is referred to as schizophrenia, disorganized type, in DSM-IV and here requires prominent symptoms of disorganized speech, disorganized behaviour and flat and inappropriate affect, whereas the ICD-10 criteria only requires either disorganized behaviour or disorganized speech in addition to flat or inappropriate affect. The catatonic subtype in ICD-10 requires 2 weeks with at least one of a number of catatonic symptoms, whereas DSM-IV requires that the clinical

picture is dominated by at least two of similar catatonic symptoms. Undifferentiated schizophrenia in DSM-IV does not meet the criteria for any of the previous mentioned types, but in ICD-10, schizophrenic disorders with symptoms that meet the criteria for more than one of the previously mentioned types are also included under the heading "undifferentiated schizophrenia". For residual schizophrenia type, DSM-IV requires the presence of negative symptoms together with two or more characteristic symptoms in an attenuated form, whereas ICD-10 requires negative symptoms to have been present for 12 months in patients who have met the criteria for schizophrenia earlier, but not within the previous 12-month period. ICD-10 has a subtype of post-schizophrenic depression with a depressive syndrome of at least mild severity together with at least one schizophrenic symptom, which, however, must not be Schneiderian or bizarre delusions, so that it simultaneously fulfils the criteria for schizophrenia, which must have been met earlier within the previous 12 months. This subtype has no equivalent in DSM-IV, but has been included in appendix B under the heading of "postpsychotic depressive disorder of schizophrenia". Finally, ICD-10 includes simple schizophrenia as a subtype, defined by the progressive development over a period of 1 year of a significant and consistent change in the overall quality of personal behaviour, gradual appearance and deepening of negative symptoms and marked decline in social, scholastic or occupational performance (one of the exceptions to the principle of avoidance of social criteria in ICD-10). At the same time, the patient must never have had psychotic symptoms. This subtype does not have an equivalent in the DSM-IV classification either, but is included in the appendix B under the heading "simple deteriorative disorder".

The schizophreniform disorder of DSM-IV corresponds to ICD-10's schizophrenia F20.0, F20.1, F20.2, F20.3, F20.8 and F20.9 if the active-phase symptoms have lasted for at least 1 month irrespective of treatment. If the active phase has a duration of less than 1 month due to successful treatment, the criteria nevertheless require that the total duration is at least 1 month, also including prodromal or residual symptoms, and schizophreniform disorder therefore does not correspond to F23.2 (acute schizophrenia-like psychotic disorder) or F23.1 (acute polymorphic psychotic disorder with symptoms of schizophrenia).

The main difference between the schizophrenia concepts in the two systems is therefore that the ICD-10 definition of schizophrenia is purely descriptive and phenomenological, without taking into consideration course or social criteria. The duration of 1 month irrespective of treatment was primarily chosen to separate schizophrenia from acute, transient schizophrenia-like psychotic disorder and from substance-induced schizophreniform disorder, which usually remits within 1 month of abstinence from substance use. DSM-IV changed the active-phase duration from 1 week to 1 month for the same reason and in order to be consonant with ICD-10, but retained the chronic course criterion and added the social criterion. ICD-10's schizophrenia therefore seems to cover a broader area in spite of the more prominent stress on Schneiderian first-rank symptoms and the more narrow definition of bizarreness. The ICD-10 concept is further broadened by the extra subtypes, particularly by the specially defined "simple schizophrenia".

In ICD-10, schizotypal disorder (F21) is included among the schizophrenia-related disorders, whereas DSM-IV has it as a personality disorder. The diagnostic criteria sets essentially define the same condition.

For persistent delusional disorder (F22), ICD-10 requires a minimum duration of 3 months, whereas the similarly defined delusional disorder in DSM-IV only requires 1 months' duration and further allows tactile or olfactory hallucinations to be present if related to the delusional theme. ICD-10 has a subgroup of "other persistent delusional disorders" (F22.8), which include delusions accompanied by persistent hallucinatory voices or other schizophrenic symptoms that do not meet the criteria for schizophrenia. Durations shorter than 3 months are categorized under "acute and transient psychotic disorder" (F23).

The ICD-10 category of "acute and transient psychotic disorders" is defined by an acute onset of fully developed psychotic symptoms within 2 weeks, excluding delirious states, mood disorders and organic aetiology. They are subdivided into acute polymorphic psychotic disorder without or with symptoms of schizophrenia, acute schizophrenia-like psychotic disorder and other acute predominantly delusional psychotic disorder. Polymorphic disorders require at least two of the following: (a) emotional turmoil, (b) perplexity or misidentification and/or (c) markedly increased or decreased motility. In this way, they correspond to the cycloid psychosis of Kleist and Leonhard and to the *bouffée délirante* of the French traditional classification (Berner et al. 1992; Sartorius et al. 1990). A fifth character allows subtyping into disorders without and with associated acute stress, the latter corresponding to the concept of reactive psychosis. ICD-9's reactive confusion may now be classified as acute polymorphic psychotic disorder or among the other acute and transient psychotic disorders (F22.8) if the confusion is only transient and does not fulfil the symptomatic criteria for delirium. Some cases of reactive confusion, however, will more readily correspond to dissociative disorders such as dissociative fugue or trance and possession disorders. The

duration must not exceed 1 month for polymorphic psychotic disorder with symptoms of schizophrenia or acute schizophrenia-like psychotic disorder in order to distinguish the disorder from schizophrenia. The other acute and transient psychotic disorders have a duration limitation of 3 months to delineate them from delusional disorders. DSM-IV's brief psychotic disorder with psychotic symptoms of less than 1 months' duration may correspond to any ICD-10 acute and transient psychotic disorder of short duration (less than 1 month). DSM-IV also provides a subdivision relating to the presence of a marked stressor.

Schizoaffective disorders differ substantially in their definitions in the two systems. The ICD-10 definition requires the presence of a full affective syndrome of moderate or severe degree together with symptoms characteristic of schizophrenia for at least 2 weeks, including first-rank symptoms, bizarre delusions, disorganized speech or catatonic behaviour. Both the mood and the schizophrenic symptoms must be prominent and appear concurrently for at least a part of the episode. DSM-IV requires furthermore a period of at least 2 weeks with delusions or hallucinations in the absence of prominent mood symptoms. In DSM-IV, psychotic symptoms which occur only in the presence of a mood episode are diagnosed as a mood disorder with psychotic features, regardless of the characteristics of the psychotic symptoms. ICD-10's schizoaffective disorders are thus a much broader concept that includes many cases of DSM-IV mood disorders with psychotic features. The ICD-10 diagnostic guidelines mention that their relationship to typical mood disorders and schizophrenic disorders is uncertain, but that they are given a separate category because they are too common to be ignored. It further mentions that the cross-sectional diagnosis of schizoaffective disorder, appearing occasionally in a series of typical affective disorders, should not invalidate the diagnosis of bipolar disorder or recurrent depressive disorder.

4.3 Mood Disorders

The ICD-10 and DSM-IV systems both group together disorders with affective symptoms. They are primarily subdivided according to the nature of the symptoms into manic, depressive and mixed episodes. The episodes are further subdivided according to severity and the presence of psychotic symptoms. A manic episode requires in both systems the presence of expansive elevated or irritable mood for at least 1 week or less if severe enough to require hospital admission, accompanied by three manic symptoms, or four if the mood is irritable, out of a number of almost identical manic symptoms grouped slightly differently in the two systems. DSM-IV further requires severe interference with social functioning, whereas ICD-10 stresses interference with personal daily-life functioning. Hypomania is defined in DSM-IV by precisely the same symptoms, except that the required duration is only at least 4 days. The disorder should be distinct and observable by others, but should not cause severe interference with social functioning or necessitate hospitalization. ICD-10's hypomania requires elevated or irritable mood for at least 4 consecutive days accompanied by three symptoms among a number of hypomanic or manic symptoms, leaving out inflated self-esteem or grandiosity and flight of ideas and stressing that it should only lead to some interference with personal functioning in daily living. As to the presence of psychotic symptoms, ICD-10 includes delusions or hallucinations other than those listed as typically schizophrenic, i.e., first-rank symptoms and bizarre delusions, whereas DSM-IV allows any psychotic symptoms, regardless of their nature, as long as they are confined to the mood episode. This makes the DSM-IV concept much broader, covering a number of the schizoaffective disorders in ICD-10. Both systems have a further subdivision according to congruence of the psychotic symptoms with the mood state.

The depressive episode in ICD-10 and the major depressive episode in DSM-IV have different thresholds. ICD-10 mentions three nuclear depressive symptoms: (1) depressed mood, present for the most of the day, almost every day for at least 2 weeks, (2) loss of interest or pleasure and (3) decreased energy or increased fatigability. In addition to the nuclear symptoms, seven accessory or additional depressive symptoms are mentioned. To qualify as a depressive episode, four or more depressive symptoms are required, including at least two of the nuclear symptoms. DSM-IV has a different algorithm for major depressive episodes, requiring five or more depressive symptoms including either depressed mood or loss of interest or pleasure. Severity in ICD-10 is correlated with the number of depressive symptoms, whereas DSM-III relies on clinical judgement. Moderate or severe depression fulfils the criteria in both systems, but mild depressive episodes have a lower threshold in ICD-10. A linguistic problem with the use of the word "major" may contribute to the higher threshold in DSM-IV. In non-English-speaking countries, the word "major" might be understood as "severe", restricting the use of the diagnosis in clinical practice. Recommendations for anti-depressant treatment are furthermore often linked to "major depression" due to the extensive use of DSM diagnostics in psychopharmacological treatment literature, thus also including mild cases, recommendations which, on the other hand, make the pharmaceutical companies happy. As

to the presence of psychotic symptoms, the same differences apply as for the manic episodes with psychotic symptoms. Depressive episodes of mild or moderate severity in ICD-10 have a further subdivision according to the presence of a "somatic syndrome", corresponding to the melancholic features in DSM-IV.

A mixed affective episode in ICD-10 requires either a mixture or a rapid alternation of hypomanic–manic and depressive symptoms during an episode lasting for at least 2 weeks, whereas DSM-IV defines a mixed episode as one which meets the criteria for a manic episode and the symptomatic criteria for a major depressive episode nearly every day for at least a 1-week period, causing marked interference with social functioning or necessitating hospitalization.

The mood disorders in ICD-10 are further subdivided as to single hypomanic, manic, depressive or mixed affective episodes or recurrent episodes in the form of bipolar disorder or recurrent depressions. A bipolar disorder is defined by the presence of at least two episodes, of which one is hypomanic, manic or mixed. Episodes are limited from each other either by a shift of polarity or a period of at least 2 months' remission without prominent mood symptoms. DSM-IV includes single manic, hypomanic or mixed episodes in the bipolar disorders and thus has a more simplistic main division between bipolar and depressive disorders. DSM-IV has further subdivisions into bipolar I and bipolar II disorders, which are considered to be various combinations of affective syndromes in ICD-10's bipolar disorders. DSM-IV further subdivides by specifiers for catatonic or atypical features, post-partum onset, longitudinal course, seasonal pattern and rapid cycling, which are considered as variations in course and symptomatology in ICD-10. The presence of atypical or catatonic features, particularly if they reach a state of extreme catatonic excitement, previously described as acute delirium in psychotic disorders, may in ICD-10 lead to consider the use of the diagnostic category F3x.8 ("other").

In addition to the episodic disorders, both systems include persistent mood disorders, cyclothymia and dysthymia, which in DSM-IV are grouped together with bipolar and depressive disorders, respectively. Both systems require numerous episodes of hypomanic or mild depressive symptoms for cyclothymia or sub-threshold depressive symptoms alone for dysthymia with only short-lasting remissions, not exceeding 2 months, occurring within a total period of at least 2 years with no manic, mixed or depressive episodes, which in DSM-IV may occur later as superimposed on a persistent mood disorder. ICD-10 provides a list of hypomanic or mild depressive symptoms, of which patients should have at least three during some of the episodes as well as the elevated or depressed mood, but, as mentioned, not enough to fulfil the criteria for a manic or depressive episode. For dysthymic disorders, DSM-IV has a list of six depressive symptoms, of which only two or more should be present together with depressed mood.

Finally, ICD-10 includes recurrent brief depressive disorder in the category of "other mood disorders" (F38.1). DSM-IV has no equivalent, but includes the disorder in appendix B in the criteria sets and axes provided for further study. Both systems contain many subcategories, particularly DSM-IV with all its subspecifications (possibly too many, because extensive use of a very high number of subcategories may cause reliability problems, especially in research).

4.4 Anxiety Disorders

Anxiety in ICD-10 is always autonomic anxiety, i.e., requires the presence of at least one of the autonomic arousal symptoms (palpitations, sweating, trembling or dry mouth), in contrast to DSM-IV, which merely mentions anxiety, fear or discomfort without specific qualification. The difference may, however, be more theoretical than practical except in mild or very mild cases. ICD-10 subdivides the anxiety disorders into phobic disorders, including agoraphobia, social phobia and specific phobias, and other anxiety disorders, including panic disorder and generalized anxiety disorder.

Agoraphobia appears in situations or places in which it may be difficult to escape or get help, such as in crowds, public places, when travelling alone or away from home, and at least two of these situations should be implied. Social phobia appears in social situations with fear of being the focus of attention or of behaving embarrassing or humiliating. Specific phobias appear in the presence of specific objects or in certain situations other than the previously mentioned ones. The phobic disorders are manifested by fear or avoidance of the situations or objects, and they must have appeared at least once with a minimum of at least two anxiety symptoms, one being autonomic, out of a list of 14 anxiety symptoms, accompanied in social phobia by either blushing or shaking, fear of vomiting or of micturation or defecation. The individual must realize that the fear or avoidance is excessive or unreasonable. In addition to the presence of organic aetiology, exclusion criteria are those of psychotic disorders and mood disorders, i.e., the simultaneous presence of a depressive episode will give preference to the diagnosis of a depressive disorder. This is in contrast to the co-morbidity rule in DSM-IV.

Panic disorder is defined by the occurrence of at least four panic attacks in a 4-week period, each attack including at least four anxiety symptoms, of which one must be autonomic. The DSM-IV criteria for panic attacks also require four or more anxiety symptoms out of an almost identical list of symptoms, but they do not have to be autonomic, and for panic disorder, the attack only needs to be recurrent without a specified minimum number of attacks. In DSM-IV, the diagnosis of agoraphobia is linked together with panic disorder as the primary disorder which may appear together with or without agoraphobia. Agoraphobia can also appear alone (agoraphobia without history of panic disorder). ICD-10 deals with this the other way round, with a subdivision of agoraphobia according to the presence or absence of panic disorder. ICD-10's panic disorder has a subdivision into moderate and severe panic disorder, the severe form requiring four panic attacks per week over a 4-week period.

Generalized anxiety disorder in both systems requires a period of 6 months with prominent tension, worry and apprehension about everyday events and problems. In ICD-10, at least four anxiety and tension symptoms are required, one being autonomic, out of a list of 22 symptoms, including the previous 14 anxiety symptoms and an additional eight tension and nonspecific symptoms. This corresponds to six symptoms mentioned in DSM-IV, of which three are required to be present; however, only one symptom is required in children, for which ICD-10 has a specific category, "generalized anxiety disorder of childhood" (F93.80).

ICD-10 further includes a category for mixed anxiety and depressive disorder, which includes very mild subthreshold anxiety or depression with no specific criteria provided. This was prompted by a strong desire on the part of primary health care physicians. It has no equivalent in DSM-IV, but it is included in appendix B for further study.

4.5
Obsessive Compulsive Disorder

DSM-IV also includes obsessive–compulsive disorders among the anxiety disorders, whereas in ICD-10 they appear as an independent type of neurotic disorders (F42). ICD-10 requires obsessions or compulsions for at least 2 weeks and differentiates between predominantly obsessional, compulsive or mixed disorders. DSM-IV requires practically the same symptoms and characteristics, but stresses the relation to anxiety and distress, caused by the symptoms and by the individual's attempts to ignore, suppress or neutralize the obsessional thoughts or impulses by other thoughts or actions leading to compulsion. If these attempts are purely mental acts, the DSM-IV's "cognitive compulsions" are equivalent to ICD-10 "obsessions".

4.6
Reaction to Severe Stress and Adjustment Disorders

DSM-IV further includes acute stress reaction and post-traumatic stress disorder among the anxiety disorders, whereas ICD-10 places them in the subsection F43 (reaction to severe stress and adjustment disorders). The criteria for acute stress disorder differ markedly between the two systems. ICD-10 requires the presence of anxiety and tension symptoms, which may be sufficient for mild cases, whereas moderate and severe cases require at least two or four symptoms, respectively, of a list of seven symptoms: (1) withdrawal from social interaction, (2) narrowing of attention, (3) apparent disorientation, (4) anger or aggression, (5) despair or hopelessness, (6) inappropriate or purposeless overactivity and/or (6) uncontrollable and excessive grief. For severe cases, dissociative stupor may be present instead of the four symptoms. Symptoms must appear immediately after exposure to an exceptional mental or physical stressor (within 1 h) and should diminish after not more than 8 h, or not more than 48 h for continued stressors. DSM-IV does not require any time frames for onset or duration and requires three or more of five dissociative symptoms and repeated experiences of the traumatic event with avoidance of stimuli that arouse recollections of the trauma, in addition to symptoms of anxiety or increased arousal. For post-traumatic stress disorder, the criteria are quite similar, although they are presented in a different way, the main difference being that ICD-10 may substitute persistent symptoms of increased arousal by inability to recall important aspects of the traumatic exposure and that ICD-10 does not require a minimum duration of 1 month as DSM-IV does.

Adjustment disorder has a separate section in DSM-IV just before the personality disorders. ICD-10's adjustment disorder requires onset of symptoms within 1 month of exposure to an identifiable psychosocial stressor, not of an unusual or catastrophic type, in contrast to DSM-IV, which requires the onset of symptoms within 3 months of any stressor so long as the criteria for acute stress disorder or post-traumatic stress disorder are not met. Both systems mention depressive, anxiety and behavioural symptoms, which must not meet the criteria for any other disorders. ICD-10 has a duration limitation of 6 months, except for brief depressive reaction, which should not exceed 1 month, and prolonged depressive reaction with a

duration not exceeding 2 years. DSM-IV gives a subspecification into acute and chronic adjustment disorder according to whether the duration was less or more than 6 months.

4.7 Somatoform Disorders

Somatoform disorders in both systems include somatization disorder, undifferentiated somatoform disorder, hypochondriacal disorder and pain disorder. ICD-10's somatization disorder requires a history of at least 2 years of complaints of multiple and variable physical symptoms that cannot be explained by a detectable physical disorder, with preoccupation with the symptoms causing distress and leading to repeated consultation or investigations, with a persistent refusal to accept medical reassurance. A total number of at least six symptoms from two organ groups are required out of a total of 14 gastrointestinal, cardiovascular, genitourinary or skin and pain symptoms. The preoccupation should be with symptoms alone and not with a belief of any serious physical disease as in hypochondriacal disorder. DSM-IV requires a combination of four pain symptoms, two gastrointestinal symptoms, one sexual symptom and one pseudo-neurological symptom occurring over a period of several years beginning before the age of 30, resulting in consultations for treatment or impaired social functioning. As in ICD-10, the symptoms and their effects cannot be explained by a known general medical condition or as the effect of any substance or medication. DSM-IV further mentions that they must not be intentionally produced or feigned.

Undifferentiated somatoform disorder has almost identical criteria in the two systems, only requiring a duration of at least 6 months and a lower number of somatoform symptoms.

Hypochondriacal disorder in ICD-10 corresponds to DSM-IV's hypochondriasis and body dysmorphic disorder, and they have similar criteria.

Pain disorder, referred to in ICD-10 as persistent somatoform pain disorder, requires a period of at least 6 months' persistent, severe and distressing, unexplained pain, which has become the main focus of the patient's attention. DSM-IV is more circumscript in the pain disorder criteria, requiring that psychological factors are judged to play an important role and that the pain symptoms are not intentionally produced or feigned. DSM-IV provides a subspecification into pain disorder associated with psychological factors and pain disorder associated with both psychological factors and a general medical condition and makes a further subdivision into acute and chronic pain disorders with a duration of less or more than 6 months.

ICD-10 has a category of somatoform autonomic dysfunction for somatoform symptoms accompanied by autonomic arousal, attributed to a physical disorder in at least one out of five organ systems: the cardiovascular system, the upper and lower gastrointestinal system, the respiratory system and the genito-urinary system. DSM-IV has no specific equivalent, but includes the disorder under "undifferentiated somatoform disorder".

DSM-IV includes among the somatoform disorders conversion disorder, which in ICD-10 is included among the dissociative disorders. Conversion disorder corresponds to ICD-10's dissociative motor disorders, dissociative convulsions, dissociative anaesthesia and sensory loss, and mixed dissociative disorders, with almost the same diagnostic criteria except that for the main group of dissociative disorders ICD-10 requires a convincing temporal association between the onset of symptoms and stressful events, problems or needs. DSM-IV stresses a delimitation from intentionally produced or feigned symptoms.

Among the dissociative disorders, ICD-10 further includes dissociative amnesia, fugue, trance and possession disorders, and other dissociative disorders with Ganser's syndrome and multiple personality disorder. The requirement of association with a psychosocial stressor is general for all dissociative disorders, and the term "dissociative disorder" is defined as a partial or complete loss of the normal integration between memories of the past, awareness of identity and immediate sensations and control of bodily movements. Dissociative disorders in DSM-IV are defined as a dissociation of the integrated functions of consciousness, memory, identity and perception of the environment, but not including control of bodily movements. DSM-IV's dissociative disorders include dissociative amnesia, fugue, identity disorder (formerly multiple personality disorder) and depersonalization disorder, the equivalent of depersonalization–derealization syndrome (F48.1) in ICD-10. The criteria for the subtypes are quite similar in the two systems.

ICD-10 includes one further "neurotic" disorder diagnosed in various parts of the world, i.e., neurasthenia, for which DSM-IV has no specific equivalent but includes the disorder under undifferentiated somatoform disorder.

As far as the dissociative disorders are concerned, DSM-IV has a section on factitious disorders, which in ICD-10 are placed in a rest group of other disorders of adult personality and behaviour (F68), with the long title of "intentional production or feigning of symptoms or disabilities either physical or psychological (factitious disorder)" (F68.1). However, both systems use almost identical criteria sets.

4.8
Eating Disorders

ICD-10 includes eating disorders under the heading "behavioural syndromes associated with physiological disturbances and physical factors" (F5). Eating disorders are subdivided into anorexia nervosa, atypical anorexia nervosa, bulimia nervosa, atypical bulimia nervosa and overeating or vomiting associated with other psychological disturbances. DSM-IV differentiates between only three categories: anorexia nervosa, bulimia nervosa and eating disorder not otherwise specified. For anorexia nervosa, the main criteria are similar, but ICD-10 specifically requires that weight loss must be self-induced by avoidance of fattening foods and excludes bulimic episodes, in contrast to DSM-IV, which give anorexia nervosa precedence over bulimia nervosa, with a subspecification into the restricting type and the binge-eating/purging type. For bulimia nervosa, ICD-10 excludes anorexia nervosa, but otherwise has almost the same criteria. Combinations of anorexia and bulimia are included under atypical bulimia nervosa in ICD-10, whereas atypical anorexia nervosa corresponds to DSM-IV's eating disorder not otherwise specified The classification of eating disorders in ICD-10 is one of the subclassifications which has caused most dissatisfaction, and psychiatrists treating these disorders seem to prefer the DSM-IV classification.

4.9
Sleep Disorders

Sleep disorders are defined in essentially the same way in the two systems, except that DSM-IV includes narcolepsy and breathing-related sleep disorder in the classification, which in ICD-10 are classified in Chap. VI (diseases of the nervous system).

4.10
Sexual and Gender Identity Disorders

Sexual and gender identity disorders are grouped together in DSM-IV, whereas in ICD-10 they are divided into non-organic sexual dysfunction (F52), gender identity disorder (F64) and disorders of sexual preference (F65). Sexual dysfunction has the same subcategories and similar diagnostic criteria, except that ICD-10 requires a minimum duration of 6 months and provides separate criteria for males and females when appropriate. DSM-IV has subspecifications according to whether the disorder is lifelong or acquired, generalized or situational and whether psychological or combined factors are present. DSM-IV's sexual dysfunction due to a general medical disorder or to substance use is included in ICD-10 under other specified mental disorders due to brain damage and dysfunction and to physical disease (F06.8) or under other mental or behavioural disorders due to psychoactive substance use (F1x.8), both with associated codes for particular syndromes from Chap. XIV in ICD-10 (diseases of the genito-urinary system).

Gender identity disorder is subdivided in ICD-10 into transsexualism, dual-role transvestism and gender identity disorder of childhood, with differentiated diagnostic criteria contained in the criteria for the single DSM-IV category "gender identity disorder".

Disorders of sexual preference, referred to in DSM-IV as paraphilias, are defined and subcategorized in similar ways in the two systems. DSM-IV includes frotteurism and subdivides sadomasochism into sexual masochism and sexual sadism.

ICD-10's psychological and behavioural disorders associated with sexual development and orientation, including sexual maturation disorder, egodystonic sexual orientation and sexual relationship disorder (F66) have no specific equivalents in DSM-IV, which includes them under sexual disorder not otherwise specified.

4.11
Personality Disorders

The general diagnostic criteria for personality disorders are almost identical in the two systems, requiring an inflexible, pervasive, enduring pattern of deviating inner experience and behaviour affecting two or more of the areas of cognition, affectivity, impulse control and interpersonal functioning, leading to distress or impaired social functioning, in ICD-10 also including adverse impact on the social environment. They are lifelong with onset in late childhood or adolescence and are therefore difficult to diagnose before the age of 18, except for antisocial personality disorder, which often is a continuation of childhood conduct disorder. The criteria are restrictive and demanding, and the category is not easy to use in a clinical setting. They are subdivided into specific personality disorders, which in addition require a number of typical personality traits for each disorder. DSM-IV presents the specific disorders in three clusters: Cluster "A" includes paranoid, schizoid and schizotypal personality disorder, of which the latter is grouped with schizophrenia and related disorders in ICD-10. Cluster "B" with antisocial, borderline, histrionic and narcissistic personality disorder corresponding to ICD-10's dyssocial, histrionic and emotionally unstable personality disorder of borderline type. ICD-10 has no equivalent to narcissistic personality disorder, which, however, is included in an

annex with provisional (DSM-IV) criteria for selected disorders. DSM-IV cluster "C" includes avoidant, dependent and obsessive-compulsive personality disorder, corresponding to ICD-10's anxious, dependent and anankastic personality disorders. ICD-10 further includes emotionally unstable personality disorder of impulsive type, which has no equivalent in DSM-IV. The corresponding personality disorders have differently formulated criteria in the two systems, but nevertheless essentially define the same conditions. ICD-10 further includes subcategories for mixed personality disorder with a mixture of features not fulfilling the criteria for any specific personality disorders; troublesome personality changes regarded as secondary to co-existing affective or anxiety disorders; and enduring personality change after catastrophic experiences or after psychiatric illness due to the impact of the catastrophic events or the experience of psychiatric illness from which the patient has recovered completely without residual symptoms. DSM-IV has no equivalents, but includes these disorders under personality disorder not otherwise specified.

4.12 Other Disorders

Finally, DSM-IV has a section with impulse-control disorders not elsewhere classified, including intermittent explosive disorder, kleptomania, pyromania, pathological gambling and trichotillomania. The corresponding category in ICD-10 (F63, habit and impulse disorders) includes pathological gambling, fire-setting, stealing and trichotillomania. Pathological gambling is presented differently in ICD-10, requiring two or more episodes within a period of 1 year, with mental preoccupation with gambling and an intense urge to gamble which is difficult to control, accompanied with lack of profitable outcome, personal distress and interference with personal functioning. DSM-IV requires at least five out of a total number of ten characteristic features which roughly reflect the same condition as defined in ICD-10. The other impulse disorders have essentially the same criteria. Intermittent explosive disorder has no specific equivalent in ICD-10, but has to be included among other habit and impulse disorders. If lifelong, it may correspond to ICD-10's emotionally unstable personality disorder of impulsive type.

ICD-10 finally includes a few categories with no specific equivalents in DSM-IV: mental and behavioural disorders associated with the puerperium, not elsewhere classified (F53), psychological and behavioural factors associated with disorders or diseases classified elsewhere (F54) and elaboration of physical symptoms for psychological reasons (F68.0).

4.13 Disorders of Infancy, Childhood or Adolescence

The disorders of infancy, childhood or adolescence in both systems include mental retardation, development disorders and behavioural and emotional disorders. Developmental disorders (ICD-10 F8) include specific disorders affecting various areas of speech, language, motor or scholastic skills and pervasive developmental disorders such as autistic disorder, atypical autism, Rett's syndrome and other childhood disintegrative disorder, covering corresponding diagnostic categories in the two systems.

ICD-10's behavioural and emotional disorders with onset usually occurring in childhood and adolescence (F9) include hyperkinetic disorders, conduct disorders, emotional disorders, social functioning disorders and other disorders with onset specific to childhood or adolescence. In DSM-IV, attention deficit or hyperactivity disorder covers a broader area, including cases with attention deficit alone. DSM-IV conduct disorder has almost the same diagnostic criteria as ICD-10 conduct disorder, but a different subcategorization into childhood-onset or adolescence-onset type and subspecification according to severity, whereas ICD-10 subdivides the disorder into conduct disorder confined to the family context, unsocialized and socialized conduct disorder and further includes oppositional defiant disorder, which is separately categorized in DSM-IV.

Among the emotional disorders, ICD-10 provides criteria for separation anxiety disorder similar to the corresponding DSM-IV category. It further provides diagnostic criteria for the subcategories phobic social disorder, generalized anxiety disorder of childhood and sibling rivalry disorder.

The ICD-10 disorders of social functioning include elective mutism and reactive and disinhibited attachment disorder of childhood, which correspond to DSM-IV's selective mutism and reactive attachment disorder of infancy or early childhood.

Finally, both categories include a number of other childhood psychiatric disorders such as tic disorders, feeding or elimination disorders and speech disorders.

ICD-10 also has a nonspecific residual category (F99, mental disorder not otherwise specified) for disorders which do not fulfil any specific criteria and cannot even be placed in the main section's subcategories for unspecified disorders. DSM-IV has no equivalent, but may use unspecified mental disorder for non-psychotic disorders.

DSM-IV on the other hand has a section for other conditions that may be a focus of clinical attention, including psychological factors affecting medical con-

ditions, which corresponds to ICD-10's psychological and behavioural factors associated with disorders or diseases classified elsewhere (F54). DSM-IV further includes medication-induced movement disorders or adverse effects of medication corresponding to ICD-10 conditions from the neurological chapter on problems related to abuse or neglect, other relational problems, conditions or factors with corresponding specific T or Z categories in ICD-10. DSM-IV provides additional codes for deferred or provisional diagnoses indicating diagnostic uncertainty.

4.14 Multiaxial Presentation

DSM-IV has continued the multiaxial presentation introduced by DSM-III with five axes: axis I for clinical disorders, axis II for personality disorders and mental retardation, axis III for general medical conditions, axis IV for psychosocial and environmental problems and axis V for global assessment of functioning (American Psychiatric Association 1994). ICD-10 has introduced a corresponding multiaxial presentation for adult psychiatric disorders with three axes: axis I for clinical diagnoses of any type, including personality disorders, developmental disorders or mental retardation and general medical conditions, axis II for impairments, disabilities and handicaps due to mental disorders affecting various areas of social functioning and, finally, axis III containing a number of contextual factors influencing health status and contact with health services, derived from the ICD-10 Z categories and comprising problems related to social environment and circumstances, negative life events in childhood, problems related to upbringing or to primary support groups and problems related to lifestyle or difficulties in life management (World Health Organization 1997).

No criteria are provided for axis II or III. For axis II, the WHO revision of the *International Classification of Impairments Disabilities and Handicaps* is awaited.

For child psychiatry, a *Multiaxial Classification of Child and Adolescent Psychiatric Disorders* is provided with six axes: clinical psychiatric syndromes, developmental disorders, intellectual level, physical disorders, psychosocial factors and global assessment of psychosocial disability (World Health Organization 1996).

5 Diagnostic Instruments

In collaboration with the U.S. National Institutes of Health (NIH), WHO has provided instruments for the assessment of symptoms required for the diagnostic criteria in both the ICD-10 and the DSM systems. Schedules for Clinical Assessment in Neuropsychiatry (SCAN; World Health Organization 1992c) and the International Personality Disorder Examination (IPDE; Loranger et al. 1997; World Health Organization 1992d) cover the majority of adult psychiatric disorders and are designed for use in clinical research with computer-assisted diagnostic processing. In epidemiological research, the Composite International Diagnostic Interview (CIDI; World Health Organization 1990, 1993b) is a fully structured interview schedule to be used by lay interviewers for both ICD-10 and DSM diagnoses. The schedules have been updated to also cover DSM-IV diagnoses (World Health Organization 1994, 1995, 1997).

6 ICD-10 and DSM-IV General Status

The majority of mental and behavioural disorders are included in both classification systems, although for some disorders with different systematic distribution and under different headings. A few categories appear only in one of the classifications, and not in the other. Thus ICD-10 includes mild cognitive disorder, post-concussional syndrome, simple schizophrenia, post-schizophrenic depression, recurrent brief depressive disorder, somatoform autonomic dysfunction, neurasthenia, enduring personality change after catastrophic experience or psychiatric illness and generalized anxiety disorder of childhood. They have no equivalents in DSM-IV, which, however, has placed several of them in the appendix B in the criteria sets and axes provided for further study under slightly different headings. DSM-IV, on the other hand, includes a variety of subspecifications regarding course and special features in the mood disorders, narcolepsy, breathing-related sleep disorder, intermittent explosive disorder and narcissistic personality disorder and a special section for other conditions that may be a focus of clinical attention, including medication-induced movement disorders, relational problems, problems related to abuse or neglect and other additional problems such as non-compliance with treatment, malingering, borderline intellectual functioning, age-related cognitive decline, bereavement, acculturation problems, phase of life problems and religious or spiritual problems, which are not included in ICD-10's mental and behavioural disorder section, but may find equivalent codes in other ICD-10 chapters.

The diagnostic criteria for the majority of categories are almost identical or at least quite similar, defining essentially the same conditions. For some disorders,

however, they may differ substantially as to the threshold for inclusion in the diagnostic category in question. This applies to important main categories such as schizophrenia and major depressive episode. For moderate or severe disorders, however, the correspondence is almost complete, which is the case for the majority of patients in clinical psychiatry. For disorders of mild severity, just above the threshold for inclusion in the diagnostic categories, differences may be substantial and will primarily affect epidemiological research or primary health care. The DSM-IV use of social criteria may contribute to this differentiation.

ICD-10 and DSM-IV both rely on demonstrated satisfactory reliability for main categories and subcategories. Further validation is still needed to provide evidence for which criteria sets have the most valid delineations with the most valid thresholds for inclusion. The WHO ICD-10 Advisory Committee has called for a moratorium of at least 10 years before the next round of revisions begin in order to allow time for such validation studies. Classificatory areas in special need of validation would be schizoaffective disorders in comparison with schizophrenia and psychotic mood disorders, some of the special ICD-10 categories such as mild cognitive disorder, simple schizophrenia and neurasthenia, and the specific personality disorders. Candidates for validation studies in DSM-IV are found in appendix B (criteria sets and axes provided for further study), including some of the categories corresponding to the special ICD-10 categories. The major psychiatric journals should be encouraged to accept contributions using either of the two classification systems or both, in order to promote publication of validation studies essential for future revisions of both systems. WHO has not yet achieved its goal of one common language in international psychiatry. Future revisions will hopefully approximate the classifications further with the ultimate goal of having one common classification system to the benefit of psychiatric patients all over the world.

7 References

American Psychiatric Association (1968) Diagnostic and statistical manual of mental disorders, 2nd edn. American Psychiatric Association, Washington, DC

American Psychiatric Association (1980) Diagnostic and statistical manual of mental disorders, 3rd edn (DSM-III). American Psychiatric Association, Washington, DC

American Psychiatric Association (1987) Diagnostic and statistical manual of mental disorders, 3rd edn revised (DSM-III-R). American Psychiatric Association, Washington, DC

American Psychiatric Association (1994) Diagnostic and statistical manual of mental disorders, 4th edn (DSM-IV). American Psychiatric Association, Washington, DC

American Psychiatric Association (1995) Diagnostic and statistical manual of mental disorders, 4th edn (DSM-IV). International version with ICD-10 codes. American Psychiatric Association, Washington, DC

Berner P, Gabriel E, Katschnig H, Kieffer W, Koehler K, Lenz G, Nutzinger D, Schanda H, Simhandl C (1992) Diagnostic criteria for functional psychoses. Cambridge University Press, Cambridge

Cooper J, Kendell RE, Gurland BJ, Sharpe L, Copeland JRM, Simon R (1972) Psychiatric diagnosis in New York and London. Institute of Psychiatry, Maudsley Monographs 20. Oxford University Press, New York

Cooper J (1988) The structure and presentation of contemporary psychiatric classifications with special reference to ICD-9 and ICD-10. Br J Psychiatry 152[Suppl 1]: 21–28

Feighner JP, Robins E, Guze SB, Woodruff RA, Winokur G, Munoz R (1972) Diagnostic criteria for use in psychiatric research. Arch Gen Psychiatry 26: 57–63

Jablensky A (1988) Methodological issues in psychiatric classification. Br J Psychiatry 152[Suppl 1]: 15–20

Jaspers K (1973) Allgemeine Psychopathologie, 9th edn. Springer, Berlin Heidelberg New York

Kramer M, Sartorius N, Jablensky A, Gulbinat W (1979) The ICD-9 classification of mental disorders: a review of its development and contents. Acta Psychiatr Scand 59: 241–262

Loranger AW, Janca A, Sartorius N (1997) Assessment and diagnosis of personality disorders. Cambridge University Press, Cambridge

Sartorius N (1988) International perspectives of psychiatric classification. Br J Psychiatry 152[Suppl 1]: 9–14

Sartorius N (1991) The classification of mental disorders in the tenth revision of the International Classification of Diseases. Eur Psychiatry 6: 315–322

*Sartorius N (1993) WHO's work on epidemiology of mental disorders. Soc Psychiatry Psychiatr Epidemiol 28: 147–155

Sartorius N (1995) Understanding the ICD-10 Classification of Mental Disorders. A pocket reference. Science, London

**Sartorius N, Jablensky A, Regier DA, Burke JP, Hirschfeld RMA (eds) (1990) Sources and traditions of classification in psychiatry. Hogrefe and Huber, Toronto

*Sartorius N, Kaelbe CT, Cooper JE, Roper MT, Rae DS, Gublinat W, Üstün TB, Regier DA (1993) Progress toward achieving a common language in psychiatry. Results from the field trials of the Clinical Guidelines accompanying the WHO Classification of Mental and Behavioural Disorders in ICD-10. Arch Gen Psychiatry 50: 115–124

Sartorius N, Üstün TB, Korten A, Cooper J, Drimmelen J van (1995) Progress toward achieving a common language in psychiatry II. Results from the international field trials of the ICD-10 Diagnostic Criteria for Research for mental and behavioural disorders. Am J Psychiatry 152(10): 1427–1437

Schneider K (1967) Klinische Psychopathologie. Thieme, Stuttgart

Spitzer RL, Endicott J, Robins E (1977) Research diagnostic criteria for a selected group of functional disorders, 3rd edn. New York State Psychiatric Institute, New York

Stengel E (1959) Classification of mental disorders. Bull World Health Org 21: 601–663

World Health Organization (1965) Manual of the international statistical classification of diseases, injuries and

causes of death, 8th revision. World Health Organization, Geneva

World Health Organization (1974) Glossary of mental disorders and guide to their classification, for use in conjunction with the international classification of diseases, 8th revision. World Health Organization, Geneva

World Health Organization (1975) International classification of diseases, 9th revision. World Health Organization, Geneva

World Health Organization (1978) Mental disorders: glossary and guide to their classification in accordance with the ninth revision of the international classification of diseases. World Health Organization, Geneva

World Health Organization (1990) The composite international diagnostic interview (CIDI) core version 1.0. World Health Organization, Geneva

World Health Organization (1992a) International statistical classification of diseases and related health problems, tenth revision, vol I. World Health Organization, Geneva

World Health Organization (1992b) The ICD-10 classification of mental and behavioural disorders, clinical descriptions and diagnostic guidelines. World Health Organization, Geneva

World Health Organization (1992c) Schedules for clinical assessment in neuropsychiatry (SCAN). World Health Organization, Geneva

World Health Organization (1992d) The international personality disorder examination (IPDE). World Health Organization, Geneva

World Health Organization (1993a) The ICD-10 classification of mental and behavioural disorders, diagnostic criteria for research. World Health Organization, Geneva

World Health Organization (1993b) The composite international diagnostic interview (CIDI) core version 1.1. American Psychiatric Press, Washington, DC

World Health Organization (1994) Schedules for clinical assessment in neuropsychiatry (SCAN) version 2.0. Annual Psychiatric Press, Washington, DC

World Health Organization (1995) The composite international diagnostic interview (CIDI) core version 2.0. World Health Organization, Geneva

World Health Organization (1996) Multiaxial classification of child and adolescent psychiatric fisorders. Cambridge University Press, Cambridge

World Health Organization (1997) The multiaxial presentation of the ICD-10 for use in adult psychiatry. Cambridge University Press, Cambridge

CHAPTER

4

J. van Drimmelen-Krabbe, A. Bertelsen, C. Pull

Conversion Tables for ICD-10 and DSM-IV

Table 1. Conversion table: ICD-10 to DSM-IV (ICD-9-CM)

ICD-10	DSM-IV (ICD-9-CM)
F00–F99 Organic, including symptomatic, mental disorders	
A fifth character may be added to specify dementia in F00–F03, as follows: 0 Without additional symptoms 1 Other symptoms, predominantly delusional 2 Other symptoms, predominantly hallucinatory 3 Other symptoms, predominantly depressive 4 Other mixed symptoms	
F00 Dementia in Alzheimer's disease	(For all subtypes also code 331.0 Alzheimer's disease on axis III)
F00.0 Dementia in Alzheimer's disease with early onset (also code G30.0 Alzheimer's disease with early onset)	290.xx Dementia of the Alzheimer's type, with early onset .10 Uncomplicated .12 With delusions .13 With depressed mood
F00.1 Dementia in Alzheimer's disease with late onset (also code G30.1 Alzheimer's disease with late onset)	290.xx Dementia of the Alzheimer's type, with late onset .0 Uncomplicated .20 With delusions .21 With depressed mood
F00.2 Dementia in Alzheimer's disease, atypical or mixed type	No equivalent
F00.9 Dementia in Alzheimer's disease, unspecified	No equivalent
F01 Vascular dementia	290.xx Vascular dementia .40 Uncomplicated .42 With delusions .43 With depressed mood (also code vascular condition on axis III; no specifiers for subtypes)
F01.0 Vascular dementia of acute onset	
F01.1 Multi-infarct dementia	
F01.2 Subcortical vascular dementia	
F01.3 Mixed cortical and subcortical vascular dementia	
F01.8 Other vascular dementia	
F01.9 Vascular dementia, unspecified	
F02 Dementia in other diseases classified elsewhere	
F02.0 Dementia in Pick's disease (also code G31.0 Pick's disease)	290.10 Dementia due to Pick's disease (also code 331.1 Pick's disease on axis III)
F02.1 Dementia in Creutzfeldt-Jakob disease (also code A81.0 Creutzfeldt-Jacob disease)	290.10 Dementia due to Creutzfeldt-Jakob disease (also code 046.1 Creutzfeldt-Jakob disease on axis III)
F02.2 Dementia in Huntington's disease (also code G10 Huntington's disease)	294.1 Dementia due to Huntington's disease (also code 333.4 Huntington's disease on axis III)
F02.3 Dementia in Parkinson's disease (also code G20 Parkinson's disease)	294.1 Dementia due to Parkinson's disease (also code 332.0 Parkinson's disease on axis III)
F02.4 Dementia in human immunodeficiency virus (HIV) disease (also code B22.0 HIV dementia)	294.9 Dementia due to HIV disease (also code 043.1 HIV infection affecting central nervous system on axis III)
F02.8 Dementia in other specified diseases classified elsewhere	294.1 Dementia due to other general medical conditions (also code aetiological code on axis III)
F03 Unspecified dementia	294.8 Dementia NOS
F04 Organic amnesic syndrome, not induced by alcohol and other psychoactive substances	294.0 Amnestic disorder due to a general medical condition
F05 Delirium, not induced by alcohol and other psychoactive substances	
F05.0 Delirium, not superimposed on dementia, so described	293.0 Delirium due to a general medical condition
F05.1 Delirium, superimposed on dementia	290.11 Dementia of the Alzheimer's type with early onset, with delirium 290.3 Dementia of the Alzheimer's type with late onset, with delirium 290.41 Vascular dementia with delirium
F05.8 Other delirium	–.– Delirium due to multiple aetiologies
F05.9 Delirium, unspecified	780.09 Delirium NOS

ICD-10	DSM-IV (ICD-9-CM)
F06 Other mental disorders due to brain damage and dysfunction and to physical disease	
F06.0 Organic hallucinosis	293.82 Psychotic disorder due to a general medical condition, with hallucinations
F06.1 Organic catatonic disorder	293.89 Catatonic disorder due to a general medical condition
F06.2 Organic delusional (schizophrenia-like) disorder	293.81 Psychotic disorder due to a general medical condition, with delusions
F06.3 Organic mood (affective) disorders	293.83 Mood disorder due to a general medical condition
.30 Organic manic disorder	- With manic features
.31 Organic bipolar disorder	- With depressive features
.32 Organic depressive disorder	- With major depressive-like episode
.33 Organic mixed affective disorder	- With mixed features
F06.4 Organic anxiety disorder	293.89 Anxiety disorder due to a general medical condition
F06.5 Organic dissociative disorder	*No equivalent, use*: 293.9 Mental disorder NOS due to a general medical condition
F06.6 Organic emotionally labile (asthenic) disorder	*No equivalent, use*: 293.9 Mental disorder NOS due to a general medical condition
F06.7 Mild cognitive disorder	294.9 Cognitive disorder NOS (*see also Appendix B*: Mild neurocognitive disorder)
	-.- Sexual dysfunction due to a general medical condition
F06.8 Other specified mental disorders due to brain damage and dysfunction and to physical disease	780.5x Sleep disorder due to a general medical condition
F06.9 Unspecified mental disorder due to brain damage and dysfunction and to physical disease	293.9 Mental disorder not otherwise specified due to a general medical condition
F07 Personality and behavioural disorders due to brain disease, damage and dysfunction	
F07.0 Organic personality disorder	310.1 Personality change due to a general medical condition
F07.1 Post-encephalitic syndrome	294.4 Cognitive disorder, NOS
F07.2 Post-concussional syndrome	294.4 Cognitive disorder, NOS (*see also Appendix B*: Post-concussional disorder)
F07.8 Other organic personality and behavioural disorders due to brain disease, damage and dysfunction	310.1 Personality change due to a general medical condition, other type
F07.9 Unspecified organic personality and behavioural disorder due to brain disease, damage and dysfunction	310.1 Personality change due to a general medical condition, unspecified
F09 Unspecified organic or symptomatic mental disorder	293.9 Mental disorder not otherwise specified due to a general medical condition
F10–F19 Mental and behavioural disorders due to psychoactive substance use	
F10.– Mental and behavioural disorders due to use of alcohol	Alcohol-related disorders
F11.– Mental and behavioural disorders due to use of opioids	Opioid-related disorders
F12.– Mental and behavioural disorders due to use of cannabinoids	Cannabis-related disorders
F13.– Mental and behavioural disorders due to use of sedatives or hypnotics	Sedative-, hypnotic- or anxiolytic-related disorders
F14.– Mental and behavioural disorders due to use of cocaine	Cocaine-related disorders
F15.– Mental and behavioural disorders due to use of other stimulants, including caffeine	Amphetamine (or amphetamine-like)- and caffeine-related disorders
F16.– Mental and behavioural disorders due to use of hallucinogens	Hallucinogen-related disorders
F17.– Mental and behavioural disorders due to use of tobacco	Nicotine-related disorders
F18.– Mental and behavioural disorders due to use of volatile solvents	Inhalant-related disorders
F19.– Mental and behavioural disorders due to multiple drug use and use of other psychoactive substances	Phencyclidine (or phencyclidine-like)-related disorders Polysubstance-related disorder Other (or unknown) substance-related disorders

ICD-10	DSM-IV (ICD-9-CM)
Four- and five-character categories may be used to specify the clinical conditions, as follows:	
.0 Acute intoxication	303.00 Alcohol intoxication 305.90 Caffeine intoxication 292.89 Amphetamine, cannabis, cocaine, hallucinogen, inhalant, opioid, phencyclidine, sedative, hypnotic, anxiolytic or other (or unknown) substance intoxication
.00 Uncomplicated	No equivalent
.01 With trauma or other bodily injury	No equivalent
.02 With other medical complications	No equivalent
.03 With delirium	291.0 Alcohol intoxication delirium 292.81 Amphetamine, cannabis, cocaine, hallucinogen, inhalant, opioid, phencyclidine, sedative, hypnotic, anxiolytic or other (or unknown) substance intoxication delirium
.04 With perceptual distortions	292.89 Amphetamine, cannabis, cocaine, opioid, phencyclidine or other (or unknown) substance intoxication with perceptual disturbances
.05 With coma	No equivalent
.06 With convulsions	No equivalent
.07 Pathological intoxication	No equivalent
.1 Harmful use	305.00 Alcohol abuse 305.70 Amphetamine abuse 305.20 Cannabis abuse 305.60 Cocaine abuse 305.30 Hallucinogen abuse 305.50 Opioid abuse 305.40 Sedative, hypnotic or anxiolytic abuse 305.90 Inhalant, phencyclidine and other (or unknown) substance abuse
.2 Dependence syndrome	303.90 Alcohol dependence
.20 Currently abstinent	305.10 Nicotine dependence
.21 Currently abstinent, but in a protected environment	304.40 Amphetamine dependence
.22 Currently on a clinically supervised maintenance or replacement regime (controlled dependence)	304.30 Cannabis dependence 304.20 Cocaine dependence
.23 Currently abstinent, but receiving treatment with aversive or blocking drugs	304.60 Inhalant dependence 304.20 Opioid dependence
.24 Currently using the substance (active dependence)	304.50 Hallucinogen dependence
.25 Continuous use	304.90 Phencyclidine dependence
.26 Episodic use (dipsomania)	304.10 Sedative, hypnotic or anxiolytic dependence 304.80 Polysubstance dependence 304.90 Other (or unknown) substance dependence *Specifiers*: – With physiological dependence – Without physiological dependence – Early full remission – Early partial remission – Sustained full remission – Sustained partial remission – On agonist therapy – In a controlled environment
.3 Withdrawal state	291.8 Alcohol withdrawal 292.0 Cocaine, amphetamine, nicotine, opioid, sedative, hypnotic, anxiolytic or other (or unknown) substance withdrawal
.30 Uncomplicated	No equivalent
.31 With convulsions	No equivalent
.4 Withdrawal state with delirium	291.0 Alcohol withdrawal delirium 292.81 Sedative, hypnotic or anxiolytic withdrawal delirium 292.81 Other (or unknown) substance-induced delirium

ICD-10	DSM-IV (ICD-9-CM)
.40 Without convulsions	No equivalent
.41 With convulsions	No equivalent
.5 Psychotic disorder	291.x Alcohol-induced psychotic disorder
.51 Predominantly delusional	.5 With delusions
.52 Predominantly hallucinatory	.3 With hallucinations
.50 Schizophrenia-like	292.xx Amphetamine-, cannabis-, cocaine-, hallucinogen-,
.53 Predominantly polymorphic	inhalant-, opioid-, phencyclidine-, sedative-, hypnotic-, anxiolytic- or other (or unknown) substance-induced psychotic disorder
	.11 With delusions
	.12 With hallucinations
.54 Predominantly depressive symptoms	291.8 Alcohol-induced mood disorder or
.55 Predominantly manic symptoms	292.84 Amphetamine-, cannabis-, cocaine-, hallucinogen-,
.56 Mixed	inhalant-, opioid-, phencyclidine-, sedative-, hypnotic-, anxiolytic- or other (or unknown) substance-induced mood disorder
	– With depressive features
	– With manic features
	– With mixed features
.6 Amnesic syndrome	291.1 Alcohol-induced persisting amnestic disorder
	292.82 Sedative-, hypnotic-, anxiolytic or other (or unknown) substance-induced persisting amnestic disorder
.7 Residual and late-onset psychotic disorder	
.70 Flashbacks	292.89 Hallucinogen persisting perception disorder
.71 Personality or behaviour disorder	No equivalent
.72 Residual affective disorder	292.84 Amphetamine-, cannabis-, cocaine-, hallucinogen-, inhalant-, opioid-, phencyclidine-, sedative-, hypnotic-, anxiolytic- or other (or unknown) substance-induced mood disorder
	– With depressive features
	– With manic features
	– With mixed features
.73 Dementia	291.2 Alcohol-induced persisting dementia
	292.83 Sedative-, hypnotic-, or anxiolytic, Other (or unknown) substance-induced persisting dementia
.74 Other persisting cognitive impairment	No equivalent
.75 Late-onset psychotic disorder	No equivalent
.8 Other mental and behavioural disorders	291.8 Alcohol-induced anxiety, sleep disorder and sexual dysfunction
	292.89 Amphetamine-, caffeine-, cannabis-, cocaine-, hallucinogen-, inhalant-, phencyclidine-, sedative-, hypnotic-, anxiolytic- or other (or unknown) substance-induced anxiety disorder
	292.89 Amphetamine-, cocaine-, opioid-, sedative-, hypnotic-, anxiolytic- or other (or unknown) substance-induced sexual dysfunction
	292.89 Amphetamine-, caffeine-, cocaine-, opioid-, sedative-, hypnotic-, anxiolytic- or other (or unknown) substance-induced sleep disorder
.9 Unspecified mental and behavioural disorder	291.9 Alcohol-related disorder NOS
	292.9 Amphetamine-, caffeine-, cannabis-, cocaine-, hallucinogen-, inhalant-, phencyclidine-, sedative-, hypnotic-, anxiolytic- or other (or unknown) substance-related disorder NOS
F20–F29 Schizophrenia, schizotypal and delusional disorders	
F20 Schizophrenia	295.xx Schizophrenia or 295.40 Schizophreniform disorder
A fifth character may be used to classify course:	(see note below F20)

ICD-10	DSM-IV (ICD-9-CM)
0 Continuous	– Continuous
1 Episodic with progressive deficit	– Episodic with inter-episode residual symptoms
2 Episodic with stable deficit	
3 Episodic remittent	– Episodic with no inter-episode residual symptoms
4 Incomplete remission	– Single episode in partial or full remission
5 Complete remission	
8 Other	– Other or unspecified pattern
9 Course uncertain, period of observation too short	
F20.0 Paranoid schizophrenia[a]	295.30 Schizophrenia, paranoid type
F20.1 Hebephrenic schizophrenia[a]	295.10 Schizophrenia, disorganized type
F20.2 Catatonic schizophrenia[a]	295.20 Schizophrenia, catatonic type
F20.3 Undifferentiated schizophrenia[a]	295.90 Schizophrenia, undifferentiated type
F20.4 Post-schizophrenic depression	*See Appendix B*: Post-psychotic depressive disorder of schizophrenia
F20.5 Residual schizophrenia	295.60 Schizophrenia, residual type
F20.6 Simple schizophrenia	*See Appendix B*: Simple deteriorative disorder
F20.8 Other schizophrenia[a]	No equivalent
F20.9 Schizophrenia, unspecified[a]	No equivalent
F21 Schizotypal disorder	301.22 Schizotypal personality disorder
F22 Persistent delusional disorders	*For all subtypes*:
F22.0 Delusional disorder	297.1 Delusional disorder
F22.8 Other persistent delusional disorders	
F22.9 Persistent delusional disorder, unspecified	
F23 Acute and transient psychotic disorders	*For all subtypes*:
A fifth character may be used to identify the presence or absence of associated acute stress:	298.8 Brief psychotic disorder (only if duration is less than 1 month; see note below F23)
0 Without associated acute stress	– Without marked stressor(s)
1 With associated acute stress	– With marked stressor(s)
F23.0 Acute polymorphic psychotic disorder without symptoms of schizophrenia[b]	
F23.1 Acute polymorphic psychotic disorder with symptoms of schizophrenia	
F23.2 Acute schizophrenia-like psychotic disorder	
F23.3 Other acute predominantly delusional psychotic disorders	*If duration is more than 1 month, use*: 297.1 Delusional disorder
F23.8 Other acute and transient psychotic disorders[b]	
F23.9 Acute and transient psychotic disorders, unspecified[b]	
F24 Induced delusional disorder	297.3 Shared psychotic disorder
F25 Schizoaffective disorders	*For all subtypes*:
F25.0 Schizoaffective disorder, manic type	295.70 Schizoaffective disorder
F25.1 Schizoaffective disorder, depressive type	– Bipolar type
F25.2 Schizoaffective disorder, mixed type	– Depressive type
F25.8 Other schizoaffective disorders	– Bipolar type
F25.9 Schizoaffective disorder, unspecified	
F28 Other non-organic psychotic disorders	No equivalent
F29 Unspecified non-organic psychosis	298.9 Psychotic disorder NOS
F30–F39 Mood affective disorders	
F30 Manic episode	296.0x Bipolar I disorder, single manic episode
F30.0 Hypomania	No equivalent
F30.1 Mania without psychotic symptoms	.01 Mild, .02 moderate and .03, severe without psychotic features
F30.2 Mania with psychotic symptoms	.04 Severe with psychotic features
F30.8 Other manic episodes	– With catatonic features
F30.9 Manic episode, unspecified	296.80 Bipolar disorder NOS

[a]*Note: The conversion for F20.0–F20.3, F20.8 and F20.9 applies for a duration of 6 months or more. For duration less than 6 months, use*: 295.40 Schizophreniform disorder.

[b]*Note: The conversion for F23.0–F23.2, F23.8 and F23.9 applies for a duration of less than 1 month. For a duration of more than 1 month use*: 298.9 Psychotic disorder NOS.

ICD-10	DSM-IV (ICD-9-CM)
F31 Bipolar affective disorder	296.xx Bipolar I disorder or 296.89 Bipolar II disorder
F31.0 Bipolar affective disorder, current episode hypomanic	296.40 Bipolar I disorder or 296.89 Bipolar II disorder, most recent episode hypomanic
F31.1 Bipolar affective disorder, current episode manic without psychotic symptoms	296.4x Bipolar I disorder .41, .42, .43 Most recent episode manic, mild, moderate, or severe without psychotic features
F31.2 Bipolar affective disorder, current episode manic with psychotic symptoms	296.44 Bipolar I disorder or 296.89 Bipolar II disorder, most recent episode manic, severe with psychotic features
F31.3 Bipolar affective disorder, current episode mild or moderate depression	296.51 Bipolar I disorder or 296.89 Bipolar II disorder, most recent episode depressed, mild
.30 Without somatic syndrome	296.52 Bipolar I disorder or 296.89 Bipolar II disorder, most recent episode depressed, moderate
.31 With somatic syndrome	– With melancholic features
F31.4 Bipolar affective disorder, current episode severe depression without psychotic symptoms	296.53 Bipolar I disorder or 296.89 Bipolar II disorder, most recent episode depressed, severe without psychotic features
F31.5 Bipolar affective disorder, current episode severe depression with psychotic symptoms	296.54 Bipolar I disorder or 296.89 Bipolar II disorder, most recent episode depressed, severe with psychotic features
F31.6 Bipolar affective disorder, current episode mixed	296.6x Bipolar I disorder, most recent episode mixed .61 mild, .62 moderate, .63 severe without psychotic symptoms, .64 severe with psychotic symptoms
F31.7 Bipolar affective disorder, currently in remission	296.5 Bipolar I disorder or 296.89 Bipolar II disorder, in partial remission
	296.6 Bipolar I disorder or 296.89 Bipolar II disorder, in full remission
F31.8 Other bipolar affective disorders	296.xx Bipolar I disorder or 296.xx Bipolar II disorder
	– With catatonic features
	– With atypical features
F31.9 Bipolar affective disorder, unspecified	296.7 Bipolar I disorder or 296.89 Bipolar II disorder, most recent episode unspecified
	296.80 Bipolar disorder NOS
F32 Depressive episode	296.2x Major depressive disorder, single episode
F32.0 Mild depressive episode	.21 Mild
.00 Without somatic syndrome	
.01 With somatic syndrome	– With melancholic features
F32.1 Moderate depressive episode	.22 Moderate
.10 Without somatic syndrome	
.11 With somatic syndrome	– With melancholic features
F32.2 Severe depressive episode without psychotic symptoms	.23 Severe without psychotic features
F32.3 Severe depressive episode with psychotic symptoms	.24 Severe with psychotic features
F32.8 Other depressive episodes	– With catatonic/atypical features
F32.9 Depressive episode, unspecified	.20 Unspecified
F33 Recurrent depressive disorder	296.3x Major depressive disorder, recurrent
F33.0 Recurrent depressive disorder, current episode mild	.31 Mild
.00 Without somatic syndrome	
.01 With somatic syndrome	– With melancholic features
F33.1 Recurrent depressive disorder, current episode moderate	.32 Moderate
.10 Without somatic syndrome	
.11 With somatic syndrome	– With melancholic features
F33.2 Recurrent depressive disorder, current episode severe without psychotic symptoms	.33 Severe without psychotic features
F33.3 Recurrent depressive disorder, current episode severe with psychotic symptoms	.34 Severe with psychotic features
F33.4 Recurrent depressive disorder, currently in remission	.35 In partial remission or .36 In complete remission
F33.8 Other recurrent depressive disorders	– With catatonic/atypical features
F33.9 Recurrent depressive disorder, unspecified	.30 Unspecified
F34 Persistent mood (affective) disorders	
F34.0 Cyclothymia	301.13 Cyclothymic disorder
F34.1 Dysthymia	300.4 Dysthymic disorder

ICD-10	DSM-IV (ICD-9-CM)
F34.8 Other persistent mood (affective) disorders	300.4 Dysthymic disorder with atypical features
F34.9 Persistent mood (affective) disorder, unspecified	No equivalent
F38 Other mood (affective) disorders	
F38.0 Other single mood (affective) disorders	296.0x Bipolar I disorder, single mixed episode
.00 Mixed affective episode	.01 Mild, .02 moderate, .03 severe without psychotic symptoms, .04 severe with psychotic symptoms
F38.1 Other recurrent mood (affective) disorders	
.10 Recurrent brief depressive disorder	*See Appendix B*: Recurrent brief depressive disorder
F38.8 Other specified mood (affective) disorders	No equivalent
F39 Unspecified mood (affective) disorder	296.90 Mood disorder NOS
F40–F48 Neurotic, stressrelated and somatoform disorders	
F40 Phobic anxiety disorders	
F40.0 Agoraphobia	
.00 Without panic disorder	300.22 Agoraphobia without history of panic disorder
.01 With panic disorder	300.21 Panic disorder with agoraphobia
F40.1 Social phobias	300.23 Social phobia
F40.2 Specific (isolated) phobias	300.29 Specific phobia
F40.8 Other phobic anxiety disorders	No equivalent
F40.9 Phobic anxiety disorder, unspecified	300.00 Anxiety disorder NOS
F41 Other anxiety disorders	
F41.0 Panic disorder (episodic paroxysmal anxiety)	300.01 Panic disorder without agoraphobia
F41.1 Generalized anxiety disorder	300.02 Generalized anxiety disorder
F41.2 Mixed anxiety and depressive disorder	*See Appendix B*: Mixed anxiety–depressive disorder
F41.3 Other mixed anxiety disorders	No equivalent
F41.8 Other specified anxiety disorders	No equivalent
F41.9 Anxiety disorder, unspecified	300.00 Anxiety disorder NOS
F42 Obsessive–compulsive disorder	*For all subtypes*:
F42.0 Predominantly obsessional thoughts or ruminations	300.3 Obsessive–compulsive disorder
F42.1 Predominantly compulsive acts (obsessional rituals)	
F42.2 Mixed obsessional thoughts and acts	
F42.8 Other obsessive–compulsive disorders	
F42.9 Obsessive–compulsive disorder, unspecified	
F43 Reaction to severe stress and adjustment disorders	
F43.0 Acute stress reaction	308.3 Acute stress disorder
F43.1 Post-traumatic stress disorder	309.81 Post-traumatic stress disorder
F43.2 Adjustment disorders	309.xx Adjustment disorder
.20 Brief depressive reaction	.0 With depressed mood (excludes bereavement)
.21 Prolonged depressive reaction	
.22 Mixed anxiety and depressive reaction	.28 With mixed anxiety and depressed mood
.23 With predominant disturbance of other emotions	.24 With anxiety
.24 With predominant disturbance of conduct	.3 With disturbance of conduct
.25 With mixed disturbance of emotions and conduct	.4 With mixed disturbance of emotions and conduct
.28 With other specified predominant symptoms	No equivalent
F43.8 Other reactions to severe stress	No equivalent
F43.9 Reaction to severe stress, unspecified	No equivalent
F44 Dissociative (conversion) disorder	
F44.0 Dissociative amnesias	300.12 Dissociative amnesia
F44.1 Dissociative fugue	300.13 Dissociative fugue
F44.2 Dissociative stupor	300.15 Dissociative disorder NOS
F44.3 Trance and possession disorders	300.15 Dissociative disorder NOS (*see also Appendix B*: Dissociative trance disorder)
	300.11 Conversion disorder
F44.4 Dissociative motor disorders	– With motor symptom or deficit
F44.5 Dissociative convulsions	– With seizures or convulsions
F44.6 Dissociative anaesthesia and sensory loss	– With sensory symptom or deficit

ICD-10	DSM-IV (ICD-9-CM)
F44.7 Mixed dissociative (conversion) disorders	– With mixed presentation
F44.8 Other dissociative (conversion) disorders	300.15 Dissociative disorder NOS
.80 Ganser's syndrome	
.81 Multiple personality disorder	300.14 Dissociative identity disorder
.82 Transient dissociative (conversion) disorders occurring in childhood and adolescence	300.15 Dissociative disorder NOS
.88 Other specified dissociative (conversion) disorders	
F44.9 Dissociative (conversion) disorder, unspecified	300.15 Dissociative disorder NOS
F45 Somatoform disorders	
F45.0 Somatization disorder	300.81 Somatization disorder
F45.1 Undifferentiated somatoform disorder	300.81 Undifferentiated somatoform disorder
F45.2 Hypochondriacal disorder	300.7 Hypochondriasis or 300.7 Body dysmorphic disorder
F45.3 Somatoform autonomic dysfunction	*No equivalent, use*: 300.81 Undifferentiated somatoform disorder
.30 Heart and cardiovascular system	
.31 Upper gastrointestinal tract	
.32 Lower gastrointestinal tract	
.33 Respiratory system	
.34 Genitourinary system	
.38 Other organ or system	
F45.4 Persistent somatoform pain disorder	307.80 Pain disorder, associated with psychological factors
F45.8 Other somatoform disorders	300.81 Somatoform disorder NOS
F45.9 Somatoform disorder, unspecified	300.81 Somatoform disorder NOS
F48 Other neurotic disorders	
F48.0 Neurasthenia	*No equivalent, use*: 300.81 Undifferentiated somatoform disorder
F48.1 Depersonalization–derealization syndrome	300.6 Depersonalization disorder
F48.8 Other specified neurotic disorders	No equivalent
F48.9 Neurotic disorder, unspecified	300.9 Unspecified mental disorder (non-psychotic)
F50–F59 Behavioural syndromes associated with physiological disturbances and physical factors	
F50 Eating disorders	
F50.0 Anorexia nervosa	307.1 Anorexia nervosa
F50.1 Atypical anorexia nervosa	307.50 Eating disorder NOS
F50.2 Bulimia nervosa	307.51 Bulimia nervosa
F50.3 Atypical bulimia nervosa	307.1 Anorexia nervosa, binge-eating type
F50.4 Overeating associated with other psychological disturbances	No equivalent
F50.5 Vomiting associated with other psychological disturbances	No equivalent
F50.8 Other eating disorders	No equivalent
F50.9 Eating disorder, unspecified	307.50 Eating disorder NOS
F51 Non-organic sleep disorders	
F51.0 Non-organic insomnia	307.42 Primary insomnia
F51.1 Non-organic hypersomnia	307.44 Primary hypersomnia
F51.2 Non-organic disorder of the sleep–wake schedule	307.45 Circadian rhythm sleep disorder
F51.3 Sleepwalking (somnambulism)	307.46 Sleepwalking disorder
F51.4 Sleep terrors (night terrors)	307.46 Sleep terror disorder
F51.5 Nightmares	307.47 Nightmare disorder
F51.8 Other non-organic sleep disorders	307.47 Dyssomnia NOS or 307.47 Parasomnia NOS
F51.9 Non-organic sleep disorder, unspecified	307.47 Dyssomnia NOS or 307.47 Parasomnia NOS
F52 Sexual dysfunction, not caused by organic disorder or disease	
F52.0 Lack or loss of sexual desire	302.71 Hypoactive sexual desire disorder
F52.1 Sexual aversion and lack of sexual enjoyment	
.10 Sexual aversion	302.79 Sexual aversion disorder
.11 Lack of sexual enjoyment	*No equivalent, use*: 302.70 Sexual dysfunction NOS

ICD-10	DSM-IV (ICD-9-CM)
F52.2 Failure of genital response	302.72 Female sexual arousal disorder or 302.72 Male erectile disorder
F52.3 Orgasmic dysfunction	302.73 Female orgasmic disorder or 307.74 Male orgasmic disorder
F52.4 Premature ejaculation	302.75 Premature ejaculation
F52.5 Non-organic vaginismus	306.51 Vaginismus (not due to a general medical condition)
F52.6 Non-organic dyspareunia	302.76 Dyspareunia (not due to a general medical condition)
F52.7 Excessive sexual drive	*No equivalent, use*: 302.70 Sexual dysfunction NOS
F52.8 Other sexual dysfunction, not caused by organic disorders or disease	302.70 Sexual dysfunction NOS
F52.9 Unspecified sexual dysfunction, not caused by organic disorder or disease	302.70 Sexual dysfunction NOS
F53 Mental and behavioural disorders associated with the puerperium, not elsewhere classified	No equivalents
F53.0 Mild mental and behavioural disorders associated with the puerperium, not elsewhere classified	
F53.1 Severe mental and behavioural disorders associated with the puerperium, not elsewhere classified	
F53.8 Other mental and behavioural disorders associated with the puerperium, not elsewhere classified	
F53.9 Puerperal mental disorder, unspecified.	
F54 Psychological and behavioural factors associated with disorders or diseases classified elsewhere	316 Psychological factors affecting medical condition
F55 Abuse of non-dependence-producing substances	*For all subtypes*:
F55.0 Antidepressants	305.90 Other substance abuse
F55.1 Laxatives	
F55.2 Analgesics	
F55.3 Antacids	
F55.4 Vitamins	
F55.5 Steroids or hormones	
F55.6 Specific herbal or folk remedies	
F55.8 Other substances that do not produce dependence	
F55.9 Unspecified	
F59 Unspecified behavioural syndromes associated with physiological disturbances and physical factors	316 Other or unspecified psychological factors affecting medical condition
F60–F69 Disorders of adult personality and behaviour	
F60 Specific personality disorders	
F60.0 Paranoid personality disorder	301.0 Paranoid personality disorder
F60.1 Schizoid personality disorder	301.20 Schizoid personality disorder
F60.2 Dyssocial personality disorder	301.7 Antisocial personality disorder
F60.3 Emotionally unstable personality disorder	No equivalent
.30 Impulsive type	
.31 Borderline type	301.83 Borderline personality disorder
F60.4 Histrionic personality disorder	301.50 Histrionic personality disorder
F60.5 Anankastic personality disorder	301.4 Obsessive–compulsive personality disorder
F60.6 Anxious (avoidant) personality disorder	301.82 Avoidant personality disorder
F60.7 Dependent personality disorder	301.6 Dependent personality disorder
F60.8 Other specific personality disorders	301.81 Narcissistic personality disorder or 301.9 Personality disorder NOS (*see also Appendix B*: Passive–aggressive personality disorder)
F60.9 Personality disorder, unspecified	301.9 Personality disorder NOS
F61 Mixed and other personality disorders	301.9 Personality disorder NOS
F61.0 Mixed personality disorders	
F61.1 Troublesome personality changes	
F62 Enduring personality changes, not attributable to brain damage and disease	*No equivalents, use*: 301.9 Personality disorder NOS
F62.0 Enduring personality change after catastrophic experience	

ICD-10	DSM-IV (ICD-9-CM)
F62.1 Enduring personality change after psychiatric illness	
F62.8 Other enduring personality changes	
F62.9 Enduring personality change, unspecified	
F63 Habit and impulse disorders	
F63.0 Pathological gambling	312.31 Pathological gambling
F63.1 Pathological fire-setting (pyromania)	312.33 Pyromania
F63.2 Pathological stealing (kleptomania)	312.32 Kleptomania
F63.3 Trichotillomania	312.39 Trichotillomania
F63.8 Other habit and impulse disorders	312.30 Impulse-control disorder NOS
F63.9 Habit and impulse disorder, unspecified	312.30 Impulse-control disorder NOS
F64 Gender identity	
F64.0 Transsexualism	302.85 Gender identity disorder in adolescents or adults
F64.1 Dual-role transvestism	302.85 Gender identity disorder in adolescents or adults
F64.2 Gender identity disorder of childhood	302.6 Gender identity disorder in children
F64.8 Other gender identity disorders	302.6 Gender identity disorder NOS
F64.9 Gender identity disorder, unspecified	302.6 Gender identity disorder NOS
F65 Disorders of sexual preference	
F65.0 Fetishism	302.81 Fetishism
F65.1 Fetishistic transvestism	302.3 Transvestic fetishism
F65.2 Exhibitionism	302.4 Exhibitionism
F65.3 Voyeurism	302.82 Voyeurism
F65.4 Paedophilia	302.2 Paedophilia
F65.5 Sadomasochism	302.83 Sexual masochism or 302.84 Sexual sadism
F65.6 Multiple disorders of sexual preference	302.9 Paraphilia NOS
F65.8 Other disorders of sexual preference	302.9 Paraphilia NOS
F65.9 Disorder of sexual preference, unspecified	302.9 Paraphilia NOS
F66 Psychological and behavioural disorders associated with sexual development and orientation	*For all subtypes*: 302.9 Sexual disorder NOS
F66.0 Sexual maturation disorder	
F66.1 Egodystonic sexual orientation	
F66.2 Sexual relationship disorder	
F66.8 Other psychosexual development disorders	
F66.9 Psychosexual development disorder, unspecified	
A fifth character may be used to indicate association with: .0 Heterosexuality .1 Homosexuality .2 Bisexuality .8 Other, including prepubertal	
F68 Other disorders of adult personality and behaviour	
F68.0 Elaboration of physical symptoms for psychological reasons	*No equivalent, use*: 300.81 Undifferentiated somatoform disorder
F68.1 Intentional production or feigning of symptoms or disabilities, either physical or psychological (factitious disorder)	300.1 Factitious disorder .16 With psychological signs and symptoms .19 With physical signs and symptoms .19 With combined psychological and physical signs and symptoms .19 NOS
F68.8 Other specified disorders of adult personality and behaviour	301.9 Personality disorder NOS
F69 Unspecified disorder of adult personality and behaviour	301.9 Personality disorder NOS
F70–F79 Mental retardation	
F70 Mild mental retardation	317 Mild mental retardation
F71 Moderate mental retardation	318.0 Moderate mental retardation
F72 Severe mental retardation	318.1 Severe mental retardation
F73 Profound mental retardation	318.2 Profound mental retardation

ICD-10	DSM-IV (ICD-9-CM)
F78 Other mental retardation	319 Mental retardation, severity unspecified
F79 Unspecified mental retardation	319 Mental retardation, severity unspecified
A fourth character may be used to specify the extent of associated behavioural impairment: F7x.0 No, or minimal, impairment of behaviour F7x.1 Significant impairment of behaviour requiring attention or treatment F7x.8 Other impairments of behaviour F7x.9 Without mention of impairment of behaviour	(No specifiers for associated behavioural impairments)
F80–F89 Disorders of psychological development	
F80 Specific developmental disorders of speech and language	
F80.0 Specific speech articulation disorder	315.39 Phonological disorder
F80.1 Expressive language disorder	315.31 Expressive language disorder
F80.2 Receptive language disorder	315.31 Mixed receptive–expressive language disorder
F80.3 Acquired aphasia with epilepsy (Landau–Kleffner syndrome)	307.9 Communication disorder NOS
F80.8 Other developmental disorders of speech and language	307.9 Communication disorder NOS
F80.9 Developmental disorder of speech and language, unspecified	307.9 Communication disorder NOS
F81 Specific developmental disorders of scholastic skills	
F81.0 Specific reading disorder	315.00 Reading disorder
F81.1 Specific spelling disorder	*No equivalent, use*: 315.9 Learning disorder NOS
F81.2 Specific disorder of arithmetical skills	315.1 Mathematics disorder
F81.3 Mixed disorder of scholastic skills	315.9 Learning disorder NOS
F81.8 Other developmental disorders of scholastic skills	315.9 Learning disorder NOS
F81.9 Developmental disorder of scholastic skills, unspecified	315.9 Learning disorder NOS
F82 Specific developmental disorder of motor function	315.4 Developmental coordination disorder
F83 Mixed specific developmental disorders	No equivalent
F84 Pervasive developmental disorders	
F84.0 Childhood autism	299.00 Autistic disorder
F84.1 Atypical autism	299.80 Pervasive developmental disorder NOS (includes atypical autism)
F84.2 Rett's syndrome	299.80 Rett's disorder
F84.3 Other childhood disintegrative disorder	299.10 Childhood disintegrative disorder
F84.4 Overactive disorder associated with mental retardation and stereotyped movements	*No equivalent, use*: 299.80 Pervasive developmental disorder NOS
F84.5 Asperger's syndrome	299.80 Asperger's syndrome
F84.8 Other pervasive developmental disorders	299.80 Pervasive developmental disorder NOS
F84.9 Pervasive developmental disorder, unspecified	299.80 Pervasive developmental disorder NOS
F88 Other disorders of psychological development	313.9 Disorder of infancy, childhood or adolescence NOS
F89 Unspecified disorder of psychological development	313.9 Disorder of infancy, childhood or adolescence NOS
F90–F98 Behavioural and emotional disorders with onset usually occurring in childhood and adolescence	
F90 Hyperkinetic disorders	
F90.0 Disturbance of activity and attention	314.0 Attention-deficit/hyperactivity disorder
F90.1 Hyperkinetic conduct disorder	*No equivalent, use both* 314.0 and 312.8 Conduct disorder
F90.8 Other hyperkinetic disorders	314.9 Attention-deficit/hyperactivity disorder NOS
F90.9 Hyperkinetic disorder, unspecified	314.9 Attention-deficit/hyperactivity disorder NOS
F91 Conduct disorders	
F91.0 Conduct disorder confined to the family context	312.8 Conduct disorder

ICD-10	DSM-IV (ICD-9-CM)
F91.1 Unsocialized conduct disorder	312.8 Conduct disorder
F91.2 Socialized conduct disorder	312.8 Conduct disorder
F91.3 Oppositional defiant disorder	313.81 Oppositional defiant disorder
F91.8 Other conduct disorders	312.9 Disruptive behaviour disorder NOS
F91.9 Conduct disorder, unspecified	312.9 Disruptive behaviour disorder NOS
F92 Mixed disorders of conduct and emotions	*No equivalents, use*: 312.8 Conduct disorder plus another appropriate category
F92.0 Depressive conduct disorder	
F92.8 Other mixed disorders of conduct and emotions	
F92.9 Mixed disorder of conduct and emotions, unspecified	
F93 Emotional disorders with onset specific to childhood	
F93.0 Separation anxiety disorder of childhood	309.21 Separation anxiety disorder
F93.1 Phobic anxiety disorder of childhood	300.29 Specific phobia or 300.22 Agoraphobia
F93.2 Social anxiety disorder of childhood	300.23 Social phobia
F93.3 Sibling rivalry disorder	V61.8 Sibling relational problem
F93.8 Other childhood emotional disorders	313.9 Disorder of infancy, childhood or adolescence NOS
F93.9 Childhood emotional disorder, unspecified	313.9 Disorder of infancy, childhood or adolescence NOS
F94 Disorders of social functioning with onset specific to childhood and adolescence	
F94.0 Elective mutism	313.23 Selective mutism
F94.1 Reactive attachment disorder of childhood	313.89 Reactive attachment disorder of infancy or early childhood – Inhibited type
F94.2 Disinhibited attachment disorder of childhood	313.89 Reactive attachment disorder of infancy or early childhood – Disinhibited type
F94.8 Other childhood disorders of social functioning	313.9 Disorder of infancy, childhood or adolescence NOS
F94.9 Childhood disorders of social functioning, unspecified	313.9 Disorder of infancy, childhood or adolescence NOS
F95 Tic disorders	
F95.0 Transient tic disorder	307.21 Transient tic disorder
F95.1 Chronic motor or vocal tic disorder	307.22 Chronic motor or vocal tic disorder
F95.2 Combined vocal and multiple motor tic disorder (Gilles de la Tourette's syndrome)	307.23 Tourette's disorder
F95.8 Other tic disorders	307.20 Tic disorder NOS
F95.9 Tic disorder, unspecified	307.20 Tic disorder NOS
F98 Other behavioural and emotional disorders with onset usually occurring in childhood and adolescence	
F98.0 Non-organic enuresis	307.6 Enuresis (not due to a general medical condition)
F98.1 Non-organic encopresis	787.6 Encopresis with constipation and overflow or 307.7 Encopresis without constipation and overflow
F98.2 Feeding disorder of infancy and childhood	307.59 Feeding disorder of infancy or early childhood
F98.3 Pica of infancy and childhood	307.52 Pica
F98.4 Stereotyped movement disorders	307.3 Stereotypic movement disorder
F98.5 Stuttering (stammering)	307.0 Stuttering
F98.6 Cluttering	307.9 Communication disorder NOS
F98.8 Other specified behavioural and emotional disorders with onset usually occurring in childhood and adolescence	313.9 Disorder of infancy, childhood or adolescence NOS
F98.9 Unspecified behavioural and emotional disorders with onset usually occurring in childhood and adolescence	313.9 Disorder of infancy, childhood or adolescence NOS
F99 Unspecified mental disorder	*No equivalent, use*: 300.9 Unspecified mental disorder (non-psychotic)

NOS, not otherwise specified.

Table 2. Conversion table: DSM-IV (ICD-9-CM) to ICD-10

DSM-IV (ICD-9-CM)	ICD-10
If criteria are currently met, one of the following severity specifiers may be noted after the diagnosis: mild, moderate or severe. If criteria are no longer met, one of the following specifiers may be noted: in partial remission, in full remission or prior history.	
Disorders usually first diagnosed in infancy, childhood or adolescence	
Mental retardation	
Note: These are coded on axis II.	*Note: The fourth character .9 (without mention of impairment of behaviour) is used, because* DSM-IV *has no specifier for associated impairment of behaviour.*
317 Mild mental retardation	F70.9 Mild mental retardation
318.0 Moderate mental retardation	F71.9 Moderate mental retardation
318.1 Severe mental retardation	F72.9 Severe mental retardation
318.2 Profound mental retardation	F73.9 Profound mental retardation
319 Mental retardation, severity unspecified	F79.9 Unspecified mental retardation
Learning disorders	
315.00 Reading disorder	F81.0 Specific reading disorder
315.1 Mathematics disorder	F81.2 Specific disorder of arithmetical skills
315.2 Disorder of written expression	F81.8 Other developmental disorders of scholastic skills
315.9 Learning disorder NOS	F81.9 Developmental disorder of scholastic skills, unspecified
Motor skills disorder	
315.4 Developmental coordination disorder	F82 Specific developmental disorder of motor function
Communication disorders	
315.31 Expressive language disorder	F80.1 Expressive language disorder
315.31 Mixed receptive–expressive language disorder	F80.2 Receptive language disorder
315.39 Phonological disorder	F80.0 Specific speech articulation disorder
307.0 Stuttering	F98.5 Stuttering (stammering)
307.9 Communication disorder NOS	F80.9 Developmental disorder of speech and language, unspecified
Pervasive developmental disorders	
299.00 Autistic disorder	F84.0 Childhood autism
299.80 Rett's disorder	F84.2 Rett's syndrome
299.10 Childhood disintegrative disorder	F84.3 Other childhood disintegrative disorder
299.80 Asperger's disorder	F84.5 Asperger's syndrome
299.80 Pervasive developmental disorder NOS (includes a typical autism)	F84.9 Pervasive developmental disorder, unspecified, or F84.1 Atypical autism
Attention-deficit and disruptive behaviour disorder	
314.xx Attention-deficit/hyperactivity disorder	
.01 Combined type	F90.0 Disturbance of activity and attention
.00 Predominantly inattentive type	F98.8 Attention deficit disorder without hyperactivity
.01 Predominantly hyperactive–impulsive type	F90.0 Disturbance of activity and attention
314.9 Attention-deficit/hyperactivity disorder NOS	F90.9 Hyperkinetic disorder, unspecified
312.8 Conduct disorder	F91.8 Other conduct disorders
Specify type: childhood-onset type/adolescent-onset type	
313.81 Oppositional defiant disorder	F91.3 Oppositional defiant disorder
312.9 Disruptive behaviour disorder NOS	F91.9 Conduct disorder, unspecified
Feeding and eating disorders of infancy or early childhood	
307.52 Pica	F98.3 Pica of infancy and childhood
307.53 Rumination disorder	F98.2 Feeding disorder of infancy and childhood
307.59 Feeding disorder of infancy or early childhood	F98.2 Feeding disorder of infancy and childhood
Tic disorders	
307.23 Tourette's disorder	F95.2 Combined vocal and multiple motor tic disorder (Gilles de la Tourette's syndrome)
307.22 Chronic motor or vocal tic disorder	F95.1 Chronic motor or vocal tic disorder
307.21 Transient tic disorder	F95.0 Transient tic disorder
Specify if: single episode/recurrent	
307.20 Tic disorder NOS	F95.9 Tic disorder, unspecified

DSM-IV (ICD-9-CM)		ICD-10	
Elimination disorders			
–.–	Encopresis	F98.1	Non-organic encopresis
787.6	With constipation and overflow incontinence		.12 Soiling that is associated with excessively fluid faeces such as with retention with overflow
307.7	Without constipation and overflow incontinence	F98.1	Non-organic encopresis .11 Adequate bowel control with normal faeces deposited in inappropriate places .10 Failure to acquire physiological bowel control
307.6	Enuresis (not due to a general medical condition) *Specify type*: Nocturnal only Diurnal only Nocturnal and diurnal	F98.0	Non-organic enuresis .00 Nocturnal enuresis only .01 Diurnal enuresis only .02 Nocturnal and diurnal enuresis
Other disorders of infancy, childhood or adolescence			
309.21	Separation anxiety disorder *Specify if*: early onset	F93.0	Separation anxiety disorder of childhood
313.23	Selective mutism	F94.0	Elective mutism
313.89	Reactive attachment disorder of infancy or early childhood *Specify type*: Inhibited type Disinhibited type	F94.1 F94.2	Reactive attachment disorder of childhood Disinhibited attachment disorder of childhood
307.3	Stereotypic movement disorder *Specify if*: with self-injurious behaviour	F98.4	Stereotyped movement disorders .41 Self-injurious
313.9	Disorder of infancy, childhood or adolescence NOS	F98.9	Unspecified behavioural and emotional disorders with onset usually occurring in childhood and adolescence
Delirium, Dementia, Amnestic and other Cognitive Disorders			
Delirium			
293.0	Delirium due to ... (indicate the general medical condition) – If superimposed on dementia	F05.0 F05.1	Delirium, not superimposed on dementia, so described Delirium superimposed on dementia
---.–	Substance intoxication delirium (refer to substance-related disorders for substance-specific codes)	F1x.03	Acute intoxication with delirium due to psychoactive substances
---.–	Substance withdrawal delirium (refer to substance-related disorders for substance-specific codes)	F1x.4	Withdrawal state with delirium
---.–	Delirium due to multiple aetiologies (code each of the specific aetiologies)	F05.8	Other delirium
780.09	Delirium NOS	F05.9	Delirium, unspecified
Dementia			
290.xx	Dementia of the Alzheimer's type, with early onset (also code 331.0 Alzheimer's disease on axis III) .10 Uncomplicated .11 With delirium .12 With delusions .13 With depressed mood *Specify if*: with behavioural disturbance	F00.0	Dementia in Alzheimer's disease with early onset (also code G30.0 Alzheimer's disease with early onset) .00 Without additional symptoms Code F00.0 plus F05.1 Delirium superimposed on dementia .01 Other symptoms, predominantly delusional .03 Other symptoms, predominantly depressive .04 Other mixed symptoms
290.xx	Dementia of the Alzheimer's type, with late onset (also code 331.0 Alzheimer's disease on axis III) .0 Uncomplicated .3 With delirium .20 With delusions .21 With depressed mood *Specify if*: with behavioural disturbance	F00.1	Dementia in Alzheimer's disease with late onset (also code G30.1 Alzheimer's disease with late onset) .10 Without additional symptoms Code F00.1 plus F05.1 Delirium superimposed on dementia .11 Other symptoms, predominantly delusional .13 Other symptoms, predominantly depressive .14 Other mixed symptoms
290.xx	Vascular dementia .40 Uncomplicated .41 With delirium	F01.9	Vascular dementia, unspecified .90 Without additional symptoms Code F01.9 plus F05.1 Delirium superimposed on dementia

DSM-IV (ICD-9-CM)	ICD-10
.42 With delusions .43 With depressed mood *Specify if*: with behavioural disturbance	.91 Other symptoms, predominantly delusional .93 Other symptoms, predominantly depressive .94 Other mixed symptoms
294.9 Dementia due to HIV disease (also code 043.1 HIV infection affecting central nervous system on axis III)	F02.4 Dementia in human immunodeficiency virus (HIV) disease (also code B22.0 HIV dementia)
294.1 Dementia due to head trauma (also code 854.00 head injury on axis III)	F02.8 Dementia in other specified diseases classified elsewhere (also code S06.9 Intracranial injury unspecified)
294.1 Dementia due to Parkinson's disease (also code 332.0 Parkinson's disease on axis III)	F02.3 Dementia in Parkinson's disease (also code G20 Parkinson's disease)
294.1 Dementia due to Huntington's disease (also code 333.4 Huntington's disease on axis III)	F02.2 Dementia in Huntington's disease (also code G10 Huntington's disease)
290.10 Dementia due to Pick's disease (also code 331.1 Pick's disease on axis III)	F02.0 Dementia in Pick's disease (also code G31.0 Pick's disease)
290.10 Dementia due to Creutzfeldt-Jakob disease (also code 046.1 Creutzfeldt-Jakob disease on axis III)	F02.1 Dementia in Creutzfeldt-Jakob disease (also code A81.0 Creutzfeldt-Jacob disease)
294.1 Dementia due to ... (indicate the general medical condition not listed above; also code the general medical condition on axis III)	F02.8 Dementia in other specified diseases classified elsewhere
---.- Substance-induced persisting dementia (refer to substance-related disorders for substance-specific codes)	F1x.73 Dementia due to psychoactive substance abuse
---.- Dementia due to multiple aetiologies (code each of the specific aetiologies)	F02.8 Dementia in other specified diseases classified elsewhere
- Mixed Alzheimer and vascular dementia	F00.2 Dementia in Alzheimer's disease, atypical or mixed
294.8 Dementia NOS	F03 Unspecified dementia
Amnestic disorders	
294.0 Amnestic disorder due to ... (indicate the general medical condition) *Specify if*: transient/chronic	F04 Organic amnesic syndrome, not induced by alcohol and other psychoactive substance
---.- Substance-induced persisting amnestic disorder (refer to substance-related disorders for substance-specific codes)	F1x.6 Amnesic syndrome due to psychoactive substance use
294.8 Amnestic disorder NOS	No equivalent
Other cognitive disorders	
294.9 Cognitive disorder NOS	F06.7 Mild cognitive disorder
Mental disorders due to a general medical condition not elsewhere classified	
293.89 Catatonic disorder due to ... (indicate the general medical condition)	F06.1 Organic catatonic disorder
310.1 Personality change due to ... (indicate the general medical condition) *Specify type*: labile type/disinhibited type/aggressive type/apathetic type/paranoid type/other type/combined type/unspecified type	F07.0 Organic personality disorder
293.9 Mental disorder NOS due to ... (indicate the general medical condition)	F09 Unspecified organic or symptomatic mental disorder
Substance-related disorders	
[a]The following specifiers may be applied to substance dependence:	Currently using the substance
- With physiological dependence	F1x.241 With physical symptoms
- Without physiological dependence	F1x.240 Without physical symptoms
Code course of dependence in fifth character:	F1x.2 Dependence syndrome
- Early full remission/early partial remission/sustained full remission/sustained partial remission	.20 Currently abstinent
- In a controlled environment	.21 Currently abstinent, but in a protected environment
- On agonist therapy	.22 Currently on a clinically supervised maintenance or replacement regime (controlled dependence)

DSM-IV (ICD-9-CM)	ICD-10
– Mild/moderate/severe	.24 Currently using the substance (active dependence) or .25 continuous use
The following specifiers apply to substance-induced disorders as noted: – [I]With onset during intoxication – [W]With onset during withdrawal	
Alcohol-related disorders	F10.– Mental and behavioural disorders due to use of alcohol
Alcohol use disorders	
303.90 Alcohol dependence[a]	.2 Dependence
305.00 Alcohol abuse	.1 Harmful use
Alcohol-induced disorders	
303.00 Alcohol intoxication	.0 Acute intoxication
291.8 Alcohol withdrawal	.3 Withdrawal state
Specify if: with perceptual disturbances	
291.0 Alcohol intoxication delirium	.03 Acute intoxication with delirium
291.0 Alcohol withdrawal delirium	.4 Withdrawal state with delirium
291.2 Alcohol-induced persisting dementia	.73 Dementia
291.1 Alcohol-induced persisting amnesic disorder	.6 Amnesic disorder
291.x Alcohol-induced psychotic disorder	.5 Psychotic disorder
.5 With delusions[I,W]	.51 Predominantly delusional
.3 With hallucinations[I,W]	.52 Predominantly hallucinatory
291.8 Alcohol-induced mood disorder[I,W]	.5 Psychotic disorder .54 Predominantly depressive psychotic symptoms .55 Predominantly manic psychotic symptoms .56 Mixed .72 Residual affective disorder
291.8 Alcohol-induced anxiety disorder[I,W]	.8 Other mental and behavioural disorders
291.8 Alcohol-induced sexual dysfunction[I]	.8 Other mental and behavioural disorders
291.8 Alcohol-induced sleep disorder[I,W]	.8 Other mental and behavioural disorders
291.9 Alcohol-related disorder NOS	.9 Unspecified mental and behavioural disorder
Amphetamine (or amphetamine-like)-related disorders	F15.– Mental and behavioural disorders due to use of other stimulants, including caffeine
Amphetamine use disorders	
304.40 Amphetamine dependence[a]	.2 Dependence
305.70 Amphetamine abuse	.1 Harmful use
Amphetamine-induced disorders	
292.89 Amphetamine intoxication	.0 Acute intoxication
Specify if: with perceptual disturbances	.04 With perceptual disturbances
292.0 Amphetamine withdrawal	.3 Withdrawal state
292.81 Amphetamine intoxication delirium	.03 Acute intoxication with delirium
292.xx Amphetamine-induced psychotic disorder	.5 Psychotic disorder
.11 With delusions[I]	.51 Predominantly delusional
.12 With hallucinations[I]	.52 Predominantly hallucinatory
292.84 Amphetamine-induced mood disorder[I,W]	.5 Psychotic disorder .54 Predominantly depressive psychotic symptoms .55 Predominantly manic psychotic symptoms .56 Mixed .72 Residual affective disorder
292.89 Amphetamine-induced anxiety disorder[I]	.8 Other mental and behavioural disorders
292.89 Amphetamine-induced sexual dysfunction[I]	.8 Other mental and behavioural disorders
292.89 Amphetamine-induced sleep disorder[I,W]	.8 Other mental and behavioural disorders
292.9 Amphetamine-related disorder NOS	.9 Unspecified mental and behavioural disorder
Caffeine-related disorders	F15.– Mental and behavioural disorders due to use of other stimulants, including caffeine
Caffeine-induced disorders	
305.90 Caffeine intoxication	.0 Acute intoxication
292.89 Caffeine-induced anxiety disorder[I]	.8 Other mental and behavioural disorders
292.89 Caffeine-induced sleep disorder[I]	.8 Other mental and behavioural disorders
292.9 Caffeine-related disorder NOS	.9 Unspecified mental and behavioural disorder

DSM-IV (ICD-9-CM)	ICD-10
Cannabis-related disorders	F12.– Mental and behavioural disorders due to use of cannabinoids
Cannabis use disorders	
304.30 Cannabis dependence[a]	.2 Dependence
305.20 Cannabis abuse	.1 Harmful use
Cannabis-induced disorders	
292.89 Cannabis intoxication	.0 Acute intoxication
Specify if: with perceptual disturbances	.04 With perceptual disturbances
292.81 Cannabis intoxication delirium	.03 With delirium
292.xx Cannabis-induced psychotic disorder	.5 Psychotic disorder
.11 With delusions[I]	.51 Predominantly delusional
.12 With hallucinations[I]	.52 Predominantly hallucinatory
292.89 Cannabis-induced anxiety disorder[I]	.8 Other mental and behavioural disorders
292.9 Cannabis-related disorder NOS	.9 Unspecified mental and behavioural disorders
Cocaine-related disorders	F14.– Mental and behavioural disorders due to use of cocaine
Cocaine use disorders	
304.20 Cocaine dependence[a]	.2 Dependence
305.60 Cocaine abuse	.1 Harmful use
Cocaine-induced disorders	
292.89 Cocaine intoxication	.0 Acute intoxication
Specify if: with perceptual disturbances	.04 With perceptual disturbances
292.0 Cocaine withdrawal	.3 Withdrawal
292.81 Cocaine intoxication delirium	.03 Acute intoxication with delirium
292.xx Cocaine-induced psychotic disorder	.5 Psychotic disorder
.11 With delusions[I]	.51 Predominantly delusional
.12 With hallucinations[I]	.52 Predominantly hallucinatory
292.84 Cocaine-induced mood disorder[I,W]	.5 Psychotic disorder
	.54 Predominantly depressive psychotic symptoms
	.55 Predominantly manic psychotic symptoms
	.56 Mixed
	.72 Residual affective disorder
292.89 Cocaine-induced anxiety disorder[I,W]	.8 Other mental and behavioural disorders
292.89 Cocaine-induced sexual dysfunction[I]	.8 Other mental and behavioural disorders
292.89 Cocaine-induced sleep disorder[I,W]	.8 Other mental and behavioural disorders
292.9 Cocaine-related disorder NOS	.9 Unspecified mental and behavioural disorder
Hallucinogen-related disorders	F16.– Mental and behavioural disorders due to use of hallucinogens
Hallucinogen use disorders	
304.50 Hallucinogen dependence[a]	.2 Dependence
305.30 Hallucinogen abuse	.1 Harmful use
Hallucinogen-induced disorders	
292.89 Hallucinogen intoxication	.0 Acute intoxication
292.89 Hallucinogen persisting perception disorder (flashbacks)	.70 Flashbacks
292.81 Hallucinogen intoxication delirium	.03 Acute intoxication with delirium
292.xx Hallucinogen-induced psychotic disorder	.5 Psychotic disorder
.11 With delusions[I]	.51 Predominantly delusional
.12 With hallucinations[I]	.52 Predominantly hallucinatory
292.84 Hallucinogen-induced mood disorder[I]	.5 Psychotic disorder
	.54 Predominantly depressive psychotic symptoms
	.55 Predominantly manic psychotic symptoms
	.56 Mixed
	.72 Residual affective disorder
292.89 Hallucinogen-induced anxiety disorder[I]	.8 Other mental and behavioural disorders
292.9 Hallucinogen-related disorder NOS	.9 Unspecified mental and behavioural disorder
Inhalant-related disorders	F18.– Mental and behavioural disorders due to use of volatile solvents
Inhalant use disorders	

DSM-IV (ICD-9-CM)	ICD-10
304.60 Inhalant dependence[a]	.2 Dependence
305.90 Inhalant abuse	.1 Harmful use
Inhalant-induced disorders	
292.89 Inhalant intoxication	.0 Acute intoxication
292.81 Inhalant intoxication delirium	.03 Acute intoxication with delirium
292.82 Inhalant-induced persisting dementia	.73 Dementia
292.xx Inhalant-induced psychotic disorder	.5 Psychotic disorder
.11 With delusions[I]	.51 Predominantly delusional
.12 With hallucinations[I]	.52 Predominantly hallucinatory
292.84 Inhalant-induced mood disorder[I]	.5 Psychotic disorder
	.54 Predominantly depressive psychotic symptoms
	.55 Predominantly manic psychotic symptoms
	.56 Mixed
	.72 Residual affective disorder
292.89 Inhalant-induced anxiety disorder[I]	.8 Other mental and behavioural disorders
292.9 Inhalant-related disorder NOS	.9 Unspecified mental and behavioural disorders
Nicotine-related disorders	F17.– Mental and behavioural disorders due to use of tobacco
Nicotine use disorder	
305.10 Nicotine dependence[a]	.2 Dependence
Nicotine-induced disorder	
292.0 Nicotine withdrawal	.3 Withdrawal
292.9 Nicotine-related disorder NOS	.9 Unspecified mental and behavioural disorder
Opioid-related disorders	F11.- Mental and behavioural disorders due to use of opioids
Opioid use disorders	
304.0 Opioid dependence	.2 Dependence
305.50 Opioid abuse	.1 Harmful use
Opioid-induced disorders	
292.89 Opioid intoxication	.0 Acute intoxication
Specify if: with perceptual disturbances	.04 With perceptual disturbances
292.0 Opioid withdrawal	.3 Withdrawal
292.81 Opioid intoxication delirium	.03 Acute intoxication with delirium
292.xx Opioid-induced psychotic disorder	.5 Psychotic disorder
.11 With delusions[I]	.51 Predominantly delusional
.12 With hallucinations[I]	.52 Predominantly hallucinatory
292.84 Opioid-induced mood disorder[I]	.5 Psychotic disorder
	.54 Predominantly depressive psychotic symptoms
	.55 Predominantly manic psychotic symptoms
	.56 Mixed
	.72 Residual affective disorder
292.89 Opioid-induced sexual dysfunction[I]	.8 Other mental and behavioural disorders
292.89 Opioid-induced sleep disorder[I,W]	.8 Other mental and behavioural disorders
292.9 Opioid-related disorder NOS	.9 Unspecified mental and behavioural disorder
Phencyclidine (or phencyclidine-like)-related disorders	F19.– Mental and behavioural disorders due to multiple drug use and use of other psychoactive substances
Phencyclidine use disorders	
304.90 Phencyclidine dependence	.2 Dependence
305.90 Phencyclidine abuse	.1 Harmful use
Phencyclidine-induced disorders	
292.89 Phencyclidine intoxication	.0 Acute intoxication
Specify if: with perceptual disturbances	.04 With perceptual disturbances
292.81 Phencyclidine intoxication delirium	.03 With delirium
292.xx Phencyclidine-induced psychotic disorder	.5 Psychotic disorder
.11 With delusions[I]	.51 Predominantly delusional
.12 With hallucinations[I]	.52 Predominantly hallucinatory
292.84 Phencyclidine-induced mood disorder[I]	.5 Psychotic disorder
	.54 Predominantly depressive psychotic symptoms
	.55 Predominantly manic psychotic symptoms
	.56 Mixed

DSM-IV (ICD-9-CM)	ICD-10
	.72 Residual affective disorder
292.89 Phencyclidine-induced anxiety disorder[I]	.8 Other mental and behavioural disorders
292.9 Phencyclidine-related disorder NOS	.9 Unspecified mental and behavioural disorder
Sedative-, hypnotic- or anxiolytic-related disorders	F13.– Mental and behavioural disorders due to use of sedatives or hypnotics
Sedative, hypnotic or anxiolytic use disorders	
304.10 Sedative, hypnotic or anxiolytic dependence[a]	.2 Dependence
305.40 Sedative, hypnotic or anxiolytic abuse	.1 Harmful use
Sedative-, hypnotic- or anxiolytic-induced disorders	
292.89 Sedative, hypnotic or anxiolytic intoxication	.0 Intoxication
292.0 Sedative, hypnotic or anxiolytic withdrawal	.3 Withdrawal
Specify if: with perceptual disturbances	
292.81 Sedative, hypnotic or anxiolytic intoxication delirium	.03 Acute intoxication with delirium
292.81 Sedative, hypnotic or anxiolytic withdrawal delirium	.4 Withdrawal state with delirium
292.82 Sedative-, hypnotic- or anxiolytic-induced persisting dementia	.73 Dementia
292.83 Sedative-, hypnotic- or anxiolytic-induced persisting amnestic disorder	.6 Amnesic disorder
292.xx Sedative-, hypnotic- or anxiolytic-induced psychotic disorder	.5 Psychotic disorder
.11 With delusions[I]	.51 Predominantly delusional
.12 With hallucinations[I]	.52 Predominantly hallucinatory
292.84 Sedative-, hypnotic- or anxiolytic-induced mood disorder[I,W]	.5 Psychotic disorder
	.54 Predominantly depressive psychotic symptoms
	.55 Predominantly manic psychotic symptoms
	.56 Mixed
	.72 Residual affective disorder
292.89 Sedative-, hypnotic- or anxiolytic-induced anxiety disorder[W]	.8 Other mental and behavioural disorders
292.89 Sedative-, hypnotic- or anxiolytic-induced sexual dysfunction[I]	.8 Other mental and behavioural disorders
292.89 Sedative-, hypnotic- or anxiolytic-induced sleep disorder[I,W]	.8 Other mental and behavioural disorders
292.9 Sedative-, hypnotic- or anxiolytic-related disorder NOS	.9 Unspecified mental and behavioural disorder
Polysubstance-related disorder	F19.– Mental and behavioural disorders due to multiple drug use and use of other psychoactive substances
304.80 Polysubstance dependence[a]	.2 Dependence
Other (or unknown) substance-related disorders	F19.– Mental and behavioural disorders due to multiple drug use and use of other psychoactive substances
Other (or unknown) substance use disorders	
304.90 Other (or unknown) substance dependence[a]	.2 Dependence
305.90 Other (or unknown) substance abuse	.1 Harmful use
Other (or unknown) substance-induced disorders	
292.89 Other (or unknown) substance intoxication	.0 Intoxication
Specify if: with perceptual disturbances	.04 With perceptual disturbances
292.0 Other (or unknown) substance withdrawal	.3 Withdrawal
Specify if: with perceptual disturbances	
292.81 Other (or unknown) substance-induced delirium	.03 Acute intoxication with delirium or .4 Withdrawal state with delirium
292.82 Other (or unknown) substance-induced persisting dementia	.73 Dementia
292.83 Other (or unknown) substance-induced persisting amnestic disorder	.6 Amnesic disorder
292.xx Other (or unknown) substance-induced psychotic disorder	.5 Psychotic disorder

DSM-IV (ICD-9-CM)	ICD-10
.11 With delusions[I,W]	.51 Predominantly delusional
.12 With hallucinations[I,W]	.52 Predominantly hallucinatory
292.84 Other (or unknown) substance-induced mood disorder[I,W]	.5 Psychotic disorder
	.54 Predominantly depressive psychotic symptoms
	.55 Predominantly manic psychotic symptoms
	.56 Mixed
	.72 Residual affective disorder
292.89 Other (or unknown) substance-induced anxiety disorder[I,W]	.8 Other mental and behavioural disorders
292.89 Other (or unknown) substance-induced sexual dysfunction[I]	.8 Other mental and behavioural disorders
292.89 Other (or unknown) substance-induced sleep disorder[I,W]	.8 Other mental and behavioural disorders
292.9 Other (or unknown) substance-related disorder NOS	.9 Unspecified mental and behavioural disorder or F55 Abuse of non-dependence-producing substances
Schizophrenia and other psychotic disorders	
295.xx Schizophrenia	F20 Schizophrenia
The following classification of longitudinal course applies to all subtypes of schizophrenia:	A fifth character may be used to classify course:
– Episodic with inter-episode residual symptoms (*specify if*: with prominent negative symptoms)	.x1 Episodic with progressive deficit
– Episodic with no inter-episode residual symptoms	.x2 Episodic with stable deficit
– Continuous (specify if: with prominent negative symptoms)	.x3 Episodic remittent
	.x0 Continuous
– Single episode in partial remission (*specify if*: with prominent negative symptoms/single episode in full remission)	.x4 Incomplete remission
	.x5 Complete remission
– Other or unspecified pattern	.x8 Other
	.x9 Course uncertain, period of observation too short
.30 Paranoid type	F20.0 Paranoid schizophrenia
.10 Disorganized type	F20.1 Hebephrenic schizophrenia
.20 Catatonic type	F20.2 Catatonic schizophrenia
.90 Undifferentiated type	F20.3 Undifferentiated schizophrenia
.60 Residual type	F20.5 Residual schizophrenia
295.40 Schizophreniform disorder *Specify if*: without good prognostic features/with good prognostic features	F20.x Schizophrenia, with duration less than 6 months
295.70 Schizoaffective disorder	F25 Schizoaffective disorders
Specify type:	
Bipolar type	.0 Manic type
Depressive type	.1 Depressive type
297.1 Delusional disorder *Specify type*: erotomanic type/grandiose type/jealous type/persecutory type/somatic type/mixed type/unspecified type	F22.0 Delusional disorder
298.8 Brief psychotic disorder	F23.1 Acute polymorphic psychotic disorder with symptoms of schizophrenia or F23.0, F23.2, F23.3, F23.8 and F23.9 with a duration of less than 6 months
Specify if:	
With marked stressor(s)	.x1 With associated acute stress
Without marked stressor(s)	.x0 Without associated acute stress
With postpartum onset	plus O99.3 Mental disorders and diseases of the nervous system complicating pregnancy, childbirth and the puerperium
297.3 Shared psychotic disorder	F24 Induced delusional disorder
293.xx Psychotic disorder due to ... (indicate the general medical condition)	
.81 With delusions	F06.2 Organic delusional disorder
.82 With hallucinations	F06.0 Organic hallucinosis

DSM-IV (ICD-9-CM)	ICD-10
---.- Substance-induced psychotic disorder (refer to substance-related disorders for substance-specific codes) *Specify if*: with onset during intoxication/with onset during withdrawal	F1x.5 Psychotic disorders due to psychoactive substance use
298.9 Psychotic disorder NOS	F29 Unspecified non-organic psychosis
Mood disorders *Code current state of major depressive disorder or bipolar I disorder in fifth digit*: 1=Mild 2=Moderate 3=Severe without psychotic features 4=Severe with psychotic features *Specify if*: mood-congruent psychotic features/mood-incongruent psychotic features 5=In partial remission 6=In full remission 0=Unspecified *The following specifiers apply (for current or most recent episode) to mood disorders as noted*: [a]Severity/psychotic/remission specifiers/[b]chronic/[c]with catatonic features/[d]with melancholic features/[e]with atypical features/[f]with postpartum onset *The following specifiers apply to mood disorders as noted*: [g]With or without full inter-episode recovery/[h]with seasonal pattern/[i]with rapid cycling	
Depressive disorders	
296.xx Major depressive disorder	
.2x Single episode[a–f]	F32.x Depressive episode .0 Mild .1 Moderate .2 Severe without psychotic symptoms .3 Severe with psychotic symptoms .4 In remission .8 Other (catatonic or atypical features) .9 Unspecified
.3x Recurrent[a–h]	F33.x Recurrent depressive disorder .0 Mild .1 Moderate .2 Severe without psychotic symptoms .3 Severe with psychotic symptoms .4 In remission .8 Other (catatonic or atypical features) .9 Unspecified
300.4 Dysthymic disorder *Specify if*: early onset/late onset *Specify if*: with atypical features	F34.1 Dysthymia
311 Depressive disorder NOS	F32.9 Depressive episode, unspecified
Bipolar disorders	
296.xx Bipolar I disorder	
.x Single manic episode[a,c,f]	F30.x Manic episode .1 Without psychotic symptoms .2 With psychotic symptoms
Specify if: mixed	F38.00 Mixed affective episode
	F31.x Bipolar affective disorder, current episode
.40 Most recent episode hypomanic[g–i]	.0 Hypomanic
.4x Most recent episode manic[a,c,f,g–i]	.1 Manic without psychotic symptoms or .2 Manic with psychotic symptoms
.6x Most recent episode mixed[a,c,f,g–i]	.6 Mixed

DSM-IV (ICD-9-CM)	ICD-10
.5x Most recent episode depressed[a–i]	.3 Mild or moderate depression or .4 Severe depression without psychotic symptoms or .5 Severe depression with psychotic symptoms
.7 Most recent episode unspecified[g–i]	.7 In remission or F31.9 Bipolar affective disorder, unspecified
296.89 Bipolar II disorder[a–i]	F31.x Bipolar affective disorder
Specify if: (current or most recent episode): hypomanic/depressed	.0 Current episode hypomanic
	.3 Current episode mild or moderate depression
	.4 Current episode severe depression without psychotic symptoms
	.5 Current episode severe depression with psychotic symptoms
– Catatonic or atypical features	F31.8 Other bipolar affective disorder (catatonic or atypical features)
301.13 Cyclothymic disorder	F34.0 Cyclothymia
296.80 Bipolar disorder NOS	F31.9 Bipolar affective disorder, unspecified
Other mood disorders	
293.83 Mood disorder due to ... (indicate the general medical condition)	F06.3 Organic mood (affective) disorder
Specify type	
With depressive features/with major depressive-like episode	.32 Organic depressive disorder
With manic features	.30 Organic manic disorder
With mixed features	.33 Organic mixed affective disorder
---.– Substance-induced mood disorder (refer to substance-related disorders for substance-specific codes)	F1x.5x Psychotic disorder due to psychoactive substance use
Specify type: with depressive features/with manic features/with mixed features	.54 Predominantly depressive symptoms
	.55 Predominantly manic symptoms
	.56 Predominantly mixed symptoms
	F1x.72 Residual affective disorder
Specify if: with onset during intoxication/with onset during withdrawal	
296.90 Mood disorder NOS	F39 Unspecified mood (affective) disorder
Anxiety disorders	
300.01 Panic disorder without agoraphobia	F41.0 Panic disorder
300.21 Panic disorder with agoraphobia	F40.01 Agoraphobia with panic disorder
300.22 Agoraphobia without history of panic disorder	F40.00 Agoraphobia without panic disorder
300.29 Specific phobia	F40.2 Specific (isolated) phobias
Specify type: animal type/natural environment type/blood-injection-injury type/situational type/other type	
300.23 Social phobia	F40.1 Social phobias
Specify if: generalized	
300.3 Obsessive–compulsive disorder	F42.x Obsessive–compulsive disorder
Specify if: with poor insight	.0 Predominantly obsessional thoughts or ruminations
	.1 Predominantly compulsive acts
	.2 Mixed obsessional thoughts and acts
	.8 Other obsessive–compulsive disorders
	.9 Obsessive–compulsive disorder, unspecified
309.81 Post-traumatic stress disorder	F43.1 Post-traumatic stress disorder
Specify if: acute/chronic	
Specify if: with delayed onset	
308.3 Acute stress disorder	F43.0 Acute stress disorder
300.02 Generalized anxiety disorder	F41.1 Generalized anxiety disorder
293.89 Anxiety disorder due to ... (indicate the general medical condition)	F06.4 Organic anxiety disorder
Specify if: with generalized anxiety/with panic attacks/with obsessive–compulsive symptoms	
---.– Substance-induced anxiety disorder (refer to substance-related disorders for substance-specific codes)	F1x.8 Other mental and behavioural disorders due to use of psychoactive substances

DSM-IV (ICD-9-CM)	ICD-10
Specify type: with generalized anxiety/with panic attacks/with obsessive-compulsive symptoms/with phobic symptoms *Specify if*: with onset during intoxication/with onset during withdrawal	
300.00 Anxiety disorder NOS	F41.9 Anxiety disorder, unspecified
Somatoform disorders	
300.81 Somatization disorder	F45.0 Somatization disorder
300.81 Undifferentiated somatoform disorder	F45.1 Undifferentiated somatoform disorder
300.11 Conversion disorder	F44.x Dissociative (conversion) disorders
Specify type:	
With motor symptom or deficit	.4 Dissociative motor disorders
With sensory symptom or deficit	.6 Dissociative anaesthesia and sensory loss
With seizures or convulsions	.5 Dissociative convulsions
With mixed presentation	.7 Mixed dissociative (conversion) disorders
307.8x Pain disorder	F45.4 Persistent somatoform pain disorder
.80 Associated with psychological factors	
.89 Associated with both psychological factors and a general medical condition	
Specify if: acute/chronic	
300.7 Hypochondriasis	F45.2 Hypochondriacal disorders
Specify if: with poor insight	
300.7 Body dysmorphic disorder	F45.2 Hypochondriacal disorders
300.81 Somatoform disorder NOS	F45.9 Somatoform disorder, NOS
Factitious disorders	
300.xx Factitious disorder	F68.1 Intentional production or feigning of symptoms or disabilities, either physical or psychological (factitious disorder)
.16 With predominantly psychological signs and symptoms	
.19 With predominantly physical signs and symptoms	
.19 With combined psychological and physical signs and symptoms	
300.19 Factitious disorder NOS	
Dissociative disorders	
300.12 Dissociative amnesia	F44.0 Dissociative amnesia
300.13 Dissociative fugue	F44.1 Dissociative fugue
300.14 Dissociative identity disorder	F44.81 Multiple personality disorder
300.6 Depersonalization disorder	F48.1 Depersonalization–derealization syndrome
300.15 Dissociative disorder NOS	F44.9 Dissociative (conversion) disorder, unspecified
Sexual and gender identity disorders	
Sexual dysfunctions *The following specifiers apply to all primary sexual dysfunctions*: lifelong type/acquired type/generalized type/situational type/due to psychological factors/due to combined factors	
Sexual desire disorders	
302.71 Hypoactive sexual desire disorder	F52.0 Lack or loss of sexual desire
302.79 Sexual aversion disorder	F52.10 Sexual aversion
Sexual arousal disorders	
302.72 Female sexual arousal disorder	F52.2 Failure of genital response
302.72 Male erectile disorder	F52.2 Failure of genital response
Orgasmic disorder	
302.73 Female orgasmic disorder	F52.3 Orgasmic dysfunction
302.74 Male orgasmic disorder	
302.75 Premature ejaculation	F52.4 Premature ejaculation
Sexual pain disorders	
302.76 Dyspareunia (not due to a general medical condition)	F52.6 Non-organic dyspareunia

DSM-IV (ICD-9-CM)	ICD-10
306.51 Vaginismus (not due to a general medical condition)	F52.5 Non-organic vaginismus
Sexual dysfunction due to a general medical condition	F06.8 Other specified mental disorders due to brain damage and dysfunction and to physical disease
625.8 Female hypoactive sexual desire disorder due to ... (indicate the general medical condition)	Plus N94.8 Other specified conditions associated with female genital organs and menstrual cycle
608.89 Male hypoactive sexual desire disorder due to ... (indicate the general medical condition)	Plus N50.8 Other specified disorders of male genital organs
607.84 Male erectile disorder due to ... (indicate the general medical condition)	Plus N48.4 Impotence of organic origin
625.0 Female dyspareunia due to ... (indicate the general medical condition)	Plus N94.1 Dyspareunia
608.89 Male dyspareunia due to ... (indicate the general medical condition)	Plus N50.8 Other specified disorders of male genital organs
625.8 Other female sexual dysfunction due to ... (indicate the general medical condition)	Plus N94.8 Other specified conditions associated with female genital organs and menstrual cycle
608.89 Other male sexual dysfunction due to... (indicate the general medical condition)	Plus N50.8 Other specified disorders of male genital organs
302.70 Sexual dysfunction NOS	F52.9 Unspecified sexual dysfunction, not caused by organic disorder or disease
Paraphilias	
302.4 Exhibitionism	F65.2 Exhibitionism
302.81 Fetishism	F65.0 Fetishism
302.89 Frotteurism	F65.8 Other disorders of sexual preference
302.2 Paedophilia	F65.4 Paedophilia
Specify if: sexually attracted to males/sexually attracted to females/sexually attracted to both	
Specify if: limited to incest	
Specify if: exclusive type/non-exclusive type	
302.83 Sexual masochism	F65.5 Sadomasochism
302.84 Sexual sadism	F65.5 Sadomasochism
302.3 Transvestic fetishism	F65.1 Fetishistic transvestism
Specify if: with gender dysphoria	
302.82 Voyeurism	F65.3 Voyeurism
302.9 Paraphilia NOS	F65.9 Disorder of sexual preference, unspecified
Gender identity disorders	
302.xx Gender identity disorder	
.6 In children	F64.2 Gender identity disorder of childhood
.85 In adolescents or adults	F64.0 Transsexualism or F64.1 Dual-role transvestism
Specify if: sexually attracted to males/sexually attracted to females/sexually attracted to both/sexually attracted to neither	
302.6 Gender identity disorder NOS	F64.9 Gender identity disorder, unspecified
302.9 Sexual disorder NOS	No equivalent
Eating disorders	
307.1 Anorexia nervosa	F50.0 Anorexia nervosa
Specify type:	
Restricting type/purging type	
Binge-eating type	F50.3 Atypical bulimia nervosa
307.51 Bulimia nervosa	F50.2 Bulimia nervosa
Specify type: purging type/non-purging type	
307.50 Eating disorder NOS	F50.9 Eating disorder, unspecified
Sleep disorders	
Primary sleep disorders	
Dyssomnias	
307.42 Primary insomnia	F51.0 Non-organic insomnia

DSM-IV (ICD-9-CM)	ICD-10
307.44 Primary hypersomnia	F51.1 Non-organic hypersomnia
Specify if: recurrent	
347 Narcolepsy	F51.8 Other non-organic sleep disorders plus G47.4 Narcolepsy
780.59 Breathing-related sleep disorder	F51.8 Other non-organic sleep disorders plus G47.3 Sleep apnoea
307.45 Circadian rhythm sleep disorder	F51.2 Non-organic disorder of sleep–wake schedule
Specify type: delayed sleep phase type/jet lag type/shift work type/unspecified type	
307.47 Dyssomnia NOS	F51.9 Non-organic sleep disorder, unspecified
Parasomnias	
307.47 Nightmare disorder	F51.5 Nightmares
307.46 Sleep terror disorder	F51.4 Sleep terrors
307.46 Sleepwalking disorder	F51.3 Sleepwalking (somnambulism)
307.47 Parasomnia NOS	F51.9 Non-organic sleep disorders, unspecified
Sleep disorders related to another mental disorder	
307.42 Insomnia related to ... (indicate the axis I or axis II disorder)	F51.0 Non-organic insomnia
307.44 Hypersomnia related to ... (indicate the axis I or axis II disorder)	F51.1 Non-organic hypersomnia
Other sleep disorders	
780.xx Sleep disorder due to ... (indicate the general medical condition)	F06.8 Other specified mental disorders due to brain damage and dysfunction and to physical disease plus G47.x Sleep disorders
.52 Insomnia type	G47.0 Insomnias
.54 Hypersomnia type	G47.1 Hypersomnia
.59 Parasomnia type	G47.8 Other sleep disorders
.59 Mixed type	G47.8 Other sleep disorders
---.- Substance-induced sleep disorder (refer to substance-related disorders for substance-specific codes)	F1x.8 Other mental and behavioural disorders due to psychoactive substance use
Specify type: insomnia type/hypersomnia type/parasomnia type/mixed type	
Specify if: with onset during intoxication/with onset during withdrawal	
Impulsecontrol disorders not elsewhere classified	
312.34 Intermittent explosive disorder	F63.8 Other habit and impulse disorders
312.32 Kleptomania	F63.2 Pathological stealing (kleptomania)
312.33 Pyromania	F63.1 Pathological fire-setting (pyromania)
312.31 Pathological gambling	F63.0 Pathological gambling
312.39 Trichotillomania	F63.3 Trichotillomania
312.30 Impulse-control disorder NOS	F63.9 Habit and impulse disorder, unspecified
Adjustment disorders	
309.xx Adjustment disorder	F43.2 Adjustment disorders
.0 With depressed mood	.20 Brief depressive reaction or .21 Prolonged depressive reaction
.24 With anxiety	.23 With predominant disturbance of other emotions
.28 With mixed anxiety and depressed mood	.22 Mixed anxiety and depressive reaction
.3 With disturbance of conduct	.24 With predominant disturbance of conduct
.4 With mixed disturbance of emotions and conduct	.25 With mixed disturbance of emotions and conduct
.9 Unspecified	F43.2 Adjustment disorders
Specify if: acute/chronic	
Personality disorders	
Note: These are coded on axis II.	
301.0 Paranoid personality disorder	F60.0 Paranoid personality disorder
301.20 Schizoid personality disorder	F60.1 Schizoid personality disorder
301.22 Schizotypal personality disorder	F21 Schizotypal disorder
301.7 Antisocial personality disorder	F60.2 Dyssocial personality disorder
301.83 Borderline personality disorder	F60.31 Emotionally unstable personality disorder, borderline type

DSM-IV (ICD-9-CM)	ICD-10
301.50 Histrionic personality disorder	F60.4 Histrionic personality disorder
301.81 Narcissistic personality disorder	F60.8 Other specific personality disorders
301.82 Avoidant personality disorder	F60.6 Anxious (avoidant) personality disorder
301.6 Dependent personality disorder	F60.7 Dependent personality disorder
301.4 Obsessive–compulsive personality disorder	F60.5 Anankastic personality disorder
301.9 Personality disorder NOS	F60.9 Personality disorder, unspecified
Other conditions that may be a focus of clinical attention	
Psychological factors affecting medical condition	
316 ... (Specified psychological factor) affecting ... (indicate the medical condition) Choose name based on nature of factors: – Mental disorder affecting medical condition – Psychological symptoms affecting medical condition – Personality traits or coping style affecting medical condition – Maladaptive health behaviours affecting medical condition – Stress-related physiological response affecting medical condition – Other or unspecified psychological factors affecting medical condition	F54 Psychological and behavioural factors associated with disorders or diseases classified elsewhere
Medication-induced movement disorders	
332.1 Neuroleptic-induced parkinsonism	G21.1 Other drug-induced secondary parkinsonism
333.92 Neuroleptic malignant syndrome	G21.0 Malignant neuroleptic syndrome
333.7 Neuroleptic-induced acute dystonia	G24.0 Drug-induced dystonia
333.99 Neuroleptic-induced acute akathisia	G21.1 Other drug-induced secondary parkinsonism
333.82 Neuroleptic-induced tardive dyskinesia	G24.0 Drug-induced dystonia
333.1 Medication-induced postural tremor	G25.1 Drug-induced tremor
333.90 Medication-induced movement disorder NOS	G25.9 Extrapyramidal and movement disorder, unspecified
Other medication-induced disorder	
995.2 Adverse effects of medication NOS	T88.7 Unspecified adverse effect of drug or medicament
Relational problems	
V61.9 Relational problem related to a mental disorder or general medical condition	Z63.7 Other stressful life events affecting family and household
V61.20 Parent–child relational problem	Z63.8 Other specified problems related to primary support group
V61.1 Partner relational problem	Z63.0 Problems in relationship with spouse or partner
V61.8 Sibling relational problem	F93.3 Sibling rivalry disorder
V62.81 Relational problem NOS	Z63.9 Problem related to primary support group, unspecified
Problems related to abuse or neglect	
V61.21 Physical abuse of child (code 995.5 if focus of attention is on victim)	T74.1 Physical abuse
V61.21 Sexual abuse of child (code 995.5 if focus of attention is on victim)	T74.2 Sexual abuse
V61.21 Neglect of child (code 995.5 if focus of attention is on victim)	T74.0 Neglect or abandonment
V61.1 Physical abuse of adult (code 995.5 if focus of attention is on victim)	T74.1 Physical abuse
V61.1 Sexual abuse of adult (code 995.5 if focus of attention is on victim)	T74.2 Sexual abuse
Additional conditions that may be a focus of clinical attention	
V15.81 Non-compliance with treatment	Z91.1 Personal history of non-compliance with medical treatment and regimen
V65.2 Malingering	Z76.5 Malingerer (conscious simulation)
V71.01 Adult antisocial behaviour	Z72.8 Other problems related to lifestyle
V71.02 Child or adolescent antisocial behaviour	Z72.8 Other problems related to lifestyle
V62.89 Borderline intellectual functioning	R41.8 Other and unspecified symptoms and signs involving cognitive functions and awareness
Note: *This is coded on axis II*	

DSM-IV (ICD-9-CM)	ICD-10
780.9 Age-related cognitive decline	R41.8 Other and unspecified symptoms and signs involving cognitive functions and awareness
V62.82 Bereavement	Z63.4 Disappearance and death of family member
V62.3 Academic problem	Z55.8 Other problems related to education and literacy
V62.2 Occupational problem	Z56.7 Other and unspecified problems related to employment
313.82 Identity problem	F93.8 Other childhood emotional problems
V62.89 Religious or spiritual problem	Z71.8 Other specified counselling
V62.4 Acculturation problem	Z60.3 Acculturation difficulty
V62.89 Phase of life problem	Z60.0 Problems of adjustment to life-cycle transitions
Additional codes	
300.9 Unspecified mental disorder (non-psychotic)	F99 Mental disorder, not otherwise specified
V71.09 No diagnosis or condition on axis I	Z03.2 Observation for suspected mental and behavioural disorder
799.9 Diagnosis or condition deferred in axis I	Z03.2 Observation for suspected mental and behavioural disorder
V71.09 No diagnosis on axis II	Z03.2 Observation for suspected mental and behavioural disorder
799.9 Diagnosis deferred on axis II	Z03.2 Observation for suspected mental and behavioural disorder
Multiaxial system	**Multiaxial presentation**
Axis I Clinical disorders	Axis I Clinical diagnoses
Other conditions that may be a focus of clinical attention	
Axis II Personality disorders	Axis I Clinical diagnoses
Mental retardation	Axis I Clinical diagnoses
Axis III General medical conditions	Axis I Clinical diagnoses
Axis IV Psychosocial and environmental problems	Axis III Contextual factors
Axis V Global assessment of functioning	Axis II Disability

Codes that precede the DSM-IV category are ICD-9-CM codes (*International Classification of Diseases*, 9th revision, clinical modification, 2nd edn, 1980. U.S. Government Printing Office. Washington DC, 20402). There are several differences between ICD-9 and ICD-9-CM, mainly on the level of the fourth digit. A conversion table between ICD-9 and ICD-9-CM is available on the Internet at www.who.ch: WHO Headquarters, Major Programmes; Division of Mental Health and Prevention of Substance Abuse; ICD-10 Conversion Tables.

NOS, not otherwise specified.

CHAPTER

5

A. Sims, S. Curran

Examination of the Psychiatric Patient

1 Underlying Principles

1.1 Purpose

Psychiatric treatment, using any recognised intervention, is based upon information concerning the present state and past history of the individual patient, described as the psychiatric examination. This has two essential components which are described in this chapter: history-taking and examination of mental state. Any medical examination is potentially experienced by the patient as unpleasant, and psychiatric examination is no exception. It is also time-consuming, and each item should therefore be carried out for a specific purpose. It is also important to ask the patient's permission and explain the procedure and its purpose beforehand. Repeated examinations of the same patient by different members of the clinical team should be discouraged, as this is unpleasant for the patient and will tend to undermine the value of the procedure.

1.2 Relationship Between Patient and Doctor

An important consequence of carrying out a psychiatric examination, although not its primary justification, is the forging of a therapeutic relationship between patient and doctor. This is both explicit and implicit in the contract that is established between them during history-taking and mental state examination. Explicit is the enquiry by the doctor for detailed information; implicit is that the doctor knows what questions need to be asked and is compassionate in demonstrating concern for the patient's distress.

These attributes of the doctor (competence, knowledge of the patient's problems and concern) help establish a therapeutic relationship and are equally important whatever the nature of treatment. Patients are sensitive to what they construe as the doctor being judgmental, and this is especially true at the initial interview.

1.3 Diagnosis

The diagnosis consists not of a single word or phrase encompassing everything about the patient, but rather a diagnostic formulation (see Vol. 1, Part 2, Chap. 2). The formulation states the problem and outlines the differential diagnosis, aetiology, further investigations, plan of treatment and prognosis.

To make a competent psychiatric diagnosis, the interviewer needs theoretical knowledge of psychiatry and an ability to elicit psychopathological phenomena with precision. A checklist of diagnostic criteria is never enough on its own; "the proper use of these criteria requires specialised clinical training that provides both a body of knowledge and clinical skills" (American Psychiatric Association 1994).

1.4 Formulating Management

Medical diagnosis is the process of ascertaining the nature of the problem presented and categorising it in such a way that those problems may be solved from a menu of options. This will usually involve pharmacological, psychological and social intervention rather than a single method of treatment.

1.5 Specific Purposes

Obtaining information may be distressing for patients and is time-consuming for the doctor. The interviewer should therefore be able to give a reason for each question asked. As well as diagnosis, forging relationships and formulating treatment, the interview may be used for other purposes:

- *Monitoring change.* The doctor will wish to assess regularly the condition of those under treatment by enquiring about their current state.
- *Medico-legal report.* Although the general principles remain similar, there are specific considerations which are discussed later.
- *Research.* Data collection must be comprehensive and consistent in its collection so that comparisons may be made between subjects.

1.6 Structuring the Interview

The purpose of a particular interview and the information that is required should have been decided and discussed with the patient beforehand. A framework for history-taking and examination is used so that significant areas of information are not omitted and patients are not irritated by repeated questions.

While proceeding through the predetermined schedule, the clinician gives emphasis to those parts of the history and examination that are particularly signifi-

cant in the particular case. Clinicians should develop interviewing structures that suit their particular practice and style, and any published framework is a guideline, and not a directive.

1.6.1 Interviewing and Communication

Putting the patient at ease from the start necessitates attention to the setting, careful planning of the interview and confidence in carrying out the examination. The patient is usually anxious, and this should be recognised by the doctor. The patient will be more comfortable if it is explained how long the interview will take and what areas will be covered.

A meaningful psychiatric history requires skilful use of *open-ended* questions, *direct* questions in significant areas and clarification of clinically important topics while watching carefully for verbal and visual clues revealing the patient's emotional state. The interviewer should have prepared strategies for dealing with frequent situations, e.g., a garrulous patient or a patient who becomes upset during interview.

Examination of the mental state requires the interviewer to find out as precisely as possible what is the subjective experience of the patient. The interviewer's own capacity for subjective experience is enlisted in understanding the patient's description and recounting this back to the patient: the method of empathy (Sims 1995a).

Psychiatrists must be able to communicate with their patients, and this implies knowledge of the stages through which an interview goes and the attitudes and expectations of the patient at different stages (apprehensive withholding at the beginning, garrulous irrelevancy in the middle and significant disclosure at the end). Clinicians should be aware that their own attitudes, prejudices and expectations can modify information given by the patient.

1.6.2 Training for Interviewing

On commencing psychiatric practice, trainees need to talk with patients and to learn interviewing skills, including how to structure the interview, with examples of good practice to emulate from more experienced clinicians. For example, a senior psychiatrist may elicit delusional perception or a skilful social worker may demonstrate a comprehensive family history. Trainees should rehearse interview technique and have feedback on their performance under non-threatening conditions, preferably with colleagues using video techniques to record and monitor progress. Attention should be paid to the wording of questions, strategies for alleviating anxiety and timing, both for the whole interview and its constituent parts. A hospital involved in the training of psychiatrists should have a designated senior clinician, knowledgeable in educational methods and interview techniques, acting as supervisor for this training.

1.6.3 Potential Pitfalls

Some skills used in ordinary conversation have a perverse effect when used in psychiatric interviewing. Although an odd response received in normal conversation would usually be ignored and the questioner would pass on to another topic, in the psychiatric interview this could result in missing significant psychopathology. In addition, distress should not be played down, as this will prevent further disclosure. Reassuring the patient before they have fully explained their predicament and falsely reassuring when the prognosis is poor are both harmful approaches. Furthermore, changing from a topic of discussion before patients have fully explained themselves discourages them from proceeding further with their account. All these gambits may be used by an inexperienced interviewer to keep the consultation at a "safe" level which may make the interviewer feel more comfortable, but does not help the patient.

1.7 Facilitating the Interview

Some apparently minor considerations can make the process of carrying out an interview comfortable and relatively stress free for both patient and doctor, but if neglected can result in the examination failing to meet its objectives.

1.7.1 Setting

It is important to consider the comfort of the patient and the needs of the interview. The room should be clean, comfortable, well appointed and without unnecessary distractions, such as extraneous noise. The interview should start promptly, and the patient should be seated comfortably at about the same level as the interviewer and on one side, not across a desk or table.

1.7.2 Mental State of the Interviewer

The posture adopted by the doctor conveys a message to the patient. Slouching back in a chair may be interpreted as disrespectful, while craning forward on

one's elbows may communicate tension or aggression. The interviewer must concentrate totally and ideally should be tranquil, free of immediate emotional distraction and unhurried.

1.7.3 Duration

The interview should start on time, progress steadily, cover the required ground and end at the specified time with a satisfying conclusion. If the process is likely to be prolonged, e.g., with a new in-patient, it is better to use separate periods of about 1 h rather than have a lengthy session which becomes arduous for the patient.

1.7.4 Order of Questions

The order of questions can influence the answers given to them. It is best to order the interview according to a recognised structure and to persist with this, thus making serial interviews with different patients comparable.

1.7.5 Direct Questions

Early in the interview or when starting on any new topic, it is best to use open-ended questions. These should then be followed by direct questions, which should not be "leading questions", i.e., directing an answer expected by the questioner. The skill of asking questions is to combine open enquiry that will allow unexpected comment with obtaining specific information required by the interviewer.

2 Method of Examination

The following section describes in detail the separate steps in history-taking and examination of psychiatric patients; a framework suitable as a basis for clinical practice is given in Table 1.

2.1 Source and Reasons for Referral

The referral may come from the patient (self-referral), family or hospital doctor, social worker, solicitor, probation officer, employer or another person, and this influences what the patient says and expects from the interview. The reason for referral (e.g., family doctor

Table 1. Basic framework for psychiatric history and examination

Informant and source of referral
Informant
Source and reason for referral
Psychiatric history
Present illness (history of presenting complaint)
Nature
Time of onset
Development over time
Precipitating or relieving factors
Help given to date
Impact of the problem
Availability of support
Patient's view of his or her problems
Past history of similar psychiatric problems
Screening questions
Family history
Medical history
Psychiatric history
Personal history
Early life and development
Schooling
Behaviour problems
Occupation
Services or war experience
Personal relationships
Menstrual history
Sexual history
Children
Present social situation
Personality before present illness (premorbid personality)
Examination of mental state
Appearance and behaviour
Speech
Mood
Depersonalisation and derealisation
Obsessional phenomena
Delusions
Illusions and hallucinations
Orientation
Attention and concentration
Memory
Insight
Reaction to the patient

requesting recommendation for treatment) should be made explicit.

It is generally helpful to have an account from another person, especially when the patient is too disturbed to give a history or the facts are uncertain because of the mental state. Details concerning the informant should be recorded and comment made of both the patient's and the informant's reliability as historians.

2.2 Psychiatric History

A good psychiatric history must compromise between obtaining all pertinent information and common sense. For more detailed questioning, the interviewer should select those areas that are likely to be important. Verbatim quotation from the patient is usually better than psychiatric terminology.

2.2.1 Present Illness

The history of the presenting complaint is the key to understanding what is wrong with the patient and subsequently formulating treatment. It also helps forge the relationship between patient and doctor.

Nature
Complaints should be listed in the order of significance and described with precision, preferably in the patient's own words, as verbatim statements are both more graphic and more likely to convince clinicians who consult the case notes at a later date. Comment should be made upon severity and the degree of distress and social disability each symptom causes.

Timing
Important issues are the time of onset of each symptom, how long each episode lasts, whether symptoms are continuous or intermittent and what provokes these symptoms. The order of occurrence does not necessarily imply cause, but time relationships may help to confirm the diagnosis.

Development
Symptoms may change over time in their severity, prominence, the meaning they hold for the patient and their significance for the clinical picture. Sometimes change may be of diagnostic importance.

Precipitating or Relieving Factors
Most symptoms are influenced to some extent by outside circumstances; perhaps by distraction from other perceptions, by the time of day or by an appreciative audience. Violent behaviour becomes more likely with anxiety, uncertainty or constraint. Sometimes the attribution for change demonstrates abnormality of mental state, e.g., patients who believe that their thinking is disordered by an electronic machine.

Help Received So Far
Previous treatment, both pharmacological and psychosocial, should be recorded with its duration, dosage, any adverse reactions and relative effectiveness.

Impact of the Complaint
The impact of the complaint concerns the effects of symptoms upon work, social functioning and relationships and their association with physical functions.

Availability of Support
The availability of support should be assessed both from the practical aspect, i.e., whether people are available to help in a crisis, and the emotional one, i.e., whether supportive relationships are reassuring and encourage involvement in normal activity or whether they are restrictive and tend to collude with ideas of incapacity.

Patient's Attitudes Towards the Complaint
The patient's attitudes towards symptoms vary from neglect and denial to exaggerated anxiety and will be expressed differently to a psychiatrist than they will to relatives and friends. The meaning of symptoms for the patient should be explored, such as guilty ideas of incurability or heredity. Exploring such misconceptions is necessary in order to rectify them.

2.2.2 History of Similar Psychiatric Problems

A history of similar psychiatric problems gives valuable information about the natural history and therefore diagnosis of the illness.

Some conditions are episodic and occur in "attacks" against a background of normality; others are cyclical and may represent a pathological response to biological events such as menstruation, the puerperium and the time of day or season of the year. Psychiatric disorders may also be reactive to external social circumstances and the way they are perceived by the patient.

2.2.3 Family History

Enquiry about the family not only informs the clinician about the patient's inheritance and the possibility of genetic illness, but also about the environment and atmosphere in which he or she developed and now lives. This involves enquiry about family history of disease and evaluation of relationships. Illnesses, and deaths, with causes are recorded for close relatives. It is important to ascertain whether family membership is

biological or adoptive. Family history of psychiatric disorder, personality problems, epilepsy and other neurological disease is directly relevant. Relationships and separations between parents and siblings should be noted, and the clinician should assess whether family relationships are close and emotional support is forthcoming.

2.2.4 Previous Medical History

Significant illnesses and operations and the patient's reaction to these should be recorded.

2.2.5 Previous Psychiatric History

Other psychiatric disorders from which the patient has suffered may indicate the nature and course of the present condition. Full details of previous treatment should be noted, including the setting and the doctor who provided treatment.

2.2.6 Personal History

As the personal history provides the background from which the patient came and the context in which he or she now lives, it is clearly crucial in understanding the patient's state and response to illness.

Early Life and Development
Any significant features concerning the mother's pregnancy, the patient's birth, habit training and achieving milestones should be noted.

Childhood Problems
Social, physical and behavioural variations from normal during childhood, including parental separations, events concerning other siblings, emotional and behavioural problems and any notable illnesses of the patient are recorded.

Schooling
The interviewer should enquire about achievement, intellectual ability and capacity for relationships with peers and authority figures during the patient's schooling.

Occupation
The patient's occupations with dates at each place of work, reasons for changing, present financial situation and satisfaction with work are listed. The duration and geographical location of military service, highest rank achieved and any problems should be recorded.

Marital and Personal Relationships
When enquiring about marital and personal relationships, the age of patient and his or her spouse at marriage, relevant matters of courtship, previous relationships and the quality of the relationship should be noted.

Menstrual History
The patient's menstrual history includes the age of menarche, regularity of periods, any significant menstrual symptoms and attitudes of the patient towards menstruation.

Sexual History
Sexual history includes attitudes to sex, heterosexual and homosexual experience, use of contraception and any past experience of sexual abuse.

Children
Names, ages, gender, development, temperament and behaviour of a patient's children should be recorded with a note on the quality of relationships of the patient with each child and between siblings. Evidence of recent change in the relationship between the patient and the children is relevant.

2.2.7 Present Social Situation

The patient's present social situation concerns his or her financial, residential and legal situation and has a bearing on assessment of diagnosis and treatment plan. The interviewer must decide how much should be asked about each of the topics within the psychiatric history at the initial interview and what should be left until later when the doctor–patient relationship has become more securely established. The pursuit of useful information should be tempered by the importance of maintaining a relationship.

2.3 Premorbid Personality

Assessment of personality is a valuable part of the psychiatric examination, because, if carried out accurately, it can be predictive of future behaviour, including the response to mental illness. The interviewer should use an accepted classification of abnormal personality and its disorder, such as the International Classification of Diseases, 10th edition (World Health Organisation 1992). Personality assessment requires skilled observation of the patient during the interview and enquiry of another informant; it is predictive of future behaviour and responses patients might make to their illness. The interviewer looks for distinctive traits that indicate personality type and style of functioning in everyday life, assessing whether these traits have developed to an extent which would fall within a broad range of normal or whether they are

excessive. Abnormal personality, which does not of itself denote disorder, is present when a personality trait considered to be clinically significant is developed to an extent that does not conform with the majority. Personality disorder is present when this abnormality of personality causes either the patient or other people to suffer.

2.3.1 Relationships

The capacity for establishing and maintaining relationships, both intimate and companionable, is assessed. This includes the capacity for relationship with the same and the other sex and the ability to relate to superiors, equals and subordinates.

2.3.2 Use of Leisure

Leisure activities are examined, i.e., whether they are solitary or social, sedentary or physically active, creative, studious or aesthetic. Note is taken of hobbies, preferred activities and membership of societies, clubs and political or religious organisations.

2.3.3 Predominant Mood

The following points need to be considered: What are the characteristics of this individual's prevailing mood? Are they present long-term or do they fluctuate? If there are alterations in mood, how rapidly do they change and in response to what provocation? How is the patient's prevailing mood perceived by the patient and what effect does it have upon other people? Are there problems resulting from the predominant mood?

2.3.4 Character

Character is the consistent aspect of personality which determines usual behaviour, i.e., whether people can form and maintain close relationships, whether they can express feelings of love, anger, frustration and sadness and whether they lose their temper or lose control of their feelings.

2.3.5 Attitudes and Standards

Moral, religious and health attitudes are significant. Moral standards are concerned not only with judgements about right and wrong, but also the latitude one will allow oneself, the guilt associated with wrongdoing, and whether standards for one's own behaviour are different from the way one judges others. Religious values not only include religious affiliation but also beliefs and the effect these have upon emotions and external behaviour (Sims 1994). Patients' attitudes towards their own health, minor symptoms and sensations and their body should be assessed.

2.3.6 Habits

Habits concerning ingestion of food, alcohol, tobacco and other drugs should be assessed. More detailed enquiry should be undertaken if relevant.

2.4 Examination of the Mental State

Mental state examination not only explores the subjective experience of the patient, but also investigates the contents of the mind as signs for the diagnosis of psychiatric entities.

In examining the mental state, the interviewer learns a comprehensive scheme and then modifies it appropriately for each patient. Although at times history-taking and examination of mental state overlap (e.g., when delusional material intrudes into the history), conceptually and in the record they should be kept *separate*. The examination of mental state should also include observations of nursing staff for an in-patient or relatives for a patient at home. Not every phenomenon will be revealed at the first interview, and sometimes significant material only becomes apparent after several interviews. Examples of psychopathology should be recorded using verbatim statements of the patient. Evaluating the mental state involves comparing observed speech of the patient with what might be expected in that situation; this requires both accurate observation and a knowledge of normality.

2.4.1 Appearance and Behaviour

When assessing appearance and behaviour, the interviewer should concentrate on general appearance, the face, posture, movement and social behaviour. General appearance includes both the way patients present themselves in terms of clothing, personal hygiene, tidiness and any idiosyncrasies and also physical characteristics such as body build, evidence of recent weight gain or loss and generally looking well or ill. Facial appearance usually reveals current mood state such as irritability, perplexity. Posture and movement may also reveal mood or the presence of neuropsychiatric disturbances. Social behaviour assesses the

patient and his or her context, e.g., jocularity during a psychiatric interview is inappropriate. A patient with dementia may demonstrate loss of contact with reality through inappropriate behaviour.

2.4.2 Speech

Speech is the only reliable way of obtaining access to thought processes of the patient. Analysis of speech, of the words used and their inaccuracies and innuendoes, is particularly valuable in exploring psychopathology (see Vol. 1, Part 1, Chap. 1; Sims 1995b). Articulation and the stream of speech should be noted. The form of thought is evaluated by assessing the relevance, precision and appropriateness of answers given. Abnormal speech should be recorded.

2.4.3 Mood

The interviewer makes a sophisticated assessment of mood by observing the affective state and describing whether mood is appropriate or not appropriate, profound or superficial, sustained or labile, etc. The patient is also asked about subjective mood, consistent feelings, and changes with time and in response to outside circumstances. Incongruity between the patient's description of mood and the interviewer's observation is noted. For rapport, the interviewer assesses the patient's ability to communicate his or her own internal state to the interviewer, measuring the patient's capacity against a notional standard of normality.

2.4.4 Depersonalisation and Derealisation

Depersonalisation describes a peculiar change in the awareness of self in which the individual feels as if he or she is unreal (Sedman 1970). Depersonalisation is common, difficult for the patient to describe and frequently accompanied by the symptom of derealisation, which denotes a similar change in the awareness of the external world. These symptoms are experienced as extremely unpleasant by patients.

2.4.5 Obsessional Phenomena

According to Lewis (1936), obsessional thoughts have three essential elements: (1) feeling of subjective compulsion, (2) resistance to it and (3) the preservation of insight. It has become conventional to make a distinction between obsessional thoughts, ideas, images or ruminations and compulsive acts.

2.4.6 Delusion

A delusion is an unshakeable false idea or belief out of keeping with the patient's educational, cultural or social background and held with extraordinary conviction and subjective certainty. Psychopathological enquiry should explore for delusion, but it cannot be enquired about directly, as the patient does not distinguish between delusion and normal ideas. The cultural context of the belief and the patient's evidence for holding that belief is assessed.

2.4.7 Illusion and Hallucinations

Three types of false perception are encountered in examining the mental state: illusion, hallucination and pseudo-hallucination; it is important both to search for them and to distinguish between them, because identifying hallucination usually indicates the diagnosis and therefore has implications for management (see Vol. 1, Part 1, Chap. 1).

2.4.8 Orientation

In assessing orientation, the patient is asked about awareness of time, place and person. Temporal disorientation may occur with milder degrees of organic disorder, while orientation of person is only lost at a late stage. Standard questions about orientation in time and place are used and the patient is also observed, e.g., finding his or her way about the hospital ward. Examining for orientation of person starts with questions about other people and their role; if this shows disturbance, enquiry is made concerning the patient's own identity.

2.4.9 Attention and Concentration

Attention is the active or passive focusing of consciousness upon an experience. Concentration is the maintaining of that focusing of consciousness on the task in hand. For formal testing, the "serial sevens" test is most often used. As always in the mental state examination, the observed result for this test is compared with what would be expected from this person in normal health.

2.4.10 Memory

In the course of history-taking, the interviewer will already have obtained information about the patient's

memory. During examination, tests are given for immediate, recent and remote memory. When abnormality is suggested, more precise cognitive testing is needed. *Immediate memory* is usually tested using digit span, forwards and backwards, which must be administered rigorously. Concentration is also involved in this test, especially in the reversed digits test. *Short-term memory* is often assessed by repeating a name and address immediately after the interviewer and then testing for its retention 5 min later; most of the name and address should be remembered correctly. Other testing for recent events will involve asking the patient about recent news, both nationally and in his or her own life. *Long-term memory* is tested by asking about events that happened further back in time; dates and names are particularly useful. The patient's social and cultural context must be taken into account. Possible abnormality in memory testing should be an indication for further standardised and structured memory testing (see Vol. 1, Part 1, Chap. 13).

2.4.11 Insight

David (1990) considered that insight "is not an 'all' or 'nothing' phenomenon but is composed of three distinct, overlapping dimensions, namely, the recognition that one has of mental illness, compliance with treatment, and the ability to re-label unusual mental events (delusions and hallucinations) as pathological." The doctor will have obtained an indication about the patient's capacity for insight during the history-taking and earlier part of the examination. Any illness of moderate to severe degree will alter patients' perspectives about themselves and their world. Insight is the understanding the individual has about his or her own state of health, capacity and worth; it also relates this assessment of personal state to other people and the world outside.

2.4.12 Reaction to the Patient

It is helpful for the interviewer to record the impression the patient makes. This can be used subsequently in evaluating prognosis and expected response to treatment; it is also useful in evaluating the credibility of the history and findings on mental state examination.

2.5 Physical Examination

The physical examination provides information about other concurrent conditions and also constitutional features. Poor physical health and psychiatric morbidity frequently co-exist, with over 50% of acute psychiatric patients having at least one physical disorder (Porter 1996). Physical examination should be conducted after the history and mental state examination, and the patient should be told what this will involve and consent should be obtained. The environment should have adequate lighting and privacy, an examination couch and appropriate, reliable instruments for the examination. In addition, there should be a chaperone for examination by a male doctor of a female patient.

2.6 Collecting Information from Other Informants

Obtaining further information from another person can sometimes be essential (see Sect. 3) and is almost always helpful in providing a different perspective. After the history has been taken, it usually becomes clear who the significant informants are in terms of providing both factual information and observational comment. However, caution is necessary when obtaining information from relatives and friends; issues of confidentiality are important, and sometimes the patient specifically forbids contact with others. The informant, most often a relative, should have known the patient over an extended period of time. A separate interview is organised, preferably face to face, with the purpose of gathering additional information, determining how disabling the condition is and how it affects the patient and other people and discussing treatment plans. The doctor should emphasise that confidential information given by the patient will not be passed on to relatives, and the patient should consent to this interview. In addition, any confidential information divulged by relatives should not be disclosed to the patient. This process also allows the reliability of statements made by both patients and informant to be verified. It also permits an assessment of the interaction between patient and informant to be assessed, which may have diagnostic significance.

2.7 Physical Investigations

2.7.1 Routine Laboratory Investigations

Routine laboratory investigations include blood and urine tests. Patients should undergo a thorough assessment as described above, and subsequent simple investigations should be chosen depending upon symptoms, signs and the differential diagnosis. No single set of routine investigations is relevant for every case, and only those investigations should be

undertaken for which there is a clinical indication. Before undertaking any investigation, patients should be told what the procedures involve and what will be undertaken, and they should be told the results. Investigations include full blood count (haemoglobin, white cell count and platelets), plasma viscosity, urea and electrolytes, thyroid function tests, liver function tests, random blood sugar and syphilis screen. Urine testing, using commercially available kits, is carried out to detect protein, ketones and sugar.

2.7.2 Neuroimaging and Electroencephalography

Neuroimaging and electroencephalographic techniques include anatomical measures, such as skull X-ray, computed tomography (CT) and magnetic resonance imaging (MRI), and functional measures, such as positron emission tomography (PET), single photon emission computed tomography (SPECT) (see Vol. 1, Part 1, Chap. 11), electroencephalography (EEG) and evoked potentials (EPS). It is important to realise that, although special investigations are useful, in 90% of cases the diagnosis can be made on the basis of the history (Gawel 1992), rising to over 95% if the physical examination is included. The EEG measured on the surface of the cerebral cortex is the spatial average of the underlying dendritic fields of the cells in the superficial layer of the cortex (Fenwick 1992), and since its introduction in the 1920s, the EEG has become an important investigation in psychiatry. EPS were first reported in the 1940s and have proved to be useful for the evaluation of certain clinical conditions. A flash of light or an auditory click produces an electrical response in the corresponding cortex, and these responses can be recorded (*evoked potentials*), showing changes in the wave form, e.g., in schizophrenia and Alzheimer's disease (Fenwick 1992; see Vol. 1, Part 1, Chap. 12).

The skull X-ray is useful for identifying skull fractures, areas of thickening and sclerosis and other abnormalities such as copper beating, vascular markings, calcification of tumours and erosion of the sella. CT scan is the primary investigation for cerebral and spinal pathology, but the posterior fossa is poorly visualised and some lesions may be isodense and thus missed. MRI provides outstanding anatomical quality. However, patients must be free of all metal, and many patients, especially children and the elderly, find the procedure very frightening. PET provides information about the central nervous system (CNS) and allows both cerebral blood flow and regional metabolism to be assessed. SPECT provides similar information to PET scans, but with lower resolution. Cerebral angiography was a relatively common procedure in the past, but has serious complications and, with the advent of superior modern neuroimaging techniques, it is now less commonly used.

2.8 Psychological Assessment

2.8.1 Simple Assessment

The patient will have undergone "informal" psychological assessment both in the assessment of personality and mental state examination. Additional simple psychological assessments that can be undertaken include observing the patient's behaviour in a variety of settings (alone, interactions with members of staff, other patients and members of the public), ability to perform activities of daily living (dressing, personal hygiene, shopping), compliance with medication, the effects of psychoactive drugs, and sleep and eating behaviour. These simple observations provide valuable information about the patient with regard to diagnosis and management, but are not standardised.

2.8.2 Special Assessment

Rating scales are extremely numerous (Royal College of Psychiatrists 1994) and include both observer- and self-rating scales. They can be used to assess behaviour, mood, speech and language, thought processes and cognitive function. Any scales used should have a high validity and a high test–retest and inter-rater reliability. Neuropsychological tests will usually be undertaken by a clinical psychologist. These instruments assist in localising regional brain damage and measure psychomotor performance, visual and auditory memory, short- and long-term memory and verbal fluency (Lezak 1983). Personality inventories may be useful in clinical research, but they have not proved to be as useful in clinical practice (see Vol. 1, Part 2, Chap. 6).

2.9 Social Assessment

Man is a social animal and usually lives in a family, neighbourhood and/or working group. Exploring the social context provides useful information on aetiology and for treatment. Social assessment reviews the patient's occupation, home environment, social class, ethnicity and religion.

2.9.1 Occupation

Occupational history gives an indication of premorbid intelligence and functioning, and any change in

employment status will help reveal the impact of the psychiatric disorder on the individual's life. Certain occupations are more likely to be associated with a particular disorder, e.g., individuals involved with the sale of alcohol, who have an increased risk of alcohol dependence. Assessment of the patient for occupation may help patients to return to work as quickly as possible. The best occupational assessment is one which takes place in the environment where the patient works, since this will be directly relevant to the individual. Many hospitals also have workshops where individuals can be assessed prior to returning to work; this is especially important if working involves using dangerous equipment or driving heavy goods vehicles (Department of Health 1995).

2.9.2 Home Environment

Although every psychiatric patient might benefit from home assessment, it is particularly important in old age psychiatry. It may give useful insight into the patient's mental state (e.g., windows covered to repel "X-rays") and clues to aetiology (e.g., empty bottles of spirits). Factors likely to exacerbate poor mental health such as inadequate heating and diet and the general layout of the house may be identified. The patient's safety at home can be assessed, e.g., presence of gas appliances or psychotropic drugs that may contribute to the mental disorder. The safety of other family members, such as children, can also be assessed from the perspective of either neglect or physical abuse. It may be more efficient for another member of the multidisciplinary team to carry out the home assessment.

2.9.3 Legal Aspects

A criminal record is relevant, especially when preparing a psychiatric report on an individual who has committed a serious offence. It may sometimes be necessary to obtain official records about previous convictions from other agencies. A history of violence is one of the best predictors of future violence (Scott 1977), and a gradual escalation in the seriousness of offences also provides invaluable information.

2.9.4 Ethnicity

Knowledge of the patient's ethnic background may be helpful in understanding the patient's psychopathology. It may also help in establishing a rapport with the patient and in maintaining the therapeutic relationship. Some conditions are commoner in certain ethnic groups, and relevant information may help with diagnosis, management and prognosis (Cox and Jorsh 1992).

2.9.5 Religious Affiliation and Belief

Assessment of a patient's religious group and beliefs is often linked to their cultural and ethnic background and can thus be important when trying to interpret and understand a particular patient's symptoms. The individual's religious beliefs may profoundly affect presentation, prognosis and response to treatment.

3 Specific Issues

Many of the particular difficulties in psychiatric examination discussed in this section are caused by barriers to communication – verbal, non-verbal and emotional. Problems occur when the patient is acutely disturbed, has limited comprehension, as in patients from other cultures, the blind and the deaf or those with a learning disability, or is incapacitated for expression, as in some neurological disorders, especially motor dysphasia.

3.1 Child and Adolescent Psychiatry

Assessment of children and adolescents differs from examination of adults in several respects. The consultation is seldom initiated by the child, and the problem is usually present as either abnormal development or behaviour. As disturbance in other family members is usual, psychiatric examination involves assessment of the whole family. Although a full psychiatric history should be taken, the focus is very much on the developmental and family history, and normal milestones for development must be known. A child has less language; other methods of assessment should therefore be utilised, such as observing play and social interactions with other children. Many different individuals contribute to the assessment of children and adolescents, including parents, the family doctor, teachers, paediatricians, social workers and educational psychologists, hence forging a working relationship is essential in the assessment. In assessing adolescents, one needs to bear in mind that this is a period of rapid physical, psychosexual, social and emotional change (Graham 1986; Steinberg 1982).

In children and adolescents with a learning disability, there may be a complicated mixture of medical, psychological, social and developmental problems in

addition to intellectual impairment and, frequently, sensory deficits, which make assessment more complicated. Assessment should be unrushed and thorough and should include input from other health care professionals and agencies to give a broad picture of the clinical problem. The assessment is aimed at targeting resources more specifically to enable the patient to reach his or her full potential, and it is necessary to involve the family, who may be under considerable stress (Clarke et al. 1985).

3.2 Acute Psychiatric Disorders

Acute psychiatric disorders may occur unexpectedly, often with seriously disruptive behaviour, and often it is impossible to undertake detailed history. In addition, assessments are frequently made in unsatisfactory situations, such as police cells or casualty departments. One should obtain as full a history as possible from the patient, family, old case notes and information from others, such as the general practitioner or police, including presenting symptoms (onset, course and severity), past psychiatric history and medical history (Macpherson et al. 1996). Mental state examination is always undertaken, however uncooperative the patient, and should include attention to hostility, aggression and withdrawal when possible physical examination is conducted. Once this has been acheived, a management plan can then be implemented (McGrath and Bowker 1987).

3.2.1 Assessing Violence

Violent incidents in hospitals are common and appear to be increasing (Haller and Deluty 1988). Assessment of such incidents is usually undertaken as an emergency and should have preceded administration of psychoactive drugs. Information should be collected from observers, and a mental state examination performed. Other staff should be available to deal with violent incidents (but out of sight), and the interview should be conducted in a calm, non-confrontational and reassuring manner, preferably by someone the patient trusts. The environment is also important. Where possible, heavy objects should be removed that could be easily thrown by the patient, and the psychiatrist should be close to the exit and have access to a personal alarm.

It is important in the assessment of potentially violent, personality disordered patients to pay attention to engagement (Norton 1996). Mutual respect and trust is vital for carrying out the clinical task professionally.

3.2.2 Assessing Suicide Risk

Direct questions about suicide do not increase the probability that an individual will commit suicide (Hawton 1987). On the contrary, such questions are an essential part of the assessment. Risk factors should be noted; these include male gender, old age, loneliness, chronic painful conditions and direct statements of intent and hopelessness (Beck et al. 1985). It is also important to evaluate factors which may reduce the risk of suicide, such as concern for one's children and religious convictions.

3.2.3 Assessing Victims of Acute Psychological Distress

Acute reactions to stress (World Health Organisation 1992) are transient disorders, usually lasting several hours to a few days in response to exceptionally stressful events, such as natural catastrophes or following a crisis in a relationship. It may not be possible to obtain a full account from the individual, but registering personal details and a subsequent offer of help, given in a calm, reassuring manner, may be all that is required.

3.3 Psychiatry of Neurology

Neuropsychiatry provides a bridge between neurology and psychiatry (Reynolds and Trimble 1989). It has been suggested that as many as 49% of neuropsychiatric referrals have a clear organic aetiology (Lishman 1992), and psychiatrists frequently assess patients with neurological problems. An integrated approach to assessment is needed, and this may be best undertaken in a combined clinic, which enables joint assessment and treatment to be undertaken. The range of psychiatric and neurological disorders seen in a neuropsychiatric setting is large (Scheepers et al. 1995), requiring knowledge of neurology and psychiatry for adequate assessment and treatment.

3.4 Old Age Psychiatry

Assessment of the elderly patient differs in some important respects from general psychiatry. Cognitive and psychomotor decline, failing physical health, decreased psychosexual function, loss of children, bereavements, retirement, loss of income and loss of role may have occurred. It is important to distinguish between change due to normal ageing and that due to

pathology. This can be difficult, especially with memory impairment, but is usually indicated by quantitative differences and rate of decline over time. Home assessment is often helpful as this gives invaluable information. Varying degrees of sensory impairment may result in the assessment taking longer. It is vital to involve relevant family members and other health care professionals (Jacoby and Oppenheimer 1995).

3.5 Forensic Examination

Forensic assessment requires both psychiatric skills and knowledge of relevant law (Faulk 1988). Two important aspects are the assessment of dangerousness and preparation of medico-legal reports. Although patient confidentiality should be respected wherever possible, it will be necessary to prepare reports, e.g., court reports, which might be discussed in public. Under such circumstances, the patient should always be informed about this *before* the interview begins.

The best predictor of future violence is previous violence. Although dangerousness is difficult to predict, Gelder et al. (1996) have suggested factors that may be associated with dangerousness (Table 2).

Table 2. Factors associated with dangerousness

Type	Factor
History	One or more previous episodes of violence
	Repeated impulsive behaviour
	Evidence of difficulty in coping with stress
	Previous unwillingness to delay gratification
	Sadistic or paranoid traits
Offence	Bizarre violence
	Lack of provocation
	Lack of regret
	Continuing major denial
Mental state	Morbid jealousy
	Paranoid beliefs plus a wish to harm others
	Deceptiveness
	Lack of self-control
	Threats to repeat violence
	Attitude to treatment
Circumstances	Provocation or precipitant likely to recur
	Alcohol or drug abuse
	Social difficulties and lack of support

3.6 Substance Misuse

To assess patients referred with drug-related problems, full psychiatric history, mental state and physical examination looking for specific signs will be required. A further discussion can be found elsewhere (see Vol. 2, Part 2, Chap. 11; Vol. 3, Part 2, Chaps. 15, 18, 20).

3.6.1 Smoking History and Examination

All patients should be asked about their smoking habits, but under-reporting of the amount is common. It is important to record the type of tobacco consumed (e.g., cigarettes), quantity per day, whether inhaled and any recent changes; physical signs such as stained fingers should be observed.

3.6.2 Alcohol History and Examination

Alcohol history includes detailed social history, development of the habit, craving, loss of control and dependence, adverse consequences, both physical and psychosocial, and an assessment of the quantity consumed in units. Both psychiatric and physical examination will be required. A detailed social assessment is also essential, since families frequently suffer emotionally, physically and financially (Edwards 1982).

3.6.3 Drug History and Examination

In terms of their drug history, patients may mislead; corroboration should therefore be sought from informants and from physical examination. The development of an illicit drug habit, drugs consumed, their quality and evidence for craving and loss of control should be recorded. In addition, landmarks (both social and psychological ones) should be noted. Enquiry and examination concerning drugs should always be included in the assessment. Urine and hair analysis may prove useful.

3.7 History and Examination Prior to Psychotherapy

Before undergoing psychotherapy, patients should be carefully assessed in terms of personal relationships, personality disorder, motivation for treatment, capacity for psychological understanding and intelligence. They require sufficient ego strength to deal with the stressful aspects of therapy and must be free from major psychiatric disorders. These preconditions for psychotherapy, especially with regard to psychiatric disorders, are primarily valid for classic psychoanalytic therapy. They are less valid for other forms, including supportive and behavioural psychotherapy. The psychiatric assessment should include a full psychiatric

history and focus principally on the issues described above (Brown and Pedder 1979; Bloch 1986).

3.8
Transcultural Issues

Cultural background is always significant, including information about customs, religion and language. Using an interpreter will take much longer than a standard interview, and the interpreter is unlikely to be familiar with psychiatric terminology. When a family member acts as interpreter, there may be issues of confidentiality and the patient may be reluctant to discuss personal details through a family member (Cochrane 1977; Littlewood and Lipsedge 1985).

3.9
Community-Based Psychiatry

Community-based assessments have the advantage of convenience for the patient. However, the patient should have some choice concerning where the interview will take place, provided circumstances are satisfactory. Confidentiality may be a consideration, since relatives or neighbours are sometimes intrusive. Safety issues for the doctor require attention, especially when assessments are undertaken in large cities: someone should be informed where the doctor is going; he or she should not carry valuables; a mobile phone is a sensible precaution; and a system needs to be arranged pending failure to return by an expected time (Parkman and Bixby 1996).

3.10
Liaison Psychiatry: Interviewing in Hospital Settings

Particular matters concerning interviews on medical or surgical wards are that patients may not have been informed about the nature of their referral (merely told that a "specialist" is coming to see them); there may be issues of confidentiality in an open ward, and conditions for interview are often difficult. It may therefore be difficult to gain the patient's trust. Many patients will have anxieties about seeing a psychiatrist, and these will need to be discussed. The notes should be consulted, laboratory investigations reviewed and discussion carried out with medical and nursing staff. As only one assessment is usual, this must be as thorough as possible (Lipowski 1985; see Vol. 1, Part 2, Chap. 13).

3.11
Examination of Professional Colleagues

Although the assessment is essentially no different to any other psychiatric examination, there are particular problems associated with interviewing medical colleagues. Having a written letter of referral from the doctor's general practitioner helps to maintain both people in a doctor–patient relationship. Psychiatrists should avoid treating relatives, friends and close colleagues, as these relationships impede the necessary psychiatric objectivity. It is essential that confidentiality is ensured, but at the same time there is a duty to assess fitness to practice and the interests of patients should be paramount. All these considerations strongly indicate that this type of assessment is preferably carried out by an experienced practitioner.

4
Utilising the Examination

Once a full psychiatric history has been taken, a mental state and physical examination performed and additional information collected, a formulation of the case should be produced. This information needs to be accurately recorded and any decisions communicated to the patient, relatives and other health care professionals.

4.1
Formulation

The formulation is a concise review of the case and includes a brief summary, the working diagnosis and differential diagnoses with arguments for and against each of these. This is followed by a review of the aetiological factors, a discussion of the treatment options and the prognosis is outlined. A good formulation is a balanced appraisal of the psychiatric assessment, but should be very firmly based on the facts of the case rather than on speculation. It also encapsulates the main issues in the case, which then facilitates communication with other health care professionals (Tantam and Greenberg 1987).

4.2
Recording Data

Recording information about the consultation is important for several reasons: it is difficult to remember

details about every patient one sees; notes recorded by one individual can be accessed by another; the record enables the patient's condition to be viewed longitudinally. In addition, notes are a legal document which the psychiatrist, by virtue of the treatment contract he or she has entered into, is obliged to preserve and, if necessary, produce; in particular, they may be demanded by a court of law.

4.2.1 Case Notes

Case notes allow information from different sources to be brought together (e.g., medical, nursing, psychology, social work) and thus form a comprehensive document about the patient. Notes should be carefully ordered, with a clear system, to make access to information both logical and simple. Entries should clearly indicate the location of assessment, e.g., in casualty, case summaries, appropriate formulations and life charts. Increasingly, patients are given access to their own notes and their content may be influenced by this (Kosky and Burns 1995).

4.2.2 Confidentiality

For a trusting therapeutic relationship, confidentiality of the psychiatric assessment is essential. Within health care teams, it is necessary to discuss patients, and patients should be informed about such discussions. Discussions should only be made with other individuals, including relatives, with the *explicit* consent of the patient, except where failure to disclose information would place the patient or someone else at risk of death or serious harm. Where information about the patient might be disclosed (perhaps in a court), the patient should be advised at the start of the assessment so that the patient may choose not to disclose this information (O'Brien 1995; General Medical Council 1995).

4.2.3 Computerised Information

Computers are increasingly used to store information about patients, and they have a number of advantages: such records are easier to read; large quantities of information can be stored in a relatively small space; information can usually be retrieved rapidly; access to patient records can be facilitated from different sites. Computers also have disadvantages: information may be lost; the clinical service could be disrupted by computer crime; issues of confidentiality are more problematic with the use of computers for information storage.

4.3 Planning Management

The most important use of the information gained from psychiatric examination is to plan management: *short term* (first few days), *medium term* (duration of hospitalisation) and *long term* (after discharge). Assessment allows decisions to be made concerning physical treatments, psychological therapies and social interventions.

Following the psychiatric assessment and formulation, it must be decided to whom relevant findings will be communicated and how this will be done: verbally or by letter, briefly or in detail, but with the *explicit* consent of the patient.

The multidisciplinary team (MDT) requires information from different sources to be brought together and ideas exchanged to produce a single, unified management plan.

Patients frequently complain that they are poorly informed about their illness. They should be given as much information as possible concerning diagnosis, management plan and prognosis. Involving patients in their own assessment and management empowers them to reduce anxiety and enhances the doctor–patient relationship, leading to improved compliance.

It is vital that the patient's family practitioner (and any other doctors involved with care) are fully informed about all aspects of the patient's problem. Written information should be communicated as concisely and as quickly as possible.

Relatives should usually be informed about the patient's progress, as they are likely to be anxious and their involvement after discharge will be important for long-term care. However, if the patient requests that relatives should not be informed, this *must* be respected.

5 References

American Psychiatric Association (1994) Diagnostic and statistical Manual of mental disorders, 4th edn. American Psychiatric Association, Washington

Beck AT, Steer RA, Kovacs M, and Garrison B (1985) Hopelessness and eventual suicide: a 10 year prospective study of patients hospitalised with suicidal ideation. Am J Psychiatry 145: 559–563

Bloch S (1986) An introduction to the psychotherapies, 2nd edn. Oxford University Press, Oxford

Brown D, Pedder J (1979) Introduction to psychotherapy; an outline of psychodynamic principles and practice. Tavistock, London

Clarke AM, Clarke ADB, Berg JH (1985) Mental deficiency. The changing outlook, 4th edn. Methuen, London

Cochrane R (1977) Mental illness in immigrants to England and Wales. An analysis of mental hospital admissions 1971. Social Psychiatry 12: 23

Cox JL, Jorsh MS (1992) Transcultural psychiatry. In: Weller M, Eysenck M (eds) The scientific basis of psychiatry, 2nd edn. Saunders, London, pp 469–490

David AS (1990) Insight and psychosis. Br J Psychiatry 156: 789–808

Department of Health (1995) Chief Medical Officer's update, no. 5, 3 March 1995. Department of Health, London

Edwards G (1982) The treatment of drinking problems. McIntyre, London

Faulk M (1988) Basic forensic psychiatry. Blackwell, Oxford

Fenwick PBC (1992) Some aspects of the use of the EEG in psychiatry. In: Weller M, Eysenck M (eds) The scientific basis of psychiatry, 2nd edn. Saunders, London, pp 192–212

Gawel MJ (1992) Some issues in clinical neurology. In: Weller M, Eysenck M (eds) The scientific basis of psychiatry, 2nd edn. Saunders, London, pp 213–239

Gelder M, Gath D, Mayou R (1996) Oxford textbook of psychiatry, 3rd edn. Oxford University Press, Oxford

General Medical Council (1995) Confidentiality; guidance from the General Medical Council. GMC, London

Graham P (1986) Child psychiatry: a developmental approach. Oxford University Press, Oxford

Haller RM, Deluty RH (1988) Assaults on staff by psychiatric inpatients: a critical review. Br J Psychiatry 152: 174–179

Hawton KE (1987) Assessment of suicide risk. Br J Psychiatry 150: 145–153

Jacoby R, Oppenheimer C (1995) Psychiatry in the elderly. Oxford University Press, Oxford

Kosky N, Burns T (1995) Patient access to psychiatric records present in an in-patient unit. Psychiatr Bull 19: 87–90

Lewis AJ (1936) Problems of obsessional illness. Proc R Soc Med 29: 325–336

Lezak MD (1983) Neuropsychological assessment, 2nd edn. Oxford University Press, Oxford

Lipowski ZJ (1985) Psychosomatic medicine and liaison psychiatry. Plenum, New York

Lishman WA (1992) What is neuropsychiatry? J Neurol Neurosurg Psychiatry 55: 983–985

Littlewood R, Lipsedge M (1985) Culture bound syndromes. In: Granville-Grossman KL (ed) Recent advances in clinical psychiatry. Churchill Livingstone, Edinburgh

Macpherson R, Anstee B, Dix R (1996) Guidelines for the management of acutely disturbed patients. Adv Psychiatr Treat 2: 194–201

McGrath G, Bowker M (1987) Common psychiatric emergencies. Wright, Bristol

Norton K (1996) Management of difficult personality disorder patients. Adv Psychiatr Treat 2: 202–210

O'Brien J (1995) GMC's guidance on confidentiality. Psychiatr Bull 19(2): 115

Parkman S, Bixby S (1996) Community interviewing: experiences and recommendations. Psychiatr Bull 20: 72–74

Porter I (1996) Is routine physical examination of psychiatric inpatients really necessary? Psychiatr Bull 20: 218–220

Reynolds EH, Trimble MR (1989) The bridge between neurology and psychiatry. Churchill Livingstone, Edinburgh

Royal College of Psychiatrists (1994) Psychiatric instruments and rating scales, 2nd edn (OP23). Royal College of Psychiatrists, London

Scheepers BDM, Bird JM, Rogers DG (1995) Neuropsychiatry: a different approach or a different clientele? Psychiatr Bull 19: 77–81

Scott PD (1977) Assessing dangerousness in criminals. Br J Psychiatry 131: 127–142

Sedman G (1970) Theories of depersonalisation: a reappraisal. Br J Psychiatry 117: 1–14

Sims ACP (1994) 'Psyche' – spirit as well as mind? Br J Psychiatry 165: 441–446

Sims ACP (1995a) Symptoms in the mind: an introduction to descriptive psychopathology, 2nd edn. Saunders, London

Sims ACP (1995b) Speech and language disorder in psychiatry. Gaskell, London

Steinberg D (1982) The clinical psychiatry of adolescence. Wiley, Chichester

Tantan D, Greenberg M (1987) The formulation. In: Rix KJB (ed) A handbook for trainee psychiatrists. Bailliere Tindall, London

World Health Organization (1992) The ICD 10 classification of mental and behavioural disorders: clinical description and diagnostic guidelines. WHO, Geneva

CHAPTER

6

H.J. Möller, R.R. Engel, D.R. Hemsley

Standardised Measurement Instruments in Psychiatry

Translator: S. Johnson

1 Aims and Methods

Standardised methods of examination are used in psychiatry to assess objectively, and, in some cases, quantify psychopathological phenomena and other clinically relevant domains, making it easier to communicate these, to verify their status and to analyse them statistically (Stieglitz and Baumann 1994; Möller et al. 1996). They are essential to develop models of psychopathology.

Major areas in which standardised procedures are applied in psychiatry include the following:

- Cross-sectional quantitative description of psychopathological abnormalities
- Assignment by a standardised method of individual cases to diagnostic categories
- Quantitative assessment of change over time in psychopathological abnormalities (with or without therapeutic interventions)

Those who favour intuitive phenomenological methods have expressed the concern that applying standardised methods cannot take sufficient account of each patient's individuality (Huber 1976). However, this criticism seems largely unfounded. There are individual characteristics which such standardised methods will not fully capture; this is necessarily the case, as standardised measurement instruments tend to be constructed on the principle that a symptom only qualifies for inclusion if it is present in at least a specified minimum proportion of the populations for which the instrument is intended. However, when required, this deficit can be remedied by using additional methods of investigation aimed at capturing distinctive characteristics of individuals. Indeed, reports of positive experiences with measurement methods aimed specifically at the investigation of individual cases may be advanced as a counter-argument (Frey et al. 1979). Such methods have a long history (see Shapiro 1966) and continue to be applied, e.g. to delusional beliefs (e.g. Brett-Jones et al. 1987).

Standardised measurement procedures can be categorised on the basis of their methodologies into standardised assessment instruments, systematic behavioural analysis and objective tests in the narrower sense of the word (based on von Zerssen and Möller 1980). The terms standardised assessment instrument or rating scale are applied to structured methods of assessing current and/or past behaviour and/or experience, based on lists of characteristics and, in some cases, descriptions of these characteristics. The extent of standardisation varies from a simple list of symptoms filled in on the basis of a freely structured exploratory interview to semi- or fully structured interview schedules. These standardised assessment procedures are especially suitable to examine the full spectrum of psychiatric symptomatology; in addition, as they are less restrictive than other procedures, they are particularly practicable. A variety of interview schedules are available and in general use.

Systematic behavioural analysis involves using a fixed set of categories to classify the quantity and type of various forms of behaviour (including speech and actions) occurring during a fixed observation frame (methods involving sampling fixed periods of time or particular events). This usually focuses on manifest behaviour, and systems of categorisation are often developed specifically to fit the particular question being asked. This method has found particular favour in the areas of behaviour therapy and research about individual communication and interactions.

Objective tests measure reactions to standardised and fixed "stimulus material". They allow analysis of specific particular psychological functions such as perception, concentration, attention and intelligence, usually from the point of view of performance. This category includes tests of attention and concentration, intelligence tests and a variety of psychophysiological indices. These tests are said to be objective as they cannot really be falsified by the examiner or the subject and there are fixed assessment criteria with corresponding methods of data analysis and fixed norms.

Because they are very practicable, rating scales are often preferred to the other methods we have discussed if the results of patient examinations are to be documented in the context of routine professional care (see Vol. 1, Part 2, Chap. 5). They are also frequently applied in clinical psychiatric research, such as clinical trials of drugs, studies of longitudinal course, in routine clinical documentation or in epidemiological studies (Cronholm and Daly 1982; Möller et al. 1983; see Vol. 1, Part 1, Chap. 2), even though, in terms of their level of precision, standardised assessment measures are methodologically inferior to objective tests and systematic behavioural analysis. Despite the methodological superiority of these latter methods, they tend to be included in clinical psychiatric research only as supplementary measures for the sake of completeness. An exception to this can be found in the investigation of specific aspects of cognitive functioning, e.g. the debate over the potential of clozapine to reduce schizophrenics' cognitive impairment (e.g. Lee et al. 1999) and of certain questions of differential diagnosis. This limited use results not only from the amount of time and effort involved in applying these tests, but also, particularly in the case of objective tests, from the fact that the constructs which they measure are rather more remote from the psychiatric approach than the more complex phenom-

ena issues which can be described using rating scales. However, some would argue that they are closer to the "core" of the disturbance.

Investigations of the relationship between the three domains of assessment form the basis of much research in psychopathology. Because standardised assessment methods are so widely applied in psychiatric practice and research, they will be the focus of the following discussion.

2 Scale Construction, Scoring Methods and Quality Criteria

Standardised methods of assessment, or rating scales, allow description in terms of numerical values of psychological abnormalities of various characteristic forms. Different measurement scales allow the degree of abnormality to be quantified to varying extents. In the simplest instances, such as symptom checklists, scales simply allow for a rating of 0 or 1 to be made for each symptom or complex of symptoms, indicating whether or not it is present. More precise assessment becomes possible if the construction of the scale allows the severity of phenomena to be described using a scale consisting of a series of levels. As there is a danger that different assessors will base their evaluations on different standards, it is important to establish a framework for the assessment by providing anchor points, e.g. by giving examples of situations which would be characteristic for each point on the scale. Overly detailed assessment using an excessively broad scale is not meaningful, as differences at the extreme end of the scale cease to reflect real and significant variations in the phenomena being examined, so that the differentiations being made are not real ones.

In some assessment instruments, the values at which points on the scale are fixed may be varied as required, so that a scale may be constructed which is as finely differentiated as required; examples include visual analogue methods of assessing subjective well-being (Luria 1975). As the measurement of psychological phenomena is essentially imprecise, a relatively coarse scale is usually adequate, especially for comparisons between individuals. A fine scale may have advantages for intra-individual comparisons. However, any improvement in measurement precision will generally be achieved not by refining the construction of the scale, but rather by improving methods of measurement (von Zerssen 1977).

The values for characteristics which belong together (e.g. individual symptoms within a syndrome) can be added together to produce a summary score. The extent to which characteristics belong together to make up a syndrome is determined during the process of test construction (see below) by applying multivariate statistical procedures such as factor and cluster analysis. In some cases, before adding together the figures for each characteristic to produce a summary score, these figures will be weighted to indicate the relative importance of each characteristic within the syndrome. However, if the characteristics have been shown to be relatively independent, theoretical or practical justification is required for any such summation (see Garety and Hemsley 1987).

Measures of psychopathology obtained from standard rating scales generally have the level of measurement of ordinal scales, i.e. they give only a rank order and do not possess the measurement level of an interval scale, in which there are equal intervals between points on the scale. A fundamental problem in measurement is that measurement instruments with more detailed scales and higher levels of precision tend to bring with them greater restrictions regarding the phenomena which can be measured. This normally means that increasing quality of measurement is accompanied by increasing abstraction from the theoretical or conventional understanding of the characteristic which is the starting point (the reliability/validity dilemma).

Standardised assessment instruments should meet the following quality criteria derived from test theory as far as possible (Lienert 1969; Fischer 1974; Sarris and Rey 1981):

1. *Objectivity*: The results should not depend on who carries out the assessment and analyses the results. Procedure, analysis and interpretation should be standardised so that, as far as possible, the same results are obtained regardless of who administers the instrument, analyses it or interprets it.
2. *Reliability*: This refers to the reliability with which a standardised assessment instrument records a characteristic. When the measurement is repeated, the same result should be obtained.
3. *Validity*: This is the extent to which the instrument records what it is intended to record. The connection between the results of measurement and any external criteria available for assessing what is to be measured should be as close as possible.
4. *Establishment of norms*: Reference values for different clinical groups and varying groups of normal probands and, where applicable, a representative sample of the general population should be available.
5. *Practicability*: The amount of resources required for administering standardised assessment instruments in terms of time, staff and material should be as low as possible.

For a particular test, these criteria will not necessarily be highly correlated with one another. For example, a test of concentration may give a reliable, but not a valid measurement of individual differences in performance when a test designed for intellectually normal individuals is applied to people with learning difficulties; in this group, it no longer functions as a measure of concentration, but rather of intelligence (Sarris and Lienert 1974). There is a partial incompatibility inherent in the relationship between reliability and validity (the reliability/validity dilemma): improvement in reliability is often accompanied by a reduction in validity and vice versa.

While for psychometric tests in the narrower sense the availability of norms is largely taken for granted, this has been approached with a great deal less rigour for clinical rating scales. Thus, for example, the Inpatient Multidimensional Psychiatric Scale (IMPS; Lorr 1974) is almost the only observer-rated scale measuring psychopathology for which norms for a representative sample of the general population are available (Hiller et al. 1986). For a number of rating scales, reference values are available for particular diagnostic groups. Referring to such norms or, more precisely, reference values has a substantial impact on the interpretation of results. For example, moderately high scores for the domain of paranoid syndromes have quite a different significance from moderately high scores for depressive symptoms, in that depressive symptoms are common in the general population, whereas paranoid symptoms are not.

In producing norms for a standardised assessment instrument, the usual starting point is the normal distribution of values. There has to be a relation to the normal distribution in order for it to be possible to derive confidence intervals (see below) and to apply particular statistical tests, such as Pearson's product–moment correlation. Two values need to be known to characterise a particular normal distribution: (1) the mean of all scores obtained for the test and (2) a measure of the extent of dispersion of these values, generally expressed in the form of standard deviation. Once these values are known, the standard properties of the Gaussian (normal) distribution allow the proportion of subjects who will have a particular test score to be calculated. Thus, for example, 68% of patients will have a test value which falls within one standard deviation either side of the mean, and around 95% a value which is no more than two standard deviations from the mean. On the basis of the norm values, it will therefore be possible to calculate where a proband's score lies in relation to a reference population (Fig. 1).

Norm values for a particular test can be straightforwardly expressed by giving the mean and standard deviation. Once this information is available, a statement may be made about the position of the proband in relation to the reference population. However, a disadvantage of referring to the numerical value of the standard deviation for a particular test is that it is difficult to compare the results obtained by a particular proband for several different tests. To allow comparisons of this sort to be made, a z-value can be calculated; this is the result obtained in a particular test expressed in terms of units of the standard deviation for that test. Results obtained by a particular proband in different tests may also be compared using percent-

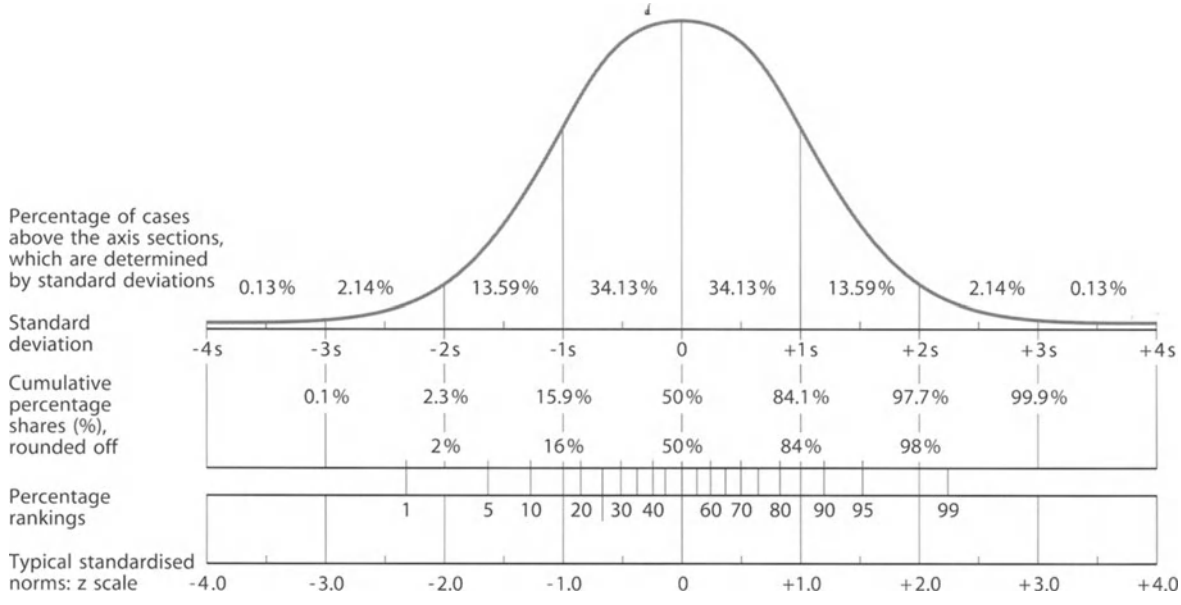


Fig. 1. Relationship between some frequently used standardised scales and the normal distribution curve

age rankings, by specifying for each test what proportion of a reference population has higher or lower scores for the test.

Various empirical methods may be used to test whether the test quality criteria specified above have been met. Appropriate ways of examining the reliability of a test include test–retest reliability, inter-rater reliability, the split-half correlation coefficient and internal consistency. To determine test–retest reliability, the same test is given to the same group of people at two different time points. The time between the two applications of the test depends on the interval to which the test is intended to apply. For tests in which the aim is to record enduring personality traits, an interval between applications of the test of between 14 days and 1 year is recommended. For tests where the aim is to record rapidly fluctuating characteristics (such as mood or subjective well-being), a time span between several minutes and a few hours is appropriate. Ideally, identical results should be obtained for each measurement, but of course this is not the case in practice, as measurement errors necessarily occur (related to strong influences caused by the test situation, practice effects, etc.). The correlation between the two values gives the test–retest reliability coefficient. Deciding whether the reliability of a test is sufficiently high depends very much on the purpose of administering the test (Hofstätter 1957; Meili 1961; Lienert 1969). As a rule, a reliability coefficient in excess of 0.8 is required. Methods for which the test–retest reliability is below 0.5 are not generally useful. The measurement accuracy of a test may be different for different diagnostic groups (differential reliability).

Several different procedures also exist for determining the validity of a test, e.g. examination of consensual validity, predictive validity, construct validity and content validity. Consensual validity is determined by correlating the results of applying the tests to a sample of probands with comparable data obtained by methods other than the application of the test (external criteria). For example, results for the test may be correlated with corresponding scores obtained for the same subjects for another test examining the same psychological characteristics. Whereas with consensual validity, test values and external criteria are measured at the same time, predictive validity is determined by investigating whether events predicted on the basis of the test results have actually happened. A classic example is the correlation of test results from an intelligence test with assessment at a later date of actual success at school.

A requirement which needs to be emphasised is that following translation of a scale from one language into another, new validity tests must be carried out with the translated version. This is also true if the scale is modified in any way.

3 Standardised Instruments for the Description of Psychopathology

Standardised assessment instruments relate to past or current behaviour and experience. The extent of psychological abnormalities is rated using fixed scales. These rating scales may focus on a single aspect, e.g. anxiety (unidimensional scales) or on several aspects (multidimensional scales) of psychopathology. For each aspect of psychopathology, assessment may be based on a global rating or on different elements within the aspect being assessed, e.g. on individual symptoms of the depressive syndrome. In this latter case, the overall score on the instrument is obtained by adding together values for these different elements.

The level of standardisation of standardised assessment scales, also known as rating scales, falls between that of unstructured clinical assessment and that of objective tests. In some of these instruments, standardisation is limited to providing guidelines describing items and the categories used for assessing them and to specifying a method of analysis (generally one or more summary scores are calculated). In other scales, a time frame is also stipulated for the assessment, and in some the framework in which observation takes place is also fixed. In the latter case, the instrument is referred to as a fully structured or standardised interview. The more extensive the standardisation procedures undertaken, the greater the reliability of an assessment instrument generally becomes. However, a highly standardised instrument tends to become less practicable. For this reason, both in everyday clinical use and in research, where resources are constrained, simpler scales such as the Present State Examination (PSE; Wing et al. 1978) tend to be preferred to fully structured instruments. The latter require a fully structured interview, whereas the simpler rating scales can be completed following a routine psychiatric interview. Particularly for the simpler rating scales, inter-rater reliability for observer-rated instruments can be improved by systematic joint training of the raters (Heimann et al. 1977). In principle, fully structured interview methods with extensive individual interviewing ought to produce high inter-rater reliability, and they therefore have particular advantages in multicentre, multinational studies in which great discrepancies need to be taken into account not only in terms of how mental states are assessed, but also in the psychiatric interview techniques used.

Standardised assessment instruments may be classified on the basis of who carries out the assessment into self-rated and observer-rated instruments. In observer-rated instruments, psychopathological abnormalities

are identified by trained assessors (e.g. doctors, psychologists, care staff, lay people trained to administer the instrument) or by significant others (e.g. partner, relatives, friends). The assessment concerns the behaviour and/or experience of the patient and is based on the assessor's own observations and/or information given by the patient. Observer-rated scales need to be constructed so that they are appropriate to the level of training of the particular types of interviewer to be involved in their administration. Thus there are scales for doctors trained in psychiatry, e.g. the AMDP (*Arbeitsgemeinschaft für Methodik und Dokumentation in der Psychiatrie*; Study-Group for Methodology and Documentation in Psychiatry) system (Baumann and Stieglitz 1983), for clinical psychologists, e.g. the Structured Clinical Interview (SCI; Burdock and Hardesty 1969), for care staff trained in psychiatry, e.g. the Nurses' Observation Scale for Inpatient Evaluation (NOSIE; Honigfeld and Klett 1965) and also for patients' relatives, e.g. the Symptoms and Social Behaviour Rating Scale for Relatives (Katz and Lyerly 1963).

Observer-rated scales mainly focus on psychopathological state, either with the global epidemiological question of classifying each individual overall as a "case" or "non-case", e.g. the interview developed by Goldberg and colleagues (Goldberg 1972), with the recording of specific aspects of mental state such as depression or anxiety, e.g. the Hamilton Depression Scale or the Hamilton Anxiety Scale (Hamilton 1959, 1967) or with assessment of the whole spectrum of psychopathology, e.g. the AMDP questionnaire (AMDP 1995). In scales which record the whole range of psychopathology, there is usually an emphasis on the detection of symptoms of functional psychoses, while only limited account is generally taken of organic or neurotic symptomatology. For specific measurement of the latter types of symptomatology, the use of scales which focus specifically on them is recommended (Table 1).

In addition to mental state, domains such as social adjustment (Weissman et al. 1981) or quality of life may also be measured by observer-rated scales. Examples include the Social Interview Schedule (SIS;

Table 1. Overview of clinical observer-rated schedules (examples)

Domain	Procedure	Reference	Abbreviation	Distinguishing mark
Global psychopathology	Brief Psychiatric Rating Scale	Overall and Gorham (1976)	BPRS	18 symptom complexes, overall score and 5 sub-scales
	Comprehensive Psychiatric Rating Scale	Asberg et al. (1978), Kuny et al. (1982)	CPRS	65 items, 4 sub-scales, 2 overall scores
	Befundbogen des AMDP-Systems (Rating schedule for the AMDP System)	AMDP (1995)	AMDP	140 items, 9 sub-scales, 3 overall scores
	Inpatient Multidimensional Psychiatric Scale	Hiller et al. (1986)	IMPS	90 items, overall score, 12 sub-scores
Depressive symptoms	Hamilton Depression Scale	Hamilton (1976a)	HAMD	17–21 items, 2–6 sub-scales
	Montgomery–Asberg Depression Scale	Montgomery and Asberg (1979)	MADRS	10 items, overall score
Mania	Bech–Rafaelson Mania Scale	Bech et al. (1978, 1991)	BRMAS	11 items, summary score
Schizophrenia	Positive and Negative Symptom Scale	Kay et al. (1988)	PANSS	30 items, summary scores, 3 sub-scales
Anxiety disorders	Anxiety Status Inventory	Zung (1976a)	ASI	20 items, summary score
	Hamilton Anxiety Scale	Hamilton (1976b)	HAMA	14 items, summary score
Obsessive–compulsive disorders	Yale-Brown Obsessive Compulsive Scale	Goodman et al. (1989a,b)		
Dementia	Arbeitsgemeinschaft für Gerontopsychiatrie	Gutzmann et al. (1989)	AGP-System	176 items, 6 sub-scales
	Alzheimer's Dementia Assessment Scale	Ihl and Weyer (1993)	ADAS	21 items, 1 summary score, 2 sub-scores

Faltermaier et al. 1987) and the Disability Assessment Schedule (Jablensky et al. 1980). Personality characteristics (Möller and von Zerssen 1987) and personality disorders can also be described using observer-rated instruments, as in the new methods of assessment of personality developed for the rating of axis II (personality disorders) of DSM-III (Pfohl et al. 1983; Saß 1986; Stangl et al. 1985). Standardised assessment instruments may also be used to document the side-effects of treatment with psychotropic drugs, as in the scales developed by Simpson and Angus (1970) for the documentation of extrapyramidal motor side-effects, or the UKU Side Effect Rating Scale, which includes a full spectrum of psychotropic drug side-effects (Lingjaerde et al. 1987).

On the basis of multivariate statistical analysis (factor and cluster analysis), the data obtained from administering rating scales may be used to derive factors. These factors identify groups of individual symptoms which tend to occur together. If we consider that the term "clinical syndrome" generally refers to a group of symptoms which frequently occur in combination, it then becomes apparent that the factors extracted from rating scales relating to mental state are conceptually identical to clinical syndromes. Multivariate analysis of the data obtained from different multidimensional psychiatric rating scales applied to different samples of patients has tended to repeatedly generate the same factors or symptom clusters (Cairns et al. 1983a,b; Gebhardt et al. 1981; Lorr et al. 1962; Mombour 1974a,b): paranoid-hallucinatory syndrome, manic syndrome, depressive syndrome, apathetic syndrome, hypochondriacal syndrome, phobic-obsessional syndrome, amnesiac syndrome. For some well-developed observer-rated scales, it has been shown that the factor structure also remains relatively stable across different studies, and for many of the factors this is true even with repeated measurements in the course of treatment (Baumann and Stieglitz 1983; Möller and Hacker 1988; Steinmeyer and Möller 1992). This invariability of the structure of factors across different samples and time points is an important aspect of the validity of a scale (factorial validity). Different psychiatric diagnostic groups are reflected in different characteristic syndrome profiles when rating scales are applied (Möller and von Zerssen 1980).

It is important to bear in mind that identically named syndromes from different scales may vary greatly in terms of the items included, and the correlation between analogous syndrome scores is not always very great. The more syndromes that are represented in a scale, the wider the range of its potential applications will be. However, in order to address specific questions, these broadly applicable rating scales should be combined with one or more specific observer-rating scales, e.g. when studying people with schizophrenia, specific scales for precise measurement of the negative syndrome should be included (Andreasen 1982). In the interests of economy, when research questions are narrow in scope (e.g. if they concern depression), it makes sense to administer specific scales which include only a few factors. As scales which measure the same domain (e.g. depressive symptoms) sometimes focus on different aspects of this domain (Mombour 1976), certain questions may be best addressed by using a combination of them. In terms of developing models of psychopathology, there is continuing debate on the relative merits of single-symptom versus syndrome research strategies (see Bentall 1992; Lenzenweger 1999). At present, the approaches are best viewed as complementary.

When observer-rated instruments are administered by professionally trained observers, it is usually assumed that, in making the rating, the observer decides how much weight to put on the information the patient gives, e.g. an observable improvement in general behaviour and demeanour is taken into account in the rating even if the patient gives no clear report of this improvement. An advantage of this expert assessment is that it reduces the scope for inaccurate assessments resulting from the distortions in patients' perception of themselves, but on the other hand it introduces the danger of distortions which are related to the assessment (rater bias). Systematic distortion in the assessor's observations (Hasemann 1971) can result from the following factors in particular:

1. *Rosenthal effect*: The assessor's expectations influence the result of the assessment.
2. Tendency on the rater's part to systematically *over- or under-rate* the degree of disturbance.
3. *Halo effect*: The result of assessment of one characteristic is influenced by the rater's knowledge of the subject's other characteristics or by the overall impression made by the subject.
4. *Logical errors*: The result of the assessment is influenced by the assessors reporting only those detailed observations which make sense to them in the context of their theoretical and logical preconceptions.

These errors may be partially compensated for by combining observer-rated scales with self-rated scales (von Zerssen 1979, 1982; von Zerssen and Möller 1980). In self-rated instruments, patients can themselves classify past or current behaviour and experience on the basis of fixed rating scales. Self-rated scales have the further advantage that their use is very economical for the assessor and eliminates observer bias. However, their use also introduces the disadvantage that conscious or unconscious tendencies to falsify responses (e.g. tendencies to exaggerate or conceal symptoms, the positive response bias, social

desirability effects) will have a greater impact on patients and are only partially detectable through use of control scales (so-called lie detector scales). Self-assessment procedures tend to be applied especially in order to identify enduring personality dispositions (see below) or current psychiatric disturbance, e.g. the *Klinische Selbstbeurteilungs-Skalen* (Clinical Self-Rating Scales, KSbS; von Zerssen 1976a–d, 1986) or the Self-Report Symptom Inventory (SCL-90; C.R. Derogatis 1977). Self-rated scales in the form of visual analogue scales (so-called barometer scales on which particular dimensions or current experience are graphically represented) are especially useful for intra-individual studies of course over time (Luria 1975). A problem with most personality scales in clinical use, e.g. the Minnesota Multiphasic Personality Inventory (MMPI; Dahlström et al. 1975; Hathaway and McKinley 1963), is that, contrary to their goals, they do not distinguish precisely between habitual personality dispositions and current disturbances in behaviour.

As with observer-rated scales, self-rated scales can also record domains other than psychopathological abnormalities, e.g. social adjustment (Weissman and Bothwell 1976), quality of life (Möller et al. 1996) or the side-effects of psychotropic drugs (National Institute of Mental Health 1976).

Apart from a few scales measuring current mental state which, as with the Self-Report Symptom Inventory (SCL-90), record a very broad spectrum of psychopathological symptoms, most self-rated scales focus on specific aspects of disturbance of subjective experience (Table 2). Examples are inventories of physical and systemic complaints (Fahrenberg 1975; von Zerssen 1976d), depressive symptom scales (Beck et al. 1961; Zung 1965; von Zerssen 1976b) or measures of general subjective well-being (Janke and Debus 1977; von Zerssen 1976c). One of the advantages of this approach is that the quantity of items is limited, a particular strength where severely disturbed psychiatric patients are concerned. In order to obtain a sufficiently clear view of the current psychological state from a subjective point of view, it is always better not only to present a checklist of adjectives describing complaints, but also to add other symptom-oriented scales such as the paranoid-depressiveness scale (von Zerssen 1976b).

However, very precise differentiation between different aspects of "subjective state" is probably not generally meaningful (von Zerssen 1979), in contrast to the detailed measurement of psychological disturbances which may be made by observer assessment. In fact, where results from clinical self-rated scales are compared with observer-rated scales administered by specialists, it seems that the various dimensions of the subjective state which self-rated instruments describe are more similar to one another than the different aspects of psychopathology delineated by clinical observer-rated assessments. This is indicated by joint factor analysis of data from observer ratings and self-ratings of mental state, for example (von Zerssen and Cording 1978). This finding was reported from a study in which the Inpatient Multidimensional Psychiatric Scale (IMPS; Lorr 1974) was applied as an observer-rated measure, while the KSbS (von Zerssen 1976a–d) were used as a self-assessed measure. The self-assessed data were mainly represented in a single factor, the first to emerge, while the observer-rated data were

Table 2. Clinical self-rated procedures (examples)

Domains	Procedure	Abbreviation	Author(s)
Global psychopathology	Self-Report Symptom Inventory	SCL-90	L.R. Derogatis et al. (1976)
		SCL-90R	CIPS (1996)
Depression	*Depressivitäts-Skala* (Depressive Symptom Scale)	DS	von Zerssen (1976b)
	Befindlichkeits-Skala (Mental State Scale)	Bf-S	von Zerssen (1976c)
	Beck Depression Inventory	BDI	Beck et al. (1986)
Anxiety disorders	Self-Rating Anxiety Scale	SAS	Zung (1976b), see also CIPS (1996)
	State-Trait Angst-Inventar (State-Trait Anxiety Inventory)	STAI	Laux et al. (1981)
Obsessive–compulsive symptoms	Hamburger Zwangsinventar (Hamburg Compulsive Inventory)	HZI	Zaworka et al. (1983), Klepsch (1989)
Alcoholism	Münchner Alkoholismustest (Munich Alcoholism Test)	MALT	Feuerlein et al. (1979)

distributed across five further factors. However, it cannot be concluded from this secondary factor analysis (in which the primary factors derived from initial analysis of each scale were also entered as variables) that self-assessment simply produces a factor reflecting a global tendency to complain rather than a differentiated picture of subjective impairment. Primary factor analysis based on single items from the KSbS and also on other self-rated instruments indicates that it certainly can be differentiated at a subjective level between different dimensions of disturbance, such as depressiveness, paranoid tendencies and somatic complaints. However, the depressiveness factor is closely associated with each of the various other types of subjective disturbance.

The level of agreement between self-assessment and observer assessment is variable and depends among other things on the type of disturbance and on symptom severity (Heimann and Schmoker 1974; Prusoff et al. 1972a,b; White et al. 1984). Thus, for example, where depressive symptomatology is severe, as at the time of inpatient admission, agreement is substantially more limited than after partial remission of symptoms at the time of discharge. This is probably connected with greater limitation of the capacity for self-observation among the severely depressed, and probably also with the fact that observers tend to recognise very severe depressive symptoms on the basis of non-verbal evidence to a greater extent than with less severe depressive symptoms, where the patient's verbal reports are more important. Compared with patients with endogenous depressions, those with neurotic depressions show a greater tendency to overstate their symptoms. Degree of agreement between self-rating and observer rating is substantially greater for the amount of change, as measured in longitudinal studies, e.g. in the context of treatment studies, than when psychopathological phenomena are recorded at a single cross-sectional time point (von Zerssen 1986; Möller and von Zerssen 1995).

Multi-methodological diagnostic procedures in which a combination of self-rated and observer-rated scales is applied (Möller et al. 1983; Seidenstücker and Baumann 1978) offer the best guarantee of satisfactory description of both subjective and objective psychopathological state. So far, this has rarely been taken account of in quality of life research, which has mainly been based on self-rating scales (see Vol. 1, Part 2, Chap. 7). In some rating instruments, a combination such as this is a specific requirement, e.g. the *Münchner Alkoholismus-Test* (Munich Alcoholism Test, MALT; Feuerlein et al. 1979). The *Nürnberger Altersinventar* (Nuremberg Old-Age Inventory) similarly requires a combination of self-ratings and observer ratings, further supplemented by tests of performance (Oswald 1979).

Measures of subjective well-being are of particular interest in the area of treatment assessment, particularly barometer scales which measure current disturbances of psychological well-being and lend themselves especially well to repeated measurement. These methods allow a very good description at the self-assessment level of response to a therapeutic intervention. Modern methods of statistical analysis, such as some of the procedures developed for time series analysis, allow satisfactory analysis of such data (Möller et al. 1987, 1989). A full discussion of such approaches is presented by Morley (1994).

Most survey instruments allow the description of current state and, with repeated application, can also be used to examine change over time.

Generally, any multidimensional instrument for examining psychopathology allows us to try to make a diagnostic classification by applying specific algorithms, e.g. to detect characteristic symptom profiles. As would be expected, the results of nosological classification based solely on psychopathological rating scales are not very satisfactory (Möller and von Zerssen 1980), as diagnosis is generally also based on information about clinical history and on hypotheses about the cause of the illness. The Catego system of Wing et al. (1974) is based on the PSE, together with a supplementary scale describing history. This combination allows very satisfactory results to be obtained in the sphere of functional psychoses, and it has now been applied in a variety of large national and international research projects (Wing et al. 1974).

In some simpler scales, a judgement about differential diagnosis may be made simply through calculations on the basis of fixed cut-off scores. Examples are the Newcastle Scale (Carney et al. 1965), designed to differentiate between endogenous and neurotic depression, and the Hachinski Scale, designed to distinguish between multi-infarct dementia and Alzheimer-type dementia (Hachinski et al. 1975).

In connection with the development of operationalised diagnosis systems such as the Research Diagnostic Criteria (RDC) and the Diagnostic and Statistical Manual III-R and IV (DSM-IIIR, DSM-IV), standardised survey instruments have been produced to allow examination of the aspects of clinical history and psychopathology on which the operationalised diagnostic criteria are based. As the RDC encompasses only endogenous psychoses, the corresponding survey instruments can of course be briefer than for the more extensive DSM-III system, which covers all diagnoses. The Schedule of Affective Disorders and Schizophrenia (SADS) was developed specifically for the RDC (Spitzer et al. 1975). For DSM-III, which covers all diagnostic categories, the Diagnostic Interview Schedule (DIS; Robins et al. 1982) was initially developed. A particular goal in the development of the DIS was that it should

be possible for it to be applied in epidemiological field studies by non-psychiatrists (e.g. social workers), but this concept did not lead to fully satisfactory results (Wittchen et al. 1985). A series of fully structured interview schedules and diagnostic instruments developed in the last decade allow both ICD-10 and DSM-III diagnoses to be generated: the Composite Diagnostic Interview (CIDI; Wittchen and Semler 1991), the Structured Clinical Interview for DSM-III (SCID; Wittchen et al. 1991) and the PSE-based Schedules for Clinical Assessment in Neuropsychiatry (SCAN; WHO 1991). An overview of the instruments developed in cooperation with the WHO is provided by Sartorius and Janca (1996). The SCID and the SCAN are currently being partly adapted to DSM-IV. These latter instruments were conceived from the beginning as coherent interview procedures with fully formulated questions (see Vol. 1, Part 1, Chap. 2), whereas the Polydiagnostische Interview (Polydiagnostic Interview, PODI; Philipp and Maier 1986), developed in Mainz, and the Polydiagnostisches System (Polydiagnostic System; Katschnig et al. 1987), developed in Vienna, simply set out and place in context the criteria for individual diagnoses in the diagnostic systems included – not only the RDC and DSM, but also a variety of others. Overall, all these instruments seem likely to lead to a considerable increase in inter-rater reliability in mental state assessment and diagnostic classification.

4 Standardised Procedures for Personality Assessment

Most standardised assessment procedures in personality diagnosis are self-assessment instruments. One of the oldest instruments, the MMPI (Hathaway and McKinley 1943), is still one of the most frequently applied world-wide. In its long form, the MMPI contains 566 items, which through the standard method of analysis provide summary scores on ten clinical and three validity scales. In addition, there are quite a number of short forms, including a German one by Gehring and Blaser (1993). A revised and newly standardised American edition has been available since 1989 (MMPI-2; Hathaway and McKinley 1989). The content of the MMPI items ranges from general statements ("I like reading technical journals") which describe interests and character traits to statements about overt psychiatric symptoms ("Sometimes I hear voices when other people do not hear them"). This and its empirical scale construction (selection of items which distinguish between a particular clinical target group such as depressed patients and a reference group, in this example non-depressed psychiatric patients), is the basis of the MMPI's leading position in psychiatric personality diagnosis. A further reason is the diverse range of validity information, which claims to make it possible to separate deliberately falsified test responses or chance responses from those indicating true psychological cases.

In the 50 years of its existence, thousands of publications about and using the MMPI have appeared (for a comparative review, see Butcher and Rouse 1996), and there is therefore much more information available to the user than its short handbook can convey. Much of the fundamental methodological work about personality questionnaires has been carried out on the basis of the MMPI (for reviews, see Meehl 1973; Wiggins 1973; see also Angleitner and Wiggins 1986). In fact, the MMPI has also been described as a "methodological nightmare" (Rodgers 1972), particularly because individual items may be counted several times over and contribute to up to six summary scales and because of the high correlations between individual scales which result from this and also from the overlap in content between different items.

Not least because of these criticisms, personality questionnaires have increasingly been developed in the United States which, unlike the MMPI, are not based on clinical experience and concepts, but on factor analysis methodology. This is based on the assumption that a more economical and stable description of personality is possible using such orthogonal scales. However, the superiority predicted in theory has not been demonstrated in practice. There are good empirical data which indicate that the various methodological starting points lead to qualitatively similar results (Burisch 1984). Examples of questionnaires constructed by factor analysis are the 16PF by Cattell (Institute for Personality and Ability Testing, Champaign, IL), the various questionnaires by Eysenck (most recently the EPS; Eysenck 1983) and in recent years above all the Big Five Questionnaires (see below) by Costa and McCrae (1992), which have had an important impact on personality psychology in the last few years. The sphere of application of these procedures is not so much in clinical practice as in occupational psychology or research. Only very recently have new developments appeared which have gone through an extensive process of development based on contemporary techniques and explicitly take into account the area of psychological disturbances. These are the Basic Personality Inventory (BPI) by Jackson (1989), a 240-item interview with 12 clinical scales, and the Personality Assessment Inventory (PAI) of Morey (1991), in which responses to 344 statements lead to ratings on a 4-point scale for 22 non-overlapping scales, of which four are validity scales, 11 clinical scales, five treatment scales and two interpersonal scales.

In German-speaking countries, the *Freiburger Persönlichkeitsinventar* (Freiburg Personality Interview, FPI) has come to be the most widely used personality test (Schorr 1995), both in its old edition of 1970 (Fahrenberg et al. 1970) and also in the revised edition FPI-R of 1984 (Fahrenberg et al. 1994). In addition to character traits in the narrower sense (e.g. inhibitedness, excitability, aggression, openness), the scope of the FPI-R also includes psychosomatic concepts such as somatic complaints, satisfaction with life and concerns about health. It thus extends beyond the relatively narrow realm of validity of "normal" personality inventories such as the 16PF (Schneewind et al. 1994), the Big Five Questionnaires NEO-PI and NEO-FFI (Costa and McCrae 1992) or the MPT, developed primarily on the basis of clinical psychiatric personality concepts (von Zerssen et al. 1988), to mention but a few, but it does not measure the full range of psychological disturbances. The Giessen-Test (GT; Beckmann et al. 1990) covers a similar domain of application, but places social psychological concepts at the fore (scales include dominance, susceptibility to influence, social power, social responsiveness, control and prevailing mood). The test is constructed on the basis of psychoanalytic theories.

In connection with the introduction of operationalised definitions of personality disorders in DSM-III (and its successors) and ICD-10, discussions took place about the extent to which self-assessment questionnaires or specific personality questionnaires could complement assessment or even stand alone as an economical screening method in the diagnosis of personality disorders. When this was tried both using existing instruments (e.g. MMPI; Morey et al. 1985; Colligan et al. 1994) and new questionnaires constructed specifically for this purpose (MCMI; Millon et al. 1994), it emerged that too much and probably the wrong thing was being expected of personality questionnaires if categorical agreement diagnoses rather than dimensional agreement with syndromes was required (Engel 1981; Dittman and Stieglitz 1994).

In order to avoid this conflict of purpose, specific instruments for the detection of personality disorder have been developed, based on the concepts of DSM-III, DSM-III-R, DSM-IV and ICD-10. These have essentially involved inclusion of additional observer ratings (Bronisch et al. 1995; Hyler et al. 1988; Loranger et al. 1994; Pfohl et al. 1989).

The International Personality Disorder Examination (IPDE; Loranger 1996) is a procedure which has already become widespread internationally and consists of a combination of a short screening questionnaire followed by a detailed interview with an experienced rater.

An important theme in psychopathological research, which has as yet been taken up relatively little internationally, is the description of premorbid personality in psychiatric patients and its possible influence on vulnerability and course. In this area, von Zerssen (for summaries, see von Zerssen 1993, 1994) has been a particularly significant contributor of concepts and instruments (MPI; von Zerssen et al. 1988).

5 Systematic Observation of Behaviour

In comparison with standardised assessment procedures, a higher level of objectivity is achieved in systematic behavioural analysis, in which only directly observable behaviour is recorded, e.g. counting up of particular details of behaviour or complexes of behaviour during defined periods of observation (von Cranach and Frenz 1969; Fassnacht 1979; Goldfried 1976). Audiovisual recording of behaviour opens up additional possibilities for these methods (Helmchen and Renfordt 1978). Because resource expenditure is great, systematic behavioural analysis has not taken on any great significance in everyday clinical practice apart from behaviour therapy, but has been preserved almost exclusively for research purposes. Even in research, it has generally been used in projects which are especially well staffed, e.g. in research on psychotherapeutic questions.

Systematic behavioural analysis appears to be particularly useful with complex social phenomena, e.g. analysis of the doctor–patient relationship or the interactions between partners (Scholz 1982), which could not be depicted in a sufficiently detailed way through simple rating scales. Interactional analysis involves not only the recording of simpler (e.g. eye contact; Wagner et al. 1983) and more complex modes of behaviour (e.g. content of speech; Winkler and Ellgring 1981), but also analysis of their sequences and of the recurring patterns which appear in these sequences (Hahlweg et al. 1984; Hirschbrunner et al. 1981).

For the study of non-verbal aspects of behaviour (Ellgring 1981), systematic behavioural observation is the most appropriate research method. The potential field of application is wide here. For clinicians, the methods promise to be especially interesting for diagnostic and treatment studies, e.g. of patients with depression (Ellgring and Clarke 1978). These methods not only allow examination of particular aspects of gesture and expression (Ekman and Friesen 1978; Ellgring 1986; Ellgrin and Nagel 1986; Polzer et al. 1992), but can have as a goal the study of the overall functioning of the psychomotor system (Frey et al. 1979, 1981). They are particularly useful where verbal communication is not possible, such as with the severely demented (see Gaebler and Hemsley 1991).

Not only non-verbal but also verbal communication may of course be examined by systematic behavioural observation (Matarazzo and Wiens 1977; Weintraub and Aronson 1967; Winkler and Ellgring 1981). Study of speech may focus on the form of speech (volume, pitch and rate), on the length of utterances or on content. Well-known methodologies include content analysis based on the Gottschalk-Gleser procedure (Gottschalk and Gleser 1969) and the Ulm methods for computerised analysis of the form and content of psychotherapeutic communication (Kächele 1976; Mergenthaler 1985).

6 Objective Tests

The data obtained from objective tests in the narrower sense of the word are derived not from the subject's statements, but from his or her reactions to fixed "stimulus material". Such tests allow analysis of particular psychological or psychologically related functions such as perception, concentration, attention and motor functioning, generally focusing on the performance aspect of these (Jäger and Petermann 1992; Brickenkamp 1996). One of the central aims of the measurement of memory and of "higher" cognitive functions (such as word comprehension or the visual motor system) is the objective demonstration of disturbances of function resulting from organic brain disorders (Lezak 1983; von Cramon et al. 1993). Other useful functions are the measurement of the effects of drugs or psychotherapy, e.g. showing improvement in concentration after anxiety reduction through psychological treatment. Such measures are also increasingly seen as being useful as predictors of real life functioning in schizophrenia (see Wykes et al. 1992). A problem for most psychological performance tests is that no standards are available for the older age-groups with whom old-age psychiatry is concerned. In this respect, the performance tests included in the Nuremberg Old-Age Inventory (Oswald and Fleischmann 1995) represent an important advance.

According to a survey carried out by Schorr (1995) of a representative sample of members of the German Association of Psychologists, of the test procedures carried out in German psychiatric treatment settings, around 24% are intelligence tests, 6% tests of general performance, 59% personality tests and 9% tests of personality development or projective test procedures. Of the intelligence tests, the Wechsler-based procedures – all versions of the *Hamburg-Wechsler Intelligenztest für Erwachsene* (HAWIE), the German version of the Wechsler Adult Intelligence Test Scale (WAIS), and the *Hamburg-Wechsler Intelligenztest für Kinder* (HAWIK), the German version of the Wechsler Intelligence Scale for Children (WISC) – were by far the most common (40% of respondents used the test frequently), and of the personality tests it was the *Freiburger Persönlichkeitsinventar* (by a long way, 37% of respondents), followed by the Giessen-Test and the MMPI. Of the projective procedures, the Thematic Apperception Test (TAT), the Sceno-Test and the Rorschach-Test continue to be used by 13%, 13% and 9% of respondents, respectively. These figures indicate a conservative tendency to adhere to old established test procedures, even though new tests are appearing on the market almost daily.

In the clinical sphere, the best-known and most internationally widespread are the intelligence tests of Wechsler, whose current form is the Wechsler Adult Intelligence Scale-Revised (WAIS-R; Wechsler 1981). In contrast to modern intelligence tests based on factor analysis and despite a number of revisions, the WAIS has psychometric weaknesses, including relatively high correlations between some sub-tests, lack of an objective method for analysis of the open items in the test and limited standardisation of the conditions in which it is carried out (e.g. Weise 1975). However, in clinical use, this disadvantage is more than compensated for by other factors, such as high attractiveness of the test materials, variety and clinical relevance of the test tasks (Klingler and Saunders 1975) and constant contact with the tester; indeed, the limited standardisation of the test situation as a result of the use of open questions, the heterogeneous test material and the interactive way in which the test is carried out are responsible both for some of its weaknesses and for its strengths.

Particularly in the clinical sphere and for people with low intelligence, the IQ scores obtained from the HAWIE-R lie on average up to 10 points below those delivered by the old version of HAWIE (Satzger et al. 1996a). Still greater IQ shifts in the standardisation studies were found in the earlier revision of the HAWIE-R (Schallberger 1987) and clearly show the problems arising from the use of older test norms. Changing from one test to another may lead to substantial errors on repeated measurement if appropriate corrections are not made.

Similar problems with the comparability of different norms can arise when short procedures to estimate general intelligence such as the *Reduzierte Wechsler Intelligenztest für Psychiatrisch Kranke* (Short Wechsler Intelligence Test for the Mentally Ill, WIP; Dahl 1968) or vocabulary tests such as the *Wortschatztest* (WST; Vocabulary Test) (Schmidt and Metzler 1992) or the *Mehrfachwahl-Wortschatz-Intelligenztest* (MWT-B; Multiple Choice Vocabulary Intelligence Test; Lehrl 1989) are used for reasons of economy. When a single sample was examined with both the

WIP and the HAWIE, discrepancies of the order of 9 IQ points were found (Orgass and Hartje 1974), indicating that the WIP values cannot be seen as numerically almost identical to the HAWIE values. The relationships with the HAWIE of norms for the more frequently used vocabulary tests vary: the MWB-R in particular provides considerably higher estimates of IQ than the HAWIE-R, while systematic deviations in norms from the HAWIE-R were not found for the WST (Satzger et al. 1996b).

In the clinical psychiatric sphere, there have been attempts not only to apply the Hamburg-Wechsler Test as a measure of global intelligence, but also to use it to describe strengths in performance and to assess differential diagnostic questions. Profile interpretations and the calculation of various indices, e.g. a quotient of discrepancies or the relationship between verbal and performance IQ, have been used to try to describe characteristic differences in the structure of intelligence between neurotic people, schizophrenics and patients with organic brain damage and even to differentiate between different forms of organic brain disease (e.g. Hobi and Klär 1972; Köhler 1974; Scheller 1973; Scheller and Sittauer 1974; Wechsler 1956). However, other studies indicate that this endeavour should be treated with considerable scepticism (Bäumler 1969; Dahl 1968; Hahlweg 1980; Hunger and Kleim 1976; Mayer et al. 1969; Seydel 1972; Sturm et al. 1975). Most of these indices are derived from theoretical considerations or have been devised or discovered more or less ad hoc on the basis of rather small samples. On a more empirical basis, Baxa and Pakesch (1972) used standards from the HAWIE to develop their Organicity Index, which compares the age-susceptible sub-tests Number Symbol Test and Mosaic Test with the relatively stable tests of "general knowledge" and "general understanding". This index rises sharply with age and thus often allows a clinically meaningful statement to be made about "premature cognitive ageing" (Engel and Satzger 1988; see also Giambra et al. 1995; Rebok et al. 1990). In an empirical investigation of the validity of different indices in distinguishing patients with diffuse organic brain disorders from a group with a variety of neurotic disturbances, this index did hold its own (Baud and Rauchfleisch 1982), although Lindsay and Powell (1994) argue that any such indices should be interpreted cautiously.

A criticism of the Wechsler test which is frequently put forward relates to its high level of dependence on speech, which it has been suggested makes the test results very susceptible to social and educational influences. At the opposite pole, there are a variety of speech-free intelligence tests such as Raven's Matrices (Standard Progressive Matrices, Advanced Progressive Matrices, most recently published in German in Raven 1996a,b) or the Culture Fair Intelligence Tests of Cattell (Institute for Personality and Ability Testing, Champaign, IL). These tests are often very valuable in the clinic, even if they cannot always fulfil their promise of great fairness with respect to sociocultural influences.

In making diagnoses in clinical psychiatric practice, tests of general performance are also frequently applied, e.g. the Pauli Test, the *Konzentrationsleistungstest* (Concentration Performance Test, KLT) and the *d2 Aufmerksamkeits-Belastungs-Test* (d2 Attention Performance Test). In people with mental illnesses, values falling below the expected norms are generally found (Brickenkamp 1994). They allow measurement of performance ability in simple visual-motor or routine cognitive tasks. The validity of these concentration tests in differential diagnosis between organic and non-organic psychological disturbances is not established (Arnold 1975; Eich 1978; Hahlweg 1979), and as the cognitive capacities required vary from test to test, their interpretation can be very difficult even with organically impaired patients.

Differential diagnosis of organic brain disturbances has always been of particular importance in routine clinical practice. In the past, it was usual to apply single tests such as the Benton-Test (Benton 1972) or the *Diagnosticum für Cerebralschädigung* (Diagnostic Test for Brain Damage, DCS; Hillers 1993), generally in combination with a general intelligence battery such as the WAIS. This approach focused excessively on relatively global psychological functions, which are known to often be substantially influenced by non-organic psychological disturbances, as in depressive pseudodementia or the cognitive disturbances characteristically found in patients with schizophrenia (Chapman 1979; Fieguth and Gonzalves 1977). More recently, prevailing test practice has been strongly influenced by neuropsychology, one of whose undoubted effects has been to sharpen awareness of whether the right question is being asked. In psychiatric clinics, detailed neuropsychological results are increasingly collected, in which individual domains of cognitive functioning are examined with specific procedures (Keefe 1995; Reischies 1987; Hartje 1981). Unfortunately, no consensus has yet been reached in this process regarding the precise procedures which should be used. This is in part because assessments vary in their aims, and these need to be clearly specified.

Comprehensive batteries of neuropsychological tests, of which expectations were high in the 1970s and early 1980s, e.g. the *Tübinger Luria-Christensen Neuropsychologische Untersuchungsreihe* (TULUC) by Hamster et al. (1980) or the *Luria Nebraska Neuropsychologische Testbatterie* (LNNB) by Golden et al. (1980) have only rarely been used, in one case (TULUC) because of excessive complexity of administration, validation and interpretation, in another (the

LNNB) because of heavy criticism (e.g. Adams 1980) of a tendency to over-simplify, coarsen and abbreviate. Moreover, while the idea of a unitary "cognitive measurement technology", embodied in a comprehensive measurement methodology, is appealing and has many adherents, especially in the United States (e.g. Russell 1994), this approach tends to break down when attempts are made to apply comprehensive batteries in practice.

A more promising approach is a flexible one in which the individual areas of functioning to be measured are selected more precisely on the basis of the particular question being asked (Bauer 1994; Benton 1994). To a large extent, this is a return to Luria's syndrome analytic approach to description of higher cortical functions (Luria 1970), which starts from a comprehensive and systematic reduction to their components of complex modes of behaviour. The analysis of the cognitive effects of brain damage is ideally suited to single-case methodology, as discussed by Canavan (1994).

Precise measurement of cognitive deficits is especially important in people with dementia (Carlesimo and Oscar-Berman 1992; Butters et al. 1995) and people with schizophrenia (Elliott and Sahakian 1995; Frith 1992). Here, two domains of functioning have particular importance: attention and memory performance in the case of dementia and "executive functions" (planning and judgement ability, mental flexibility, decisiveness) in schizophrenia. Speed of cognitive performance (as discussed by Schuri et al. 1994) is measured using reaction time measurements in paper and pencil tests such as the *Zahlen-Verbindungs-Test* (Number-Joining Test) or the Trail Making Test (new norms in Spreen and Strauss 1991), and selective attention using the aforementioned d2 Attention Performance Test. The usefulness of the Continuous Performance Task (CPT; Cornblatt and Keilp 1994; Kathmann et al. 1996) in measuring sustained attention, particularly in schizophrenia research, has been established, but the various forms of the test may reflect different psychological processes (see Nuechterlein et al. 1994).

Memory tests which are quite commonly used include a wide range of forms of word lists, e.g. Rey's auditory verbal learning test (description and norms, e.g. in Lezak 1983 and Spreen and Strauss 1991) or the corresponding sub-tests in the Alzheimer's Disease Assessment Scale (ADAS; Rosen et al. 1984) and the *Demenztest* (Dementia Test; Kessler et al. 1988), memory batteries containing a wide variety of tasks such as the revised version of the Wechsler Memory Scale (WMS-R; Wechsler 1987) and more recent computerised test procedures which make somewhat easier the work of the tester, which may be all too monotonous (Satzger and Engel 1996). The classical Benton Test should also be mentioned here (Benton 1996); in addition to memory, this test also requires awareness, visuospatial awareness and motor abilities. These abilities are often diffusely disturbed in patients with organic brain deficits, which explains the relative sensitivity of the Benton Test as an "omnibus test" for organic brain disorders (Larrabee et al. 1985; von Kerekjarto 1961; Strunk and Faust 1967). However, it has not been clearly established whether the Benton Test is valid in making a differential diagnosis when applied to other samples of psychiatric patients (Hasse-Sander et al. 1996; Hahlweg 1979; Hegenscheit and Cohen 1972; Velkoborsky 1964). For example, memory impairment in schizophrenia is particularly well documented (e.g. Saykin et al. 1991).

To measure planning ability (executive functions), the Wisconsin Card Sorting Test (WCST; see Nelson 1976) and the book version of Halstead's Category Tests (HCT; DeFillipis and McCampbell 1976) are often used. Together with the CPT, these latter instruments are also popular as stimulation tests in functional brain examinations with electrophysiological or radiological methods, in which localised brain functions are being described (e.g. Seidman et al. 1994; Catafau et al. 1994; see also Malloy and Richardson 1994).

Measurement instruments have a particular role in the area of dementia. While psychological tests generally measure circumscribed areas of performance across a wide range of subjects, in dementia instruments focus on measurement of a wide range of cognitive abilities which cannot very clearly be separated from one another, targeting a limited specific group of subjects. Typically, they are not standardised in the usual sense, but instead reference values are provided for particular groups and particular grades of severity.

A very frequently used instrument is the Mini-Mental State Examination (MMSE) by Folstein et al. (1975). The MMSE is a simple screening instrument to measure severe cognitive disturbances in older people. In the space of 5 min, questions and tasks are administered in the areas of orientation, attention, calculation, memory, speech and performance of simple tasks, and a global score is calculated. The MMSE is probably the most frequently used instrument to describe the level of severity in samples of demented patients.

While the MMSE can be administered without difficulty by someone with quite a basic level of training (and in this respect resembles most psychological tests), this is no longer the case for other procedures. The *Strukturiertes Interview für die Diagnose vom Alzheimer-Typ, der Multiinfarktdemenz und Demenzen anderer Ätiologie nach DSM-III-R und ICD-10* (Structured Interview for the Diagnosis of Alzheimer-

Type Dementia, Multi-infarct Dementia and Dementias of other types according to DSM-III-R and ICD 10 (SIDAM) by Zaudig et al. (1990) begins by setting similarly simple cognitive tasks, but the scope of the instrument then goes beyond that of the MMSE (the MMSE rating can be calculated on the basis of the SIDAM) with the inclusion of clinical ratings, allowing greater specificity about the diagnosis of dementia. Two further commonly used instruments in the area of dementia also consist of a similar combination of tests and rating procedures.

The Alzheimer's Disease Assessment Scale (ADAS; Rosen et al. 1984) is an extended form of the MMSE; it gives more weight to deficits in memory and is primarily applied in clinical trials of anti-dementia drugs. The Cambridge Mental Disorders of the Elderly Examination (CAMDEX; Roth et al. 1988) is a more extensive procedure; in addition to 60 standardised presented cognitive items, it also calls for interviews with the patient and with an informal carer. The advantages and disadvantages of a variety of cognitive measures for use with the elderly are presented by Woods (1994).

7 References

Adams KM (1980) In search of Luria's battery: a false start. J Consult Clin Psychol 48: 511–516

*AMDP (Arbeitsgemeinschaft für Methodik und Dokumentation in der Psychiatrie) (ed) (1995) Das AMDP-System. Manual zur Dokumentation psychiatrischer Befunde. Hogrefe, Göttingen

Andreasen NC (1982) Negative symptoms in schizophrenia. Definition and reliability. Arch Gen Psychiatry 39: 784–788

*Angleitner A, Wiggins JS (1986) Personality assessment via questionnaires. Current issues in theory and measurement. Springer, Berlin Heidelberg New York

Arnold W (1975) Der Pauli-Test. Springer, Berlin Heidelberg New York

Asberg M, Montgomery SA, Peris G, Schalling D, Sedvall G (1978) A comprehensive psychopathological scale. Acta Psychiatr Scand 271[Suppl]: 5–27

Baud U, Rauchfleisch U (1982) Zur Diagnostik hirnorganischer Störungen mit Hilfe des Hamburg-Wechsler-Intelligenztests für Erwachsene. Eine Untersuchung zur differentialdiagnostischen Validität des HAWIE. Diagnostica 28: 248–262

Bauer RM (1994) The flexible battery approach to neuropsychological assessment. In: Vanderploeg RD (ed) Clinician's guide to neuropsychological assessment. Erlbaum, Hillsdale, pp 259–290

Baumann U, Stieglitz RD (1983) Testmanual zum AMDP-System. Empirische Studien zur Psychopathologie. Springer, Berlin Heidelberg New York

Bäumler G (1969) Zum altersbedingten psychischen Leistungsabbau mit Berücksichtigung der Stroop-Interferenzneigung. Psychol Beitr 11: 34–68

Baxa W, Pakesch E (1972) Mitteilung über die Verwendung eines Index am HAWIE zur Bestimmung einer sekundären Intelligenzreduzierung. Wien Z Nervenheilkd 30: 131–142

Bech P, Rafaelsen OJ, Kramp P, Bolwig TG (1978) The Mania Rating Scale: scale construction and inter-observer agreement. Neuropharmacology 17: 430–431

Bech P, Kastrup M, Rafaelsen OJ (1991) Minikompendium psychiatrischer Ratingskalen. Springer, Berlin Heidelberg New York

Beck AT, Ward CH, Mendelson M, Mock J, Erbaugh J (1961) An inventory for measuring depression. Arch Gen Psychiatry 4: 561–571

Beck AT, Rush AJ, Shaw BF, Emery G (1986) Kognitive Therapie der Depression, 2nd edn. Psychologie Verlags Union, Munich

Beckmann D, Brähler E, Richter HE (1990) Gießen-Test (GT), 4th edn. Huber, Bern

Bentall RP (1992) Reconstructuring psychopathology. Psychologist 5: 61–65

Benton AL (1972) Der Benton-Test. Handbuch. Huber, Stuttgart

Benton AL (1994) Neuropsychological assessment. Annu Rev Psychol 45: 1–23

Benton AL (1996) Der Benton-Test, 7th edn. Hogrefe, Göttingen

Borkenau P, Ostendorf F (1993) NEO-Fünf-Faktoren-Inventar. Hogrefe, Göttingen

Brett-Jones J, Garety PA, Hemsley DR (1987) Measuring delusional experiences: a method and its application. Br J Clin Psychol 26: 257–265

Brickenkamp R (1994) Test d2 Aufmerksamkeits-Belastungs-Test, 8th edn. Hogrefe, Göttingen

Brickenkamp R (1996) Handbuch psychologischer und pädagogischer Tests, 2nd edn. Hogrefe, Göttingen

Bronisch T, Hiller W, Mombour W, Zaudig M (1995) IDCL-P: Internationale Diagnose Checkliste für Persönlichkeitsstörungen nach ICD-10 und DSM-IV, Manual. Huber, Bern

Burdock EI, Hardesty AS (1969) Structured clinical interview. Springer, Berlin Heidelberg New York

Burisch M (1984) Approaches to personality inventory construction: a comparison of merits. Am Psychol 39: 214–227

Butcher JN, Rouse SV (1996) Personality: individual differences and clinical assessment. Annu Rev Psychol 47: 87–111

Butters N, Delis DC, Lucas JA (1995) Clinical assessment of memory disorders in amnesia and dementia. Annu Rev Psychol 46: 493–523

Cairns V, Faltermaier T, Wittchen HU, Dilling H, Mombour W, von Zerssen D (1983a) Some problems concerning the reliability and structure of the scales in the Inpatient Multidimensional Psychiatric Scale. Arch Psychiatr Nervenkr 232: 395–406

Cairns V, von Zerssen D, Stutte KH, Mombour W (1983b) The stability of the symptom grouping in the Inpatient Multidimensional Psychiatric Scale (IMPS). J Psychiatr Res 17: 19–28

Canavan AGM (1994) Single case method in clinical neuropsychology. In: Lindsay SJE, Powell GE (eds) Handbook of clinical adult psychology, 2nd edn. Routledge, London, pp 765–786

Carlesimo GA, Oscar-Berman M (1992) Memory deficits in Alzheimer's patients: a comprehensive review. Neuropsychol Rev 3: 119–169

Carney MWP, Roth M, Garside RF (1965) The diagnosis of depressive syndromes and the prediction of ECT response. Br J Psychiatry 111: 659–674

Catafau AM, Parellada E, Lomena FJ et al (1994) Prefrontal and temporal blood flow in schizophrenia: resting and

activation technetium-99m-HMPAO SPECT patterns in young neuroleptic-naive patients with acute disease. J Nucl Med 35: 935–941

Chapman LJ (1979) Recent advances in the study of schizophrenic cognition. Schizophr Bull 5: 568–580

*CIPS (1996) Internationale Skalen für Psychiatrie, 4th edn. Beltz, Weinheim

Colligan RC, Morey LC, Offord KP (1994) The MMPI/MMPI-2 personality disorder scales: contemporary norms for adults and adolescents. J Clin Psychol 50: 168–200

Cornblatt BA, Keilp JG (1994) Impaired attention, genetics, and the pathophysiology of schizophrenia. Schizophr Bull 20: 31–46

Costa PT Jr, McCrae RR (1992) Manual for the Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI). Psychological Assessment Resources, Odessa, FL

Cronholm B, Daly RJ (1982) Evaluation of psychiatric treatment. In: Helgason T (ed) Methodology in evaluation of psychiatric treatment. Cambridge University Press, Cambridge, pp 183–204

Dahl G (1968) Übereinstimmungsvalidität des HAWIE und Entwicklung einer reduzierten Testform. Hain, Meisenheim

Dahlström WG, Welsh GS, Dahlstrom LE (1975) An MMPI-handbook, vol II: Research applications. University of Minnesota Press, Minnesota

DeFilippis NA, McCampbell E (1979) The Booklet Category Test. Psychological Assessment Resources, Odessa, FL

Derogatis LR, Lipman RS, Covi L (1976) SCL-90. Self-Report Symptom Inventory. In: Guy W (ed) ECDEU assessment manual for psychopharmacology. National Institute of Mental Health, Rockville, pp 313–331

Derogatis CR (1977) SCL. Administration, scoring and procedures. Manual for the revised version and other instruments of the psychopathology rating scale series. Johns Hopkins University School of Medicine, Baltimore

Dittmann V, Stieglitz RD (1994) Diagnostik von Persönlichkeitsstörungen. In: Stieglitz RD, Baumann U (eds) Psychodiagnostik psychischer Störungen. Enke, Stuttgart, pp 230–244

Eich FX (1978) Verfahren zur Leistungsmessung. In: Schmidt LR (ed) Lehrbuch der klinischen Psychologie. Enke, Stuttgart, pp 325–351

Ekman P, Friesen WV (1978) Manual for the facial action code. Consulting Psychologists, Palo Alto

Ellgring H (1981) Nonverbal communication. A review of research in Germany. Ger J Psychol 5: 59–84

Ellgring H (1986) Nonverbal expression of psychological states in psychiatric patients. Eur Arch Psychiatry Neurol Sci 236: 31–34

Ellgring H, Clarke AH (1978) Verlaufsbeobachtungen anhand standardisierter Videoaufzeichnungen bei depressiven Patienten. In: Helmchen H, Renfordt E (eds) Fernsehen in der Psychiatrie. Thieme, Stuttgart, pp 68–78

Ellgring H, Nagel U (1986) Zur Funktion des mimischen Ausdrucks – Mimisches Verhalten bei Vorstellungen und Mitteilungen. In: Kolitzus H, Ellgring H (eds) Video in Psychiatrie und Psychotherapie, vol 7. Max-Planck-Institut für Psychiatrie, Munich, pp 119–134

Elliott R, Sahakian BJ (1995) The neuropsychology of schizophrenia: relations with clinical and neurobiological dimensions. Psychol Med 25: 581–594

Engel RR, Satzger W (1988) Kognitive Defizite im Alter und therapeutische Evaluation von enzephalotropen Substanzen. In: Kanowski S, Ladurner G (eds) Dementielle Erkrankungen im Alter. Pathogenetische Modelle und therapeutische Wirklichkeit. Thieme, Stuttgart New York, pp 81–85

Engel RR (1981) Psychodiagnostik zur Abgrenzung psychopathologischer Störungen. In: Rey ER (ed) Klinische Psychologie. Fischer, Stuttgart, pp 67–74

Eysenck HJ (1983) Eysenck-Persönlichkeits-Inventar (EPI), 2nd edn. Hogrefe, Göttingen

Fahrenberg J (1975) Die Freiburger Beschwerdenliste FBL. Z Klin Psychol 4: 79–106

Fahrenberg J, Selg H, Hampel R (1970) Das Freiburger Persönlichkeitsinventar (FPI). Hogrefe, Göttingen

Fahrenberg J, Hampel R, Selg H (1994) Das Freiburger Persönlichkeitsinventar (FPI), 6th edn. Hogrefe, Göttingen

Faltermaier T, Hecht H, Wittchen HU (1987) Die Social Interview Schedule (deutschsprachige modifizierte Version). Roderer, Regensburg

Fassnacht J (1979) Systematische Verhaltensbeobachtung. Reinhardt, Munich

Feuerlein W, Küfner H, Ringer C, Antons K (1979) Münchner Alkoholismus-Test (MALT). Manual. Beltz, Weinheim

Fieguth G, Gonzalves N (1977) Testleistung chronisch Schizophrener im HAWIE. Arch Psychiatr Nervenkr 223: 139–149

Fischer G (1974) Einführung in die Theorie psychologischer Tests. Huber, Bern

Folstein MF, Folstein SE, McHoug PR (1975) "Mini-Mental State": a practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res 12: 189–198

Frey S, von Zerssen D, Hansen W, Harders S (1979) Probleme der Verhaltensmessung in Einzelfalluntersuchungen. In: Petermann F, Hehl FJ (eds) Einzelfallanalyse. Urban and Schwarzenberg, Munich, pp 159–182

Frey S, Hirsbrunner HP, Pool J, Daw W (1981) Das Berner System zur Untersuchung nonverbaler Interaktion. I. Die Erhebung des Rohdatenprotokolls. In: Winkler P (ed) Methoden der Analyse von Face-to-Face-Situationen. Metzler, Stuttgart, pp 203–236

*Frith CD (1992) The cognitive neuropsychology of schizophrenia. Erlbaum, Hillsdale

Gaebler HC, Hemsley DR (1991) The assessment and short term manipulation of affect in the severely demented. Behavioural Psychotherapy 19: 145–156

Garety PA, Hemsley DR (1987) Characteristics of delusional experience. Eur Arch Psychiatry Neurol Sci 236: 294–298

Gebhardt R, Pietzcker A, Freudenthal K, Langer C (1981) Die Bildung von Syndromen im AMP-System. Arch Psychiatr Nervenkr 231: 93–109

Gehring A, Blaser A (1993) MMPI. Deutsche Kurzform für Handauswertung, 2nd edn. Huber, Bern

Giambra LM, Arenberg D, Kawas C, Zonderma AB, Costa PT (1995) Adult life span changes in immediate visual memory and verbal intelligence. Psychol Aging 10: 123–139

Goldberg DP (1972) The detection of psychiatric illness by questionnaire. Oxford University Press, London

Golden CJ, Hammeke TA, Purish AD (1980) The Luria-Nebraska-Neuropsychological Battery. Western Psychological Services, Los Angeles

Goldfried MR (1976) Behavioral assessment. In: Weiner JB (ed) Clinical methods in psychology. Wiley, New York, pp 281–330

Goodman WK, Price LH, Rasmussen SA, Mazure C, Fleischmann RL, Hill CL, Heninger GR, Charney DS (1989a) The Yale-Brown Obsessive Compulsive Scale. I. Development, use, and reliability. Arch Gen Psychiatry 46: 1006–1011

Goodman WK, Price LH, Rasmussen SA, Mazure C, Delgado P, Heninger GR, Charney DS (1989b) The Yale-Brown Obsessive Compulsive Scale. II. Validity. Arch Gen Psychiatry 46: 1012–1016

Gottschalk L, Gleser A (1969) The measurement of psychological states through the content analysis of verbal behavior. University of California Press, Berkeley

Gutzmann H, Kanowski S, Krüger H, Urban R, Ciompi L (1989) Das AGP-System. Manual zur Dokumentation gerontopsychiatrischer Befunde. Springer, Berlin Heidelberg New York

Hachinski VC, Iliff LD, Cihak E et al (1975) Cerebral blood flow in dementia. Arch Neurol 32: 632–637

Hahlweg K (1979) Validierung einer Testbatterie zur Erfassung hirnorganischer Schädigungen. Diagnostica 4: 299–313

Hahlweg K (1980) Überprüfung der differentialdiagnostischen Validität des Hamburg-Wechsler-Intelligenztests für Erwachsene (HAWIE). Schweiz Z Psychol 39: 51–62

Hahlweg K, Reisner L, Kohli G, Vollmer M, Schindler L, Revenstorf D (1984) Development and validity of a new system to analyse interpersonal communication (KPI: Kategoriensystem für partnerschaftliche Interaktion). In: Hahlweg K, Jacobson NS (eds) Marital interaction: analysis and modification. Guilford, New York

Hamilton M (1959) The assessment of anxiety states by rating. Br Med Psychol 32: 50–55

Hamilton M (1967) Development of a rating scale for primary depressive illness. Br J Soc Clin Psychol 6: 278–296

Hamilton M (1976a) HAMD. Hamilton Depression Scale. In: Guy W (ed) ECDEU assessment manual for psychopharmacology. National Institute of Mental Health, Rockville, pp 179–192

Hamilton M (1976b) HAMA. Hamilton Anxiety Scale. In: Guy W (ed) ECDEU assessment manual for psychopharmacology. National Institute of Mental Health, Rockville, pp 193–198

Hamster W, Langner W, Mayer K (1980) Tübinger Luria-Christensen neuropsychologische Untersuchungsreihe (TÜ-LUC). Beltz, Weinheim

Hartje W (1981) Neuropsychologische Diagnose cerebraler Funktionsbeeinträchtigungen. Nervenarzt 52: 649–654

Hasemann K (1971) Verhaltensbeobachtung. In: Heiss R (ed) Handbuch der Psychologie, 3rd edn, vol 6. Hogrefe, Göttingen, pp 807–836

Hasse-Sander I, Horn R, Müller H, Schröder MR, Möller HJ (1996) Zur Validität des Benton-Tests in der Diagnostik der Alzheimer-Demenz. Z Gerontopsychol Psychiatr 9: 65–77

Hathaway SR, McKinley JC (1943) The Minnesota Multiphasic Personality Schedule. University of Minnesota Press, Minneapolis

Hathaway SR, McKinley JC (1963) MMPI Saarbrücken. Handbuch zur deutschen Ausgabe des MMPI. Huber, Bern

Hathaway SR, McKinley JC (1989) MMPI-2. Minnesota Multiphasic Personality Inventory 2. University of Minnesota Press, Minneapolis

Hegenscheidt M, Cohen R (1972) Zur Erfassung der Umstellungsfähigkeit bei hirnorganisch geschädigten Personen. Z Klin Psychol 1: 1–20

Heimann H, Schmocker A (1974) Zur Problematik der Beurteilung des Schweregrades psychiatrischer Zustandsbilder. Arzneimittelforsch 24: 1004–1006

Heimann H, Obermair W, Boller W, Stoll KD (1977) Videotape training in psychiatric practice. Prog Neuropsychopharmacol Biol Psychiatry 1: 141–145

Helmchen H, Renfordt E (1978) Fernsehen in der Psychiatrie. Thieme, Stuttgart

Hiller W, von Zerssen D, Mombour W, Wittchen HU (1986) Die IMPS. Beltz, Weinheim

Hillers F (1993) Diagnosticum für Cerebralschädigung (DCS), 3rd edn. Hogrefe, Göttingen

Hirschbrunner HP, Florin A, Frey S (1981) Das Berner System zur Untersuchung nonverbaler Interaktion. II. Die Auswertung von Zeitreihen visuellauditiver Information. In: Winkler P (ed) Methoden der Analyse von Face-to-Face-Situationen. Metzlersche Verlagsbuchhandlung, Stuttgart, pp 237–268

Hobi V, Klär A (1972) Zur Frage der Struktur und des Altersabbaues der Intelligenz bei Toxikomanen. Diagnostica 18: 159–179

Hofstätter PR (1957) Psychologie. Fischer, Frankfurt

Honigfeld G, Klett CJ (1965) The nurses' observation scale for inpatient evaluation. J Clin Psychol 21: 65–71

Huber G (1976) Zur Problematik quantitativer Verlaufsbeobachtungen bei Schizophrenen. Psychopathometrie 2: 61–66

Hunger I, Kleim I (1976) Zur diagnostischen Bedeutung des Abbauquotienten im Hamburg-Wechsler-Intelligenztest. Nervenarzt 47: 198–200

Hyler SE, Rieder RO, Williams JB, Spitzer RL, Hendler J, Lyons M (1988) The personality diagnostic questionnaire: development and preliminary results. J Pers Disord 2: 229–237

Ihl R, Weyer G (1993) Alzheimer's Disease Assessment Scale. Manual. Beltz, Weinheim

Jablensky A, Schwarz R, Tomow T (1980) WHO collaborative study on impairements and disabilities associated with schizophrenic disorders. A preliminary communication: objectives and methods. Acta Psychiatr Scand 62[Suppl 286]: 152–159

Jackson D (1989) Basic Personality Inventory Manual. Sigma Assessment System, Port Huron

Jäger RS, Petermann F (ed) (1992) Psychologische Diagnostik. Ein Lehrbuch, 2nd edn. Beltz, Weinheim

Janke W, Debus G (1977) Die Eigenschaftswörterliste EWLK. Ein Verfahren zur Messung der Befindlichkeit. Hogrefe, Göttingen

Kächele H (1976) Maschinelle Inhaltsanalyse in der psychoanalytischen Prozeßforschung. Post-doctoral thesis, University of Ulm

Kathmann N, Wagner M, Satzger W, Engel RR (1996) Vigilanzmessung auf Verhaltensebene: Der Continuous Performance Test – München (CPT-M). In: Möller HJ, Engel RR, Hoff P (eds) Befunderhebung in der Psychiatrie: Lebensqualität, Negativsymptomatik und andere aktuelle Entwicklungen. Springer, Berlin Heidelberg New York, pp 331–338

Katschnig H, Lenz G, Musalek M, Nutzinger D, Schanda H, Simhandl C (1987) Das "Polydiagnostische System-2 (PS-2)". Wien Med Wochenschr (Sonderheft) 137: 3–4

Katz MM, Lyerly SB (1963) Methods of measuring adjustment and social behavior in the community. Psychol Rep 13: 503–535

Kay SR, Fiszbein A, Opler LA (1988) Positive and negative symptom scale (PANSS) for schizophrenia. Schizophr Bull 13(2): 21–76

Keefe RS (1995) The contribution of neuropsychology to psychiatry. Am J Psychiatry 152: 6–15

Kessler F, Denzler P, Markowitsch HJ (1988) Demenztest. Beltz, Weinheim

Klepsch R (1989) Entwicklung computerdialogfähiger Kurzformen des Hamburger Zwangsinventars. Deutscher Studienverlag, Weinheim

Klingler DE, Saunders DR (1975) A factor analysis of the items for nine subtests of the WAIS. Multivar Behav Res 10: 131–154

Köhler W (1974) Kriterien verstandesmäßigen Leistungsverlusts chronisch Alkoholkranker im HAWIE. Z Exp Angew Psychol 21: 103–115

Kuny S, Maurer M, Luckner MV, Woggon B (1982) Deutschsprachige Version der Comprehensive Psychological Rating Scale (CPRS). Int Pharmacopsychiatry 17: 314–337

Larrabee GJ, Kane RL, Schuck JR, Francis DJ (1985) Construct validity of various memory testing procedures. J Clin Exp Neuropsychol 7: 239–250

Laux L, Glanzmann P, Schaffner P, Spielberger CD (1981) Das State-Trait-Angstinventar (STAI). Beltz, Weinheim

Lee MA, Jayathilake K, Meltzer HY (1999) A comparison of the effects of clozapine with typical neuroleptics on cognitivie function in neuroleptic responsive schizophrenics. Schizophr Res 37(i): 1–12

Lehrl S (1989) Mehrfachwahl-Wortschatz-Intelligenztest MWT-B. Manual, 2nd edn. Perimed, Erlangen

Lenzenweger MF (1999) Schizophrenia: refining the phenotype, resolving endophenotypes. Behav Res Ther 37: 281–295

*Lezak MD (1983) Neuropsychological Assessment. Oxford University Press, New York

Lienert GA (1969) Testaufbau und Testanalyse. Beltz, Weinheim

Lindsay SJ, Powell GE (1994) Practical issues of investigation in clinical psychology. In: Lindsay SJ, Powell GE (eds) Handbook of clinical adult psychology, 2nd edn. Routledge, London, pp 1–34

Lingjaerde O, Ahlfors UG, Bech P, Dencker SJ, Elgen K (1987) The UKU side effect rating scale. Acta Psychiatr Scand 76[Suppl]: 1–100

Loranger AW (1996) IPDE: International Personality Disorder Examination; ICD-10 Modul. Huber, Bern

Loranger AW, Sartorius N, Andreoli A et al (1994) IPDE: the International Personality Disorder Examination. The WHO/ADAMHA international pilot study of personality disorders. Arch Gen Psychiatry 51: 215–224

Lorr M (1974) Assessing psychotic behavior by the IMPS. In: Pichot P, OlivierMartin R (eds) Psychological measurements in psychopharmacology. Karger, Basel (Modern problems in pharmacopsychiatry, vol 7, pp 50–63)

Lorr M, McNair M, Klett CJ, Lasky JJ (1962) Evidence of ten psychotic syndromes. J Cons Psychol 26: 185–189

Luria RE (1970) Die höheren kortikalen Funktionen des Menschen und ihre Störungen bei örtlichen Hirnschädigungen. Deutscher Verlag der Wissenschaften, Berlin

Luria RE (1975) The validity and reliability of the Visual Analogue Mood Scale. J Psychiatr Res 12: 51–57

Malloy PF, Richardson ED (1994) Assessment of frontal lobe functions. J Neuropsychiatry Clin Neurosci 6: 399–410

Matarazzo JD, Wiens AN (1977) Speech behavior as an objective correlate of empathy and outcome in interview and psychotherapy research. A review with implications for behavior modification. Behav Modif I: 453–480

Mayer K, Mayer B, Hamster W (1969) Psychodiagnostische und faktorenanalytische Untersuchungen zur sog. traumatischen Hirnleistungsschwäche. Dtsch Z Nervenheilkd 196: 331–342

Meehl PE (1973) Psychodiagnosis: selected papers. University of Minnesota Press, Minneapolis

Meili R (1961) Lehrbuch der psychologischen Diagnostik. Huber, Bern

Mergenthaler E (1985) Textbank systems. Springer, Berlin Heidelberg New York

Millon T, Millon C, Davis R (1994) Manual for the Millon Clinical Multiaxial Inventory III (MCMI-III). National Computer Systems Assessment, Minneapolis

Möller HJ, Hacker H (1988) A study concerning the sample dependency and temporal variance of the factor structure in the Inpatient Multidimensional Psychiatric Scale. Psychopathology 21: 281–290

Möller HJ, von Zerssen D (1980) Probleme und Verbesserungsmöglichkeiten der psychiatrischen Diagnostik. In: Biefang S (ed) Evaluationsforschung in der Psychiatrie. Enke, Stuttgart, pp 167–207

Möller HJ, von Zerssen D (1987) Prämorbide Persönlichkeit von Patienten mit affektiven Psychosen. In: Kisker KP, Lauter H, Meyer JE, Müller C, Strömgren E (eds) Affektive Psychosen. Psychiatrie der Gegenwart, 3rd edn, vol 5. Springer, Berlin Heidelberg New York, pp 165–179

Möller HJ, von Zerssen D (1995) Self-rating procedures in the evaluation of antidepressants. Review of the literature and results of our studies. Psychopathology 28: 291–306

Möller HJ, Barthelmes H, von Zerssen D (1983) Forschungsmöglichkeiten auf der Grundlage einer routinemäßig durchgeführten Basis- und Befunddokumentation. Psychiatr Clin 16: 45–61

Möller HJ, Leitner M, Dietzfelbinger T (1987) A linear mathematical model for computerized analyses of mood curves. An empirical investigation on mood courses in depressive and schizophrenic inpatients. Eur Arch Psychiatry Neurol Sci 236: 260–268

Möller HJ, Blank R, Steinmeyer EM (1989) Single-case evaluation of sleep-deprivation effects by means of nonparametric time-series analysis (according to the HTAKA model). Eur Arch Psychiatry Neurol Sci 239: 133–139

Möller HJ, Engel RR, Hoff P (1996) Befunderhebung in der Psychiatrie: Lebensqualität, Negativsymptomatik und andere aktuelle Entwicklungen. Springer, Berlin Heidelberg New York

Mombour W (1974a) Syndrome bei psychiatrischen Erkrankungen. Eine vergleichende Untersuchung mit Hilfe von zwei Schätzskalen für die psychopathologischen Befunde (IMPS und AMP). Arch Psychiatr Nervenkr 219: 331–350

Mombour W (1974b) Symptomhäufigkeiten bei psychiatrischen Erkrankungen. Eine vergleichende Untersuchung mit zwei Schätzskalen für den psychopathologischen Befund (IMPS und AMP-Skala). Arch Psychiatr Nervenkr 219: 133–152

Mombour W (1976) Systematik psychischer Störungen. In: Pongratz LJ (ed) Handbuch der Psychologie, vol 8/1. Hogrefe, Göttingen, pp 116–153

Montgomery SA, Asberg M (1979) A new depression rating scale designed to be sensitive to change. Br J Psychiatry 134: 382–389

Morey LC (1991) Personality Assessment Inventory: professional manual. Psychological Assessment Resources, Odessa, FL

Morey LC, Waugh MH, Blashfield RK (1985) MMPI scales for DSM-III personality disorders. J Pers Assess 49: 245–251

*Morley SJ (1994) Single case methodology in psychological therapy. In: Lindsay SJ, Powell GE (eds) Handbook of clinical adult psychology, 2nd edn. Routledge, London, pp 723–745

National Institute of Mental Health (1976) Clinical global impressions. In: Guy W (ed) ECDEU assessment manual for psychopharmacology. National Institute of Mental Health, Rockville, pp 217–222

Nelson HE (1976) A modified card-sorting test sensitive to frontal lobe defects. Cortex 12: 313–324

Nuechterlein KH, Dawson ME, Green MF (1994) Information processing abnormalities as neuropsychological vulnerability indicators for schizophrenia. Acta Psychiatr Scand 90[Suppl 384]: 71–79

Orgass B, Hartje W (1974) Bewährung einer HAWIE-Kurzform (WIP nach Dahl) bei hirngeschädigten Patienten. II. Numerische Übereinstimmung zwischen WIP und HAWIE. Diagnostica 20: 22–30

Oswald WD (1979) Psychopathometrische Verfahren und Fragebögen für gerontopsychologische Untersuchungen. Z Gerontol 12: 341–350

Oswald WD, Fleischmann UM (1995) Nürnberger-Alters-Inventar (NAI), 3rd edn. Hogrefe, Göttingen

Oswald WD, Roth E (1987) Der Zahlen-Verbindungs-Test, 2nd edn. Hogrefe, Göttingen

Overall JE, Gorham DR (1976) BPRS. Brief Psychiatric Rating Scale. In: Guy W (ed) ECDEU assessment manual for psychopharmacology. National Institute of Mental Health, Rockville, pp 157–169

Pfohl B, Stangl D, Zimmerman M (1983) SIDP, 2nd edn. I. Interview. Department of Psychiatry, University of Iowa

Pfohl B, Blum N, Zimmermann M, Stangl D (1989) Structured Interview for DSM-III-R Personality (SIDP-R). Iowa City, IA, Department of Psychiatry, University of Iowa

Philipp M, Maier W (1986) The Polydiagnostic Interview: a structured interview for the polydiagnostic classification of psychiatric patients. Psychopathology 19: 175–185

Polzer U, Juckel G, Gaebel W (1992) Emotional induzierte motorische Aktivität oro-fazialer mimischer Muskulatur bei schizophrenen Patienten: Erste Ergebnisse und ein neurobiochemischer Ausblick. In: Baumann P (ed) Biologische Psychiatrie der Gegenwart. 3. Drei-Länder-Symposium für Biologische Psychiatrie, Lausanne. Springer, Berlin Heidelberg New York, pp 89–91

Prusoff BA, Klerman GL, Paykel ES (1972a) Pitfalls in the self-report assessment of depression. Can Psychiatr Ass J 17: 101–107

Prusoff BA, Klerman GL, Paykel ES (1972b) Concordance between clinical assessment and patient selfreport in depression. Arch Gen Psychiatry 26: 546–552

Raven J (1996a) Advanced Progressive Matrices (APM), 2nd edn. Hogrefe, Göttingen

Raven J (1996b) Standard Progressive Matrices (SPM), 3rd edn. Hogrefe, Göttingen

Rebok G, Brandt J, Folstein M (1990) Longitudinal cognitive decline in patients with Alzheimer's disease. J Geriatr Psychiatry Neurol 3: 91–97

Reischies FM (1987) Neuropsychologisches Defizit-Screening. Nervenarzt 58: 219–226

Robins LN, Helzer JE, Ratcliff KS, Seyfried W (1982) Validity of the Diagnostic Interview Schedule. Version II: DSM-III diagnoses. Psychol Med 12: 855–870

Rodgers DA (1972) Minnesota Multiphasic Personality Inventory. In: Buros OK (ed) The seventh mental measurements yearbook. Highland Park, Gryphon, NJ, pp 243–250

Rosen WG, Mohs RC, Davis KL (1984) A new scale for rating Alzheimer's disease. Am J Psychiatry 141: 1356–1364

Roth M, Huppert FA, Tym E, Mountjoy CQ (1988) CAMDEX: The Cambridge Examination for Mental Disorders of the Elderly. Cambridge University Press, Cambridge

Russell EW (1994) The cognitive-metric, fixed battery approach to neuropsychological assessment. In: Vanderploeg RD (ed) Clinician's guide to neuropsychological assessment. Erlbaum, Hillsdalle, NJ, pp 211–258

Sarris V, Lienert GA (1974) Konstruktion und Bewährung von klinisch-psychologischen Testverfahren. In: Schraml WJ, Baumann U (eds) Klinische Psychologie. II. Methoden, Ergebnisse und Probleme der Forschung. Huber, Bern, pp 286–351

Sarris V, Rey ER (1981) Allgemeine Grundlagen von klinisch-psychologischen Testfaktoren. In: Rey ER (ed) Klinische Psychologie. Fischer, Stuttgart, pp 11–28

Sartorius N, Janca A (1996) Psychiatric assessment instruments developed by the World Health Organization. Soc Psychiatry Psychiatr Epidemiol 31: 55–69

Saß H (1986) Zur Klassifikation der Persönlichkeitsstörungen. Nervenarzt 57: 193–203

Satzger W, Engel RR (1996) Computerisierter Gedächtnis- und Aufmerksamkeitstest München (CGT-M), 2nd edn. Beltz, Weinheim

Satzger W, Dragon E, Engel RR (1996a) Zur Normenäquivalenz von HAWIE-R und HAWIE. Diagnostica 43: 119–138

Satzger W, Feßmann H, Dragon E, Engel RR (1996b) Ist IQ gleich IQ? Poster at the AMDP conference, 28–30 November 1996, Berlin

Saykin AJ, Gur RC, Gur RE, Mozley PD, Mozley LH, Resnick SM, Kesler DB, Stratiniak P (1991) Neuropsychological functioning in schizophrenia: selective impairment in memory and learning. Arch Gen Psychiatry 48: 618–624

Schallberger U (1987) HAWIK und HAWIK-R: Ein empirischer Vergleich. Diagnostica 33: 1–13

Scheller R (1973) Zur Brauchbarkeit des Hamburg-Wechsler-Intelligenztests für Erwachsene (HAWIE) als differentialdiagnostisches Instrument. Psychol Praxis 17: 68–80

Scheller R, Sittauer H (1974) Analytische Diskrimination dreier hirnorganischer Gruppen anhand von HAWIE-Daten. Psychol Praxis 18: 78–86

Schmidt KH, Metzler P (1992) Wortschatztest (WST). Beltz, Weinheim

Schneewind KA, Schröder G, Cattell RB (1994) Der 16-Persönlichkeits-Faktoren-Test (16-PF), 3rd edn. Huber, Bern

Scholz OB (1982) Interaktionsdiagnostik. In: Baumann U, Berbalk H, Seidenstücker G (eds) Klinische Psychologie. Trends in Forschung und Praxis, vol 5. Huber, Bern, pp 112–149

Schorr A (1995) Stand und Perspektiven diagnostischer Verfahren in der Praxis. Ergebnisse einer repräsentativen Befragung westdeutscher Psychologen. Diagnostica 41: 3–20

Schuri U, Keller I, Matthes-von Cramon G (1994) Leistungsdiagnostik aus neuropsychologischer Sicht. In: Stieglitz RD, Baumann U (eds) Psychodiagnostik psychischer Störungen. Enke, Stuttgart, pp 138–148

Seidenstücker G, Baumann U (1978) Multimethodale Diagnostik. In: Baumann U, Berbalk H, Seidenstücker G (eds) Klinische

Psychologie. Trends in Forschung und Praxis, vol 1. Huber, Bern, pp 134–183

Seidman LJ, Yurgelun Todd D, Kremen WS, Woods BT, Goldstein JM, Faraone SV, Tsuang MT (1994) Relationship of prefrontal and temporal lobe MRI measures to neuropsychological performance in chronic schizophrenia. Biol Psychiatry 35: 235–246

Seydel U (1972) HAWIE-Kurzformen und deren Kreuzvalidierung. Diagnostica 18: 121–135

*Shapiro MB (1966) The single case in clinical psychological research. J Gen Psychol 74: 3–23

Simpson GM, Angus JWS (1970) A rating scale for extrapyramidal side effects. Acta Psychiatr Scand Suppl 212: 11–19

Spitzer RL, Endicott J, Robins E (1975) Reliability of clinical criteria for psychiatric diagnosis. Am J Psychiatry 132: 1187–1192

*Spreen O, Strauss E (1991) A compendium of neuropsychological tests. Oxford University Press, New York

Stangl D, Pfohl B, Zimmerman M, Bowers W, Corenthal C (1985) A structured interview for the DSM-III personality disorders. Arch Gen Psychiatry 42: 591–596

Steinmeyer EM, Möller HJ (1992) Facet theoretic analysis of the Hamilton-D scale. J Affect Disord 25: 53–61

*Stieglitz RD, Baumann U (1994) Psychodiagnostik psychischer Störungen. Enke, Stuttgart

Strunk P, Faust VB (1967) Die Bewertung hirnorganischer Befunde bei Verhaltensstörungen im Kindesalter. Arch Psychiatr Z Neurol 210: 152–160

Sturm W, Hartje W, Kiteringham J (1975) Zur diagnostischen Brauchbarkeit einiger neuer Abbau-Indices aus dem HAWIE. Nervenarzt 46: 690–694

Velkoborsky J (1964) Der Benton-Test in der klinischen Praxis. Diagnostica 10: 91–101

*von Cramon DY, Mai N, Ziegler W (1993) Neuropsychologische Diagnostik. VCH Verlagsgesellschaft, Weinheim

von Cranach M, Frenz HG (1969) Systematische Beobachtung. In: Graumann CF (ed) Handbuch der Psychologie, vol 7. Hogrefe, Göttingen, pp 269–331

von Kerekjarto M (1961) Wahrnehmungstests zur Diagnose und Differentialdiagnose der multiplen Sklerose. Z Exp Angew Psychol 8: 369–380

von Zerssen D (1976a) Klinische Selbstbeurteilungs-Skalen (KSbS) aus dem Münchener Psychiatrischen Informationssystem (PSYCHIS München). a) Allgemeiner Teil. Beltz, Weinheim

von Zerssen D (1976b) Klinische Selbstbeurteilungs-Skalen (KSbS) aus dem Münchener Psychiatrischen Informationssystem (PSYCHIS München). b) Paranoid-Depressivitäts-Skala. Beltz, Weinheim

von Zerssen D (1976c) Klinische Selbstbeurteilungs-Skalen (KSbS) aus dem Münchener Psychiatrischen Informationssystem (PSYCHIS München). c) Befindlichkeits-Skala. Beltz, Weinheim

von Zerssen D (1976d) Klinische Selbstbeurteilungs-Skalen (KSbS) aus dem Münchener Psychiatrischen Informationssystem (PSYCHIS München). d) Beschwerden-Liste. Beltz, Weinheim

von Zerssen D (1977) Konstitutionstypologische Forschung. In: Strube G (ed) Die Psychologie des 20. Jahrhunderts, vol 5. Kindler, Zurich, pp 545–616

von Zerssen D (1979) Klinisch-psychiatrische Selbstbeurteilungs-Fragebögen. In: Baumann U, Berbalk H, Seidenstücker G (eds) Klinische Psychologie. Trends in Forschung und Praxis, vol 2. Huber, Bern, pp 130–159

von Zerssen D (1982) Personality and affective disorders. In: Paykel ES (ed) Handbook of affective disorders. Churchill Livingstone, Edinburgh, pp 212–228

von Zerssen D (1986) Clinical Self-Rating Scales (CSRS) of the Munich Psychiatric Information System (PSYCHIS München). In: Sartorius N, Ban TA (eds) Assessment of depression. Springer, Berlin Heidelberg New York, pp 270–303

von Zerssen D (1993) Normal and abnormal variants of premorbid personality in functional mental disorders. Conceptual and methodological issues. J Pers Disord 7: 116–136

von Zerssen D (1994) Diagnostik der prämorbiden Persönlichkeit. In: Stieglitz RD, Baumann U (eds) Psychodiagnostik psychischer Störungen. Enke, Stuttgart, pp 216–229

von Zerssen D, Cording C (1978) The measurement of change in endogenous affective disorders. Arch Psychiatr Nervenkr 226: 95–112

von Zerssen D, Möller HJ (1980) Psychopathometrische Verfahren in der psychiatrischen Therapieforschung. In: Biefang S (ed) Evaluationsforschung in der Psychiatrie: Fragestellungen und Methoden. Enke, Stuttgart, pp 129–166

von Zerssen D, Pfister H, Koeller DM (1988) The Munich Personality Test (MPT) – a short questionnaire for self-rating and relatives rating of personality traits: formal properties and clinical potential. Eur Arch Psychiatry Neurol Sci 238: 73–93

Wagner H, Clarke AH, Ellgring JH (1983) Eye-contact and individual looking: the role of chance. Br J Psychol 22: 61–62

Wechsler D (1956) Die Messung der Intelligenz Erwachsener. Huber, Bern

Wechsler D (1981) Manual for the Wechsler Adult Intelligence Scale – Revised. Psychological Corporation, New York

Wechsler D (1987) Wechsler Memory Scale – Revised. Psychological Corporation, New York

Weintraub W, Aronson H (1967) The application of verbal behavior analysis to the study of psychological defense mechanism. IV. Speech pattern associated with depressive behavior. J Nerv Ment Dis 144: 22–28

Weise G (1975) Psychologische Leistungstests, vol I. Hogrefe, Göttingen

Weissman MM, Bothwell S (1976) Assessment of social adjustment by patient selfreport. Arch Gen Psychiatry 33: 1111–1115

Weissman MM, Sholomskas D, John K (1981) The assessment of social adjustment: an update. Arch Gen Psychiatry 38: 1250–1258

White J, White K, Razani J (1984) Effects of endogenicity and severity on consistancy of standard depression rating scales. J Clin Psychiatry 45: 260–261

*WHO (1991) Schedule for the clinical assessment in neuropsychiatry. WHO, Geneva

*Wiggins JS (1973) Personality and prediction: principles of personality assessment. Addison-Wesley, Reading, MA

Wing JK, Cooper JE, Sartorius N (1974) Measurement and classification of psychiatric symptoms. Cambridge University Press, Cambridge

Wing JK, Cooper JE, Sartorins N (1978) Instruction manual for the Present State Exarnination and CATEGO. Institute of Psychiatry, London

Winkler B, Ellgring H (1981) Codierung und Analyse der Sprachinhalte klinischer Interviews. In: Lange-Seidl A (ed) Zeichenkonstitution (Akten des 2. Semiotischen Kolloquiums, Regensburg 1978, pp 140–141)

Wittchen HU, Semler G (1991) Composite International Diagnostic Interview (CIDI, version 1.0). Beltz, Weinheim

Wittchen HU, Semler G, von Zerssen D (1985) A comparison of two diagnostic methods. Clinical ICD diagnoses vs DSM-III and research diagnostic criteria using the Diagnostic Interview Schedule (version 2). Arch Gen Psychiatry 42: 677–684

Wittchen HU, Zaudig M, Spengler P et al (1991) Wie zuverlässig ist operationalisierte Diagnostik? – Die Test-Retest Reliabilität des Strukturierten Interviews für DSM-III-R. Z Klin Psychol 20(2): 136–153

*Woods RT (1994) Problems in the elderly: investigation. In: Lindsay SJ, Powell GE (eds) Handbook of clinical adult psychology, 2nd edn. Routledge, London, pp 413–437

Wykes T, Katz R, Sturt E, Hemsley DR (1992) Abnormalities of response processing in a chronic psychiatric group: a possible predictor of failure on rehabilitation programmes. Br J Psychiatry 160: 244–252

Zaudig M, Mittelhammer J, Hiller W (1990) SIDAM – Strukturiertes Interview für die Diagnose der Demenz vom Alzheimer-Typ, der Multiinfarkt-Demenz und Demenzen anderer Ätiologie nach DSM-III-R und ICD-10. Logomed, Munich

Zaworka W, Hand I, Jauernig G, Lünenschloss K (1983) Hamburger Zwangsinventar (HZI). Beltz, Weinheim

Zung WWK (1965) A selfrating depression scale. Arch Gen Psychiatry 12: 63–67

Zung WWK (1976a) ASI. Anxiety Status Inventory. In: Guy W (ed) ECDEU assessment manual for psychopharmacology. National Institute of Mental Health, Rockville, pp 199–204

Zung WWK (1976b) SAS. Self-rating Anxiety Scale. In: Guy W (ed) ECDEU assessment manual for psychopharmacology. National Institute of Mental Health, Rockville, pp 337–340

CHAPTER

7

M. Bullinger-Naber, D. Naber

Assessing the Quality of Life in Mental Illness

Translator: K. Schlag

1 Health-Related Quality of Life as a Subject of Psychiatry

Whenever in the past few years goals and standards for the assessment of medical management have been debated, health-related quality of life has increasingly been an issue (Najman and Levine 1981; Spilker 1996). This development originated in the question of whether the classical medical treatment outcome criteria (symptoms, survival time) provide a sufficient basis for treatment. With the definition of health as suggested by the World Health Organization (WHO) in 1947 (i.e. physical, mental and social well-being) in mind and based on a humanistic view of patients as individuals, the concept of health-related quality of life reflects our modern-day philosophy (Schölmerich and Thews 1992).

Health-related quality of life is defined by international agreement as well-being and functioning – physically, psychologically (i.e. emotionally and mentally) and socially – as reported by the patient (Bullinger 1991). The WHO working group on quality of life defined this term as an individual's subjective perception of his or her own role in life in relationship to the culture and value systems he or she lives in and taking into account the goals, expectations, standards and concerns of a person. This is a working concept that is influenced in a complex manner by physical health, the psychological state, the extent of independence, social relationships and marked characteristics of a person's surroundings (WHOQOL Group 1994). It comprises many areas – physical, psychological, functional, social, moral and spiritual – as well as minor aspects of these.

The concept of quality of life is not new in that it sets the standard for a physician's decisions, but in that it is an attempt to assess the subjective well-being and functioning of a patient, which could thus become accessible for scientific research. Consideration needs to be given to the fact that health-related quality of life cannot be measured directly and is a theoretical psychological concept that can only be assessed if the relevant aspects and components are operationalised, keeping in mind the multi-dimensionality and subjective quality of the concept (Bullinger 1996).

The assessment of quality of life can be considered as established in medicine, pioneers having been those working in the field of oncology, surgery and cardiovascular medicine (see Spilker 1996), whereas the subject of quality of life only sporadically began to be discussed in psychiatry in the 1980s and only more extensively in the past few years (Lehmann et al. 1982; Katschnig and König 1994; Möller et al. 1996).

The reasons for this relative inactivity in psychiatry in comparison with other medical disciplines are manifold (Hogan and Awad 1992; Helmchen 1990; Lehmann 1996).

The main reason seems to lie in the belief that psychiatry has always essentially and primarily been concerned with the well-being and functioning of the patient, as these areas represent symptoms of psychiatric illness. This can be countered by arguing that an assessment of the symptoms of a psychiatric disorder cannot be compared with quality of life as experienced by the patient.

A second reason can be found in the assumption that the patients' self-report will be distorted – precisely because of their psychiatric illness – and therefore will not be reliable. This applies especially for psychotic episodes in the course of schizophrenic diseases, but also for affective disorders such as depression. In these cases, symptoms and quality of life are almost identical. It remains questionable whether patients are always able to describe and assess their well-being and functioning adequately and whether a psychiatric patient's judgement of his or her own personal experience and behaviour differs in principal from that of patients with other disorders.

A third reason has to do with the sceptical attitude many have towards the validity of such a complex concept as the quality of life. Not only has the lack of a theoretical backing for quality of life of patients in general and psychiatric patients in particular been criticised, but doubts have also been expressed concerning the question of whether an interpersonal, not individually centered assessment of quality of life with standardised instruments is even possible, meaningful or desirable. This brings up the issue of the theoretical foundation and measuring models that the quality of life assessment instruments are based on, especially with respect to their applicability to psychiatric patients.

A fourth reason is to be found in ethical objections to quality of life research. The concept of health-related quality of life implies certain values and is connected with implicit standards that refer to experience and behaviour. If the term "quality of life" is used in order to postulate an ideal standard and to assess patients according to their achievement of this standard, a discussion of "worthwhile living" – including the dark German history in this matter – becomes inevitable. In this context, the issue arises of the extent to which the concept of "health-related quality of life" brings us a step forward (towards more say and an optimisation of treatment for patients) or a step backward (towards establishing outlived role concepts and classifications of psychiatric patients).

2 Concepts of Quality of Life

The concept of quality of life has played a role in sociological literature ever since the 1940s and 1950s. In the 1960s, in particular, large and even intercultural studies were conducted to explore the quality of life in different countries, "quality of life" being defined primarily in the context of socio-economic resources and people's health standard in a country and assessed by various means, e.g. in terms of the gross national product and the rate of newborn mortality. Campbell (1981) took a more individual approach to quality of life research in America in his study on the *Quality of American Life*, in which questions were answered concerning the individuals' satisfaction with life. Likewise, Glatzer and Zapf (1984) examined the quality of life in the Federal Republic of Germany. An important contribution to the discussion on quality of life has also been made in psychology, using the term "well-being" (Abele and Becker 1991).

The concept of quality of life was introduced fairly late in medicine. According to several authors, it was introduced by an early article in 1967, while others believe it entered the field of medicine when more began to be written on the subject (Spilker 1996). It did not take very long for "health-related quality of life" to become a recognised expression, in contrast to the term "quality of life", as defined in sociology, because in medicine aspects of human experience and behaviour need to be regarded with particular reference to health (Patrick and Erikson 1992). North America especially, but also England, plays a leading role in research in the area of quality of life, referred to in these countries as "health outcome research" (Stewart and Ware 1992).

In Germany, the concept of quality of life was welcomed somewhat reluctantly, and systematic and broader attempts at research in this field have been noted since the end of the 1980s.

In contrast to the literature on happiness, well-being and satisfaction, which is at least partially empirically based, it is very surprising how little theoretical work has been done on the concept of quality of life so far in medicine (and, for that matter, in sociology). An early compilation of the aspects of quality of life (Calman 1987) includes different approaches, none of which can be called theories in a scientific sense, because they tend to represent basic definitions and simple models.

Basically, three types of quality of life models can be differentiated. The first focuses on the individual and postulates that quality of life can never be assessed independently of the person in question, as it differs by definition from one person to another in its dimensions. The advocates of this approach presuppose that quality of life can only be described intra-individually.

A second category of definitions is based on the assumption that quality of life can be assessed for a limited number of aspects that are relevant to different individuals. Various authors agree that these aspects seem to be the health aspects defined by WHO, namely physical, mental and social well-being. The corresponding approaches to measurement attempt to make this dimensionality accessible to assessment and to characterise individuals according to their fulfillment of a certain aspect or patterns of different aspects.

A third group of definitions postulates that quality of life cannot be measured as a relevant aspect either intra-individually or inter-individually, but that quality of life is an implicit concept. In this case, it is presupposed that quality of life cannot be ascertained by direct questioning but can only be implicitly assessed depending on the priorities of the patient. This is the so-called health-economic or cost-utility approach to measuring quality of life; the individuals concerned assess their health outcome with the help of certain state of health scenarios or approaches taken from game theory.

Each of these roughly outlined conceptual approaches has led to the emergence of a specific methodology that has manifested itself in the development of measuring tools. As far as the basic question of the definability of quality of life is concerned, there is agreement that a nominal definition would not make sense, that an operational definition exists and that a truly scientific theoretical basis is not feasible at the present time (Bullinger 1991). The operational definition is preferred by researchers who postulate inter-individual universality of quality of life, referring mainly to the above-mentioned aspects of physical, mental, social and functional health from the perspective of the individuals concerned. The assumption is that, on a high level of abstraction, people differ less than expected in their appreciation of quality of life.

There are approaches in the literature on social psychology concerning the theoretical discussion of a general quality of life concept that show that quality of life assessments depend on complex "social comparisons" (Glatzer and Zapf 1984), on the individual relationship between expectation and achievement ("goal attainment"; Brunstein 1993) and on processes of adjustment to external circumstances (adjustment level theory).

Angermeyer (1994) employs somewhat different models in psychiatric research for the quality of life assessment of mentally ill individuals. The basis of these is the assumption that the concept of quality of life far exceeds disease-related aspects such as symptoms, impairments and disabilities and comprises the

patient's very subjective experience of his or her objective life situation. The models include the satisfaction model, the necessities model and the role function model. The *satisfaction model* postulates that a patient's quality of life results from his or her objective circumstances in different areas of life and his or her satisfaction with these circumstances (Lehmann 1983). The component "significance" needs to be included in the satisfaction model because people differ in their individual values and preferences when confronted with different objective circumstances of life (Becker et al. 1993).

In contrast to the satisfaction and significance models, the *necessities model* is explicitly based on the hierarchy of human motives as described by Maslow (1970). Bigelow et al. (1982) also express the opinion that happiness and satisfaction are based on social and environmental conditions that form a basis for the fulfillment of needs. Katschnig (1994) critically points out, however, that a multi-dimensional rather than hierarchical view of human needs has an important function and that the quality of life of psychiatric patients can be seen as a reduction of disabling factors. This approach is similar to the role function models that stress the fulfillment of socially expected roles, as is also true for the applications of the quality of life model, e.g. of rehabilitation research.

Bech (1996) postulates that patients' health-related quality of life can be compared with the subjective assessment of their own health profile, which he describes as subjective well-being. He suggests an approach which differentiates between varying levels of assessment and, in the assessment of quality of life, between an incidence-related (i.e. objective) and an assessment-related (i.e. subjective) approach. In so doing, he differentiates aspects of quality of life according to physical (P), cognitive (C), affective (A), social (S), economic (E) and ego strength-related (E) aspects. The resulting PCASEE model is both a theoretical model and a way to assess quality of life, with incidence relating to the severity of the symptom and assessment relating to the evaluation of a symptom's effect on the patient's well-being.

The question of whether quality of life depends more on objective circumstances of life or more on subjective perceptions of well-being and functioning has been much more controversial in psychiatry than in other medical disciplines. Objective circumstances of life include living situation, work and recreational possibilities, the availability of which is seen as an indicator of a high quality of life. With this approach, however, several problems come up: It is not clear (a) whether the objective circumstances are an inherent part or external determinants of quality of life, (b) whether how favourable the objective circumstances are, as assessed by the external observer, coincides with the subjective experience of the patient and (c) whether quality of life assessment should not rather focus on the subjective extent of well-being and functioning after all.

In a survey of the approaches used to date towards an assessment of quality of life in psychiatric patients, Kilian (1995) notes that there is little to no correlation between objective circumstances of life and subjective satisfaction and that treatment has only a short-lived effect on a person's subjective satisfaction with life, which returns to its original level in the course of treatment.

3 Methods for Assessing Quality of Life

The focus on quality of life in medicine has not only led to a considerably higher number of relevant publications, but also to the development of a multitude of measuring tools for quality of life assessment. In currently more than 20,000 publications, 800 instruments for the assessment of quality of life are to be found. These are usually questionnaires, but also include interviews, mainly ascertaining self-reporting by the patient, but also ratings by other people (e.g. family, nursing staff). The measuring tools available to date (see Table 1) can be roughly divided into those that assess the non-disease-specific quality of life of the respondents, i.e. that can be applied to healthy or sick individuals independently of their current health condition, and those that assess quality of life specifically for a certain illness, so-called disease-specific measuring tools. In this respect, a difference can again be made in psychiatry between methods that are applicable to chronically mentally ill patients in general and those that were developed specifically for certain disease entities (e.g. schizophrenia, depression, anxiety disorder).

3.1 Generic Methods

The generic methods are particularly well established in the American-speaking countries and were developed in the course of epidemiological, public health and health systems research (McDowell and Newell 1987; Spilker 1996). Among these are instruments such as the Nottingham Health Profile (NHP; Hunt et al. 1981), the Sickness Impact Profile (SIP; Bergner et al. 1981), the SF-36 Health Survey (Ware and Sherbourne 1992) and the General Health Rating Index (GHQ; Goldberg and Hillier 1979). The SCL-90 is also increasingly being used as an assessment instrument

Table 1. Scales for the assessment of quality of life of psychiatric patients

Name	Abbreviation	Authors	Year	Type	Duration (min)	Items (*n*)	Dimensions (*n*)	Reliability	Validity	Sensitivity	Objective	Original population
Standardized Social Schedule	SSS	Claire and Cairns	1978	INT: SR, IR	45	48	6	IR = 0.76	NR	NR	Description	221 chronic neuroses, 109 women PMS, 48 healthy subjects
Community Adjust New Form	CAF	Stein and Test	1980	SEMI-INT	45	140	12	NR	NR	NR	Evaluation	130 patients in State Hospital, 55% with schizophrenia
Quality of Life Checklist	QLC	Malm et al.	1981	INT, IR	60	93	13 (single items)	NR	NR	NR	Prognosis, description	40 out-patients
Satisfaction with Life Domains Scale	SLDS	Baker and Intagliata	1982	QUES, SR	10	15	1	NR	Construct	NR	Evaluation	118 chronically mentally ill patients, 56% with schizophrenia
Oregon Quality of Life Questionnaire	OQLQ	Bigelow et al.	1982	INT SR SEMI, IR	45 30	263 146	14	RIT = 0.84 (0.05–0.98), RTT = 0.5 (0.37–0.64)	Predictive	NR	Evaluation	Various patient groups and normal population
Quality of Life Interview	QOLI	Lehmann et al.	1982		45	143 (subjective and objective)	8	Subjective RIT = 0.85, objective RIT = 0.86	Construct, predictive	NR	Description	Various patient groups and normal population
Quality of Life Scales	QLS	Heinrichs et al.	1984	SEMI-INT, IR	45	21	4	IR = 0.84–0.97	Confirmatory factor analysis	TP	Description	Various clinical studies
Client Quality of Life Interview	CQLI	Mulkern et al.	1986	INT, SR, IR	45	46	NR	NR	NR	NR	Description	109 severely mentally ill patients
California Well-Being Project Client Interview	CWBPCI	Campbell et al.	1989	INT: Patient Family Doctors	60 30 20	151 76 77	NR	NR	NR	NR	Description	331 mentally ill patients

Table 1 (Continued)

Name	Abbreviation	Authors	Year	Type	Duration (min)	Items (n)	Dimensions (*n*)	Reliability	Validity	Sensitivity	Objective	Original population
Lancaster Quality of Life Profile	LQOLP	Oliver	1992	INT, SR, QUES	60	100	9	RTT = 0.49–0.78, RIT = 0.84–0.86	NR	NR	Evaluation	Various mentally ill patients
Wisconsin Quality of Life Index for Mental Health	QLI-MI	Becker et al.	1992	QUES, SR	60	103	9	RTT = 0.82–0.87	Intercorrelation	NR	Evaluation	40 chronically ill patients
Semistructured Interview Questionnaire	QOLIS	Holcomb et al.	1993	SEMI-INT, IR	30	87	8	RIT = 0.72–0.93	Factor analysis	NR	Evaluation	201 chronically ill patients
Quality of Life Enjoyment and Satisfaction Questionnaire	Q-LES-Q	Endicott et al.	1993	QUES, SR	50	93	8	RTT = 0.63–0.89	Convergent CGI, BDI	NR	Evaluation	83 patients with depression
Drug Attitude Inventory	DAI	Hogan and Awad	1992	QUES, SR	20	30	1	RIT = 0.89	Construct	TP	Evaluation	Clinical studies on neuroleptics
Smith-Kline-Beecham Quality of Life Scale	SBQOL	Stoker et al.	1992	QUES, SR	7	23	3	RTT = 0.66–0.83	Construct, SIP, GHQ	TP	Evaluation	129 patients with depression
Quality of Life in Depression	QLDS	Hunt et al.	1992	QUES, SR	20	24	1	RIT = 0.93	Convergent HAD	TP	Evaluation	196 patients with depression
Subjective Well-being Neuroleptics	SWN	Naber	1995	QUES, SR	15–20	38	5 scales and total score	RIT = 0.95 (0.73–0.88), RTT = 0.75–0.88	Discrimination	TP	Evaluation	280 patients with schizophrenia

INT, interview; SR, self-report; IR, interview rating; SEMI, semi-structured; QUES, questionnaire; RIT, internal consistency (Cronbach's alpha); RTT, test–retest reliability; IR, inter-rater reliability; SIP, Sickness Impact Profile; GHQ, General Health Rating Index; PMS, premenstrual tension syndrome; CGI, Global Clinical Impression Score; HAD, Hospital Anxiety and Depression Scale; BDI, Beck Depression Inventory; NR, not reported; TP, tested, present.

for quality of life (Derogatis et al. 1973). These methods are also available in German and can be employed independently of the patient's current health condition and refer to a self-report of mental, social and physical well-being as well as of behaviour in everyday life. Moreover, methods have recently been developed in German-speaking countries that explore satisfaction with and significance of certain areas of life (e.g. Huber et al. 1988), such as the Munich life quality dimension list (MLDL; Heinisch et al. 1991; Franz et al. 1996), or that assess the instrumental component of quality of life (Questionnaire on Everyday Life; Bullinger et al. 1993).

Newly developed and currently being tested in psychiatry is the quality of life questionnaire brought out by WHO (WHOQOL Group 1994). This research is particularly interesting as, in the course of developing the questionnaire, representatives of different cultures were asked to identify the essential aspects of quality of life (Sartorius 1995). The resulting discussion brought forth a matrix of relevant quality of life aspects that – by relating back to the corresponding individual culture – could then be filled with content on location by means of focus groups, i.e. the individual cultures were asked to formulate different items for the assessment of these aspects. Fifteen nations participated in the study, and the resulting item pool of over 3000 questions was translated into English and tested by an expert group for redundancy. At the same time, the aspects of quality of life were weighted independently of the different countries. The resulting matrix of over 300 questions, which referred to both subjective perception and the current condition concerning certain aspects of quality of life, was tested psychometrically in a further study on 300 people in 15 countries (on a total of 4500 individuals). The results of the psychometric testing, inter alia using structural analysis techniques, lead to the conclusion that culture-independent essential aspects exist in psychological, physical, social, spiritual, functional and economic areas, although minor differences were found in the significance of these aspects for the individual cultures.

3.2 Disease-Specific Methods

In the past few years, especially in Anglo-American countries, a series of methods have been developed that assess the disease-specific quality of life of the mentally ill.

Among these is the Standardized Social Schedule (SSS; Clare and Cairns 1978). It is performed as an interview, includes 48 questions and takes approximately 45 min. It has been tested psychometrically and is also available in German; it covers six life domains: living situation, occupation, social role, economic situation, leisure and social activities and family.

A semi-structured interview is found in the Community Adjustment Form (CAF; Stein and Test 1980). This interview assesses quality of life using 140 questions referring to 12 aspects with regard to leisure, circumstances, occupational background, etc. The Quality of Life Checklist (QLC; Malm et al. 1981) is a rating scale covering 93 items on 13 areas of life that can be performed in the context of a semi-structured interview. The Satisfaction with Life Domain Scale (SLDS; Baker and Intagliata 1982) is a self-report assessed by questionnaire exploring 15 items on satisfaction with different life domains. The Oregon Quality of Life Questionnaire (OQLQ; Bigelow et al. 1982, 1990) exists as a structured interview that assesses the patient's self-report (263 items) and as a semi-structured interview that is assessed by the interviewer (146 items). The OQLQ results in 14 scale scores and takes approximately 45 min. The psychometric characteristics of the OQLQ have been evaluated for varying patient populations. This is also true for the Quality of Life Interview (QOLI; Lehmann 1988; Lehmann et al. 1986, 1993), which consists of a structured interview version for the patient's self-report. It has 143 items and takes 45 min to perform, comprising eight aspects of life (life situation, daily activities and functions, family relationships, social relations and finances, work and school, legal and security aspects and health). An objective description of circumstances of life and a subjective assessment of satisfaction with a specific life domain are made. The Quality of Life Scale (QLS; Heinrichs et al. 1984) is a semi-structured interview that has not yet been tested psychometrically; it is filled out by a trained clinician. It includes 21 items that reflect the interviewer's assessment of the functioning of the patient in each of the 21 domains and takes 45 min to perform. The 21 items can be reduced to four scales, namely basic intrapsychic elements, interpersonal relationships, instrumental role functions and a total value. The Client Quality of Life Interview (CQLI; Mulkern et al. 1986) is a structured interview of a person's self-report and is carried out by a trained interviewer. It consists of 46 questions that assess the patients and 19 interviewer ratings with ordinal scales. For each life domain (e.g. basic necessities of life, professional training, daily activities, privacy), both the activities and the subjective feelings of the patient are explored. The California Well-Being Project Client Interview (CWBPCI; Campbell et al. 1989) was developed for patients (151 questions), for family members (76 questions) and for doctors and nurses (77 items). The questions are asked in the form of an interview, although completion and submission by post by the patient or group administration by questionnaire is

also possible. The Lancaster Quality of Life Profile (LQOLP; Oliver 1992), based on the QOLI by Lehmann, is a structured interview of the patients' self-report carried out by trained nursing staff and takes 60 min. The 100 items assess both objective and subjective aspects of the quality of life in several domains, including work, education, leisure, participation in social activities, religion, finances, life situation, legal and security aspects, family relationships, social relations and health. Among the psychometric characteristics are reliability quotients of $\alpha = 0.49–0.78$. The Wisconsin Quality of Life Index for Mental Health (QLI-MI; Becker et al. 1992) assesses nine aspects with 103 items that are evaluated according to their significance in a patient self-report; parts of the report can also be filled out by the doctor or family. Tests that have been done so far on this method (which is still under development) on 40 chronically mentally ill patients show a good test–retest reliability.

The semi-structured Quality of Life Schedule (QOLIS; Holcomb et al. 1993) consists of 87 items that explore eight aspects of quality of life in the form of an interview by expert assessment. So far, the method has been used on 201 chronically mentally ill patients and been successfully tested psychometrically.

A new tool is the Quality of Life Enjoyment and Satisfaction Questionnaire (Q-LES-Q; Endicott et al. 1993), consisting of 93 items that are transformed into eight sum scales, including physical health, subjective feelings, recreational activities, social relationships and general activities.

3.3 Comparison of Different Procedures

The older literature is characterised by lengthy procedures that are carried out in the form of an interview and often with an assessment by the interviewer. However, newer publications tend to concentrate on shorter scales that can also be obtained as self-reports from the patients in the form of questionnaires. Some procedures aim at contrasting objective circumstances of life and subjective assessments, whereas others focus more on direct self-reports by the patients concerning their experience and behaviour. An attempt to streamline procedures and a growing reliance on self-reporting by patients are evident in this development.

Some publications compare the characteristics of several different tools, e.g. the Quality of Life Interview by Lehmann et al. (1993) and the Quality of Life Scale by Heinrichs et al. (1984). For 59 patients who were assessed using both measuring tools with a 2-month interval together with SCL-90 and the BPRS (Brief Psychiatric Rating Scale), this comparison showed that the correlations between the two measuring tools in corresponding contexts were significant and the test–retest correlations of both instruments were good.

In other publications, the correlation between assessment by the patient and assessment by caregivers was investigated with reference to the quality of life of these patients. A total of 31 chronic psychiatric patients were evaluated by a measuring tool constructed ad hoc that measures nine aspects of quality of life and that can be assessed by patients and nursing staff alike. It became evident that patients can assess their quality of life consistently and that they assessed it significantly higher than the people caring for them (Thapa and Rowland 1989). A more recent publication compared assessments made by 37 schizophrenic patients and their doctors, e.g. with the help of the Wisconsin Quality of Life Index for Mental Health (Sainfort et al. 1996).

The results in this case were that the assessments made by the patients and by those that took care of them tended to correlate positively, and much more so for clinical aspects such as symptoms and functional impairments than for social or professional aspects. Accordingly, the percentage of correspondences was lowest for social support (44%) and highest for functioning (66%).

In contrast to the use of specific tools for the assessment of patients' quality of life, the application of well-standardised procedures from generic quality of life research has become established. Different subscales from existing instruments are often employed together. One example of this kind of approach is a study by Revicki et al. (1992), who employed clinical instruments and subscales from the SF-36 Health Survey and the SIP on 40 patients with clinical depression. It took the patients 10 min to fill out the whole questionnaire, and reproducibility was very good, with an intra-class coefficient of $r = .77$. The correlation between measuring tools for quality of life and the clinical symptom scales (Montgomery/Asberg scale) was substantial ($r = 0.3–0.62$); symptom changes were associated with changes in the quality of life scales. Reliability and validity of the newly created instrument were also acceptable.

In addition to the variety of scales available for the assessment of psychiatric patients' quality of life, there are other areas that should be mentioned briefly in this context. One area refers to quality of life assessment of caregivers, both family members and medical personnel (Schene et al. 1996), while another refers to the assessment of patients' satisfaction with treatment (Kelstrup et al. 1993; Leimkühler and Müller 1996).

In contrast to the internationally (methodically) well-researched generic measuring tools (such as SF-36

and NHP), only few of the disease-specific measuring tools that are available in English have been tested in their original language for psychometric characteristics such as reliability, validity and sensitivity. Exceptions are the WHO Quality of Life Scale and the Berlin Life Quality Inventory. Moreover, for the tools that were originally in English, neither an adequate translation nor any psychometric testing according to current internationally accepted criteria and procedures are available (Sartorius and Kuyken 1994).

4
Application of Quality of Life Scales in Psychiatry

In publications, life quality scales have been used to describe the quality of life of patient populations, i.e. the determinants by which to judge quality of life and to evaluate the effect of medical care strategies in the out-patient and rehabilitation sectors. There are a small number of studies in this respect that test the effects of certain kinds of care (supervised living situations, hospitalisation) on the patients' quality of life. In contrast to this, the employment of quality of life scales in clinical psychiatric studies can be described as rudimentary. Barely any regard has been given to the question of the extent to which quality of life indicators can be helpful in analyses of health economics. At the present time, only theoretical considerations about the instruments applied in this respect have been addressed in survey articles (Gudex 1996).

4.1
Characteristics and Determinants of the Quality of Life of Psychiatric Patients

In cross-sectional studies, tools have been employed for quality of life assessment in order to characterise patient groups in terms of quality of life compared to different reference populations. For psychiatric patients compared with physically ill patients and healthy controls, quality of life indicators pointed to a marked impairment not only in mood but also in role function. Correspondingly, by means of a precursor of the SF-36 Health Survey, significantly lower scale values were found for psychiatric patients than for age- and sex-matched normal populations not only in the subscales of mental well-being, but also for vitality, emotional and physical role functioning and social functioning (Spitzer et al. 1995). Satisfaction with life assessments gave similar results (Franz et al. 1996; Huber et al. 1988; Lehmann 1996).

The socio-demographic findings on quality of life are heterogeneous. In one literature survey on sex-related differences in the quality of life of schizophrenic patients, it was found that schizophrenic women report higher satisfaction ratings than men (Röder-Wanner 1995; Priebe et al. 1996). In another study, the effects of sex and age on the assessment of quality of life were investigated (Lehmann et al. 1992). The result of this was that psychiatric symptoms correlated with objective quality of life indicators markedly less than with subjective satisfaction with life indicators. In mid-life, however, chronically mentally ill men differ from women in that men are significantly more satisfied with their life than women, especially with regard to leisure. Furthermore, in 110 patients with severe psychiatric disorders, significant differences were found, with working patients having a higher quality of life (Fabian 1992).

The influence of clinical symptoms on quality of life was investigated in a more recent study on 49 chronic psychiatric patients (Corrigan and Buican 1995). The patients were examined with the Quality of Life Interview developed by Lehmann and also filled out questionnaires on social functions and support; in addition, observer-rated psychopathology was assessed by means of the BPRS. It was found that a higher quality of life index correlated positively with more social competence and intelligence and negatively with the depression score in the BPRS. In a multiple regression, it was found that depression, social network, intelligence and interpersonal contact were predictors for quality of life with a declared variance of 55%, which points to the contribution of these factors to the concept of quality of life.

The role of social support was also pointed out in a study on 729 severely ill psychiatric patients who had a significantly positive correlation between quality of life – as measured with the Satisfaction with Life Domain Scale – and social support (Baker et al. 1992).

Furthermore, in a population of 101 psychiatric patients, it was found that a high degree of self-reported quality of life was connected with fewer depressive symptoms, fewer side effects of medication and better family interaction (Sullivan et al. 1992).

In a study of 500 chronically ill patients, Mechanic et al. (1992) observed that disease attribution had a major effect on quality of life. Patients who attribute their symptoms to a physical, biological or medical cause – as opposed to those who describe themselves as insane "for no reason" – report more positive social relations and a higher quality of life. The negative effect of symptoms being attributed to "insanity" on quality of life was explained by the authors to be due to the perceived stigma, low self-esteem and a higher degree of depressive symptoms.

4.2 Schizophrenia

Quality of life publications on schizophrenic patients primarily focus on how neuroleptic treatment is perceived by patients. The Instrument for Assessment of Subjective Well-Being Under Neuroleptic Treatment (SWN; Naber 1995) measures five aspects of schizophrenic patients' well-being by means of 38 questions regarding the subjective effect of neuroleptics (15–20 min duration for completion) and has been successfully tested for its psychometric characteristics on 280 patients. In another study on 216 schizophrenic patients in remission, patients who later showed compliance had already differed from non-compliant patients at the time of discharge from hospital. This points to the fact that subjective well-being during neuroleptic treatment should be an important criterion in assessing the outcome of neuroleptic treatment (Naber et al. 1994). Moreover, it was shown that subjective well-being during treatment with neuroleptics (as measured with the SWN) correlated quite highly ($r = 0.6$) with quality of life in general (as measured with the MLDL), whereas both aspects correlated only poorly ($r = 0.3$–0.4) with objective psychopathology (as measured with PANSS).

In using the Q-Sort Procedure, in which patients rank their agreement with different statements in a certain order, a more recent study was able to distinguish four groups of patients according to how they had answered the questions on their medication; these groups were characterised as (a) "barely questioning", (b) "autonomously sceptical", (c) "balanced agreement" or (d) "autonomously positive". The groups differed in how they judged the effect of neuroleptic treatment on their quality of life, the first group reporting lower values for quality of life (Day et al. 1996).

In a group of 53 chronic schizophrenic patients undergoing treatment with depot neuroleptics, Larsen and Gerlach (1996) found that 60% of patients generally regarded their medication as positive, but that – despite the overall positive assessment – the majority felt bothered by the side effects. Abnormal movements were perceived by the patients to be the least and by the doctors to be the most bothersome, and the patients felt most impaired by the psychological side effects, which were found to be least impairing by the physicians. No correlation was found between the clinical severity of the illness and the severity as judged by the patients themselves.

Awad (1992) also points out that, in order to understand the effects of medication on the quality of life of schizophrenic patients, not only knowledge of the side effects but also of the subjective perception of the effects of neuroleptics is needed. So far, quality of life scales have not been widely employed in clinical studies (Selai and Trimble 1994).

4.3 Depression

In contrast to the state of affairs in the research of schizophrenia, patient self-reports have been included for a while now as a criterion in the assessment of treatment results in depression research. The instruments used include the Hospital Depression and Anxiety Scale, the Hopkins Symptom Checklist and the Psychological General Well-Being Index. The SF-36 Health Survey has been tested and validated in depressed patients (Wells et al. 1989; Bech 1995). The Smith-Klein and Beecham Quality-of-Life Scale (SBQOL; Stocker et al. 1992) is particularly suited to the assessment of depression; using 23 items in three domains, it measures the relative discrepancy between current and self-ideal assessments in different areas of life. The method was employed in a population of 129 depressive patients, and validity and test–retest reliability were found to be adequate. A further test developed for depressive patients is the Quality of Life in Depression Scale (QLDS; Hunt et al. 1981), which comprises 29 items measuring only one domain (albeit reliably); this test was found to correlate with the Hospital Depression Scale in a study with 196 patients (Grégoire et al. 1994).

So far, there have been studies on the effect of serotonin re-uptake inhibitors and monoamine oxidase inhibitors with regards to quality of life. In one study, the (non-disease-specific) testing of fluoxetine versus moclobemide in 209 depressive patients by means of a precursor of the SF-36 Health Survey and a dimensional measurement to assess quality of life showed a positive effect of both forms of treatment on quality of life, although there was no significant difference between the two treatment strategies (Lonnqvist et al. 1994). Using the SF-36 Health Survey and the General Health Questionnaire for assessment purposes, 651 depressive patients showed a significant improvement in quality of life during as opposed to prior to treatment with moclobemide (Walker et al. 1995).

A one-dimensional "thermometer scale" on quality of life compared 100 monopolarly and bipolarly depressed patients with two control groups comprising 50 healthy patients and 50 patients with personality disorders: in the patients with affective disorders, no significant differences were found between those treated with lithium and a

control group in respect to quality of life (Lepkifer et al. 1988).

4.4 Anxiety Disorders

Quality of life of patients with panic disorders was assessed in a large-scale epidemiological study using an earlier version of the SF-36 Health Survey from the Medical Outcome Study. A total of 18,000 adults were interviewed in the course of 18 years, 5034 of which were chosen from a specific area and 250 of which had panic disorders. It was shown that panic disorders coincide with clearly evident social and physical impairments, comparable to those of clinical depressions. These include subjective feelings of poor physical and mental health, but also alcohol and drug abuse, a higher probability of suicide attempts, limited social and personal functioning, lack of financial independence and a more frequent use of psychoactive medication or consultation of physicians because of emotional problems (Markowitz et al. 1989).

Significantly reduced quality of life, especially in the areas of role function, social function and partnerships, was also found by Massion et al. (1993), who examined 357 individuals with a current episode of panic disorder and/or generalised anxiety disorder by means of the SF-36 Health Survey. The only study that has been performed to date on treatment effects on patients' quality of life was carried out by Telch et al. (1995), who divided 156 patients with panic disorder and agoraphobia randomly into a cognitive/behavioural therapy treatment group and a waiting control group. The Social Adjustment Scale (SAS) and a scale that measures functioning disorders (SDS) were both used as self-reports. At the beginning of treatment, the patients showed strongly impaired quality of life; these impairments diminished significantly during treatment in comparison to the control group and remained lower in follow-up. Fear and phobic avoidance were significantly correlated with a lower quality of life, but the frequency of panic attacks was not.

4.5 Quality of Life in Patients After Hospitalisation

A Canadian study investigated the quality of life of 43 patients up to 3 years after discharge from psychiatric rehabilitation programmes. A total of 86% of the patients appreciated the greater independence and privacy of their life outside the programme, and 77% perceived their quality of life to be higher than during hospitalisation, which was also apparent in higher values for the quality of life domains social functioning, recreational activities and circumstances of life (Gerber et al. 1994).

An Italian observational study on 40 patients in the course of 1 year after discharge from hospital showed that those who had taken part in a vocational rehabilitation programme reported a higher quality of life (Pirfo et al. 1994).

In a British study on 62 patients, no differences in quality of life were found between flat-sharing, supervised living groups and private living (Oliver and Mohamed 1992).

A larger group of 1527 patients was questioned concerning their (generally highly assessed) quality of life; an important predictor of quality of life proved to be income and the feeling of having control over one's own life (Rosenfeld 1992). Twenty-nine patients living in the community were questioned with the help of the Quality of Life Inventory developed by Lehmann; the result was a significant improvement in circumstances, social contact and recreational activities 1 year after discharge from hospital (Barry and Crosby 1996). By means of one analytical method, Mercier (1994) found for 152 patients that more autonomy leads to a higher quality of life and that this in turn leads to integration in the community.

In a study on 61 out-patients, the standard of living after discharge proved to play a major role in the quality of life, although satisfaction with one's self played less of a role (Skantze et al. 1992).

In a comparison between American and British out-patients, comparable quality of life was found despite a greater extent of psychopathological disorders in the American group, which, according to the authors, points out how much more adequate medical care is in the United States (Warner and Huxley 1993). Dencker and Dencker (1995) also stress the necessity of quality control in medical care in order to provide a basis for a positive development of patients' quality of life.

In an American study, 53 patients were questioned on their quality of life 11 years after discharge (Okin and Pearsall 1993). The result was that socially integrated patients assessed their quality of life highest. These results not only point out the role of circumstances, but also that relevant learning and training programmes can promote patients' social competence and quality of life (Mercier 1994; Englert et al. 1994). Atkinson et al. (1996) reported on a randomised study in which a group trained in this respect benefitted significantly from what they had learned in social functioning in comparison to a waiting control group. In general, research on psychiatric care – immediate care, rehabilitation and community care – focuses strongly on the subjective experience of the patient, which is seen as an important outcome criterion

(Sartorius 1995; Perschak et al. 1994; Simmons 1994; Corten et al. 1994).

5 Prospects

Since it first began, research on quality of life has had three main issues to deal with; these were brought up by other sources, especially from the realm of medicine. The first issue has to do with a basic scepticism towards any possibility of defining quality of life and making it operationally conceivable (Kilian 1995). The second issue refers to the construction and quality of the methods intended to assess quality of life, with these methods all too often being suspected of being "too subjective", susceptible to bias and therefore worthless (Guyatt et al. 1993). The third aspect concerns the usefulness of quality of life assessment, i.e. of the extent to which the results of this research are relevant for health policies or for medical treatment on an individual basis. This is where the goals of quality of life research need to be discussed.

In quality of life research, questions remain open concerning the individual representation of a quality of life concept or rather the problem of representing individual aspects of quality of life by means of standardised instruments; in addition, there is the question of the congruence between self-reports and observer reports; the conceptual accordance between well-being, depression and quality of life; and the weighting of quality of life indicators, the clinical relevance of quality of life assessments in practice and the question of what is most important for patients: their objective circumstances or their subjective quality of life.

Although there is consensus on the fact that the quality of life concept is at least three-dimensional and should also be assessed multi-dimensionally as well as based on patients' statements, there has been much doubt as to whether this is the right solution for the application of such a complex concept. As far as the problem of measurability is concerned, the generic and disease-specific approaches existing so far present a limited, though methodically acceptable approach to quality of life research. The question of how valid a report is and to what extent the substantial criteria of the individual patient's quality of life concept are represented by it still remains to be answered.

A critical analysis of the value of quality of life as a criterion in medicine suggests that the application of research results should also be regarded with scepticism. Clinical studies may have provided information on positive treatment effects on quality of life. However, considering in particular the large differences between clinical results and results of quality of life research, the consequences that changes in quality of life have in modern-day, individually focused treatment concepts and/or will be able to have within the borders that health policies define remain questionable. If the improvement of patients' quality of life is really to become a primary criterion in medical decisions, this will become expensive, as it cannot be realised with classical measures alone and thus will be a subject of intense social discussion.

The inclusion of the concept of quality of life in international psychiatric research has progressed slowly in comparison to other medical disciplines (Lehmann et al. 1982), and this is also true for the German-speaking countries (Möller et al. 1996). The quality of life of schizophrenic patients in particular needs to be dealt with (Kaiser et al. 1996; Lauer 1994, 1996; Stieglitz 1996). In the short time in which the subject has been discussed, however, a solid basis for research has been laid, especially as far as a theoretical basis for the quality of life concept in psychiatry is concerned, particularly in Europe (Angermeyer and Kilian 1996). In contrast to other fields, psychiatry has been substantially concerned with the concept of quality of life and has made a conceptual contribution to the clarification of the term.

On the other hand, the development of disease-specific scales for quality of life assessment of mentally ill individuals has only just begun. Many of the disease-specific methods that were developed in Anglo-American countries were originally not aimed to be applications of the quality of life concept, although now they are used as such. An assessment needs to be made of the extent to which these tools are an adequate representation of the entity "quality of life" (as a multi-dimensional concept including psychological, physical, social, mental and everyday well-being and functioning as reported by the patient). It also needs to be established to what extent detailed interviews are necessary for an assessment of quality of life or whether shorter and thus more time-saving and patient-oriented scales would make sense, at least in clinical studies. On the other hand, it remains unclear how the quality of life of psychiatric patient groups could be described qualitatively and/or to what extent the experience and behaviour of patients coincide with the aspects of quality of life that are being used. Furthermore, basic psychometric work has been done concerning internationally available scales; after being adequately translated and applied to German-speaking patients, these include a calculation of characteristic test theory values. Scales have so far only been employed in a few descriptive studies and on the question of patient care after immediate treatment, although the studies usually only involve small numbers of patients. Epidemiological and clinical studies

are almost entirely lacking, although they would be necessary for the assessment of the need for medical care and of treatment effects with respect to the criterion of quality of life. The question of whether psychiatric patients can judge their quality of life themselves can be answered affirmatively. With the exception of severe clinical symptoms or mental retardation, according to the study results available, patients give consistent and adequate information on their well-being and functioning.

The fact that patients suffer from psychiatric symptoms does not mean that their subjective experience of their current state is irrelevant with their symptoms distorting their well-being. With the exception of only a few psychiatric states, such as acute psychosis or progressive dementia, patients are willing and able to supply information about their experience and their perception of the world. Self-reporting by patients needs to be taken seriously, especially with respect to the search for tolerable medication with fewer side effects.

As the interviewer's assessment of quality of life (observer rating) usually differs from the self-report made by patients (Gater et al. 1995), these data should also be dealt with separately: the interviewer's assessment should not be regarded as an external criterion of validity for the experience of the patient but as an independent source of information.

In view of the current state of research, work is still necessary as far as quality of life assessment in psychiatry is concerned – in theoretical, methodological and practical terms. This work will only be meaningful if the quality of life concept is oriented towards the patient and is accepted as relevant in its multi-dimensionality by physicians, patients and family members, i.e. as a way to describe the patient's condition and as a criterion for therapeutic action (Strauss 1996). This basic decision for (or against) psychiatric research that is oriented towards the concept of quality of life is desirable and has implications for health policies that cannot be decided by quality of life research, although a solid basis for discussion can be laid by developing concepts, methods and possibilities for application.

It is generally true for the field as a whole that quality of life is a concept that involves value judgement. Generic quality of life scales, in particular, are implicitly derived from norms set up for a physically fit, mentally healthy, socially integrated and functionally competent person who – at least in our culture – would be highly respected. If such judgements on quality of life form the basis for treatment decisions, it needs to be considered whether patient groups are not to some extent being systematically discriminated against (e.g. physically handicapped or mentally ill patients). A way out of this dilemma is certainly not to exclude the subject of quality of life from psychiatry, but to face the inherent ethical problems involved with the definition of treatment goals, something which also applies in principal to clinical data. The question that needs to be answered within society is which kind of therapy will lead to which results with what kind of financial profit – not only with regards to the traditional classically defined state of health (e.g. measurement of laboratory parameters), but in the subjective experience of the patient as well (health-related quality of life). Further efforts are necessary in research in order to include conceptual and ethical aspects above and beyond the methodological and practical aspects of research on quality of life that are already being treated.

6 References

Abele A, Becker P (1991) Wohlbefinden. Juventa, Weinheim

Angermeyer M (1994) Symptomfreiheit oder Lebensqualität: Ziele der Schizophreniebehandlung. In: Katschnig H, König P (eds) Schizophrenie und Lebensqualität. Springer, Berlin Heidelberg New York (Aktuelle Probleme der Schizophrenie, vol 5, pp 65–80)

**Angermeyer M, Kilian R (1996) Quality of life in mental illness. In: Katschnig H, Freeman H, Sartorius N (eds) Quality of life and mental disorders. Wiley, Chichester, pp 119–132

Atkinson JM, Coia DA, Harper Gilmour W, Harper JP (1996) The impact of education groups for people with schizophrenia on social functioning and quality of life. Br J Psychiatry 168: 199–204

Awad A (1992) Quality of life of schizophrenic patients on medications and implications for new drug trials. Hosp Commun Psychiatry 43: 262–265

Baker F, Intagliata J (1982) Quality of life in the evaluation of community support systems. Eval Program Plan 5: 69–79

Baker F, Jodrey D, Intagliata J (1992) Social support and quality of life of community support clients. Commun Ment Health J 28: 397–411

Barry MM, Crosby C (1996) Quality of life as an evaluative measure in assessing the impact of community care on people with long-term psychiatric disorders. Br J Psychiatry 168: 210–216

**Bech P (1995) Rating scales for psychopathology, health status and quality of life. Springer, Berlin Heidelberg New York

Bech P (1996) Quality of life measurements in major depression. Eur Psychiatry 11: 123–126

Becker M, Diamond R, Sainfort F (1992) A new patient focused index for measuring quality of life in persons with severe and persistent mental illness. Qual Life Res 2: 239–251

Bergner M, Bobbit RA, Carter WB et al (1981) The sickness impact profile: development and final revision of a health status measure. Med Care 19: 787–805

Bigelow DA, Brodsky G, Steward L et al (1982) The concept and measurement of quality of life as a dependent variable in evaluation of mental health services. In: Strahler GJ, Tash WR (eds) Innovative approaches to mental health evaluation. Academic, New York, pp 345–366

Bigelow DA, Gareau M, Young D (1990) A quality of life interview. Psychosoc Rehabil J 14: 94–98

Brunstein M (1993) Personal goals and subjective wellbeing – a longitudinal study. J Pers Soc Psychol 65: 1061–1075

Bullinger M (1991) Quality of life – definition, conceptualization and implications – a methodologist's view. Theor Surg 6: 143–149

Bullinger M (1996) Lebensqualität – ein Ziel- und Bewertungskriterium medizinischen Handelns. In: Möller HG, Engel R, Hoff P (eds) Befunderhebung in der Psychiatrie: Lebensqualität, Negativsymptomatik und andere aktuelle Entwicklungen. Springer, Berlin Heidelberg New York, pp 14–29

Bullinger M, Kirchberger J, Steinbüchl N von (1993) Der Fragebogen Alltagsleben – ein Verfahren zur Erfassung der gesundheitsbezogenen Lebensqualität. Z Med Psychol 3: 121–131

Calman KC (1987) Definition and dimensions of quality of life. In: Aaronson NK, Beckmann J, Bernheim J, Zittoun R (eds) The quality of life of cancer patients. Raven, New York, pp 88–102

Campbell J (ed) (1981) The quality of American life. Russel-Sage, New York

Campbell J, Schraiber R, Temkin T, Tuscher T (1989) The Well-Being Project: mental health clients speak for themselves. Report to the California Department of Mental Health (internal report)

Clare A, Cairns V (1978) Design, development and use of a standardized interview to assess social maladjustment and dysfunction in community samples. Psychol Med 8: 589–604

Corrigan P, Buican B (1995) The construct validity of subjective quality of life for the severely mentally ill. J Nerv Ment Dis 183: 281–285

Corten P, Mercier C, Pelc I (1994) "Subjective quality of life": clinical model for assessment of rehabilitation treatment in psychiatry. Soc Psychiatry Psychiatric Epidemiol 29: 178–183

Day J, Bentall R, Warner S (1996) Schizophrenic patients' experiences of neuroleptic medication: a Q-methodological investigation. Acta Psychiatr Scand 93: 397–402

Dencker SJ, Dencker K (1995) The need for quality assurance for a better compliance and increased quality of life in chronic schizophrenic patients. Int Clin Psychopharmacol 9[Suppl 5]: 35–40

Derogatis LR, Lipman RS, Covi R (1973) SCL 90 – an outpatient psychiatric rating scale. Psychopharmacol Bull 19: 13–27

Endicott J, Nee J, Harrison W, Blumenthal R (1993) Quality of life and enjoyment questionnaire – a new measure. Psychopharmacol Bull 29: 321–326

Englert JS, Ahrens B, Gebhardt R, Kliefoth M, Saupe R, Stieglitz RD, Unnewehr S (1994) Implications of the concepts "coping" and "quality of life" for criteria of course and outcome. Pharmacopsychiatry 27: 34–36

Fabian ES (1992) Supported employment and the quality of life: does a job make a difference? Rehabil Counsel Bull 36: 84–97

Franz M, Plüddemann K, Gruppe H, Gallhofer B (1996) Modifikation und Anwendung der Münchner Lebensqualitäts-Dimensionen-Liste bei schizophrenen Patienten. In: Möller HG, Engel R, Hoff P (eds) Befunderhebung in der Psychiatrie: Lebensqualität, Negativsymptomatik und andere aktuelle Entwicklungen. Springer, Berlin Heidelberg New York, pp 103–112

Gater RA, Kind P, Gudex C (1995) Quality of life in liaison psychiatry: a comparison of patient and clinician assessment. Br J Psychiatry 166: 515–520

Gerber GJ, Coleman GE, Johnston L, Lafave HG (1994) Quality of life of people with psychiatric disabilities 1 and 3 years after discharge from hospital. J Qual Life Res 3: 379–383

Glatzer W, Zapf W (1984) Lebensqualität in der Bundesrepublik Deutschland. Campus, Frankfurt

Goldberg DP, Hillier VF (1979) A scaled version of the general health questionnaire. Psychol Med 9: 139–149

Grégoire J, de Leval N, Mesters P, Czarka M (1994) Validation of the quality of life in depression scale in a population of adult depressive patients aged 60 and above. Qual Life Res 3: 13–19

Gudex C (1996) Measuring patient benefit in mental illness. Eur Psychiatry 11: 155–158

Guyatt GH, Feeny DH, Patrick DL (1993) Measuring health-related quality of life. Ann Int Med 118: 622–629

Heinisch M, Ludwig M, Bullinger M (1991) Psychometrische Testung der "Münchner Lebensqualitäts-Dimensionen-Liste (MLDL)". In: Bullinger M, Ludwig M, Steinbüchel N von (eds) Lebensqualität bei kardiovaskulären Erkrankungen. Hogrefe, Göttingen, pp 73–91

Heinrichs D, Hanlon T, Carpenter W (1984) The quality of life scale: an instrument for rating the schizophrenic deficit syndrome. Schizophr Bull 10: 388–398

**Helmchen H (1990) "Lebensqualität" als Bewertungskriterium in der Psychiatrie. In: Schölmerich P, Thews G (eds) "Lebensqualität" als Bewertungskriterium in der Medizin. Fischer, Stuttgart, pp 93–115

Hogan TP, Awad AG (1992) Subjective response to neuroleptics and outcome in schizophrenia: a re-examination comparing two measures. Psychol Med 22: 347–352

Holcomb W, Morgan P, Adams N, Ponder H, Farrel M (1993) Development of a structured interview scale for measuring quality of life of the severely mentally ill. J Clin Psychol 49: 830–834

Huber D, Heinrich G, Herschbach P (1988) Measuring the quality of life: a comparison between chrinically ill patients and healthy persons. Pharmacopsychiatry 21: 453–455

Hunt SM, McKenna SP, McEwen J, Williams J, Papp E (1981) The Nottingham Health Profile: subjective health status and medical consultations. Soc Sci Med 15A: 221–229

*Kaiser W, Priebe S, Hoffmann K, Isermann M (1996) Subjektive Lebensqualität bei Patienten mit chronischer Schizophrenie. Nervenarzt 67: 572–582

Katschnig H (1994) Wie läßt sich die Lebensqualität bei psychischen Krankheiten erfassen? In: Katschnig H, König P (eds) Schizophrenie und Lebensqualität. Springer, Berlin Heidelberg New York (Aktuelle Probleme der Schizophrenie, vol 5, pp 1–13)

Katschnig H, König P (eds) (1994) Schizophrenie und Lebensqualität. Springer, Berlin Heidelberg New York (Aktuelle Probleme der Schizophrenie, vol 5)

Kelstrup A, Lund K, Lauritsen B, Bech P (1993) Satisfaction with care reported by psychiatric inpatients: relationship to diagnosis and medical treatment. Acta Psychiatr Scand 87: 374–379

Kilian R (1995) Ist Lebensqualität meßbar? Probleme der quantitativen und Möglichkeiten der qualitativen Erfassung von Lebensqualität in der Psychiatrie. Psychiatr Prax 22: 97–101

Larsen E, Gerlach J (1996) Subjective experience of treatment, side-effects, mental state and quality of life in chronic schizophrenic out-patients treated with depot neuroleptics. Acta Psychiatr Scand 93: 381–388

Lauer G (1994) Bereichsspezifische subjektive Lebensqualität und krankheitsbedingte Einschränkungen chronisch schizophrener Patienten. Psychiatr Prax 21: 70–73

Lauer G (1996) Lebensqualität und Schizophrenie: Ein Überblick über empirische Ergebnisse. In: Möller HJ, Engel R, Hoff P (eds) Befunderhebung in der Psychiatrie: Lebensqualität, Negativsymptomatik und andere aktuelle Entwicklungen. Springer, Wien New York, pp 63–72

Lehmann A (1983) The well-being of chronic mental patients. Arch Gen Psychiatry 40: 369–373

Lehmann A (1988) A quality of life interview for the chronically mentally ill. Eval Program Plan 11: 51–62

**Lehmann A (1996) Measures of quality of life among persons with severe and persistent mental disorders. In: Katschnig H, Freeman H, Sartorius N (eds) Quality of life and mental disorders. Wiley, Chicester, pp 117–128

Lehmann A, Ward W, Linn L (1982) Chronic mental patients, the quality of life issue. Am J Psychiatry 139: 1271–1276

Lehmann A, Possidente S, Hawker F (1986) The quality of life of chronic patients in state hospital and in community residences. Hosp Commun Psychiatry 37: 901–907

Lehman AF, Slaughter JG, Myers CP (1992) Quality of life experiences of the chronically mentally ill: gender and stages of life effects. Eval Program Plan 15: 7–12

Lehmann A, Postrado L, Rachuba L (1993) Convergent validation of quality of life assessment for persons with severe mental illnesses. Qual Life Res 2: 327–333

Leimkühler AM, Müller U (1996) Patientenzufriedenheit – Artefakt oder soziale Tatsache? Nervenarzt 67: 765–773

Lepkifer E, Horesu N, Floru S (1988) Life satisfaction and adjustment in lithium treated patients in remission. Acta Psychiatr Scand 78: 391–438

Lonnqvist J, Sintonen H, Syvälahti E et al (1994) Antidepressant efficacy and quality of life in depression: a double-blind study with moclobemide and fluoxetine. Acta Psychiatr Scand 89: 363–369

Malm U, May P, Deneker SJ (1981) Evaluation of the quality of life of the schizophrenic outpatient: a checklist. Schizophr Bull 7: 477–487

Maslow AH (1970) Motivation and personality. Harper and Row, New York

Massion A, Warshaw M, Keller B (1993) Quality of life and psychiatric morbidity in panic disorder and generalized anxiety disorder. Am J Psychiatry 150: 600–607

Markowitz J, Weissmann M, Oulette R, Lish J, Klermann G (1989) Quality of life in panic disorder. Arch Gen Psychiatry 46: 984–992

McDowell I, Newell C (1987) Measuring health: a guide to rating scales and questionnaires. Oxford University Press, New York

*Mechanic D, McAlpine D, Rosenfield S, Davis D (1992) Effects of illness attribution and depression on the quality of life among persons with serious mental illness. Br J Psychiatry 11: 155–164

Mercier C (1994) Improving the quality of life of people with severe mental disorders. Soc Indicat Res 33: 165–192

Möller HJ, Engel R, Hoff P (1996) Befunderbebung in der Psychiatrie: Lebensqualität, Negativsymptomatik und andere aktuelle Entwicklungen. Springer, Berlin Heidelberg New York

Mulkern V, Agosta J, Ashbaugh J et al (1986) Community support program client follow-up study. Report to NIMH Rockville, Maryland, USA

*Naber D (1995) A self-rating to measure subjective effects of neuroleptic drugs, relationship to objective psychopathology, quality of life, compliance and other clinical variables. Int Clin Psychopharmacol 10[Suppl 3]: 133–138

Naber D, Walther A, Kircher T, Hayek D, Holzbach R (1994) Subjective effects of neuroleptics predict compliance. In: Gaebel W, Awad A (eds) Prediction of neuroleptic treatment outcome in schizophrenia – concepts and methods. Springer, Berlin Heidelberg New York, pp 85–98

Najman JM, Levine S (1981) Evaluating the impact of medical care and technology on the quality of life. A review and critique. Soc Sci Med 15F: 107–115

Okin R, Pearsall D (1993) Patients' perceptions of their quality of life 11 years after discharge from a state hospital. Hosp Commun Psychiatry 3: 236–240

Oliver J (1992) The social care directive: development of a quality of life profile for use in community services for the mentally ill. Soc Work Soc Sci Rev 3: 5–45

Oliver J, Mohamed H (1992) The quality of life of the chronically mentally ill. Br J Soc Work 22: 391–404

*Patrick DL, Erickson P (1992) Health status and health policy. Oxford University Press, New York

Perschak H, Suter PM, Vetter W (1994) Determinanten der Lebensqualität und des Gesundheitszustandes bei ambulanten Patienten. Schweiz Med Wochenschr 124: 1945–1947

Pirfo E, Alberg C, Asizio I, Catapano S, Cortese M, Romano C (1994) Job preparation and improvement of the quality of life of schizophrenic patients in today's Metropolis. Int J Mental Health 23: 11–22

Priebe S, Kaiser W, Huxley P (1996) Lebensqualität als Planungs- und Evaluationskriterium psychiatrischer Versorgung. Gesundheitswesen 58[Suppl 1]: 86–90

Revicki D, Turner R, Brown R, Martindale J (1992) Reliability and validity of a health-related quality of life battery for evaluating outpatient and depressant treatment. Qual Life Res 1: 257–266

Röder-Wanner U (1995) Schizophrenie und Lebensqualität – geschlechtsspezifische Aspekte. Fortschr Neurol Psychiatrie 63: 393–401

Rosenfeld S (1992) Factors contributing to the quality of life of the chronically mentally ill. J Health Soc Behav 33: 299–315

Sainfort F, Becker M, Diamond R (1996) Judgments of quality of life of individuals with severe mental disorders: patient self-report versus provider perspectives. Am J Psychiatry 153: 497–502

Sartorius N (1995) Rehabilitation and quality of life. Int J Mental Health 24: 7–13

*Sartorius N, Kuyken W (1994) Translation of health status instruments. In: Orley J, Kuyken W (eds) Quality of life assessment: international perspectives. Springer, Berlin Heidelberg New York, pp 41–57

Schene A, Tessler R, Gamache G (1996) Instruments measuring family or caregiver burden in severe mental illness. In: Katschnig H, Freeman H, Sartorius N (eds) Quality of life and mental disorders. Wiley, Chicester

Schölmerich P, Thews G (1992) "Lebensqualität" als Bewertungskriterium in der Medizin. Symposium der Akademie der Wissenschaften und der Literatur. Fischer, Stuttgart

Selai C, Trimble M (1994) The role of quality of life measures in psychopharmacology. Hum Psychopharmacol 9: 211–214

Simmons S (1994) Quality of life in community mental health care – a review. Int J Nurs Stud 31: 183–193

Skantze K, Malm U, Dencker SJ, May PRA, Corrigan P (1992) Comparison of quality of life with standard of living in schizophrenic out-patients. Br J Psychiatry 161: 797–801

**Spilker B (1996) Quality of life assessment in clinical trails. Raven, New York

Spitzer RL, Kroenke K, Linzer M et al (1995) Health-related quality of life in primary care patients with mental disorders: results from the PRIME-MD 1000 study. JAMA 274: 1511–1517

Stein L, Test M (1980) Alternative to mental hospital treatment. I. Conceptual model, treatment program and clinical evaluation. Arch Gen Psychiatry 37: 392–372

**Stewart AL, Ware J (1992) Measuring function and wellbeing. Duke University Press, Durham, NC

Stieglitz R (1996) Erfassung von Lebensqualität bei schizophrenen Patienten. In: Möller HJ, Engel R, Hoff P (eds) Befunderhebung in der Psychiatrie: Lebensqualität, Negativsymptomatik und andere aktuelle Entwicklungen. Springer, Berlin Heidelberg New York, pp 73–82

Stoker M, Dunbar G, Beaumont G (1992) The Smith-Kline-Beecham "quality of life" scale: a validation and reliability study in patients with affective disorder. Qual Life Res 1: 385–395

Strauss J (1996) Subjectivity. J Nerv Mental Dis 184: 205–212

Sullivan G, Wells KB, Leake B (1992) Clinical factors associated with better quality of life in a seriously mentally ill population. Hosp Commun Psychiatry 43: 794–798

Telch M, Schmidt N, Jaimez T, Jacquin K, Harrington P (1995) Impact of cognitive-behavioral treatment on quality of life in panic disorder patients. J Consult Clin Psychol 63: 823–830

Thapa K, Rowland LA (1989) Quality of life perspectives in long-term care: staff and patients perceptions. Acta Psychiatr Scand 80: 267–271

Walker V, Streiner D, Novosel S, Rocchi A, Levine M, Dean D (1995) Health-related quality of life in patients with major depression who are treated with moclobemide. J Clin Psychopharmacol 15[Suppl 2]: 60S–67S

Ware J, Sherbourne CD (1992) The MOS 36-items short form health survey (SF-36). Conceptual framework and item selection. Med Care 30: 473–483

Warner R, Huxley P (1993) Psychopathology and quality of life among mentally ill patients in the community: British and US samples compared. Br J Psychiatry 163: 505–509

*Wells K, Stewart H, Hay SR (1989) The functioning of wellbeing of depressed patients – results from the medical outcome study. JAMA 202: 914–919

*WHOQOL Group (eds) (1994) The development of the WHO quality of life assessment instrument (The WHOQOL). In: Orley J, Kuyken W (eds) Quality of life assessment: international perspectives. IPSEN Foundation, Paris, pp 98–105

CHAPTER

8

L. Eisenberg

Prevention of Psychiatric Disorders

1 Definitions

The public health approach to prevention distinguishes between three levels of disease prevention: primary, secondary, and tertiary prevention.

Primary prevention is designed to preclude the development of disease among susceptible populations; it employs health promotion (e.g., the teaching of hygienic practices, universal education to promote cognitive development, the provision of optimal nutrition to enhance resistance, social support for family life, peer programs in public schools to diminish the onset of health injurious habits) and specific protection (e.g., immunizations, iodinization of salt, eliminating lead from gasoline).

Secondary prevention is designed to shorten the duration of illness once it occurs, reduce the likelihood of contagion, and limit sequelae by means of early diagnosis and prompt treatment, i.e., the use of psychotropic drugs and psychosocial interventions to abort acute psychotic states. Treatment (secondary prevention) of the first disease in a causal series constitutes primary prevention for those conditions which would otherwise follow in its wake, i.e., medical control of hypertension to reduce the likelihood of a cerebrovascular accident or treatment of hypothyroidism to avoid the onset of myxedematous madness.

Tertiary prevention is directed at individuals with irreversible disease; its goals are to limit disability (i.e., reform of institutional programs to avert the chronic social breakdown syndrome), to minimize exacerbations of the underlying disease (i.e., psychosocial education for the families of schizophrenic patients), and to promote rehabilitation (i.e., social skills training, vocational guidance, sheltered workshops for the chronically mentally ill).

In the first instance, the goal is to prevent the development of disease; in the second, to shorten its duration after it has occurred; and in the third, to preserve function so far as is possible when no effective treatment for the disease itself is available. This chapter will focus on primary prevention, but will also consider the other two levels briefly.

2 Is It Possible to Prevent Mental Disorders?

The concept of prevention in psychiatry is commonly met with skepticism. It is argued that so little is known of cause and treatment that "prevention" can be no more than a pious, vague, and perhaps even meddlesome notion. However, such a response underestimates the power of public health methods and the legitimate scope of psychiatry. Effective measures of prevention need not wait upon precise knowledge of disease etiology; programs of control can be devised from epidemiologic data on susceptibility and transmission. To cite the classic example, in 1854, by studying the distribution of cases of cholera in London, Snow was able to indict the contaminated water citizens were drawing from the Broad Street pump as the source of the epidemic, 30 years before Koch isolated and cultivated the cholera vibrio. The fact is that effective prevention of *some* psychiatric disorders is not only possible, but is substantially complete for *certain* disorders in *some* countries; what is not possible now, or in the foreseeable future, however, is the prevention of *all* mental disorders.

Pellagra and paresis are telling examples of primary prevention. Patients suffering from pellagra crowded the orphanages and the mental hospitals in the United States in the early decades of this century. Well before it was shown to result from niacin deficiency, but after it was recognized as a nutritional disease (Sydenstricker 1958), pellagra was eliminated in the United States by improvements in diet (reduction in the dependence on milled corn as a food staple). In the first decades of this century, the wards of mental hospitals were crowded with patients suffering from general paralysis of the insane. Now paretics are hardly to be found in industrialized countries. Effective programs to treat syphilis with penicillin have eliminated the tertiary disease manifestations of spirochete infection. Neither of these preventive measures is "psychiatric." However, what matters is not the mode of action of the agent, the venue in which it is applied, or the academic discipline of the practitioner, but the effectiveness of the measure in preventing diseases causing disturbed mental function (APA Task Force on Prevention Research 1990).

Moreover, secondary and tertiary prevention (clinical care) has led to even more impressive gains for the recurrent and chronic mental disorders. This is reflected in the progressive reduction in the need for psychiatric beds during the last half of this century. Between the mid-1950s and the mid-1990s in the United States, the number of inpatient mental hospital beds (public and private combined) fell by more than half (public beds by more than 90%), although the U.S. population increased by more than half over that time span. The reduction in the need for hospitalization and in the average length of stay when hospitalization is necessary are a product of changes in sociomedical administrative policy (the "open hospital" and "community mental health") and the sequential development of more effective treatment methods, e.g., psychopharmacology, new methods of psychotherapy, psychoeducational groups for families, social skills

training for patients, partial hospitalization and day care, sheltered workshops, and group residences in the community.

The emphasis in this chapter is on validated means for preventing mental disorders. The measures are presented in a life cycle perspective beginning before conception and continuing into old age. Primary preventions head the list; secondary and tertiary measures applicable in clinical settings are discussed briefly; and the chapter concludes by considering health policy.

3 Preventive Interventions

3.1 Family Planning and Prenatal Care

The more numerous and the more closely spaced the pregnancies in the reproductive lives of women, the greater the risks for maternal and infant mortality and the worse the developmental outcome for the children (World Bank 1993). Studies in developed countries have demonstrated that the larger the number of children in a family (other variables such as socioeconomic status having been controlled for), the lower their educational attainment (Blake 1989). Unplanned and unwanted teenage pregnancies are associated with a high risk for mother and child (Brown and Eisenberg 1995). Taken together, these findings indicate the importance of family planning services to reduce the number of offspring and to lengthen interbirth intervals in order to optimize the ability of parents to care for their children. The health risks associated with modern contraception are far less than those associated with pregnancy and childbirth (DaVanzo et al. 1990).

Family planning provides an opportunity to improve outcomes for mother and baby by identifying and managing risks *before* conception. Ordinarily, prenatal care begins weeks or even months into the pregnancy when the conceptus has already implanted in the uterus and begun to develop its central nervous system. Pre-conception planning, in contrast, focuses on optimizing health and avoiding controllable risks before pregnancy begins. Topics typically addressed in pre-conception care include diet and weight, exercise, smoking, alcohol and drug use, reducing environmental risks, bringing vaccinations up to date, and managing sexually transmitted diseases (STDs) if they are present, as well as bringing chronic medical conditions such as diabetes, hypertension, and cardiovascular disease under as much control as is possible. Because maternal diabetes is associated with complications of pregnancy, strict metabolic control before and during pregnancy has been shown to reduce the risk to the developing fetus (Fuhrman et al. 1984).

Family planning makes it possible to minimize toxic effects on spermatogenesis by optimizing nutrition and by reducing alcohol and drug use and exposure to chemicals (Cefalo and Moos 1995). It permits the identification of vulnerability to genetic diseases through early diagnosis. Carrier detection coupled with prenatal testing and access to abortion allows families at high risk to give birth to healthy children of their own by enabling them to abort defective fetuses (Milunsky 1992; D'Alton and DeCherney 1993). This can sharply reduce the incidence of inherited disease of the central nervous system (though without affecting carrier rates). A telling example is provided by Tay-Sachs disease, a gangliosidosis resulting from a deficiency of hexosaminidase A. It is manifest by motor weakness beginning in the first 6 months of life and progressive motor and mental deterioration, with the development of feeding problems, deafness, blindness, convulsions, and spasticity; the patient usually dies of bronchopneumonia before the fourth year of life. The disorder occurs principally in a defined population group, i.e., Ashkenazi Jews (Jews from Eastern Europe and their descendants). The carrier rate in this population in the United States and Canada is about 1 in 31 (in contrast to a rate of about 1 in 280 among non-Jews). Voluntary enrollment of American and Canadian Ashkenazi Jewish populations in screening programs made it possible to reduce the number of newly diagnosed cases of Tay-Sachs disease from an average of about 60 per year in the 1960s to 13 by 1980 and to three to five cases per year in the decade that followed (Kaback et al. 1993).

Comprehensive family planning services should include education about contraceptive choices and access to contraceptives. Safe abortion must be available as a backup for contraceptive failure. Adolescents use contraception irregularly; even when used faithfully, every contraceptive method has a failure rate (lowest for the pill). Major morbidity and mortality from back-alley abortions is inevitable if legal and safe abortion is denied (Brown and Eisenberg 1995).

Inadequate nutrition, cigarette smoking, alcohol consumption, drug abuse, and inadequate prenatal care during pregnancy are all associated with increased hazards to the fetus, including higher rates of low birth weight infants. Low birth weight, in turn, is associated with higher neonatal mortality rates (NMR) and developmental impairment among survivors. The high technology of the neonatal intensive care unit (NICU) results in greater salvage for very low birth weight infants, but at much greater monetary costs and far less satisfactory developmental outcomes than that achieved by improving the physical and social

condition of the mother during pregnancy (Shiono and Behrman 1995). Thus, although NICUs in the United States have lower birth weight-specific NMRs than those in Sweden, the overall Swedish NMR is lower than the U.S. NMR because the proportion of low birth weight infants born in the United States is half again higher than that in Sweden (Guyer et al. 1982). The limits to technology are indicated by the outcomes in U.S. NICUs. The addition of pulmonary surfactant and dexamethasone therapy has doubled the rate of survival for very low birth weight infants (less than 750 g), but at the cost of survivors handicapped by subnormal cognitive function and neurosensory abnormality, including cerebral palsy, blindness, or deafness (Hack et al. 1996).

Where feasible, screening for elevated blood levels of α-fetoprotein, for chromosomal anomalies by cytogenetic methods, and for morphologic abnormalities by ultrasonography can permit the detection and abortion of abnormal fetuses carried by mothers at risk. Spina bifida and anencephaly are related neural tube defects which are major causes of morbidity and mortality in infancy and childhood. Mid-trimester screening by measuring maternal serum α-fetoprotein or by ultrasound examination can identify more than four fifths of affected pregnancies. The application of screening and abortion of affected fetuses has led to a fall in the prevalence of these defects.

There is now strong evidence that neural tube defects are caused by inadequate intake of folate in the periconceptional period (Czeizel et al. 1994; Medical Research Council 1991). Only in the case of planned pregnancies can women deliberately increase their intake of folic acid before conception. Thus it has been proposed that all fertile women take a supplement until menopause or that food for the entire population be fortified with folate. The principal concern about fortification is that individuals in the population with untreated vitamin B_{12} deficiency (but without a clinically evident anemia) may develop an irreversible neuropathy from folate supplementation if diagnosis is delayed (Bower and Stanley 1996; Dickinson 1995).

3.2
Newborn Screening

A number of congenital metabolic abnormalities can be detected by routine screening of newborns. Notable among the correctable conditions are phenylketonuria (PKU), galactosemia, and congenital disorders of thyroid function, all of which result in severe central nervous system pathology if treatment is not instituted in the first weeks of life and maintained thereafter. The clinical manifestations of the first two can be prevented by appropriate diet, and of the third by extrinsic thyroxin. The fact that they occur at low frequency in Caucasian populations – 1 in 3600 to 1 in 5000 for disorders of thyroid function, 1 in 10,000 to 1 in 25,000 for PKU, and 1 in 60,000 to 1 in 80,000 for galactosemia (American Academy of Pediatrics Committee on Genetics 1989) – makes newborn screening a practicable public health measure only in countries with highly developed health services. Although the cost of case detection is relatively high because the conditions are rare, so are the lifetime costs to the community in caring for severely affected children. Screening programs are, however, of little or no value in the absence of a comprehensive follow-up program to ensure that the infant at risk receives optimal care (Rowley and Huntzinger 1985; Holtzman et al. 1986).

3.3
Childhood Immunizations

According to UNICEF (1996), 20 million deaths in children under the age of 5 have been prevented since 1980 by means of immunization against diphtheria, pertussis, tetanus, measles, polio, and tuberculosis; however, there are still some 2 to 3 million deaths each year from vaccine-preventable diseases because of the failure to extend the vaccination program to all susceptible children. Moreover, mortality data underestimate the magnitude of the public health burden because they do not tabulate the morbidity from central nervous system pathology and from the psychosocial consequences of chronic handicap among survivors. Full implementation of the WHO Expanded Program of Immunization would not only yield enormous gains in further reduction of mortality in childhood (and thus make parents more willing to forego large family size) but spare compromised brain function and psychosocial disability among survivors. Measles encephalitis, subacute sclerosing panencephalitis, and mental retardation from congenital rubella syndrome or hemophilus B meningitis are rapidly disappearing from the clinical scene in countries with comprehensive immunization programs (Gruenberg et al. 1986).

3.4
Preventing Malnutrition

Deficits in the intake of specific micronutrients as well as in overall protein-calorie intake can impair brain development with major consequences for cognitive and emotional function.

Iodine deficiency disorders (IDD) constitute the most pressing instance of micronutrient deficiencies which lead to brain malfunction. IDD affects between

600 million and one billion people in the world (Hetzel 1989). Clinical manifestations include still births, abortions, and congenital anomalies; endemic cretinism, characterized by mental deficiency, deaf mutism, spastic diplegia, and other forms of neurologic defect; and impaired mental function associated with goiter. IDD in individuals at risk can be prevented for 3–5 years with one injection of 2–4 ml iodized poppy seed oil, a treatment that can be given by primary care workers. To prevent fetal IDD, iodized oil must be administered before conception (another benefit of planned childbirth); treatment even as early as the first trimester of pregnancy is not fully effective. Oil injections are both feasible and practical as an immediate means to control endemic IDD. For reasons of costs and convenience, the long-term goal must be the introduction of an iodized salt program for the entire population. Program success is dependent upon public education to elicit the full support of the population (Hetzel 1986, 1989).

Worm infestations in children result in retarded physical stature and cognitive ability. It is now possible to treat very effectively 19 of the 23 major human helminth infections with one of three drugs taken orally: albendazole, praziquantel, or ivermectin. In principle, it should now be possible to treat for iodine deficiency, vitamin A deficiency, and worm infestations by oral medication administered at school. Such programs are now being tested in the field (Warren 1991).

Severe protein-calorie malnutrition, life-threatening in itself, increases the likelihood that exposure to infectious agents will result in clinical disease; malnutrition impairs host defenses. Furthermore, malnourished individuals show increased systemic manifestations of diseases which are more limited in those who are well nourished. Gastrointestinal diseases, made more likely by malnutrition, increase nutritional stresses on the host by increasing caloric requirements at the same time as food intake and absorption are impaired. Traditional "treatments" for diarrhea by reducing intake of food and fluids worsen the threat. UNICEF has taken the lead in minimizing harm from diarrheal disease in children by promoting oral rehydration with an easily prepared solute solution.

The conjoining of chronic malnutrition with disadvantageous family circumstances results in retarded cognitive and social development. Studies of malnourished children indicate that it is the interacting and multiplicative effects of the simultaneous biological and social insults that result in damage (Dobbing 1987). Grantham-McGregor et al. (1978, 1991) have shown that nutrition plus social stimulation for hospitalized malnourished children, maintained after hospital discharge by parents who have been educated by home visitors, result in greater developmental gains than renourishment alone. Effective remediation must be targeted at the entire complex of social and nutritional deprivation.

Monitoring the growth of young children, a simple method well within local resources, permits the early detection of developmental failure. It is one of the four components of the UNICEF "GOBI" initiative: growth monitoring, oral rehydration, breast-feeding, and immunization (Grant 1995).

Insuring universal education for women is the common denominator in measures to improve child health. National infant mortality data show a stronger inverse correlation with years of education women receive than they do with the gross national product (Caldwell 1986). The intervening behavioral mechanisms include later age of marriage, fewer and more widely spaced children, and use of hygienic health-promoting behaviors among educated women (Hobcraft 1993).

3.5 Injury Prevention

Vehicular accidents are a major source of head and spinal cord injuries among survivors. Such injuries are preventable by vigorous enforcement of lower speed limits (Wagenaar et al. 1990), by better highway design, traffic regulations, vigorous prosecution of drunken driving, automatic seat belts, child safety seats, and air bags. Cycling is a major cause of hospital admissions for head injuries. In recent studies (Thompson et al. 1989, 1996), cyclists wearing helmets had an odds ratio for brain injury after vehicular accidents of 0.35 compared to those without helmets.

Poisonings in children can be minimized by laws requiring "child-proof" safety caps on bottles of medication and toxic chemicals for domestic use (Walton 1982). Blood lead levels in children can be reduced by effective public controls on the lead content of gasoline (Centers for Disease Control 1982a).

3.6 Home Visiting and Enriched Day Care

Olds and his colleagues (Olds et al. 1986; Olds and Kitzman 1990) have shown that pre- and postnatal home visitation, transportation for health care, and sensory and developmental screening were effective in preventing abuse and neglect among children born to socially disadvantaged primiparas. The women visited by nurses made better use of community services,

experienced greater social support, improved their diets, and reduced their smoking. Length of gestation and newborn birth weight were improved, and there were fewer verified cases of abuse among poor, unmarried teenage mothers.

Studies in both developed and developing countries have shown that children growing up under deprived circumstances exhibit deficits in cognitive development, lower levels of academic achievement, and increased rates of behavioral and antisocial disorders (Eisenberg and Earls 1975). These disastrous outcomes can be made less likely through enriched day care programs which involve parents as active participants. Several long-term outcome studies have demonstrated that children enrolled in day care exhibit better occupational history, fewer out-of-wedlock pregnancies, and lower rates of academic and behavioral pathology (Berrueta-Clement et al. 1984; Jordan et al. 1985; Lazar et al. 1982).

Day care programs can facilitate the attainment of a second goal, namely the teaching of parenting skills to adolescents by having them participate in the care of toddlers under supervision. The effectiveness of this strategy has not been formally demonstrated, but its desirability is indicated by the fact that experience in child care within the family, the traditional way these skills have been transmitted, is becoming ever less available. With smaller family size and fragmentation of the family, it can no longer be taken for granted that such "naturalistic" education is available to all children.

3.7
Community- and School-Based Programs

Pierson et al. (1983) introduced a parent education and diagnostic screening program with periodic developmental examinations from 6 months, weekly play groups from 2 years, and daily prekindergarten from 3 years of age. Classroom observation demonstrated that experimental children had less learning difficulty and fewer reading problems in the second grade. Ramey and Campbell (1984) evaluated a child-centered prevention program emphasizing language, cognitive, perceptual-motor, and social development in children aged 18–54 months. The enrolled children scored significantly higher than controls on a series of tests of mental ability. Botvin et al. (1984) assessed a school-based 12-unit curriculum delivered by peer leaders or classroom teachers (with periodic booster sessions in subsequent years). The goal was to give junior high school students the skills to resist pressures to smoke and use drugs, to help them develop self-esteem, and to cope with social anxiety. The outcome was a reduction in the onset of smoking in the experimental students, both by self-report and saliva tests.

These and other model programs have been reviewed by an American Psychological Association Task Force on Prevention (Price et al. 1989). The Task Force noted that the common features of successful programs include "careful targeting of the population, the capacity to alter life trajectory, the provision of social support and the teaching of social skills, the strengthening of existing family and community supports, and rigorous evaluations of effectiveness" (p. 57).

3.8
Can Dementia Be Prevented?

The prevalence of dementia increases with age, from a rate of less than 1% in those aged 60–64 to a rate as high as 40% in those aged 90 and older. Prevalence increases each year both because of an increase in the size of the population at risk (i.e., the "graying" of the world population), but also because of an increase in the survival time of patients with dementia. Epidemiologic studies (Katzman 1993) in countries as different as the United States, France, Italy, Sweden, Finland, Israel, and China reveal a negative correlation between the amount of schooling received in youth and the prevalence of dementia in old age. Not only is this evident in cross-sectional studies, but lack of formal education also predicts a decline in cognitive function in a community population of individuals aged 65 and over (Evans et al. 1993). If these findings are not merely an artifact of tests for dementia (and careful analysis of the data suggests that it is not), it is possible that school-related intellectual activity results in an increase in synaptic density during development, just as a stimulus-rich environment does for the brains of other mammalian species (Eisenberg 1995). The additional "brain reserve" may delay the appearance of clinical symptoms, even though the early pathological changes of Alzheimer's disease may be present (Mori et al. 1997; Alexander et al. 1997). Thus it is an appealing extrapolation from the evidence to suggest that increasing access to public education could minimize the occurrence of dementia in the developing world, an initial approach to "primary prevention of dementia." This proposal, however fanciful, is worth implementing in and for itself. Education is an engine for development, both personal and social. This chapter has already noted the reduction in infant mortality associated with greater schooling for women (Caldwell 1986). Measures such as universal public education that have great general usefulness merit implementation, even when a direct effect on the prevention of a particular disease (in this case, dementia) is only inferential.

4
Clinical Settings: Secondary and Tertiary Prevention

The iatrogenic diseases resulting from inappropriate prescription can be reduced by training primary health care workers in the recognition and treatment of psychosocial disorders. A recent WHO 19-country study has revealed the high frequency of psychiatric morbidity among patients in general practice and the fact that health workers recognize only half of these problems (Üstün and Sartorius 1995). Similar findings have been demonstrated in child health care in the developing world (Geil et al. 1981). Training in the recognition and management of psychiatric disorders in general practice can not only reduce unnecessary diagnostic examinations and inappropriate medication, but also make effective mental health care available and thus reduce morbidity (Eisenberg 1992; Goldberg and Huxley 1992).

The "treatment gap" in epilepsy (i.e., the percentage of patients with active epilepsy who are *not* in treatment) in less developed countries has been estimated to be as high as 70%–90% (Shorvon and Farmer 1988). Because untreated epilepsy is associated with increasing psychosocial impairment worsened by the stigma associated with the disorder, improving the skills of primary health workers in the recognition of the disorder and its treatment can markedly diminish psychosocial handicap (Eisenberg 1997). Furthermore, improvement in obstetrical care, more effective accident prevention, and prompt treatment of central nervous system infections can reduce the incidence of symptomatic epilepsies.

The primary prevention of the schizophrenias is beyond present capacities. If anything, according to Kramer (1989), there will be an inexorable increase in cases of schizophrenia in the developing world because the population at risk is increasing (i.e., young adults between the age of 20 and 40). However, secondary prevention through the use of neuroleptic drugs for patients, psychoeducational training for family members, and social skills training for patients can reduce the duration of treatment episodes and the likelihood of relapse (Brown et al. 1972; Leff et al. 1982; Hogarty et al. 1986; Tarrier et al. 1989). Tertiary prevention, by keeping hospital stays to a minimum, redesigning institutional programs, and providing social skills training in sheltered workshops, can avert chronic social breakdown syndromes among patients with chronic disorders (Gruenberg and Kennedy 1988). Systematic research in China has demonstrated the remarkable effectiveness of community-based interventions in enabling chronic schizophrenic patients to be maintained in the community (Phillips et al. 1994).

In similar terms, although we lack a knowledge base for the primary prevention of affective disorders, the use of tricyclics, selective serotonin reuptake inhibitors (SSRIs), lithium, and interpersonal and cognitive psychotherapy (Frank et al. 1989) can shorten morbid episodes and reduce the likelihood of recurrence (Quality Assurance Project 1983; NIMH Consensus Development Conference Statement 1985). In view of the long-term morbidity associated with depressive illness (Kiloh et al. 1988) and the evidence that the incidence of depression in successive birth cohorts has been rising (Cross-National Collaborative Group 1992), emphasis on the diagnosis and treatment of depression in primary care merits high priority in public health programs.

The U.S. National Institute of Mental Health has developed a Depression Awareness, Recognition and Treatment (DART) Program in order to improve the ability of health workers to identify and treat depression. O'Hara and colleagues (1996) evaluated a series of DART programs at 18 sites across the United States by comparing participants' knowledge about depression before and after 2-day workshops. The findings demonstrated increases in levels of knowledge as well as a high degree of satisfaction with the program on the part of the health professionals who attended. Six-month follow-up evaluations indicated a continued positive evaluation. What is still to be established is the impact of the increased knowledge on actual behavior in clinical practice once the participants have returned to their own workplaces.

Similarly to the DART Program, the World Psychiatric Association in collaboration with the International Committee for Prevention and Treatment of Depression has initiated an Educational Program on Depressive Disorders (WPA/PTD 1997; Linden 1998). It goal is to distribute worldwide present knowledge on the diagnosis and treatment of depression and to foster clinical awareness of this widespread and disabling disorder.

Indeed, suicide prevention provides the most telling psychiatric example of prevention by treating an antecedent disorder. Depression is the most potent risk factor for suicide. Given that there are effective treatments for depression, making such treatment widely available should lead to lower rates of suicide in the population. Empirical evidence of such an effect has been reported in two studies. Rutz et al. (1989) reported a statistically significant reduction in suicide among women (and hence in the total rate) on the Swedish island of Gotland 1 year after all general practitioners on the island had completed an educational program on the recognition and care of depressed patients. The sex difference may reflect

differences in the willingness to consult doctors or in greater physician diagnostic sensitivity to depression in women. Similarly, Rimer et al. (1990) reported a negative correlation between rates of suicide and rates of treated depression in administrative districts across Hungary.

Ahrens et al. (1995) reanalyzed data on more than 800 patients who suffered from affective disorders and who had been maintained on long-term lithium prophylaxis. Whereas the mortality risk for untreated patients is three times greater than that for an aged-matched population sample, the ratio of observed to expected deaths in the treated population was the same as that for the general population. The effects of lithium indicated a reduction in deaths from suicide and cardiovascular disease.

5
Is Prevention Always Preferable?

Having shown that prevention of some mental disorders is possible, we must still ask whether it is always preferable. From the standpoint of the individual who would otherwise have become ill, avoidance of illness is almost always more desirable than treatment, because it avoids the morbidity associated with illness and its care. However, when prevention requires changing habitual behavior and particularly when change causes withdrawal symptoms (as in smoking cessation), the "cost" of prevention may deter its acceptance. Moreover, the time lag between risk taking and its pathological consequences, a matter of decades, reinforces public skepticism, because smoking cessation reduces rather than altogether eliminates the probability of disease. Although the risk of lung cancer is far greater among smokers than nonsmokers, not all smokers develop cancer and lung cancer also occurs in nonsmokers. Ways of detecting those at greatest risk (presumably because of genetic susceptibility to the carcinogens in cigarette smoke) have not yet been uncovered. The relatively slow progress of antismoking campaigns attests to the strength of social forces which reinforce cigarette smoking.

From the standpoint of the community, in contrast to that of the individual, decisions about undertaking preventive measures require weighing up competing social objectives. To pursue the instance of smoking, its elimination involves the loss of jobs in tobacco farming and cigarette production, the loss of revenue from taxation, and the loss of foreign exchange from exports. These factors do not gainsay the extraordinary benefits from smoking cessation, i.e., significant reductions in rates of cancer, ischemic heart disease, chronic obstructive pulmonary disease, prematurity, and other health hazards; however, they do highlight the magnitude of the political challenge (Director General's Report 1986).

5.1
Weighing Risks and Benefits

Prevention is so attractive a concept that, all too often, little thought is given to the potential for "toxicity" that may be associated with implementing putative preventive programs. Methods to identify individuals at risk can harm those incorrectly labeled. Programs can fail. They can cause disease in those who are not enrolled as well as harm those who are included. Paradoxically, they can have an overall negative effect on health in the community even when they benefit target groups.

Population screening to identify individuals at increased risk for disease inescapably carries with it dual hazards, i.e., that of being incorrectly labeled as being at risk (false positives) and that of being incorrectly reassured of safety (false negatives). The likelihood of error is a function of test sensitivity (the proportion of affected individuals who test positive), test specificity (the proportion of unaffected indivduals who test negative), and the a priori probability of being affected (the proportion of individuals in the population with the condition).

Ineffective measures carry two kinds of costs: first, the waste of resources which could have been used to provide other services; second, a negative impact on public opinion. When unrealistic promises are made in the name of prevention, the failure to redeem these promises makes the public cynical about other public health proposals.

An effective measure can be counterproductive when it does not reach a sufficient proportion of the population. If the uptake of rubella vaccine among susceptible children is less than 90% (the level necessary for herd immunity), an immunization program can lead to an increase in rates of congenital rubella syndrome (CRS) in the next generation because it postpones the age at natural infection among the unvaccinated. Even though CRS rates among the vaccinated population will be close to zero, rates will increase disproportionately among the offspring of the unvaccinated (Knox 1984).

An incompletely implemented prevention program can harm those it is designed to benefit. Newborn screening for sickle cell anemia poses risks for the infants positive for the disease and those positive for the trait. Infants with sickle cell anemia will benefit only if parents are thoroughly informed about the significance of the finding and if appropriate medical care is available from physicians knowledge-

able about managing the disease (Rowley and Huntzinger 1985). Infants at greatest risk for harm without offsetting benefit are those with the sickle cell trait. Undue parental apprehension may lead to a vulnerable child syndrome (Green and Solnit 1964); as an adult, the individual may be denied employment or insurance if trait is mistaken for disease. Further, as in the case of every recessive genetic disorder, testing parents carries the risk of discovering that the husband is not the father of the child, information which compromises the relationship between the parents.

That premature introduction of programs purported to prevent disease can be hazardous to health is not an argument against prevention, but it does emphasize the need for careful review of potential hazards before undertaking new initiatives and rigorous evaluation to assess both negative and positive effects after programs are introduced.

5.2 Setting Priorities

Given the fact that the resources available for health programs are finite and that assigning them to one use precludes their availability for another, the challenge is to assess the competing claims of various proposals for prevention and to decide on the proper balance between prevention and treatment. Decisions must be based on a close analysis of the extent and distribution of disease on a country by country basis and the resources, both internal and external, available to each for public health measures. The criteria to be employed in weighing up competing options include the illness burden produced by the disease, effectiveness of the intervention, its toxicity, its feasibility, and the opportunity cost (see below).

The illness burden resulting from a specific disease (i.e., senile dementia of the Alzheimer type) or a category of diseases (i.e., all mental disorders) is a function of the following: (a) the prevalence of the disease or diseases, (b) the severity of the morbidity and mortality resulting these diseases, and (c) the distribution of the ages at onset. That burden includes the cost of providing medical and social services for the ill, the losses to economic productivity resulting from sickness and death, the encumbrance on the family from caring for the sick, and the pain and suffering experienced by sick individuals and their families. Because they are more easily measured, the first two indicators (health care costs and output losses) are the ones employed when attempts are made to quantitate illness burden.

Black and Pole (1975), in a pioneering study in the United Kingdom, measured burden in terms of inpatient days, outpatient referrals, sickness benefits, and mortality. On the basis of 1972 data, they reported that mental illness accounted for 31% of all inpatient days (and mental handicap for an additional 15%), 4% of outpatient referrals, almost 8% of general practitioner consultations, and almost 10% of sickness benefit days; just over 1% of life years lost was attributable to suicide.

Rice et al. (1976, 1985, 1992) have pioneered U.S. studies of illness burden; they employed an analogous set of measures, but, in addition, they computed direct (costs for care) and indirect (losses of output from sickness or premature death) economic costs of illness. As in the United Kingdom, mental disorders accounted for many inpatient days, were a major cause of disability, and exacted considerable economic costs and life years lost. In an assessment based on 1980 data, the Board on Mental Health and Behavioral Medicine of the Institute of Medicine (1984) noted that mental illness accounted for about $20 billion in direct health care costs, about 8% of all direct health care expenditures in that year. Indirect expenditures were estimated as eight times as great. A recent Institute of Medicine study of child and adolescent mental disorders (Institute of Medicine 1989) estimates that about 16% of inpatient and 23% of outpatient pediatric insurance costs incurred by dependents in the United States were attributable to psychiatric problems.

According to the World Development Report (World Bank 1993), mental health problems make up 8.1% of the global burden of disease. Notably, this figure does *not* include brain trauma, mental retardation, or neurotic disorders; thus it is a considerable underestimate. The WHO World Health Report (WHO 1995) lists the ten leading causes of morbidity and of disability. Neurotic, stress-related, and somatoform disorders together make up the third most important cause of morbidity; in terms of chronically disabled individuals, mood disorders are the most important single cause; mental retardation is fourth, epilepsy sixth, dementia seventh, and schizophrenia ninth. The most striking data of all were provided by a recent analysis completed by Murray and Lopez (1996) from the Harvard Center for Population and Development Studies. The unit they employed in their statistics was the disability-adjusted life year (DALY). This takes into account years of life lost (by subtracting age at death from remaining life expectancy) and the handicap imposed by chronic disease by discounting the value of the years so affected. In 1990, depression was the fourth leading cause of DALYs, exceeded only by lower respiratory infections, diarrheal diseases, and perinatal conditions. In the year 2020, depression will be second only to ischemic heart disease as a cause of world disease burden. However,

despite the fact that neuropsychiatric conditions account for five of the ten most important causes of long-term disability and the fact that depression alone is currently the fourth most important cause of DALYs and will be the second leading cause by 2020, mental health is largely missing from the international health agenda.

The potential toxicity of an intervention must be weighed against its efficacy in reducing the risks posed by the disease to be prevented; in other words, although every effort must be made to minimize toxicity, even substantial side effects are tolerable if the intervention is highly effective in averting serious morbidity and mortality. The decision, however, cannot rest upon expert opinion alone. The public must have accurate information about benefits and risks set before it and must be persuaded that the one outweighs the other. Distorted publicity on side effects can lead to an increase in avoidable morbidity and mortality. Such has been the case with pertussis immunization in the United Kingdom (Centers for Disease Control 1982); concern about vaccine side effects reduced parental compliance and resulted in a substantial increase in cases of pertussis.

The feasibility of a program depends upon whether the necessary infrastructure for implementation is in place, i.e., personnel with the requisite competence to administer it, appropriate geographic distribution of such personnel, and a health system able to coordinate their activities. Thus, even a potentially effective intervention may not make sense for a particular community if it requires a level of skill and an administrative system not to be had in that community. The simpler the method and the more compatible it is for use by primary health care workers, the more likely it can be implemented in a less developed country.

Finally, it is necessary to take into account the opportunity cost (i.e., those programs forgone because funding and energy have been committed to the program under review). Because health care resources are finite, a decision to embark on a prevention program will entail diversion of resources from other health programs. The allure of cost saving from prevention is so attractive to governments that treatment facilities may be closed down on the mere assumption that disease will be avoided. Such has been the case in the United Kingdom and the United States, where the illusory promise of "community" mental health, with its claim of averting iatrogenic illness arising from hospitalization, has led to the wholesale closing of psychiatric hospitals, leaving many chronically ill mental patients without any care because community alternatives were not made available (Weller 1989).

6 Need for Epidemiologic Intelligence

Policy decisions can be no better than the quality of the information on which they are based. In the absence of reliable data on the extent, distribution, and social burden of mental disorders in the population and of time trends in prevalence, decisions on the allocation of resources can be no better than shots in the dark. It is essential to develop information centers capable of collecting and feeding back data on the nature and magnitude of the health problem in each country and on the results from treatment and prevention programs. All new public health initiatives should therefore include a mandatory set-aside of funds to pay for epidemiologic surveillance in order to facilitate periodic reassessments of effectiveness and to permit midcourse corrections in resource allocation.

7 References

Ahrens B, Mueller-Oerlinghausen B, Schou M et al (1995) Excess cardio-vascular and suicide mortality of affective disorders may be reduced by lithium prophylaxis. J Affect Disord 33: 67–75

Alexander GE, Furey ML, Grady CL et al (1997) Association of premorbid intellectual function with cerebral metabolism in Alzheimer's disease: implications for the cognitive reserve hypothesis. Am J Psychiatry 154: 165–172

American Academy of Pediatrics Committee on Genetics (1989) Newborn screening fact sheets. Pediatrics 83: 449–464

APA Task Force on Prevention Research (1990) Report of the Task Force. Am J Psychiatry 147: 1701–1704

Berrueta-Clement JR, Schweinhart LJ, Barnett WS et al (1984) Changed lives: the effects of the Perry Preschool Program on youths through age 19. High Scope, Ypsilanti/MI

Black DAK, Pole JD (1975) Priorities in biomedical research: indices of burden. Br J Prev Soc Med 29: 222–227

Blake J (1989) Number of siblings and educational attainment. Science 245: 32–36

Botvin GJ, Baker E, Renick NL et al (1984) A cognitive-behavioral approach to substance abuse prevention. Addict Behav 9: 137–147

Bower C, Stanley FJ (1996) Issues in the prevention of spina bifida. J R Soc Med 89: 436–442

Brown GW, Birley JLT, Wing JK (1972) Influence of family life on the course of schizophrenic disorders: a replication. Br J Psychiatry 121: 241–258

*Brown S, Eisenberg L (eds) (1995) The best intentions: unintended pregnancy and the well-being of children and families. Washington National Academy Press, Washington

Caldwell JC (1986) Routes to low mortality in poor countries. Popul Dev Rev 12: 171–220

Cefalo RC, Moos K (1995) Preconception healthcare: a practical guide, 2nd edn. Mosby, St. Louis

Centers for Disease Control (1982a) Blood lead levels in U.S. population. MMWR 31: 132–133

Centers for Disease Control (1982b) Pertussis – England and Wales. MMWR 31: 629–632

Cross-National Collaborative Group (1992) The changing rate of major depression: cross-national comparisons. JAMA 268: 3098–3105

Czeizel AE, Dudas H, Metnaki J (1994) Pregnancy outcomes in a randomized controlled trial of periconceptional multivitamin supplementation. Arch Gynecol Obstet 255: 131–139

D'Alton ME, DeCherney AH (1993) Prenatal diagnosis. N Engl J Med 328: 114–120

DaVanzo J, Parnell AM, Foege WH (1990) Health consequences of contraceptive use and reproductive patterns. JAMA 265: 2692–2696

Dickinson CJ (1995) Does folic acid harm people with vitamin B12 deficiency? Q J Med 88: 357–364

Director General's Report (1986) Tobacco or health. Executive Board, World Health Assembly, Geneva (EB 77)

Dobbing J (ed) (1987) Early nutrition and later achievement. Academic, London

Eisenberg L (1992) Treating depression and anxiety in primary care. N Engl J Med 326: 1080–1084

**Eisenberg L (1995) The social construction of the human brain. Am J Psychiatry 152: 1563–1575

Eisenberg L (1997) Sociocultural perspectives. In: Engel J, Pedley TA (eds) Epilepsy: a comprehensive textbook. Lippincott-Raven, Philadelphia

Eisenberg L, Earls FJ (1975) Poverty, social depreciation and child development. In: Hamburg DA (ed) American handbook of psychiatry, vol 6. Basic Books, New York, pp 275–291

Evans DA, Beckett LA, Albert MS, Hebert LE, Scherr PA, Funkenstein HH, Taylor JO (1993) Level of education and change in cognitive function in a community population of older persons. Ann Epidemiol 3: 71–77

Frank E, Kupfer DJ, Perel JM (1989) Early recurrence in unipolar depression. Arch Gen Psychiatry 46: 397–400

Fuhrman K, Reiher HP, Semmler K et al (1984) The effect of intensified conventional insulin therapy before and during pregnancy on the malformation rate in the offspring of diabetic mothers. Exp Clin Endocrinol 83: 173–177

Geil R, de Arango MV, Climent CE et al (1981) Childhood mental disorders in primary health care: results of observations in four developing countries. Pediatrics 68: 677–683

**Goldberg D, Huxley P (1992) Common mental disorders: a biosocial model. Routledge, London

Grant JP (1995) The state of the world's children. Oxford University Press, Oxford

Grantham-McGregor SM, Stewart ME, Desai P (1978) A new look at the assessment of mental development in young children recovering from severe malnutrition. Dev Med Child Neurol 20: 773–778

Grantham-McGregor SM, Powell CA, Walker SP et al (1991) Nutritional supplementation, psychosocial stimulation and mental development of stunted children: the Jamaican study. Lancet 338: 1–5

Green M, Solnit A (1964) Reactions to the threatened loss of a child: a vulnerable child syndrome. Pediatrics 34: 58–66

Gruenberg EM, Kennedy C (1988) Some determinants of social disability. In: Henderson AS, Burrows GD (eds) Handbook of social psychiatry. Elsevier, Amsterdam

Gruenberg EM, Lewis C, Goldston SE (eds) (1986) Vaccinating against brain syndromes: the campaign against measles and rubella. Oxford University Press, New York

*Guyer B, Wallach LE, Rosen SL (1982) Birth-weight standardized mortality rates and the prevention of low birth weight: how does Massachusetts compare with Sweden? N Engl J Med 306: 1230–1233

Hack M, Friedman H, Fanaroff AA (1996) Outcomes of extremely low weight infants. Pediatrics 98: 931–937

Hetzel BS (1986) Mental defect due to iodine deficiency. In: Berg MJ (ed) Science and service in mental retardation. Methuen, New York

Hetzel BS (1989) The story of iodine deficiency. Oxford University Press, Oxford

**Hobcraft J (1993) Women's education, child welfare and child survival. Health Transition Rev 3: 159–175

Hogarty GE, Anderson CM, Reiss DL et al (1986) Family psychoeducation, social skills training and maintenance chemotherapy in the after care of schizophrenic patients. Arch Gen Psychiatry 43: 633–642

Holtzman C, Slazyk WE, Cordero JF et al (1986) Descriptive epidemiology of missed cases of phenylketonuria and congenital hypothyroidism. Pediatrics 78: 553–558

Institute of Medicine (1984) Research on mental illness and addictive disorders. Report of the Board on Mental Health and Behavioral Medicine. National Academy, Washington

Institute of Medicine (1989) Research on children and adolescents with mental, behavioral, and developmental disorders. National Academy, Washington

Jordan TJ, Grallo R, Deutsch M et al (1985) Long term effects of early enrichment: a twenty year perspective on persistence and change. Am J Community Psychol 13: 393–415

*Kaback M, Lim-Steele J Dabholkar D et al (1993) Tay-Sachs disease – carrier screening, prenatal diagnosis and the molecular era: an international perspective, 1970–1993. JAMA 270: 2307–2315

*Katzman R (1993) Education and the prevalence of dementia and Alzheimer's disease. Neurology 43: 14–20

Kiloh LG, Andrews G, Neilson M (1988) The long term outcome of depressive illness. Br J Psychiatry 153: 752–757

Knox EG (1984) Theoretical aspects of rubella vaccination strategies. Int J Infect Dis 7: 194–197

Kramer M (1989) Barriers to prevention. In: Cooper B, Helgason T (eds) Epidemiology and the prevention of mental disorders. Routledge, London, pp 30–55

Lazar I, Darlington R, Murray H et al (1982) Lasting effects of early education. Monogr Soc Res Child Dev 47(1–2): 1–151 (serial no 194)

*Leff J, Kuipers L, Berkowitz R et al (1982) A controlled trial of social intervention in the families of schizophrenic patients. Br J Psychiatry 141: 121–134

Linden M (1998) Fortbildungsprogramm der WPA und des PTD-Komitees zu depressiven Störungen. Münch Med Wochenschr 140: 146–149

Medical Research Council. Vitamin Study Research Group (1991) Prevention of neural tube defects. Lancet 338: 131–137

Milunsky A (1992) Genetic disorders and the fetus, 3rd edn. Johns Hopkins University Press, Baltimore

Mori E, Hirono N, Yamashita H et al (1997) Pre-morbid brain size as a determinant of reserve capacity against intellectual decline in Alzheimer's disease. Am J Psychiatry 154: 18–24

*Murray CJL, Lopez AD (1996) The global burden of disease and injury, vol 1. Harvard University Press, Cambridge/MA

NIMH Consensus Development Conference Statement (1985) Mood disorders: pharmacologic prevention of recurrences. Am J Psychiatry 142: 469–476

O'Hara MW, Gorman LL, Wright EJ (1996) Description and evaluation of the Iowa Depression Awareness, Recognition and Treatment Program. Am J Psychiatry 153: 645–649

Olds DL, Kitzman H (1990) Can home visitation improve the health of women and children at environmental risk? Pediatrics 86: 108–116

Olds DL, Henderson CR, Tatelbaum R (1986) Preventing child abuse and neglect: a randomized trial of nurse home visitation. Pediatrics 78: 65–78

Phillips M, Pearson V, Lange R (eds) (1994) Psychiatric rehabilitation in China: models for change in a changing society. Br J Psychiatry 165[Suppl 24]: 1–194

Pierson DE, Bronson MB, Dromey E et al (1983) The impact of early education: measured by classroom observations and teacher ratings of children in kindergarten. Eval Rev 7: 191–216

Price RH Cowen EL, Lorion RP et al (1989) The search for effective programs: what we learned on the way. Am J Orthopsychiatry 59: 49–58

Quality Assurance Project (1983) A treatment outline for depressive disorders. Aust N Z J Psychiatry 17: 129–146

Ramey CT, Campbell FA (1984) Preventive education for high risk children: cognitive consequences of the Carolina Abecedarian project. Am J Ment Defic 88: 515–523

Rice DP, Feldman JJ, White KL (1976) The current burden of illness in the United States. Institute of Medicine, National Academy of Sciences, Washington

Rice DP, Hodgson TA, Kopstein AN (1985) The economic costs of illness: a replication and update. Health Care Financ Rev 7: 61–80

*Rice DP, Kelman S, Miller LS (1992) The economic burden of mental illness. Hosp Community Psychiatry 43: 1227–1232

Rimer Z, Barsi J, Veg K, Katona CLE (1990) Suicide rates in Hungary correlate negatively with reported rates of depression. J Affect Disord 20: 87–91

Rowley PT, Huntzinger DJ (1985) Newborn sickle cell screening: benefits and burdens realized. Am J Dis Child 137: 341–345

Rutz W, Knorring L von, Walinder J (1989) Frequency of suicide on Gotland after systematic post-graduate education of general practitioners. Acta Psychiatr Scand 80: 151–154

Shiono PH, Behrman RE (1995) Low birth weight: analysis and recommendations. Future Child 5: 4–18

Shorvon SD, Farmer PJ (1988) Epilepsy in developing countries: a review of epidemiologic, sociocultural, and treatment aspects. Epilepsia 29[Suppl 1]: 36–54

Sydenstricker VP (1958) The history of pellagra, its recognition as a disorder of nutrition and its conquest. Am J Clin Nutr 6: 409–414

Tarrier N, Barrowclough C, Vaughn C et al (1989) Community management of schizophrenia: a two year follow up of a behavioral intervention with families. Br J Psychiatry 154: 625–628

*Thompson DC, Rivara FP, Thompson RS (1996) Effectiveness of bicycle safety helmets in preventing head-injuries. JAMA 276: 1968–1973

Thompson RS, Rivara FP, Thompson DC (1989) A case-control study of the effectiveness of bicycle helmets. N Engl J Med 320: 1361–1367

UNICEF (1996) The state of the world's children. Oxford University Press, New York

United Nations (1991) The world's women 1970–1990. United Nations, New York

Üstün TB, Sartorius N (eds) (1995) Mental illness in general health care: an international study. Wiley, Chichester

Wagenaar AC, Streff FM, Schultz RH (1990) Effects of the 65 MPH speed limit on injury morbidity and mortality. Accid Anal Prev 22: 571–585

Walton WW (1982) An evaluation of the Poison Prevention Packaging Act. Pediatrics 69: 363–370

Warren KS (1991) Helminths and health of school-aged children. Lancet 338: 686–687

Weller MPI (1989) Mental illness – who cares? Nature 339: 249–252

WHO (1995) Bridging the gaps. WHO, Geneva

World Bank (1993) World Development Report – investing in health (world development indicators). Oxford University Press, Oxford

WPA (World Psychiatric Association), PTD (International Committee for Prevention and Treatment of Depression) (eds) (1997) Educational program on depressive disorders, module I: overview and fundamental aspects. NCM, New York

CHAPTER

9

W. Böker

General Principles for Psychiatric Treatment

Translator: S. Johnson

1 Benefits of Establishing General Principles for Psychiatric Care

A chapter on general principles of care has been a notable omission from previous editions of *Psychiatrie der Gegenwart.* In the past few decades, there have been impressive developments in psychopharmacology, and psychosocial treatment programmes have become widespread. These developments seem to have focused attention so much on specific treatment strategies for particular diseases and even symptoms that there has been little place for consideration of fundamental principles which apply for all disorders.

1.1 Historical Developments

In 1861, W. Grieser's textbook on the pathology and treatment of psychiatric illnesses set out the following general principles for psychiatric treatment: the fundamental principle of respect for humanity; the recognition that one should understand "madness as an illness" and "abnormal psychological phenomena as cerebral processes"; the absolutely equal validity of psychological and physical methods of treatment, with strong adherence to the principle of "adapting the treatment of the insane to the needs of each individual"; the need for both early intervention and patience where the course of illness is long; "protection from damaging influences" and "a properly regulated balance between rest and activity"; and finally, a rapid "move to a good asylum".

In 1929, a substantial chapter of Bumke's *Handbuch der Geisteskrankheiten* ("Handbook of Mental Disorders") was still devoted by Nitsche to "General Treatment and Prevention of Mental Disorders". Interestingly, this chapter accorded the basic principles primacy over specific measures:

> As the treatment of mental illnesses . . . remains mostly symptomatic, what we can learn from general principles of psychiatric treatment still constitutes a very fundamental part of our therapeutic equipment. . . . Variations of the general principles of treatment provide the only ways of managing individual psychiatric illnesses.

Indeed, treatments such as bed-rest, prolonged bathing or even Simon's treatment programme based on activity were hardly designed for specific clinical presentations. Compared with the rest of medicine, the discovery of treatment methods targeting particular psychiatric disorders came very late (Degkwitz et al. 1982).

1.2 Problems of Contemporary Psychiatry

The reappearance of general principles of treatment in this volume seems to me to be attributable to the following needs and problems in contemporary psychiatry:

1. In National Socialist Germany, the psychiatrically ill and mentally handicapped were victims of appalling acts of murder, sacrificed to a politically fanaticised social Darwinism as "empty human shells, unworthy of life". In the former USSR and the other states of the old Eastern bloc, psychiatry was abused in the service of political oppression. Following these events, there is again a pressing need for consideration of the basic values which should underpin all psychiatric care (Heimann 1991). The current emphasis on individual civil rights embedded in recent liberalisation movements provides a further impetus to this, as does the growing emphasis on the "healthy aspects" and demands for autonomy of even the severely mentally ill. The ethical and legal problems arising from processes such as the compulsory admission and treatment of the mentally ill are discussed in Chaps. 15–18 (Vol. 1, Part 2).
2. In the course of history, there have been continual changes not only in views of the nature of mental illness, but also in the fundamental concept of the human being on which psychiatry is based (Benedetti 1959). Recently, there has been particular pressure for change in traditional concepts and expectations of human behaviour, which are now becoming less and less tenable as world-wide migrations of peoples create culturally very mixed societies, especially in Western industrial countries. There is an increasing danger of misunderstanding the ways in which psychological distress is expressed by people from cultures with which we are not familiar, or of treating them wrongly because of prejudice or unconscious defence mechanisms. This creates new challenges for doctors in fulfilling their obligation to adhere to a concept of the individual which entails a commitment to ethical treatment.
3. More and more biological, psychological and social factors which may influence human experience and behaviour have been identified. In some areas, tensions and alienation have developed within our (as yet still) shared profession between a neurobiological and scientific psychiatry, allying itself increasingly with physical medicine, and representatives of more psychotherapeutic or social

psychiatric perspectives. In this climate, principles which are generally accepted need to be clearly established and stated, so as to prevent a fissure in clinical psychiatry between biological, medication-based treatments for the severely ill and psychological therapies for less severe disorders (Gabbard and Goodwin 1996). Because of the divergence of these movements, there is a need for integrating models and for tasks in which they can be united (Andreasen 1996) so that once again communication is possible using a common language within a shared framework.

4. Clinical practice in psychiatry is now increasingly pervaded by pressure for detailed analysis of performance, a pressure which originated in industry and has grown as business management models have become influential and health costs have risen. This requires searching consideration of the relationship between costs and benefits in our treatment measures. Political committees, funders of hospitals and health insurance companies expect more detailed disclosure of the details of our daily work with patients. This performance analysis must be linked to efforts to safeguard and optimise the quality of our treatment and also to incorporate into basic and specialist training knowledge gained through this process. In reasoned debates with the funders of health care and with those with political responsibility for it, our position is strengthened by the formulation and adoption of general treatment guidelines and by establishing standardised quality requirements and using these a basis to make claims for the resources which are absolutely necessary for their fulfilment.

So that these needs can be met, two questions need first to be addressed: Which fundamental concepts and models currently underlie psychiatric treatment, and which *should* underlie it?

2 Fundamental Models Underlying Current Psychiatric Treatment

La vray science et la vray estude de l'homme c'est l'homme.
CHARRON (1601, as quoted in SCHRENK 1973)

Especially when one considers efforts to establish theoretical foundations for it, psychiatry is obviously a medical discipline which is especially susceptible to being influenced by the current cultural climate and by changing ideologies. This is a result of the endless diversity of human destinies and the accompanying variations in the forms taken by psychological disturbances, the mysteries surrounding the major mental illnesses and the powerful emotions which the mentally ill can provoke in those around them. Finally, and of particular current relevance, it results from the mind–body problem, which has not so far been satisfactorily resolved (see Vol. 1, Part 2, Chap. 1).

However, the effects of theoretical concepts and of the ideological battles between their sometimes fanatical advocates seem to me much less striking when one considers the old pragmatic methods for "care of the insane" and also the modern practice of psychiatry. Calming the agitated patient, enlivening those who have sunk into apathy or melancholy and guiding the isolated back into an active life within the community are unavoidable necessities. These remain the constant tasks of the doctors and carers of the mentally ill, as they have been for centuries.

Except in the terrible periods of political abuse of psychiatry under National Socialism or Stalinist communism, psychiatric care has – at least since the time of the Enlightenment – aimed for care of the mentally ill which is humane and not determined by ideology. The ideological battles between the psychically and the somatically oriented, between adherents of orthodox psychoanalysis or extreme social psychiatry and the pharmacologically and biologically based psychiatric units died down some 20 years ago. The last great tide of ideology in our speciality was unleashed in the 1960s and 1970s by the antipsychiatry movement, a "late Romantic" movement (Glatzel 1975) which lacked internal coherence and, apart from the justified critique of the large psychiatric hospitals, has had little impact on psychiatric practice.

The previously dominant model in medicine was very much a mechanical one, fundamentally derived from Descartes and over-emphasising structural defects and linear causality. Currently, a paradigm shift away from this can be observed to varying degrees within medicine, towards a theory and practice in which psychological and somatic processes are understood as interdependent and multiply linked aspects of the functioning of a complex living system. The subjective experience of the patient and his or her "individual reality" are further essential elements in this model (von Uexküll 1986), which has in the last few decades also found its way into psychiatry.

2.1 Biopsychosocial Model

The great majority of types of psychiatric treatment are today based on the biopsychosocial model of human health and illness, which was first proposed for medicine as a whole by Engel (1970). According to this model, our behaviour and experience develop

from a genetic basis, follow individual predispositions, are modified by learnt behaviour patterns and conscious and unconscious strivings, and are shaped by interpersonal relationships and the influences of the environment.

Thus care which seeks to alter only somatic events, only processes which can be understood in psychological terms or only the social environment must be rejected as inadequate.

Extensions of this model to the theory of psychosis have given rise to widely accepted basic models for management of schizophrenia, in the forms of the vulnerability–stress model (Zubin and Spring 1977) and a further development from it, the vulnerability–stress-coping–social competence model (Nuechterlein and Dawson 1984; Ciompi 1986; Brenner 1989). Subsequently, this concept has also been applied more and more to depression and anxiety disorders. More recent psychosomatic theorising has produced other potentially fruitful models, such as the "functional" and "situational circle model" proposed by von Uexküll (1986), the systematic–ecological conceptual framework, which has not yet been very sharply formulated conceptually but whose content is very promising (Bertalanffy 1974; Ludewig 1992; Böker and Brenner 1996), chaos theory (an der Heiden 1992; Prigogine and Stengers 1981) or also Maturana's exposition of a biology of cognitive processes (Maturana and Varela 1987). In future, the benefits of these new directions in thinking will probably be seen not so much in individual treatment methods as in the formulation of integrating approaches to psychological disturbances.

2.2 Medical Historical Foundations

When one considers the biopsychosocial model from the perspective of medical history, it becomes obvious that it is a revival of much older concepts which go back as far as Arabic and Greek antiquity. These are the rules for living found in the *Diaita* of antiquity, which extend across many aspects of human functioning and represent a quest for "good order" and balance in human existence, both within the microcosm formed by each individual and in the individual's relationship with their natural environment and with the macrocosm (Lain Entalgo 1982; Garcia-Guillen 1982). The recommendations on the conduct of life which are derived from this relate to the six *res non naturales*: (1) light and air, (2) food and drink, (3) work and rest, (4) sleep and waking, (5) excretions and secretions and (6) stimulation and balancing of the mind. These are contrasted with the *res naturales*, i.e. the biological nature of man, and the *res contra naturam*, i.e. diseases. In the Middle Ages, these ideas were further developed and popularised as the *regimen sanitatis* (Schmitt 1982). The ultimate goal of this regime is a balanced and harmonious life, and this should be "the first and noblest interest of the doctor" (Schipperges 1962). The task is the "productive moulding of the tensions between the physical forces of nature and the creative and ordering mind of man" (Schrenk 1973). The concept of the *regimen sanitatis* was also at the root of "moral treatment" in the eighteenth century, and guidance towards a regulated order in life is a theme which has continued to appear in institutional regimes up to the present day. In this connection, links between psychological treatment and education, not least through the bridge formed by learning theory and cognitive psychology, obviously take on increasing significance. In this age of electronic data analysis, it is important not to lose sight of the biographical method as a way of approaching the individual patient.

Attempts to reinstate the ancient *Diaita* for modern therapeutics and to adapt it so that it may again become a fruitful method are potentially of great value, but such attempts are still very much at an early stage (see Tellenbach 1982). Finally, a question which arises here is whether in addition to biological, psychological and social dimensions one should consider a spiritual level. This area plays scarcely any role in modern treatment, apart from the pastoral care provided within psychiatric in-patient units, generally a very marginal aspect of these services.

3 Need for a Basic Philosophy of Treatment

In common with all professionals who work with the mentally ill, confidence with scientific models and concepts and skilfulness in implementing them are not the only attributes required of good psychiatrists. Their interactions with others need to be marked by a particular professional attitude, an inner attitude which is a product of attentiveness, self-criticism and a feeling of responsibility.

Attitudes to the mentally ill have fluctuated continually in the course of history between the poles of compassion, blame ("mad or bad") and fear. All these are feelings which we still encounter in relatives, but also in ourselves:

Christian humility and brotherly love, abhorrence of cripples and monsters, fear of madness, amusement about the stupidity of the insane, concern for the moral education of those fallen into sin, bureaucratic approaches to organising care, scientific curiosity, exclusion of the chronically unproductive, and solidarity

with the mad as sacrifices to an illness-producing society (and an illness-producing hospital) can be identified as attitudes which have determined the practical management of the mentally ill in the hospital, and which in varying mixtures still do determine it (Böker 1986).

Transference and counter-transference give rise to powerful feelings, which people without adequate training often cannot cope with. There is a great danger that they may lead to unhelpful acting out of inner tensions, such as aggression, or that in psychotherapy they encourage overstepping of boundaries, even to the point of sexual contact. We need to consider what characteristics the basic professional attitude needs to have in order to be faithful to the fundamental principle of humanity which was already to be found in Griesinger's work. This section will outline a series of precepts and recommendations of which even the experienced professional should always remain aware.

3.1
Respect for the Patient

We should always approach patients with sincerity, listen to them impartially and not intentionally lie to or deceive them. Even behind manic excitement or autistic withdrawal into a world of bizarre delusions, when control has been lost or in the intellectual disintegration of a dementing illness, the worth of the patient remains intact and is to be respected (the concept of individual worth is embedded in German basic law, which upholds the fundamental rights of every individual to attention, respect and self-determination and refers to the basic rights of every individual to attention, respect and individual development, regardless of behaviour, earnings or current state).

Each episode of psychiatric care should limit the patient's right to lead his or her own life as little as possible, so that there should be no unnecessary isolation, physical restraint or other use of force and no unnecessary segregation of the sexes, or unnecessary restrictions, e.g. on visits and outings. The patient's view of life and fundamental religious ideas are to be respected.

3.2
Reinforcement of Autonomy

A major goal of treatment is to help patients to help themselves. The only restrictions on patients taking responsibility for themselves in all private matters and on their right to a degree of privacy should be those clearly required because of the illness. This is relevant, for example, to the hospital and ward regulations that apply to in-patients.

3.3
Minimal Use of Intrusive Forms of Treatment

Both in diagnostic investigations and in therapeutic interventions, attention should be paid to differences in the degree of intrusiveness between methods. Each treatment should be as unintrusive as possible, so that out-patient treatment is to be preferred to in-patient, care on open wards to be preferred to locked wards, and each intervention with medication should use the minimum effective dose. *Primum nil nocere!*

3.4
Clear Definition of Treatment Goals and Measures

We owe it to patients and relatives to explain our judgements, intentions and actions in a comprehensible manner, at an appropriate time and after rational consideration. One must therefore:

- Formulate at an early stage a clear, realistic overall treatment plan
- Strive for close collaboration between all professions and with all those who have an important role in the patient's life
- Be able to define precisely the individual indication for each therapeutic measure

3.5
Critical Self-Examination

Everyone working in psychiatry has an obligation to constant critical self-scrutiny and continuing professional education. Supervision and peer supervision provide advice, clarification and protection against boundaries being overstepped in problematic patient–therapist relationships. They should prevent therapists from losing touch with the community of their colleagues and withdrawing into an isolated ideological position.

4
Development of a Therapeutic Alliance

For a course of treatment to succeed, especially when, as often occurs in psychiatry, it is carried out over a long period, a robust patient–therapist relationship needs to develop, i.e. an explicit or unspoken treatment

alliance needs to be formed. The first contact often has a decisive influence on the success of this relationship. If a sensitive therapist is able to adopt the basic professional attitude described above, an attitude permeated by respect and concern, he or she will quickly come to set great store by the success of this encounter. However, problems can arise. The therapist and the patient do not necessarily inhabit a common world, even if they come from the same culture and linguistic community. Both patient and therapist always have been and remain compelled to adapt their way of responding to the demands of their environment and to their personal aspirations and experiences. This gives rise to endless diversity in "individual realities" and different subjective understandings of the world. von Uexküll (1986), who placed this perspective at the centre of a psychosomatically oriented treatment method, stated that:

People realise themselves not only through the development of their bodies, but also by means of a creative achievement through which the environment they encounter is transposed into an individual reality.

What happens in medicine when these different concepts of reality encounter one another? When a patient has a fairly simple, objectively demonstrable disorder of health such as a fractured wrist, an open wound or a skin abscess, doctor and patient tend to unite quite quickly to form a treatment alliance. They bend together over a piece of work, in which assessment of the condition and management give rise to little in the way of problems of interpretation.

The situation already becomes considerably less clear in the case of general medical illnesses with complex patterns of symptoms, such as heart failure or multiple organ disorders, where there are greater difficulties in achieving a clear focus on an objectively defined piece of work. Here, phenomena such as illness gain and exaggeration or dissimulation of symptoms come into play. Thus the patient's subjective evaluations and attribution of meaning influence von Uexküll's "diagnostic–therapeutic circle". Nevertheless, doctor and patient do generally manage to reach agreement on a diagnosis of an "illness", and the patient agrees to enter a treatment alliance so as to be cured of his or her malady.

The situation in psychiatry is generally more complicated, most of all in the case of psychotic patients. Here, patients' perceptions and the functioning of intrapsychic schemata are adversely affected. Cognitive disturbances or disordered emotions impose a particular interpretation on patients' view of reality, which may then deviate to an especially glaring degree from the individual reality of the therapist. "Insight into illness" and consent to treatment cannot be taken for granted. In any case, the experience of change or disturbance reported by the patient is not necessarily characterised by suffering (e.g. in mania or ecstatically uplifted states in people with schizophrenia). So as to establish a working alliance, the therapist must first get to know the individual, psychotically transformed reality of the patient and must use patient attention to attempt to find a way into this. The scientific illness concept of the doctor (who, for example, diagnoses a psychosis of a schizophrenic type) and the private logical interpretation of the psychotically ill person (who, for example, understands the shift in his experience as the result of hypnotic influence) are in no way congruent.

However, serious consequences for necessary treatment can follow from this incongruence, so that a careful piece of negotiation must be carried out with the aim of constructing a new shared reality which allows a therapeutic relationship to be established and accepted. Particularly once the acute psychotic phase of disorganisation has subsided and stability is being re-established, this work becomes essential to pave the way for rehabilitative measures and effective relapse prevention. A plausible explanatory model needs to be introduced, allowing the subjective experience of the patient to be fitted into a scientific concept (e.g. a readily comprehensible explanation of the vulnerability–stress-coping concept of schizophrenia). This allows a patient who is at risk of relapse to comprehend and integrate into his or her individual reality sensible prophylactic measures such as the prescription of neuroleptics which are effective in preventing relapse and a low-stress lifestyle, as well as early intervention for prodromal symptoms. Patients thus also learn to communicate with therapists about their illnesses in a common language.

According to the conceptual framework of the cognitive biologist Maturana (1982), different forms of life exist in a "network of structural linkages", which as humans we constantly construct through oral communication about our behaviour ("lingolaxis"). These structural linkages create "consensual areas", for whose development a "sense of mutuality" is required. Psychiatric treatment should thus be particularly successful if it is based on a "common reality", which opens up new consensual areas between therapist and patient. In the light of such observations, therapy takes on a wider significance, as one must constantly seek opportunities to fit the structures better to one another, or – in the therapist–patient relationship – to achieve better links between individual realities. Thus treatment involves something which reaches deeper than simple technical application of methods of intervention. It comes much more to involve a shared journey, taking place over a shorter or longer period of treatment, in which the various

reciprocal interactions also produce changes in the therapist. In some ways, earlier treatment concepts have already prepared the ground for such a view of treatment: examples are the concept of empathy, the psychoanalytic transference and counter-transference constructs, the concept of "co-evolution" as the "art of growing together" (Willi 1985), the "collaborative empiricism" of Perris (1989) or the aspiration of establishing a "therapy based on partnership" (Böker 1990, 1992). The problems raised in medico-legal and ethical deliberations about how to meet requirements for "informed consent" are closely related to the issues of medical practice discussed above (for an overview, see e.g. Lidz et al. 1984).

In conclusion, a general treatment principle of great importance in psychiatry is that therapists must strive to attain a framework of understanding shared between themselves and their patients within which a productive working alliance can develop.

To achieve this goal, it is not enough to devote adequate interest and time to the patient, to tell and discuss with him or her results of examination and investigations, diagnosis and treatment plans and to involve family members – all of which unfortunately is all too often not achieved satisfactorily. We must get to know precisely the subjective reality of the patient and his or her individual values and concept of reality, so that we can build a bridge over which he or she can walk into a joint reality shared by therapist and patient. Of course the patient must collaborate in this constructive communication process and is likely to do so all the more willingly if he or she is aware of our efforts. Thus patients expect us to keep them company for a stretch of their journey. To this end, we must also change. We have to adapt rigid learned schemata to the living diversity of our patients and will thus experience the widening step by step of our horizons of understanding and our creative possibilities. Currently, the training of prospective psychiatrists unfortunately takes far too little account of these requirements.

5
Fundamental Assumptions Behind the Major Approaches to Treatment

Factors identified as possible influences on psychiatric disorders and treatment programmes for such illnesses have long since become too numerous to be readily reviewed or enumerated, and considerable difficulties arise in trying to fit them all into a set of categories based on the biopsychosocial model. The problem of numbers alone means that such an enterprise is doomed to be incomplete. The close association of biological and psychological mechanisms in many or probably in the end in all methods applied and their often substantial dependence on socio-cultural context again mean that such a categorical classification is very questionable. Many complex programmes of treatment are based on several basic principles. Other treatment programmes have been eclectically assembled or are based on therapists' personal preferences, so that basic principles cannot always be clearly identified. Practical psychiatry as a "healing art" allows for as great a variety of ways of providing care to others as there are therapists.

Bearing in mind these difficulties, for the purposes of discussion and because such a division is customary, the following section adopts a classification of the very diverse range of treatments into three domains: biological (somatic) treatments, psychological treatments and social treatments.

The universally valid principal which most of all needs to be remembered is the fundamental statement in the Hippocratic Oath, *primum non nocere*, and the fundamental therapeutic attitude described in Sect. 2 is also always applicable. A further universal principle is the obligation of the therapist to respect the ethical ground rules of his profession, such as the Hippocratic Oath for doctors, and to practice according to the *lege artis*, i.e. using established methods with a clear scientific basis (evidence-based medicine). This last requirement is taking on increasing significance with recent work on quality assurance (Gaebel 1995; Haug and Stieglitz 1995) and with an improvement in the quality of life of the mentally ill (Möller et al. 1996).

5.1
Biological (Somatic) Treatments

Attempts to alter the behaviour and experiences of humans by physical means are legion and have a thousand-year-old tradition. There are probably literally no methods involving the body which have not been tried out on the mentally ill at some time.

They range from the measures which are closely related to basic care (bed rest, cold compresses, plaster, bath cures, massages, mechanical restraint), various shock treatments (immersion in water, blows, the application of hot iron, cardiazol shock and electroconvulsive therapy), regulation of food intake (diets), coarse methods involving the whole organism (vomiting cures, purges, sweating and fever cures, insulin coma), surgical interventions (leucotomy, stereotactic neurosurgery), methods altering autonomic functioning (exhaustion, sleep deprivation, light therapy) to the use of natural medicines (cabbage teas) and industrially synthesised drugs, the psychotropic drugs which in the last few decades have become particularly

sophisticated and have become the most widely applied method.

From examination of the indications and goals for the application of these methods, they can be classified into methods for dampening agitation, promoting sleep, influencing biorhythms, lifting mood and dispelling anxiety, suppressing acute psychotic symptoms, arresting epileptic fits and diminishing vulnerability to these, stimulating (stimulants) and loosening (psychedelics) of intrapsychic connections, blockade of neuronal overactivity and hyperarousal processes (neuroleptics) and many more. The difficulty of somatic approaches and the way in which they still remain unsatisfactory is their lack of specificity. Even now, the newest psychotropic drugs are still burdened with unwanted side effects. We still know too little about the natural laws underlying the development of psychopathological symptoms to be able to plan optimal somatic interventions on the basis of mechanism of action and dosage. Our treatment is not based on causes, but on symptoms or, at best, syndromes. The following general principles for somatic treatments are widely accepted as basic rules for their use.

5.1.1 Clarification of Indication

Should a sleep disturbance be treated with hypnotics, tranquillisers or low-dose neuroleptics? Which neuroleptics are to be preferred in the acute phase and which for maintenance treatment? Is it best to subdue withdrawal symptoms from illicit drugs with medication, or does a better prognosis result from a "cold withdrawal", alleviated by massage, teas and baths? Guidelines derived from international consensus conferences will in future contain more and more advice on indications and dosage. The essential foundations for these are and will continue to be delivered by carefully evaluated clinical experience (see Vol. 1, Part 2, Chap. 10).

5.1.2 Weighing Up of Effects and Side Effects

Particularly in the use of psychotropic drugs, desired therapeutic effects are often not achieved because of ignorance of the range of effective doses, lack of care in monitoring effects or inappropriate polypharmacy, so that more harm results than good. The art is not to undermine the organism's capacity for self-regulation through wrong dosage or excessively long periods of administration (e.g. in the use of hypnotics). With electroconvulsive therapy, there are "side effects" of a social kind, in the form of the widespread aversion to it among the public. Together with the biological inadequacy of explanations for its efficacy, this has led to a very marked, albeit empirically ill-founded, decline in the use of this method and to the range of accepted indications for it becoming narrower.

5.1.3 Following Rules for Application and Dosage

Should a drug be administered orally or parenterally, in immediate release or in depot form? Each of these means of administration has advantages and disadvantages. Is a sleeping drug more effective before or after midnight? The dosage determines degree of efficacy and extent of side effects. Low dosage and a shorter period of administration of neuroleptics reduce the incidence of tardive dyskinesia, but increase the relapse rate. Extra-pyramidal side effects restrict patients' compliance, etc.

5.1.4 Attention to Interactions

When given in combination, many psychotropic drugs interact with one another in desirable or in undesirable ways. For example, monoamine oxidase inhibitors cannot be given with many other anti-depressives, and clozapine should not be given with carbemazepine. In cases of multiple morbidity, drugs are prescribed for different types of indication, and attention must then be paid to their interactions. Increasingly, dependence or self-medication for psychiatric symptoms leads patients to use drugs which potentiate the intended effects of psychotropic drugs (e.g. potentiation of the effects of sedatives with alcohol) or which counteract these (e.g. amphetamine counteracts the effects of hypnotics). In order to avoid harmful effects, doctors need to elicit carefully details of self-medication which patients may not initially volunteer.

5.1.5 Attention to the Legal Framework

Lack of insight is bound at times to lead to a need to implement against the patient's will somatic interventions such as compulsory injections, restraint and forcible feeding; the legal implications must be examined before these are used.

5.2 Psychological Treatments

Psychological methods of treatment have a history just as long as that of physical forms of care. Obtaining an overview of the whole field is perhaps even more

difficult than with biological treatments. Here, one encounters a vast number of more or less clearly distinguishable methods, which are used in bewildering diversity and varying combinations by professional therapists and natural healers, but also by lay people. The spectrum extends from magical procedures (sleeping in the temple, religious dances, magic), approaches using suggestion (group attainment of ecstatic states, individual suggestion, hypnosis), relaxation methods (catharsis, autogenic training, systematic muscle relaxation), so-called exploratory methods (psychoanalysis and the various developments from it), humanistic therapy and counselling, gestalt and imaginative therapies through to practical methods of directly influencing behaviour (behaviour therapies, biofeedback) and the newer cognitive methods. Individual, group and family therapies have each established their own practitioners and preferred types of indication.

5.2.1 Definitions

Avoiding a very restricted definition of psychotherapy as the treatment of disorders of health by psychological means, a definition such as that given by Wolberg (1967) demonstrates what an extraordinarily broad range of requirements psychotherapy seeks to meet:

Psychotherapy is the treatment of emotional problems by psychological means, in which a specifically trained therapist carefully establishes a professional relationship with the patient. Possible goals are first to remove, modify or make milder current symptoms, second to change disturbed ways of relating, and third to encourage the attainment of maturity and the development of the individual (cited in Kind 1982).

Thus the ends to which psychotherapy has been applied are not only the treatment of disturbing symptoms (the so-called medical model of psychotherapy), but also the development of the individual in ways which go far beyond this, encompassing philosophical and educational values and concepts. That the sphere with which the provider of this type of care becomes concerned extends beyond the usual professional one is one of the special characteristics of psychotherapy, but also one of its dangers. Thus, in considering this category within the biopsychosocial classification of approaches to treatment, general principles of care take on a particular importance. What are the premises, characteristics and difficulties to which special attention is needed in psychotherapeutic treatment?

5.2.2 Task and Role of the Therapist

As the danger of overstepping boundaries is greater, the task of defining the working relationship between the therapist and the person seeking help before treatment begins is even more important here than with biological treatments. The therapist must view himself or herself not as a "paid friend" or as a secret partner in love or founder of a religion, but as an expert who applies conscience and knowledge as best he can to taking a position of trust and working on an assignment for a limited period of time. Training and competence must be equal to the treatment task. Where a social insurance scheme is paying for the treatment, there is an obligation not to burden this inappropriately.

5.2.3 Clarification of the Nature of the Relationship

Closely related to the above requirement is the need to establish reasonably clearly the nature of the relationship from each perspective: does the person seeking help see himself or herself as a patient, who expects from a medical therapist the alleviation of particular disturbances or disabilities, or is it more a matter of tasks such as finding meaning, maturing or getting help with unsuccessful interpersonal relationships, tasks which a non-medically qualified therapist can also undertake? Certainly, a holistic model of psychosomatic functioning makes these goals rather difficult to distinguish and might even finally remove the separation. However, in considering problems arising in the therapeutic relationship, it is helpful to devote some thought to this issue.

5.2.4 Clarification of the Indication and Choice of Method

To be able to answer the question of which method of psychotherapy should be applied for which disturbances, deficits, disabilities or needs, the therapist must have made a careful overall study of the working principles and theoretical foundation of the various methods. Obviously a few principles are effective elements in many methods (review by Grawe et al. 1994). However, experience suggests that some methods work better than others for particular disturbances, e.g. behaviour therapy for phobias and analytic treatment for long-term difficulties with relationships. In addition to factors relating to the patient's personality, the quality of the relationship between therapist and patient is of particular importance for the outcome of treatment.

The person seeking help has the right to be treated by a method which is recognised and established from the perspectives of usual practice in the therapist's profession and the current scientific consensus ("good clinical practice"). Where therapy ventures onto new territory, the therapist must inform the patient of this and seek his or her consent.

5.2.5 Clear Specification of the Setting

Clear specification of the setting includes arrangements about the place (out-patient practice, hospital unit, patient's home), form (out-patient, in-patient, individual or group therapy), duration and frequency of therapy and the context in which third parties will be involved (relatives and others with significant relationships with the patient). If these aspects of the framework for therapy are not firmly established, uncertainties arise in the therapist–patient relationship, and the danger of boundaries being overstepped (e.g. sexually) or of acting out behaviour becomes greater.

5.2.6 Awareness of Transference and Counter-transference

Not only in full psychoanalysis and psychoanalytically oriented treatment, but in all close, long-term therapist–patient relationships, feelings are evoked or reinforced whose direction and dynamics are influenced by past emotional experiences. It is an error of professional judgement to deny their existence or to underestimate their force. As discussed in Sect. 2, one of the obligations of the therapist is to use supervision and peer supervision to seek advice and help with crucial developments.

5.2.7 Clear Specification of Breaks and of the End of Therapy

Just because of the strength of the explicit or unspoken feelings which are evoked, breaks in treatment, e.g. during holidays, and the end of the therapeutic relationship are to be handled with care. Intensive weekend meetings such as take place with primal scream therapy, encounter groups or rebirthing are dangerous in that they discharge the participant, who is sometimes in an emotionally churned-up state, back into the everyday world with no follow-up care of any sort. Practical behavioural or relaxation methods with a strong educational orientation are less problematic in this respect.

5.3 Social Treatment

The deliberate use of environmental influences in the form of practical social psychiatry or specific milieu treatments has only really become prominent in this century. Before this, people did of course consider whether and in what way a particular environment might affect the development and care of the mentally ill. Isolation of the patient played a central role in this context.

5.3.1 Isolation

The view found in the work of Esquirol and his contemporaries is that the mentally ill should "be separated from his habits and usual way of life and from the people with whom he lives, and should be moved to an unknown place and entrusted to the care of people he does not know." This is a view which was accepted until quite recently and is derived from observation of the effects of pathogenic families and unfavourable living circumstances. In such an unfamiliar environment, the doctor could "more easily win the trust of the patient, in that he encounters him with no preconceptions". The author continues: "Thus consideration and experience indicate that isolation should be seen as a central method in the rational management of mental illness" (Esquirol 1869). Following the comparative study of three English psychiatric hospitals by Wing and Brown and their colleagues between 1960 and 1968 (Wing and Brown 1970), the damaging effects of years of isolation and social understimulation in often overcrowded asylums have finally been recognised. Social psychiatry, which developed as a reform movement in the post-war years, has thus worked to open the doors of psychiatric hospitals and to make these hospitals smaller in size and sometimes to close them completely, with the aim of preventing damage due to hospitalisation and returning the mentally ill to community-based treatment.

Recently, the negative effects of certain family environments for certain types of illness have been demonstrated. For example, the Expressed Emotion (EE) studies have shown the damaging effects on prognosis for rehabilitation of intrusive expression of feelings. Happily, however, this has not now led to the re-adoption of the isolation principle, but to work aimed at helping families to adopt a style of interaction less characterised by tension or to placement of patients in the care of open facilities such as day hospitals and therapeutic residential care.

What should be understood today as constituting social treatment? Are there particular methods which can be associated with this term? Clearly no interaction with mentally ill or mentally handicapped individuals and no physical or psychological treatment of any sort can happen in a complete vacuum – there is always a social milieu of some form. Making use of this milieu for treatment purposes (for a review, see Heim 1985) is a matter not so much of specific and clearly defined individual treatment methods or techniques as of overall reflection on the social context. Social treatment is a matter of learning or relearning social behaviour and social skills under the restrictions imposed by mental illness, vulnerability or disability.

5.3.2 Applications

In in-patient treatment, spheres where there are opportunities for practising such skills include the ward meeting, in which patients have the opportunity to practice conducting relevant arguments in the group and have a say in the life of the ward. In-patients may sometimes also take on particular responsibilities, strengthening their autonomy and feeling of responsibility, or they may participate in activities such as communal cooking and leisure activities, which provide training in co-operation and communication. Outside the in-patient unit, life in social care settings such as day facilities and residential care gives rise to many situations in which patients are given realistic practice in dealing with social demands (e.g. using public transport, shopping, interacting with people in official positions or job interviews). Thus initiative, social behaviour and awareness of reality are strengthened. There is currently a fluid boundary between these methods and the methods employed in behavioural therapies such as social skills training.

Social treatment reinforces the elements in the individual which are normal, well-regulated, general, everyday and healthy – that is the elements which are free and are not linked to illness. . . . The goal of all these efforts is to allow the patient to realise fully his potential in the social sphere (Dorner and Plog 1978).

Complex concepts concerning the realisation of these goals are to be found in the specialist literature on subjects including milieu therapy, occupational and social rehabilitation, occupational therapy, work training and support in employment, community-based psychiatry, work with family and network, voluntary work.

5.3.3 Treatment Principles

With such a broad set of goals, it is not easy to formulate general principles for social treatment. A few basic principles should be set out:

1. Social treatment is invariably relevant and is indicated as a basic part of every course of psychiatric treatment. No treatment takes place without a social context, but only a structured social milieu is therapeutically productive.
2. Mental illness and handicap lead to different patterns of limitations of social behaviour, which social treatment should take into account through clear targeting.
3. Group activity (community therapy) and, if desired, temporary retreat into solitude should be balanced so that they are appropriate for the individual and the illness.
4. Psychiatric work now takes place to a large extent in teams with members from various professions. Use of skills which are specific to a particular profession must be combined appropriately with the setting of joint goals, so as to avoid role diffusion or the setting up of parallel or competing treatment programmes.
5. Milieu treatment projects are sometimes in danger of degenerating into a separate pseudo-egalitarian world, and social treatment groups must therefore not lose contact with the surrounding reality to which the patient has to return.
6. Following the basic principle of "helping patients to help themselves", social treatments must not deprive patients of their right to make decisions. Care must be gradually and carefully balanced with allowing self-determination, so that each patient's potential for autonomy is realised.

6 Need for an Integrated Treatment Framework

The results of the rapidly advancing neurosciences, genetic studies which take into account environment and behaviour, and psychopharmacosocial treatment studies reinforce more and more the belief that biological, psychological and social factors interact much more closely than earlier evidence suggested (for reivews, see e.g. Gabbard 1995; Roth 1994; Roth and Menzel 1996). Thus environmental influences are of great significance for the expression of genes; the growth of synapses and the dendritic network of brain cells is a function of cognitive and emotional demands. Recurrent powerful external stimuli (psychostressors)

trigger psychopathological reactions of increasing frequency and severity, thus clearly lowering the vulnerability threshold (the so-called kindling-sensitisation concept; Post 1992):

The neuroscience revolution has now reached a level of sophistication that allows it to serve as a bridge between biology and the psychosocial environment: we are referring to the emerging understanding of the myriad ways in which even subtle changes in the environment can produce biological changes in the brain (Gabbard and Goodwin 1996).

Thus the division made above between types of psychiatric treatment - as mentioned, made mainly for purposes of clear discussion - becomes more and more questionable, and integrative thinking and treatment in psychiatry, as in modern psychosomatically orientated medicine as a whole, becomes indispensable.

6.1 Interactions Between Types of Treatment

It has long been known that there can be interactions between many therapeutic interventions, indeed between all forms of treatment. The great variety of possible interactions has, however, received insufficient attention, although their consequences are very significant. This will be illustrated with some examples.

Even at the initial interview, where the main aim is diagnosis, the foundations are established for a psychotherapeutic relationship if "consensual areas" can be discovered (sympathy can grow from empathy) and a readiness indicated to move into the "diagnostic-therapeutic cycle". The admission of acutely psychotic patients to in-patient treatment not only removes them from a family environment which may be pathologically highly charged (a "high EE" environment), but also restructures the course of their days and thus influences biorhythms such as sleeping pattern as well as capacity for social contacts. In psychotherapeutic relationships with clearly defined conceptual frameworks, the influences of other methods also have effects; thus psychoanalysis has elements of behaviour therapy, and behaviour modification "technologies" must also take account of motivational and psychodynamic factors. Neuroleptics subdue central nervous arousal and thus facilitate patients' efforts at cognitive restructuring and promote the coping abilities which they still retain. Even psychotropic medications are a convincing example of the principle of interaction of biological and psychological influences; in his critique of modern pharmacotherapy, Blankenburg (1982) rightly points out that a sharp distinction cannot always be drawn between the "purely pharmacological", chemically and physically understood "efficacy" of a drug and the much broader "effectiveness" of a medication, to which, additional placebo effects, for instance, may contribute. A further question to be considered is

. . . how far counter-regulation, which occurs in an organism to which a particular substance is administered should be taken fully into account in assessing the 'efficacy' of that organism. In prophylaxis against relapse in schizophrenic psychosis the combination of neuroleptics and social skills training has proved to be clearly more effective than the administration of medication alone. One would also suspect that the amount of medication required in programmes which aim at low neuroleptic dosages provided in combination with psychoeducational interventions would vary with the intensity of a concurrent family therapy intervention, and that variations in requirement for medication might also show clear differences related to premorbid adjustment and prevailing family atmosphere (Blankenburg 1982).

Thus, from these considerations, one may conclude that the effects of the simultaneous use of a variety of types of intervention are not simply additive, but that the whole (a therapeutic system) is more than the sum of its parts (the treatment components). A systemic way of thinking therefore leads to new insights and possibilities for therapeutic improvements, a subject which cannot be further discussed within the scope of this chapter.

6.2 General Treatment Principles

An overall general principle of treatment which should be promoted is that we should convert our knowledge about narrowly defined biopsychosocial factors which influence human development into a multi-dimensional, integrated therapeutic framework. So that this ambitious project does not lead to conceptual confusion and fuzzy classification, it seems necessary, as before, to distinguish between different levels of treatment and interpretation. However, the artificially separated treatment components must always be brought together in the context of the fundamental biopsychosocial model and considered jointly. Psychiatrists' training should already have equipped them with the knowledge required to manage such a synthesis. In this process, they must beware of the dangers of grandiosity as much as those of biological and psychological reductionism.

7 References

an der Heiden U (1992) Chaos in health and disease – phenomenology and theory. In: Tschacher W, Schiepek G, Brunner EJ (eds) Self-organization and clinical psychology. Springer, Berlin Heidelberg New York, pp 55–87

Andreasen NC (1996) Body and soul. Am J Psychiatry 153: 5

Benedetti G (1959) Wandlungen des Menschenbildes in der Psychiatrie. Schweiz Med Wochenschr 89/29: 75–755

Bertalanffy L (1974) General system theory and psychiatry. In: Arieti S (ed) American handbook of psychiatry, 1st edn, vol 3. Basic, New York, pp 705–721

Blankenburg W (1982) Kritik der modernen Pharmakotherapie. In: Tellenbach H (ed) Psychiatrische Therapie heute. Enke, Stuttgart, pp 89–119

Böker W (1986) Krankenhaus (psychiatrische). In: Müller C (ed) Lexikon der Psychiatrie. Springer, Berlin Heidelberg New York, pp 410–414

Böker W (1990) Patient, Angehörige und Arzt auf dem Weg zu einer Behandlungspartnerschaft – Kasuistik eines 19-jährigen schizoaffektiven Krankheitsverlaufs. Nervenarzt 61: 65–568

Böker W (1992) Möglichkeiten partnerschaftlicher Stabilisierungsarbeit mit Schizophrenen im Lichte der Bewältigungsforschung. In: Brenner HD, Böker W (eds) Verlaufsprozesse schizophrener Erkrankungen. Dynamische Wechselwirkungen relevanter Faktoren. Huber, Bern, pp 280–284

Böker W, Brenner HD (1996) Stand systemischer Modellvorstellungen zur Schizophrenie und Implikationen für die Therapieforschung. In: Böker W, Brenner HD (eds) Integrative Therapie der Schizophrenie. Huber, Bern, pp 17–32

Brenner HD (1989) Behandlung schizophrener Psychosen. Enke, Stuttgart

Ciompi L (1986) Auf dem Weg zu einem kohärenten multidimensionalen Krankheits- und Therapieverständnis der Schizophrenie: konvergierende neue Konzepte. In: Böker W, Brenner HD (eds) Bewältigung der Schizophrenie. Huber, Bern, pp 47–61

Degkwitz R, Hoffmann SO, Kindt H (1982) Psychisch krank – Einführung in die Psychiatrie für das klinische Studium. Urban and Schwarzenberg, Munich

Dorner K, Plog U (1978) Irren ist menschlich oder Lehrbuch der Psychiatrie/Psychotherapie. Psychiatrischer Verlag, Wunstorf

**Engel G (1970) Psychisches Verhalten in Gesundheit und Krankheit. Ein Lehrbuch für Aerzte, Psychologen und Studenten. Huber, Bern

Esquirol JED (1816) Folie, dictionnaire des sciences médicales. (German translation published in 1968: Von den Geisteskrankheiten. Huber, Bern)

Gabbard GO (1995) Mind and brain in psychiatric treatment. In: Gabbard GO (ed) Treatments of psychiatric disorders, 2nd edn, vol 1. American Psychiatric Press, Washington, DC, pp 21–34

*Gabbard GO, Goodwin FK (1996) Integrating biological and psychosocial perspectives. In: Dickstein L, Riba M, Oldham J (eds) Review of psychiatry, vol 15. American Psychiatric Press, Washington, DC, pp 527–548

Gaebel W (ed) (1995) Qualitätssicherung im psychiatrischen Krankenhaus. Springer, Berlin Heidelberg New York

Garcia-Guillen D (1982) Diaita im frühen Christentum. In: Tellenbach H (ed) Psychiatrische Therapie heute. Enke, Stuttgart, pp 12–30

Glatzel J (1975) Antipsychiatrie. Fischer, Stuttgart

Grawe K, Konati R, Bernauer F (1994) Psychotherapie im Wandel. von der Konfession zur Profession. Hogrefe, Göttingen

Grieser W (1861) Die Pathologie und Therapie der psychischen Krankheiten, 2nd edn. Krabbe, Stuttgart

Haug JH, Stieglitz RD (eds) (1995) Qualitätssicherung in der Psychiatrie. Enke, Stuttgart

Heim E (1985) Praxis der Milieutherapie. Springer, Berlin Heidelberg New York

Heimann H (1991) Die Psychiatrie am Ende des 20. Jahrhunderts. In: Ciompi L, Heimann H (eds) Psychiatrie am Scheideweg – was bleibt? Was kommt? Springer, Berlin Heidelberg New York, pp 115–124

Kind H (1982) Psychotherapie und Psychotherapeuten – Methoden und Praxis. Thieme, Stuttgart

Lain Entalgo P (1982) Der Sinn der Diaita in der Antike. In: Tellenbach H (ed) Psychiatrische Therapie heute. Enke, Stuttgart, pp 1–11

Lidz CW, Meisel A, Zernbavel E, Carter M, Sestak RM, Roth LH (1984) Informed consent. A study of decision making in psychiatry. Guilford, New York

Ludewig K (1992) Systemische Therapie. Grundlagen klinischer Theorie und Praxis. Klett-Cotta, Stuttgart

Maturana HR (1982) Erkennen: Die Organisation und Verkörperung von Wirklichkeit. Vieweg, Wiesbaden

*Maturana HR, Varela FJ (1987) Der Baum der Erkenntnis. Wie wir die Welt durch unsere Wahrnehmung erschaffen – Die biologischen Wurzeln des menschlichen Erkennens. Scherz, Bern

Möller HJ, Engel RR, Hoff P (eds) (1996) Befunderhebung in der Psychiatrie: Lebensqualität, Negativsymptomatik und andere aktuelle Entwicklungen. Springer, Berlin Heidelberg New York

Nitsche P (1929) Allgemeine Therapie und Prophylaxe der Geisteskrankheiten. In: Bumke O (ed) Handbuch der Geisteskrankheiten, vol IV. Allgemeiner Teil IV. Springer, Berlin Heidelberg New York, pp 1–131

Nuechterlein KH, Dawson ME (1984) A heuristic vulnerability/stress model of schizophrenic episodes. Schizophr Bull 10: 300–312

Perris C (1989) Cognitive therapy with schizophrenic patients. Guilford, London

Post RM (1992) Transduction of psycho-social stress into the neurobiology of recurrent affective disorder. Am J Psychiatry 149: 999–1010

Prigogine J, Stengers J (1981) Dialog mit der Natur. Neue Wege natuwissenschaftlichen Denkens. Piper, Munich

Roth G (1994) Das Gehirn und seine Wirklichkeit. Kognitive Neurobiologie und ihre philosophischen Konsequenzen. Suhrkamp, Frankfurt/Main

Roth G, Menzel R (1996) Neuronale Grundlagen kognitiver Leistungen. In: Dudel J, Menzel R, Schmidt RF (eds) Neurowissenschaft – vom Molekül zur Kognition. Springer, Berlin Heidelberg New York, pp 539–559

Schipperges H (1962) Lebendige Heilkunde. Von großen Ärzten und Philosophen aus drei Jahrtausenden. Olten, Freiburg

Schmitt W (1982) Das Regimen sanitatis des Mittelalters. In: Tellenbach H (ed) Psychiatrische Therapie heute. Enke, Stuttgart, pp 51–63

Schrenk M (1973) Über den Umgang mit Geisteskranken. Springer, Berlin Heidelberg New York

Tellenbach H (ed) (1982) Psychiatrische Therapie heute – Antike Diaita und moderne Therapeutik. Enke, Stuttgart

*Uexküll T von (1986) Psychosomatische Medizin, 3rd edn. Urban and Schwarzenberg, Munich

Willi J (1985) Koevolution. Die Kunst gemeinsamen Wachsens. Rowohlt, Hamburg

*Wing JK, Brown GW (1970) Institutionalism and schizophrenia–comparative study of three mental hospitals 1960–68. Cambridge University Press, London

Wolberg LR (1967) The technique of psychotherapy, 2nd edn. Grune and Stratton, New York

**Zubin H, Spring B (1977) Vulnerability – a new view of schizophrenia. J Abnorm Psychol 86: 103–126

CHAPTER

10

G. Andrews

Evaluation of Psychiatric Treatments

1 Introduction

In an ideal world, no new treatments should be introduced into medicine until they have been shown to be more effective, safer or less expensive than existing treatments (Cochrane 1971). In the real world, once new treatments have been shown to be effective, they are quickly promoted by their makers. The methods of the marketplace, rather than scientific argument about effectiveness, safety and cost, are used to encourage clinicians to adopt the new treatments. Before 1936, psychiatry, like much of general medicine, had few treatments that were specifically effective. Even so, the advice of clinicians was highly prized – at best, people did well with sage advice and encouragement, and at worst, people with serious disorders could be shown how to lessen the impending tragedy, while people with disorders likely to remit could be shown how to manage until the remission occurred. Even today, when we have proven and effective treatments for most mental disorders, providing considered advice about prognosis and the management of disability is an important part of medical practice. Skill in providing good clinical care (Andrews 1993), especially care for disorders for which there are, as yet, no proven or specific treatments, distinguishes good clinicians from those who are only mediocre. However, this chapter is not so much about such non-specific good clinical care. It is more about identifying specific treatments which have been proven to be effective for particular disorders and then about how to incorporate those treatments into good clinical practice.

2 Identifying Effective Treatments

A treatment is a procedure, which may be pharmacological or psychotherapeutic, prescribed to a patient in the belief that it will lessen the patient's symptoms, disability or pathology due to a mental disorder. The problem lies in deciding which treatments are of specific benefit and which are merely palliative. If a new treatment were shown to be effective in an open study with a large number of carefully diagnosed subjects who were typical of people with that disorder, would that be sufficient evidence? If the subjects were randomly allocated either to the new treatment or to treatment with placebo or standard treatment, and symptoms and disablement were reliably measured at the end of treatment and the new treatment were still superior, would that evidence be more satisfactory? If, in addition, the new treatment were sufficiently specified so that others could either make or obtain the compound, or replicate the procedure, and the study had been independently replicated, would that be sufficient evidence? There certainly would be less argument, particularly if the superiority of the new treatment over placebo was substantial. Yet there is always discussion about the effectiveness of a new treatment, simply because the above criteria are seldom adhered to. The essential steps in investigating a new treatment are discussed below.

2.1 Specification of the Disorder to Be Treated

No treatment is good for all disorders. The first step in evaluating a treatment is to specify which type of disorder and patient for whom this treatment is designed. Usually this is specified in terms of the *International Classification of Diseases*, Tenth Revision (ICD-10) or the *Diagnostic and Statistical Manual of Mental Disorders*, Fourth Edition (DSM-IV), but sometimes a greater specificity is indicated. For example, the evidence about clozapine comes from trials with patients with schizophrenia who have not responded to other treatments. It is important that studies specify the characteristics of the people treated, and it is equally important that clinicians do not generalise beyond the evidence. For example, the studies that show that paroxetine is of benefit in F32.1 and F32.2 (moderate and severe depressive episodes) do not automatically mean that it will be of benefit for depressive symptoms occurring in F41.2 (mixed anxiety and depressive disorder). Paroxetine might be effective in mixed and depressive disorder, but these studies do not demonstrate this. Studies should therefore specify the reliability and validity of the diagnostic process. "Patients with depression seen at the clinic" is not enough. Clinical judgment, plus evidence from a structured diagnostic interview or rating scale, would be preferable. For instance, "patients who met criteria for moderate and severe depression, both clinically and according to the Composite International Diagnostic Interview (CIDI)" would be sufficient.

2.2 Specification of the Elements of Treatment

If a report of an effective treatment is to inform clinicians, it must specify both the characteristics of the new treatment and the implementation regime. If the new treatment is a medication, then the compound or components of the medication must be specified, together with the levels of purity and the form of the

final preparation. In most countries, this level of specification is required by the licensing authority before the medication can be marketed. Non-drug treatments, such as psychosocial interventions and other psychotherapies, are not subject to such constraints, and it is essential that the procedures to be undertaken are fully specified. "The therapist should facilitate the resumption of growth and increase the understanding of self and other . . .", for instance, does not specify what the therapist should actually do. Many psychotherapies, especially the cognitive behavioural therapies, are now fully specified in manuals which describe exactly what the therapist should say, e.g.: "As part of this programme, we are going to examine the way you think about yourself and others. It is likely that you have some unrealistic and therefore unhelpful beliefs which . . . ." Only if a treatment is fully specified can it be replicated by others. Treatments which are effective, but unspecifiable must necessarily die with their originators.

2.3 Specification of How the Treatment Was Administered

Specification of the elements of the treatment is only one part of the story. Each description of a treatment must also specify the process of implementation: dose, route and frequency for medication, frequency and duration of sessions for a psychotherapy. It is also important to specify the setting of treatment and the use of ancillary treatments, e.g. "prescribed as part of a comprehensive in-patient rehabilitation programme" is clearly different to "prescribed monthly in outpatients". Specification has two purposes: (1) to allow others to use a proven treatment with their patients and (2) to allow the elements in the treatment to be identified if there is discussion about which element in the programme is the essential element.

2.4 Measuring the Change in Symptoms

The purpose of treatment is to reduce symptoms and complaints, reduce disability and improve the quality of life and, if one is really fortunate, reduce risk factors so that relapse becomes less likely. Measures need to be applicable, reliable, valid and sensitive to change. A list of common measures that meet these criteria is displayed in Table 1.

2.4.1 Applicability

The measures chosen must meet the broad aims of the study; certainly all clinically relevant outcomes must be measured and reported. Thus a study of depression needs to use a symptom measure that identifies the symptoms of depression, and a general symptom measure alone will not suffice. If the study is a long-term study, then the number of people who commit suicide should also be reported, because depressive symptoms and suicide would both be clinically relevant outcomes. A study in which the primary aim is the maturation of personality needs to use measures of personality maturity, and again a general measure of well-being will not suffice if it is the only measure used. Measures should be matched to the aim of the treatment. While symptom measures must be specific to the diagnosis of interest, disability and quality of life measures can be more general, applicable to all mental disorders or even to all disorders.

2.4.2 Reliability

A measure should be a reliable measure of the characteristic being measured, i.e. it should return the same value when re-administered on two occasions or by two different people to subjects whose status has not changed. The more reliable a measure is, the better. There are a number of ways of measuring reliability. *Internal consistency* refers to the assumption that items in a scale should be measuring the same general phenomenon and should be correlated. Cronbach's alpha is the statistic used, and values greater than 0.8 are desired. *Inter-rater reliability* refers to the consistency with which two raters record the information being elicited at one interview. Agreement should be close to perfect, and the kappa statistic should be close to 1.0. *Test-retest reliability* refers to the consistency of the scores on two administrations at different times when there has been no change in the state of the individual being measured and is also estimated using the kappa statistic. Test-retest reliability is always less than the inter-rater reliability, because it includes that element, and values of 0.8 are to be desired.

2.4.3 Validity

Validity refers to the extent that a measure accurately measures what is being measured. Reliability sets an upper limit on the validity of a measure and if, for instance, the kappa statistic for test-retest reliability is 0.8, the kappa statistic for validity must be less than this. Estimating validity should be quite simple: we merely have to estimate the degree to which the measure correlates with some perfect measure of the characteristic of interest. Unfortunately, there are very

Table 1. Outcome measures

Title	Time (min)	Description	Reference
Beck Anxiety Inventory	10–40	Self-rated questionnaire 21 symptoms, rated on a 4-point scale; symptoms covered include: numbness or tingling, unable to relax, unsteady, terrified, hands trembling, fear of dying, indigestion	Beck et al. (1988)
Beck Depression Inventory	5–10	Self-rated questionnaire 21 symptoms rated on a 4-point scale, with behavioural anchors provided; symptoms covered include: pessimism, sense of failure, self-dissatisfaction, guilt, self-dislike, suicidal ideas, social withdrawal, indecisiveness, body image change, insomnia, fatigability, weight loss, somatic preoccupation, loss of libido	Beck et al. (1961)
Behaviour and Symptom Identification Scale	20–30	Interviewer-administered interview, or self-administered questionnaire 32 items with five subscales: relation to self and others, daily living and role functioning, depression and anxiety, impulsive and addictive behaviour, psychosis	Eisen et al. (1994)
Brief Psychiatric Rating Scale	18	Clinician-rated scale following interview 16 symptom constructs rated on a 7-point scale: somatic concern, anxiety, emotional withdrawal, conceptual disorganisation, guilt, tension, mannerisms and posturing, grandiosity, depressive mood, hostility, suspiciousness, hallucinatory behaviour, motor retardation, uncooperativeness, unusual thought content, blunted affect	Overall and Gorham (1962)
General Health Questionnaire	6–8	Self-rated questionnaire Initially developed with 60 items covering four areas: depression, anxiety, objectively observable behaviour, hypochondriasis. Revised versions have 30, 28, 20 and 12 items	Goldberg (1972)
Health of the Nation Outcome Scales	15–30	Clinician-rated rating scale 12 scales: aggression, self-harm, alcohol and drugs, memory/orientation, physical problems, mood disturbance, hallucinations and delusions, other mental, social relationships, social environment (housing and finance), overall severity	Wing (1994)
Hamilton Anxiety Rating Scale	15 (estimate)	Clinician-rated semi-structured interview 14 symptoms rated as none (0) to severe (4): anxious mood, tension, fears, insomnia, intellectual, depressed mood, somatic (muscular), somatic (sensory), cardiovascular, respiratory, gastrointestinal, genito-urinary, autonomic, behaviour at interview	Hamilton (1959)
Hamilton Depression Rating Scale	20 (estimate)	Clinician-rated semi-structured interview 21 items with 3-, 4- and 5-point scales used; behavioural anchors provided; symptoms covered include: feelings of guilt, suicide, insomnia (early, middle, late), agitation, genital symptoms, loss of weight	Hamilton (1967)

Medical Outcomes Study Short Form-36	5–10	Self-rated questionnaire (can also be given to an informant or administered at an interview) Eight multi-item scales: physical functioning, physical and emotional role limitations, bodily pain, mental health, social functioning, vitality, general health perceptions, reported health transition	Ware and Sherbourne (1992)
Mental Health Inventory	10–15	Self-rated questionnaire 38 items with five subscales: anxiety, depression, behavioural/emotional control, general positive affect, emotional ties	Veit and Ware (1983)
Role Functioning Scale	Few minutes	Clinician-rated rating scale Four single scales: working, independent living and self-care, immediate social network relationships, extended social network relationships	Goodman et al. (1993)
Scale for the Assessment of Negative Symptoms	20–30	Clinician-rated rating scale 24 items with five subscales: alogia, affective flattening, avolition-apathy, anhedonia-asociality, attentional impairment	Andreason (1982)
Scale for the Assessment of Positive Symptoms	20–30	Clinician-rated rating scale 34 items with four subscales: hallucinations, delusions, bizarre behaviour, formal thought disorder	Andreason and Olsen (1982)
State-Trait Anxiety Inventory	10–20	Self-rated questionnaire Two scales with 20 items each: state anxiety (A-state) and trait anxiety (A-trait)	Spielberger (1983)
Symptom Checklist 90 (revised)	10–20	Self-rated questionnaire 90 items covering nine primary symptom dimensions: somatisation, obsessive compulsive, interpersonal sensitivity, depression, anxiety, hostility, phobic anxiety, paranoid ideation, psychoticism; three global indices: lobal severity index, positive symptom distress index, positive symptom total	Derogatis (1977)
WHO Psychological Impairments Rating Schedule	5–10 after interview	Clinician-rated rating scale 97 items covering ten areas: psychic tempo, attention, fatigability, initiative, communication by facial expression and by bodily language, affect display, conversation skills, self-presentation, cooperation. Item behaviours rated as absent, present, or present in a more severe form. Cost unknown; extensive training required	Jablensky et al. (1980)

few perfect or "gold standard" measures, and one must compare the measure with the best of the other measures. Ideally, the correspondence between a measure and the "gold standard" for estimating the characteristic should be very high; however, the validity of a measure will be limited by unreliability in both the criterion measure and the present measure.

Clinical judgment is often used as the gold standard for psychiatric measures, but this is inherently unreliable and kappa statistics for this type of validity are often quite low. We must remember that if two clinicians disagree, one or both must be wrong; both cannot be right. *Face validity* refers to whether the questions on a scale look as though they are measuring what they are supposed to be measuring (this is really a component of acceptability, not validity). *Content validity* refers to the extent that the measure covers the domain of information to be measured adequately. Agreement with a panel of experts is one method for assessing this characteristic. *Criterion-related validity* refers to whether or not the measure is able to predict scores on an independent and direct measure of the characteristic. *Construct validity* refers to the extent to which the measure may be said to estimate a theoretical construct or abstract attribute such as health or quality of life. Demonstration of construct validity depends on the accumulation of information from a number of sources, both those measuring the same construct (convergent validity) and those measuring different constructs (discriminant validity).

2.4.4 Sensitivity to Change

Measures used to identify effective treatments must be capable of registering change due to treatment in the target symptoms or disorder. Measures with multiple response options are more likely to be sensitive to change, but there are two additional issues, i.e., whether the measure covers the new range of scores and whether the characteristic has had time to change. For example, measures designed to cover severe psychosis commonly fail when the person has recovered, because there are no questions to describe being well. Secondly, different characteristics take different times to change, and the timing of measurement must allow for this. Symptoms may change within a week and can be measured and show change after that time. Changes in disablement require that the patient has time to experience the lessened handicap and cannot be measured for some time after the improvement in symptoms has taken place. Changes in risk factor measures take even longer to become manifest.

2.5 Ensuring that Improvement Is Due to Treatment

One might think that if reliable and valid measures showed a significant change in symptoms when a well-specified treatment was applied to a carefully diagnosed group of patients, then it could be safely assumed that the treatment was responsible for the improvement. It might be, but we cannot be certain, because three other factors that are very likely to be present can mimic the beneficial effects of a treatment. These are spontaneous remission, regression to the mean and the placebo response.

Some disorders run a phasic course. Major depression is a good example, and symptoms remit spontaneously some months after the onset of the disorder. As people seldom come for treatment until the disorder has been present for a few months, remission can be expected within some months whatever is done.

Regression to the mean is a special variety of the spontaneous remission phenomenon and applies to disorders that are chronic and do not remit. Even in such disorders, the severity of symptoms fluctuate, being worse at times and better at times. People try to manage their chronic diseases as best they can. When the symptoms have lessened in intensity, they are less likely to seek help, being more likely to go for help when their symptoms seem to have worsened. Thus, among patients who consult doctors, even if no treatment is given, it is more likely than not that the period of severe symptoms will be followed by a return to the average level of symptoms or even a move to mild symptoms. In both these circumstances, when spontaneous remission is likely or regression to the mean probable, the response may mimic a true treatment response. "Waiting list" and "no treatment" control groups can measure the change due to these phenomena, i.e., the spontaneous remission occurring in remitting conditions and regression to the mean occurring in chronic conditions.

"Waiting list" and "no treatment" control groups do not, however, control for another important confounding factor known as the placebo effect. When people believe they are being treated by an effective remedy, when in fact they are being treated by an inert remedy, a response related to the expectation of getting better may mimic the effect of treatment. People in treatment are encouraged to capitalise on each improvement trend and to minimise each setback. Improvement is experienced as progress due to treatment and setbacks as mere random phenomena of minor import. Placebo-controlled trials, in which one group is treated by a treatment which the experimenters but not the patients

know to be inert, can estimate the extent of both the placebo effect and the other two phenomena, spontaneous remission and regression to the mean. In some disorders, such as obsessive compulsive disorder and schizophrenia, the magnitude of these non-specific treatment effects is small, while in other disorder, such as depression or generalised anxiety disorder, the changes seen in the placebo group can be quite considerable and equal in magnitude to the changes resulting from the specific treatment itself.

When a group of patients are treated with any effective treatment, the resulting improvement can be decomposed into two elements: that due to the non-specific effects of being in treatment (i.e. to the result of spontaneous remission, regression to the mean and the placebo effect) and that due to the specific power of the treatment to lessen the severity of the disorder. Thus the observed clinical change in any patient who is being effectively treated will be greater than that due to the specific treatment being used, although in clinical practice it is conventional to attribute the change to the specific treatment being used and to ignore the contributions that the non-specific effects of being in treatment may be making. Sometimes, as when carrying out cost–benefit calculations, for example, it is helpful to distinguish the two components, and this is another reason why placebo-controlled trials are so helpful.

There is another element of treatment that is helpful to define. Some treatments are designed to relieve or control symptoms. Antipsychotic drugs in schizophrenia would be a good example. Treatment is thus designed to be long term and should be continued throughout the natural history of the disorder. Other treatments actually seem to end the disorder, e.g. graded exposure in agoraphobia, and treatment can be discontinued once the patient is better because they will then stay better in the absence of any contact with the therapist. Follow-up studies in agoraphobia show stability of improvement for 5 and 8 years following behaviour therapy, as though the person were cured. Cure, of course, not only means that the symptoms are relieved and disability lessened, but also that the person's risk of once again developing the condition is similar to that of the normal population. Follow-up measurement should thus specify whether the measurements were made while treatment was continuing or after treatment had been concluded. Again, such a differentiation between the control and cure effects of treatment is important in costing studies, because treatments that cure are intrinsically less expensive in the longer run, even if they might appear to be more expensive in the shorter term.

2.6 Randomised Controlled Trials

We began this chapter with the idea that no new treatments should be introduced into medicine unless they had been shown to be more effective, or as effective but safer and cheaper, than existing treatments. The randomised, controlled trial is the key to this set of decisions. Conceptually very simple, such trials are very difficult to conduct once they involve more than the short-term comparison of one drug against another. It is essential that such trials be conducted on all specific treatments used in psychiatry, and this is increasingly being done. They should also be conducted on all treatment programmes, but these are very seldom tested (see Kluiter and Wiersma 1996).

The idea behind a randomised, controlled trial is that a group of people with a specific disorder are carefully assessed and, if deemed suitable for treatment, they are informed that there are two possible treatments: usually one that has been in use for some time and a new treatment that is believed to be equally safe, but is possibly superior. They are asked to consent to either treatment and are randomly assigned to one or the other. They are then measured again prior to treatment commencing and at the end of treatment if treatment is believed to "cure" or when the effect of treatment is deemed to be maximal if the treatment is believed to control symptoms. Finally, they are measured at follow-up. Ideally, all people who entered the trial are followed up, even those who may have dropped out of treatment for one reason or another. The difference in scores of the two groups at the end of treatment or at follow-up are compared. Any difference is attributed to the differences in the two treatments, as all other sources of variation are believed to be controlled.

It is ethically unsound not to determine the efficacy of a treatment or treatment programme. If public or private money is being used to support it, and if people who are sick are accepting it on the basis that it is the best treatment available, then it must be evaluated, preferably by a randomised controlled trial against placebo or other treatment. Awkward as it may be, clinicians have a duty to evaluate all unproven treatments unless the evidence is as compelling as the examples listed below (see Sect. 2.7). Nevertheless, many proponents of unusual therapies claim that their therapies are so distinctive that unbiased assessment is impossible. Evaluation of those therapies will therefore have to depend on patient self-reports, not so much of symptoms, but of a lessening in the disability and handicap in those receiving the treatment as against

those receiving treatment as usual. There are no treatments that should be exempt from scientific scrutiny.

Getting patients to agree to enter trials is often very slow and time-consuming. Most have been referred for specialist treatment, and the fact that they might now "be experimented upon" is of real concern to them. Sometimes inducements are offered – treatment is made free or immediately available – but the effect of inducements is to narrow the type of people entering the study to those that cannot afford to go elsewhere. There is no easy answer, one can only present the truth in either of the two scenarios. The patient information for the randomised, comparison-controlled trial should state that there is a proven treatment and another about which there is reason to believe that it will prove superior (otherwise the trial would not be carried out), although it cannot be guaranteed that the effect will not be the same or worse than the standard treatment. In contrast, the patient information for a randomised, placebo-controlled trial should state that there is no proven treatment for the patient's disorder but that there is reason to believe that a new treatment might work, but again it cannot be guaranteed that the new treatment will be better than a continuation of good clinical care.

The purpose of randomisation is simply to ensure that all factors likely to influence outcome are equally distributed across the two treatment groups. As we cannot know which factors might prove of prognostic import with a new treatment, people are randomised without knowledge of their characteristics or the characteristics of their illness. While the idea is simple, the process is complicated, and tossing a coin is not enough. Experimenters have been known to bias the allocation of people with a good prognosis to the favoured treatment (Sackett et al. 1985). Thus it is essential that randomisation is carried out by a process that does not allow the allocation to treatment group to be biased by the experimenter. When the trial is complete, it is necessary to demonstrate that randomisation resulted in groups that were equivalent in relevant characteristics; otherwise, in the final calculations, some statistical corrections for the imbalance will be necessary, even if this method is not as satisfactory.

If experimenters with full knowledge of the allocation to group are asked to complete rating scales, it is very difficult for them not to be swayed in the direction of their hypothesis. Judges completing open rating scales should therefore be blind as to which patient is allocated to which group; neither therapists nor patient should know which treatment the patient is receiving, and raters of the outcome should not know either. It is possible to achieve this level of blinding in short-term drug trials, whether comparison or placebo-controlled, but impossible to manage the latter steps when the treatment involves a psychological or social intervention. Having unblinded raters who are ignorant of the purpose of the study is one solution, but the more usual solution is to use highly structured patient questionnaires and interviews that minimise the subjective element in the ratings. If allocation to group is truly random, then the only information required is the status of each patient at the end of treatment, although it is usual to measure prior to randomisation, when treatment begins, at the end of treatment and at follow-up. The pre-treatment information is valuable should randomisation not be effective and some correction by analysis of covariance be required.

As both groups are being treated, one by the experimental treatment and one by placebo or standard treatment, and as both groups are being measured at the same time after treatment begins, then spontaneous remission, regression to the mean and the placebo response should be the same in both groups. However, it is very easy for the patients in the control groups to realise that their treatment is not deemed important by the experimenter and to reduce their expectation of improvement. Unless the identity of the experimental treatment is clearly masked, as in a placebo-controlled drug trial, it is essential that each patient's perception of the adequacy of treatment be measured to identify any such bias. When programmes of treatment are being trialled, such as in-patient versus day-patient care, the views of the staff can be very influential in controlling the adequacy and the perceived adequacy of treatment. It is therefore important to ensure that treatments are carried out as intended. It is usual to measure this by tape recordings or ratings of the process of treatment. While it is important to specify treatments, it is equally important to demonstrate treatment integrity.

The analysis of a randomised, controlled trial is relatively simple. Having decided before the study begins which is the key outcome measure, the simple question is whether the scores on that measure are significantly different in the two groups. The probability of a significant finding depends on three things: the size of the difference, the variability of the scores and the number of subjects in the trial. No study should be commenced unless a power analysis shows that there is at least an 80% chance of showing a significant difference ($p < 0.05$) given the expected effect size in standard deviation units (SD). The effect size is the difference between the group mean scores at the end of treatment divided by standard deviation of the control group scores. If there are large numbers in the trial, then it is likely that the difference will be

significant even if it is of no clinical utility – essentially a false-positive finding. If the numbers in the trial are small, there may be a clinically significant difference between the groups that does not reach significance because of lack of power – essentially a false-negative finding. The relation between the number of subjects required in a study and the effect size is curvilinear; only seven subjects per group are required if the effect size is likely to be 1.5, 14 subjects if the effect size is 1.0, and more than 50 subjects per cell if the effect size is likely to be 0.5 SD. The clinical significance of a measure also depends on the nature of the measure. A one-third standard deviation difference in scores on a measure of depression is probably not clinically significant, but the same change in suicide rate would be. Lastly, all the subjects who were randomly allocated should be included in the analysis – the intention to treat rule – simply because drop-outs are more likely to occur when the treatment is more onerous or less immediately effective and hence, if ignored, would inflate the result associated with the less satisfactory treatment.

2.7
In the Absence of Randomised Controlled Trials

Sometimes a trial is not necessary to know that a treatment is effective. Streptomycin in children with tuberculous meningitis eliminated certain death, and no trial was required. Few improvements are so obvious, but this example does highlight the obvious issue: if the effect of treatment is sufficiently strong for all to see, it is unlikely to be due to faulty measurement or to the other confounding factors. Many new treatments are first evaluated by doing an open trial in which patients with a defined illness are given the treatment and measured at the beginning and the end of treatment. This will demonstrate whether the improvement might be statistically or clinically significant. If it is, then examination of the magnitude of the placebo effect, spontaneous improvement and regression to the mean that are likely in this disorder can be judged from other trials. After subtracting the placebo effect size from that obtained in the pre-post trial, it will be obvious whether the change due to the specific treatment is still likely to be clinically significant. It is then that the investigator should proceed to a randomised, controlled trial.

There are three circumstances in which evidence from open trials in a disorder should be taken seriously. The first is if there is a substantial body of evidence from pre-post trials in which the change due to treatment is large while the evidence from other trials makes it likely that the response to placebo is small. As always, independent replication of results reduces the likelihood that selection, measurement or experimenter biases could account for the positive result. Behaviour therapy for obsessive compulsive disorder was almost totally reliant on evidence from pre-post trials, yet the efficacy of this treatment is well accepted. Fortunately, there is now evidence from randomised controlled trials that supports the extensive data from the open trials. The second attribute is a dose–response curve in which small doses or hours of treatment produce only a little improvement and increasing improvement is associated with increasing doses or hours of treatment. Theoretically, the placebo response should also show a dose–response curve, but there seems to be a plateau after a certain time with no further benefit thereafter. The third circumstance in which open trial data can be convincing is when there is replicated evidence of the effectiveness of treatment from baseline pre-post trials with a long follow-up, i.e. trials in which the duration of the baseline and follow-up are both many times longer than the period of treatment. Such long baseline and long follow-up trials are common in cognitive behaviour therapy and are respected because a long baseline serves to measure the effect of spontaneous remission and regression to the mean, while the long post-treatment follow-up without continued contact with the therapist means that any sustained improvement is unlikely to be due to a placebo response. Nevertheless, despite this type of evidence, critics will only finally be convinced by evidence from randomised placebo- or comparison-controlled trials.

3
Aggregating the Results

3.1
Systematic Reviews of the Literature

There are three traditional sources of information when reviewing the evidence of treatment effectiveness: information from experts, from current practice and from treatment trials. All sources can be valuable. When all else fails, the opinions of experts are very valuable in delineating what is the best practice and in identifying the shortcomings in research. Provided the experts are representative of the profession, then their considered opinion is important; "three wise men" have always been better than a single ill-considered opinion. In like vein, a careful delineation of what is current practice among the majority of clinicians is likely to be more balanced that the view of zealots. Nevertheless, despite the value of expert opinion or a consensus among practitioners, there is a profound respect for the evidence represented by the

scientific literature. Sometimes this respect can be misplaced.

Reviews of the literature have very low status. They are merely expressions of opinion supported by a selective review of the literature. A systematic review of the literature defines how the articles were identified, the basis on which they were selected for review and the system used for collating their findings. Sometimes a complex field can be simplified if attention is confined to replicated findings. This can be further systematised if the degree of replication is identified. For example, multiple replications in independent clinics with no negative findings are more convincing than independent replications with some negative findings (the reviewer should seek to explain the differences), which are in turn more likely to be stable than evidence based on replications confined to one centre. This ranking of the confidence that can be placed in a "fact" allows the reviewer to confine his or her attention to such facts and to relate the findings to a causative theory (see Andrews et al. 1983). While useful in demonstrating the reliability of the finding, such reviews fail to estimate the strength of the association, information which is especially important when treatment studies are being reviewed.

3.2 Meta-analysis

It has always been possible to aggregate the results of studies that used the same measure of outcome. One simply reported that the average improvement to be expected across the trials on a certain measure was such and such a value. The problem for psychiatry is that there are few commonly agreed measures, so that even in one disorder, such as depression, trials will very probably use one of the ten or so reliable and valid depression measures. Aggregating the results of these trials was difficult until Glass (1976) introduced the Cohen effect size estimation to clinical practice. Using this procedure, the difference between the experimental and control group mean scores on any outcome measure is standardised by dividing the difference by the standard deviation of the control group scores. The resulting measure of benefit or effect size is, theoretically, independent of the scale from which the effect size was derived.

The rationale for meta-analysis is as follows: symptoms of patients coming for treatment will range from mild to severe and be distributed about the scores of the average patient. Random allocation of patients to experimental and control groups should result in groups with similar distributions of symptoms. After effective therapy, the condition of the experimental group will have changed more than the control group, and a measure of this effectiveness is the distance between the distributions of the experimental and control group symptom scores at this time. This can be measured in standard deviation units and is called the effect size. From the area under the normal curve, it can be calculated that, given an effect size of 1.0, the average person in the experimental group will be better than 84% of people in the control group, a gain of 34 percentile ranks. Effect size superiority over placebo of less than 0.5 is associated with weak treatments, whereas an effect size superiority of more than 1.5 is associated with strong treatments. Most established treatments in psychiatry show effect size superiority over placebo of between 0.5 and 1.5 SD.

Given that the effect size is theoretically independent of the scaling properties of the measure used, be it symptom checklist, rating of severity of disorder or evidence of disability, then the comparability of the effect size derived from different outcome indicators allows studies to be compared statistically. Furthermore, details of patients, treatments, measures and study design can be assigned numerical values and related to the effect size. Calculating the effect size is simple when means and standard deviations are reported by the authors of a study. All authors should report these statistics, but only some do. The effect size can often be calculated in the other cases, but the procedures are complicated. Nevertheless, clinicians should become used to making effect size calculations, if only to be able to standardise their understanding of the power of a treatment.

3.3 The Cochrane Collaboration

Cochrane (1971), whose ideas were used to begin this chapter, also called for a regularly updated critical summary of all relevant randomised, controlled trials. While the first meta-analysis of interest to psychiatry, and indeed to medicine, was Glass's study of all reports of psychotherapy (Glass 1976), the first meta-analyses of clinically relevant material were the re-analysis of the Glass material now confined to clinical cases (Andrews and Harvey 1981) and the meta-analysis of the treatment of depression (Quality Assurance Project 1983) published as part of an early best practice protocol initiative. Ten years later, the first Cochrane Centre was established in the United Kingdom, and there are now centres in many developed countries. This network of international centres is called the Cochrane Collaboration. The centres have established registers of randomised, controlled trials and regularly publish reviews of trials in electronic format. At present, the principal task is to hand-search hundreds of journals for trials, because MEDLINE and PsychLIT

databases do not allow all randomised controlled trials to be identified. Nevertheless, the Cochrane Collaboration has already published reviews of important and difficult topics, such as family intervention and case management in schizophrenia, comparative efficacy of anti-depressant classes and the efficacy of critical incident stress debriefing. These are being continually extended and updated.

3.4 Evidence-Based Medicine

On the basis that good practice usually means more than the application of a single treatment agent, many countries have convened meetings of experts to determine what best practice for particular disorders might be. Usually the experts review the literature, acquaint themselves with current practice and make consensus recommendations that are science based in terms of identifying which treatments have been shown to be effective in clinical trials, and practice based in terms of which treatments and clinical procedures are likely to prove efficient in a particular treatment setting. Most clinicians should now have access to a wide range of such treatment protocols and manuals. Many clinics have taken the standard protocols and modified them to suit both the practice conventions of the clinic concerned and of their patient clientele, thus ensuring that they use proven effective treatments in the most efficient manner.

The move towards consensus-based treatment protocols began 15 years ago with two initiatives, one when the U.S. National Institutes of Mental Health convened a number of conferences on clinical problems of the day, and another when an Australian group began writing treatment guidelines for the ten most common mental disorders (see Quality Assurance Project 1983). These initiatives were initially resisted, because many practitioners feared that guidelines for treatment of certain disorders would limit a physician's freedom to do what was best for a patient. An additional worry was that the physician would be legally liable for any mishap when or if such guidelines were departed from. In effect, even physicians who use such guidelines find that they are constantly changing the actual treatment guideline to meet the needs of each individual patient. Thus the guidelines inform and expand the clinician's competence; they do not limit or restrict. The change in atmosphere about guidelines means that both government agencies and professional associations are currently developing guidelines for practitioners. Both the U.S. Agency for Health Care Policy and Research and the New Zealand Department of Health are producing guidelines for general practitioners, while the American Psychiatric Association, the American Psychological Association and the British Royal College of Psychiatrists are producing guidelines for their members. That this is a general trend is evident from the World Health Organization's initiative to review the effectiveness of treatments for mental disorders (Sartorius et al. 1993), which followed the publication of ICD-10, but whether such best practice protocols will result in preferred funding for practitioners who adopt them remains to be seen.

4 Applying New Treatment Data in Practice

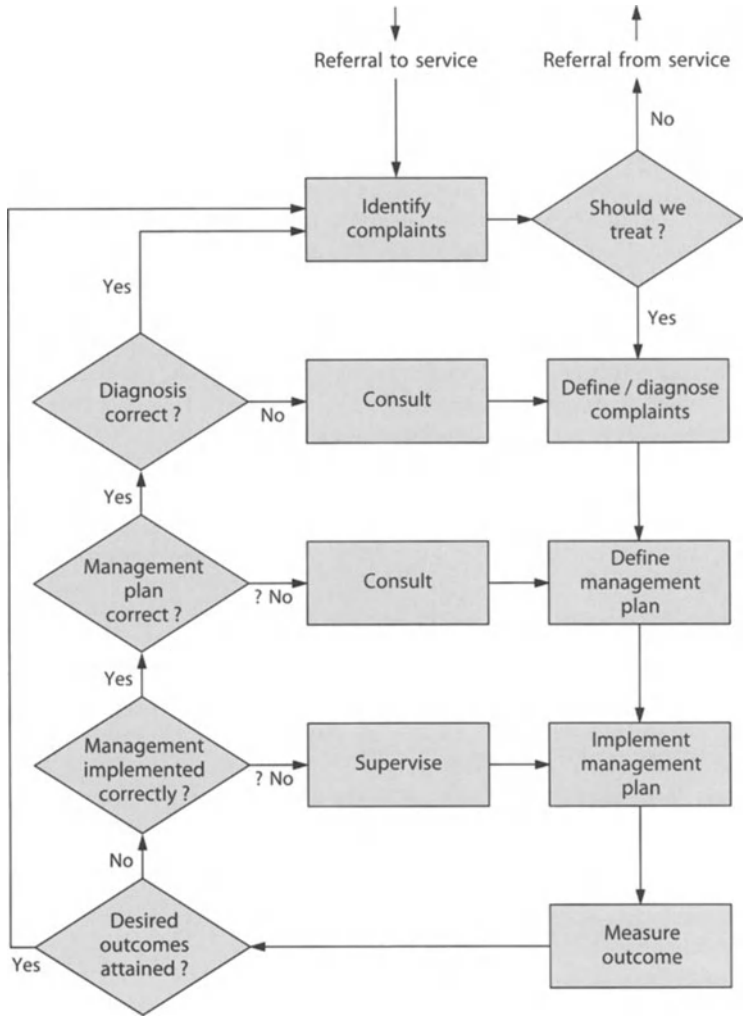
4.1 Learning About New Treatments

Good clinical practice is about absorbing the best information from the literature and then applying it to the care of an individual patient. Sackett et al. (1985) published strategies for busy clinicians to use when reading clinical journal articles. While the general issues are dealt with in the preceding sections, the specific questions that an individual clinician should ask about any article reporting an advance in treatment are somewhat different. According to their strategies, an article about treatment should not be read unless there is clear evidence of random allocation of patients to treatments, the results are statistically and clinically significant and all patients who entered the study are accounted for in the conclusion. If these provisos are met, the next set of questions concerns applicability to one's own practice, i.e. whether all clinically relevant outcomes were reported, whether the study patients were recognisably similar to one's own and whether the therapeutic manoeuvre is feasible in one's own practice with the training one has or could obtain. If the answers to these questions are positive, then one should consider using this therapy with the next suitable patient, which is a serious responsibility. There is a belief that the patients included in randomised, controlled trials are not typical of those seen in clinical practice. This is true in two ways. Patients in clinical practice tend to span a greater range; some have milder variants of the disorder, and their need for treatment would not cause them to volunteer for a trial, while others have more complex disorders and would often be excluded from a trial because of the presence of these exclusion criteria. However, on balance, there is no evidence that the patients in clinical trials are especially easy to treat. This is certainly a topic that should be investigated, because on present evidence it seems likely that the

results of randomised controlled trials are atypical in that both mild and complex cases will be underrepresented but the results will suffice for the average run of cases. Clinicians, however, are adroit at personalising therapy for the exceptional patient.

Because of clinicians' ability to personalise therapy, some discussion of the steps in good clinical care is in order whereby the results of the literature are integrated into the care of an individual patient. In fact, in order to understand the role of any specific treatment, it is helpful to have a model for good clinical practice (see Fig. 1). When patients go to a doctor or attend a mental health service, they have the right to be seen by a skilled clinician who can assess whether the service is competent to treat their complaints. People whose complaints are mild can be advised about self-care, while those with complaints that are outside the competence of the clinic can be referred elsewhere. Patients whose complaints are consistent with a disorder that the clinic is competent to treat should then be further investigated, the disorder diagnosed and symptoms, disability and risk factors measured. Once the diagnostic formulation is complete, the management plan is selected and implemented. During and at the end of treatment, the symptom, disability and risk factor measurements should be repeated, the change in scores indicating the outcome. If the desired outcome has been obtained, then the patient can be discharged from care. If the desired outcome has not been attained, then the decision tree makes it clear how to determine whether the diagnosis, management plan or implementation procedure is likely to need revision so that the desired outcome can be achieved. Paradoxically, measurement of outcome can also be useful for identifying people who do not improve even after the best clinical investigation and treatment. Normally, such people are excluded from the system, whereas regular follow-up can lead to good care and considered advice about how to manage the chronic disorder until specific treatments do become available.

Fig. 1. Model of good clinical practice



4.2 Establishing the Diagnosis

The first step in treatment is always the identification of the correct diagnosis, and this is especially important when treatments have a specific indication or a narrow therapeutic range. ICD-10 lists the criteria that must be satisfied before a diagnosis is made, yet few clinicians ask about each criterion in a systematic way. After listening to the patient's complaints at presentation, most quickly establish a probable diagnosis and spend the remainder of the interview asking questions to prove or disprove their initial hypothesis. This is a good use of scientific method, but because each clinician has idiosyncratic ways of developing the initial hypothesis, of asking the subsequent questions and of weighting the diagnostic criteria according to the answers received, it is not surprising that such clinical diagnoses are unreliable, both between clinicians and when made by the same clinician over time.

If effective treatment depends on an accurate diagnosis, how can accuracy be improved? The traditional method is to ask another clinician to review the case before starting treatment, but given the scarcity of skilled clinical staff, increasing use is being made of structured diagnostic interviews and standardised assessment questionnaires. The Composite International Diagnostic Interview (CIDI) is composed of questions that have been designed to illuminate each diagnostic criterion in Chap. V of ICD-10 (see Chap. 2, this volume, Part 2). It can be administered by non-clinical staff or, in the computerised version (CIDI-Auto), it can be self-administered. The results are then processed by a scoring programme and the exact ICD-10 diagnoses determined. It is reliable and valid and is available in most European languages (World Health Organization 1997).[1]

4.3 Measuring Symptoms, Disability and Risk Factors

There are many symptom scales, both interviewer-administered rating scales and self-completion questionnaires. While global symptom scales exist, it is preferable to use scales designed to measure the symptoms associated with a specific diagnosis (see Fig. 1). Measures of disability should be more general and cover a wide range of activities, a range that includes impairment in productive work, self-care and personal and social relationships. Examples of such general scales are listed in Table 1. In some disorders, it is also possible to measure risk factors such as neuroticism and coping in anxiety and depressive disorders. Information from measures of symptoms, disability and risk factors is important collateral information which can support the diagnosis and which, when compared with data obtained when treatment has been concluded, can produce measures of outcome.

[1] See http://www.unsw.edu.au/clients/crufad/cidi/htm

4.4 Selecting a Management Plan

Good clinical practice presumes that most management plans will comprise more than one single effective treatment. Most begin with education about the nature of the disorder and a discussion of the prognosis and the likely response to treatment. The clinician will then select a particular treatment and discuss the reasons for choosing this with the patient, ensuring that the patient understands the rationale behind the main treatment, the importance of compliance and the likely side-effects and factors that could interfere with efficiency of the treatment. The clinician should then discuss the use of any additional therapies to improve outcome, whether ancillary drug or psychological therapies, interventions to support relatives or training in social skills or rehabilitation. Finally, the clinician should make clear the expected time course of the improvement and be explicit about whether the treatment is intended to control the symptoms, in which case treatment may have to be prolonged, or to actually treat the condition to the point at which the patient can expect to be well, and remain well, after treatment has been concluded. Many texts on the management of mental disorders have sections that can be photocopied and handed to patients (Treatment Protocol Project 2000).

4.5 Ensuring Implementation

Prescribing a management plan is only half the task. One has to ensure that it is correctly implemented by patient and staff. Patients are more likely to comply with treatment if they understand the rationale behind the treatment and have a reasonable awareness of the likely side-effects or difficulties so that they can persevere with treatment when these occur. They should also have a detailed understanding of the duration of treatment and of the changes in their status that will lead to a revision of the dose or a change in the treatment. In New Zealand, best practice protocols for depression are published in popular magazines in

the hope that patients will know what is expected and that they will ensure that doctors comply with the details of the protocols. The simpler the dosage regime, the more likely patients are to comply. Similarly, the simpler the treatment regime, the more likely doctors are to administer it. Therefore, complex treatment programmes need to be written down in patient treatment manuals which both patient and doctor can follow.

4.6 Measuring Outcome in the Individual Patient

The measures of symptoms, disability and risk factors which were taken at the first assessment should be repeated at points during treatment, certainly at the end of an episode of care and at regular points during treatment in the same setting if the treatment is long term and the disorder chronic. The difference in the scores – especially changes in symptom scores – represent the outcome of treatment. These scores are of the greatest value if the difference in score is available during the consultation, because, when compared with benchmark data for this type of patient with this diagnosis and treatment combination, it will be clear whether the expected degree of improvement has occurred. If it has, then the treatment plan can be continued until the patient is relieved or cured. If the expected degree of improvement has not occurred, then the clinician can consider why this is so, e.g., whether the prescribed treatment was correctly implemented and complied with. If not, measures to improve compliance can be implemented. If compliance was not the issue, the next question is whether the treatment plan was appropriate for this patient. If review of the diagnosis and assessment information suggests that the treatment plan was not appropriate, then a more appropriate plan should be implemented. In the absence of improvement, one should be prepared to consider alternative treatments. Not everyone responds to the same treatment for their condition. If the treatment plan does match the diagnosis, then the evidence for the diagnosis should be reviewed; if no errors are discovered in that formulation, then a second opinion should be obtained from an experienced clinician. Second opinions are commonly requested in medicine, and they should also be a regular feature of psychiatric practice.

The routine measurement of outcome serves two purposes. Firstly, it informs the clinician about the progress of treatment in individual cases, which is very useful in difficult patients who are slow to improve and with whom so much time is spent. The second purpose is to allow the clinician to evaluate his or her effectiveness across a range of patients. Clinicians commonly believe they are less effective than they really are. This "clinician's illusion" results from the fact that they judge their effectiveness on the basis of the patients they are currently seeing. Clinicians spend the majority of their time with patients who are difficult and slow to improve and quite forget the patients who responded quickly and did not need to be seen again. Thus the routine review of outcome will show the clinician that, while their typical work-load is with people who are chronic, their typical patient is someone who responds quickly.

5 Conclusions

The methods for identifying effective treatments are well established. They include specifying the disorder to be treated, specifying the elements of the treatment, measuring the change in symptoms and ensuring that the observed improvement is actually due to the treatment, either by using a randomised, controlled trial or by carefully aggregating evidence from open trials. Information from trials is commonly aggregated, by systematic review, by meta-analysis and especially by the work of international consortia such as the Cochrane Collaboration. The ability to aggregate the evidence of treatment research has led to the development of best practice treatment protocols and encouragement for all clinicians to practice evidence-based medicine.

A model of good clinical care is outlined. The steps used in the care of an individual patient parallel the steps used in evaluating a treatment. It is important to be systematic about diagnosis and assessment, to use and properly implement proven treatments and to measure outcome appropriately. Such behaviour will accomplish two things. Firstly, it will inform the clinician about the progress of the individual patient, and secondly, it will allow any necessary modification of the treatment plan. Aggregating outcome data from all patients will demonstrate that the clinician's level of effectiveness is greater than that which is apparent from the progress of the patients who form the principal work-load. The routine evaluation of the outcome of treatment is an important clinical tool.

6 References

Andreason NC (1982) Negative symptoms in schizophrenia. Arch Gen Psychiatry 39: 784–788

Andreason NC, Olsen S (1982) Negative versus positive schizophrenia: definition and validation. Arch Gen Psychiatry 39: 789–794

Andrews G (1993) The essential psychotherapies. Br J Psychiatry 162: 447–451

Andrews G, Harvey R (1981) Does psychotherapy benefit neurotic patients? A reanalysis of the Smith, Glass and Miller data. Arch Gen Psychiatry 38: 1203–1208

Andrews G, Craig A, Feyer A-M, Hoddinott S, Howie PM, Neilson M (1983) Stuttering: a review of research findings and theories circa 1982. J Speech Hear Disord 48: 226–246

Beck AT, Ward CH, Mendelson M, Mock J, Erbaugh J (1961) An inventory for measuring depression. Arch Gen Psychiatry 4: 561–571

Beck AT, Epstein N, Brown G, Steer RA (1988) An inventory for measuring clinical anxiety: psychometric properties. J Cons Clin Psychol 56: 893–897

Cochrane AL (1971) Effectiveness and efficiency. Random reflections on the health service. Nuffield Provincial Hospitals Trust, Oxford

Derogatis LR (1977) SCL-90-R: administration, scoring and procedures manual-I. Johns Hopkins University, Maryland

Eisen SV, Dill DL, Grob MC (1994) Reliability and validity of a brief patient-report instrument for psychiatric outcome evaluation. Hosp Community Psychiatry 45: 242–247

**Glass GV (1976) Primary, secondary and meta-analysis of research. Educational Researcher 10: 3–8

Goldberg D (1972) The detection of psychiatric illness by questionnaire. Oxford University Press, London

Goodman SH, Sewell DR, Cooley EL, Leavitt N (1993) Assessing levels of adaptive functioning: the Role Functioning Scale. Community Ment Health J 29: 119–131

Hamilton M (1959) The assessment of anxiety states by rating. Br J Med Psychol 32: 50–55

Hamilton M (1967) Development of a rating scale for primary depressive illness. Br J Social Clin Psychol 6: 278–296

*Kluiter H, Wiersma D (1996) Randomized controlled trials of programmes. In: Knudsen HC, Thornicroft G (eds) Mental health service evaluation. Cambridge University Press, Cambridge

Jablensky A, Schwartz R, Tomov T (1980) WHO collaborative study on impairments and disabilities associated with schizophrenic disorders. Acta Psychiatr Scand [Suppl] 285: 152–163

Overall J, Gorham D (1962) The Brief Psychiatric Rating Scale. Psychol Rep 10: 799–812

Quality Assurance Project (1983) A treatment outline for depressive disorders. Aust N Z J Psychiatry 17: 129–148

*Sackett DL, Haynes RB, Tugwell P (1985) Clinical epidemiology. Little and Brown, Boston

Sartorius N, de Girolamo G, Andrews G, German A, Eisenberg L (eds) (1993) Treatment of mental disorders: a review of effectiveness. American Psychiatric Press, Washington

Spielberger C (1983) Manual for the State-Trait Anxiety Inventory. Consulting Psychologists Press, Palo Alto

Treatment Protocol Project (2000) Management of mental disorders. World Health Organisation Collaborating Centre for Mental Health and Substance Abuse, Sydney, Australia

Veit CT, Ware Jr JE (1983) The structure of psychological distress and well-being in general populations. J Consult Clin Psychol 51: 730–742

Ware JE, Sherbourne CD (1992) The MOS 36-item Short-Form Health Survey (SF-36): I. Conceptual framework and item selection. Med Care 30: 473–483

Wing J (1994) Health of the Nation Outcome Scales: HoNOS Field Trials. Royal College of Psychiatrists Research Unit, London

World Health Organization (1997) CIDI-Auto 2.1 Administrator's guide and reference. World Health Organisation Collaborating Centre for Mental Health and Substance Abuse, Darlinghurst

CHAPTER

11

R. Jenkins, R. Kessler, P. Leaf, J. Scott

Systems of Psychiatric Care: Principles and Desiderata of Good Services

1 Introduction

Systems of psychiatric care can be conceived of as complex, multi-dimensional frameworks or matrices which do not exist in isolation but are firmly rooted in and influenced by prevailing mores, political climate and culture. Any analysis of the principles and desiderata of good services therefore needs to examine the relationships between the different layers of national policy, regional support and local implementation of service delivery; between the different sectors of health care, social care, housing, social security benefits and the criminal justice system; between the complicated infrastructure of primary, secondary and tertiary care; between research, evidence-based medicine and education and training; and between the different strands of resource allocation, particularly public and sector funding and contributions from the voluntary sector. Adequate interfacing arrangements between components of the infrastructure are as crucial for successful functioning of mental health care as adequate inputs into each component to ensure maximum effectiveness of the service and high standards of efficiency and equity.

Countries around the world face a number of common challenges, whether they be in the industrialised or the developing world. These include the high burden of mental disorders, the general shift in ethos from institutional to community care, relative resource constraints and a fluctuating balance between primary and secondary care.

It is important to recognise that individual outcome is affected by public policy and social norms at a national and international level, by how care is organised, financed and delivered and by how access and treatment are provided at the level of the individual. The definition of "need" is influenced by factors at all levels and is not an absolute, but rather reflects cultural norms, resources available, alternative demands on intellectual and fiscal resources, and the real and perceived benefits. Thus it depends not only on societal definitions and public policy, but also on the services that are potentially available.

Any discussion of systems of psychiatric care need to acknowledge that the quality of the care provided is a critical issue, because efficiently linking together multiple ineffective services will not result in positive patient outcomes. That said, this chapter will focus on issues of particular relevance at the level of the service system.

2 Overarching Principles

There are a number of overarching principles which should be outlined at this stage.

2.1 Needs-Led Service

It is critical to ensure that decision-making about services is led by needs rather than supply. Thus it is important not to start with existing services but to base estimates of needs on absolute levels of disease, disability, severity, chronicity and risk in the context of some assessment of social and political concerns, real or perceived efficacy and option appraisals of the alternative use of resources, i.e., that it is predicated on accurate information about the epidemiology of mental disorders, their prevalence, course, outcome and accompanying disability, social and economic costs and the opportunity for treatment rather than simply on pre-existing service use (Stevens and Gabbay 1991). This is often quite difficult, especially in systems that are underfunded in which demand for existing services exceed supply. It is an easy, but incorrect, conclusion to draw in cases of this sort that the currently available services are the best ones to provide based on the fact that demand is high and excess resources to start up new service options are scarce or even nonexistent (Goldberg and Gater 1991). Epidemiologic surveys show clearly that demand is not a good indicator of need. Other services than those provided might well be more needed or effective or efficient. Segments of the population that are hard to reach might have greater need than those who are more likely to seek treatment. It is critical that procedures be developed to monitor need for services independent of demand for currently offered services if a treatment system is to evolve (Griffiths et al. 1992; Jenkins and Griffiths 1991). It is a challenge to develop national monitoring procedures, but models do exist. Besides collecting national mortality data, it is possible to carry out repeat national health surveys. Furthermore, a number of countries have developed a system of reporting the need for specific health problems, e.g., the DAWN system in the United States whereby a national sample of emergency rooms reports every drug overdose case, allowing the federal government to monitor trends in serious drug problems.

2.2 Explicit Values

Mental health policy also needs to take account of – although not necessarily follow – public beliefs, assumptions and attitudes. Likewise, local decision-making will need to involve a wide constituency of user and carer, professional and public interests. It is therefore important that mental health care is based on an agreed set of values and principles which are explicit and open to scrutiny. The Reed principles (Department of Health 1994b,c; Royal College of Psychiatrists 1996) provide an excellent example. Patients should be cared for:

- With regard to the quality of care and proper attention to the needs of individuals
- As far as possible, in the community, rather than in institutional settings
- Under conditions of no greater security than is justified by the degree of danger they present to themselves or others
- In such a way as to maximise rehabilitation and their chances of sustaining an independent life
- As near as possible to their own homes if they have them
- With respect for their rights as citizens

At the same time, the mental health care system can have an agenda-setting function in changing values and perceptions. This is especially important in the case of mental health care, as stigma is a serious barrier to seeking care for psychiatric problems around the world. Furthermore, stigma affects objective assessments of need at all levels, including public policy, attribution by organisations and individual assessments. There is evidence that this situation is changing in developed countries in response to major public education and outreach campaigns in conjunction with unprecedented attention in the mass media over the past decade of new pharmacological treatments. Even so, needs assessment surveys in even the most "enlightened" countries continue to find that the major self-reported reason for not seeking professional treatment among people who have a psychiatric disorder is that these people believe that this is a private matter, the kind of problem they should handle themselves. Breakthroughs in our understanding of the genetic basis of many mental disorders and the development of new and effective pharmacotherapies create an opportunity to correct this misconception by making people recognise that mental disorders are illnesses, not personal failings, and that effective and safe treatment is available.

2.3 Range and Severity of Disorders

Policy needs to address both the range and severity of disorder. Firstly, policy needs to encompass not only people with severe mental illness, but also the more common mental disorders and the sub-threshold disorders which also cause personal dysfunction, social problems and economic costs and alcohol and drug abuse; it should also focus on the promotion of mental health more generally in the whole population, including in schools (Bond and Compass 1989) and workplaces (Jenkins 1994b). Otherwise the care of people with severe mental illness by specialist teams will be compromised by the large burden of the more common disorders if such patients are not adequately dealt with in primary care, workplace employee assistance programmes, school-based services and housing- or social service-based programmes. The prevalence, chronicity and severity of minor to moderate mental disorders is so great that, unless their needs are adequately met in primary care, patients with such disorders will be referred in huge numbers to secondary care and impede attempts to target specialists to the more severe disorders (Jenkins 1992; Lloyd and Jenkins 1995; Lloyd et al. 1996; Paykel and Jenkins 1996). This means that primary care teams, occupational health teams, social service professionals and teachers need to be appropriately trained and supported.

Secondly, it needs to be remembered, in cautioning against neglect of non-severe disorders, that many severe disorders start out as comparatively mild and much more tractable conditions. Indeed, the typical person with a mental disorder does not seek treatment until many years after onset and a number of episodes (Kessler et al. 1998). Given that initial treatment-seeking is slow and that many psychiatric disorders begin early in life, we need vigorous outreach to young people still at school. There are many young people with clinically significant disorders that are risk factors for the subsequent development of severe disorders in adult life. Such outreach programmes in schools, if they exist, often target juvenile delinquents, but more emphasis is needed on shy/depressed/anxious children. These are not children who cause trouble in school, so they are often overlooked; targeting them early could, however, be a very powerful and cost-effective way of reducing morbidity in later life (Rotheram 1982; Bierman 1986; Kellam and Rebock 1992; Hawkins et al. 1992; Mubbashar 1997; Sampaio Faria et al. 1997).

3
Principles Used in Decision-Making

3.1
Equitable Allocation of Resources

Resource allocation to local areas should be proportionate to the health and social care needs of that area, and hence it will be important to develop funding formulae which take account of those needs on an equitable basis. Such formulae will need to include factors such as prevalence rates of illness, social deprivation, homelessness and high rates of high-risk groups, e.g., refugees and immigrants.

As noted above, it is important that this allocation formula is not determined primarily by differential demand for services, as such an approach will lead to a cumulation of inequity by resources flowing disproportionately to areas that already have well-developed systems of care that are visible to their communities and successful in attracting people in need of services.

Investment in areas that have underdeveloped systems of care is critical, even though this allocation of resources will necessarily be comparatively inefficient in the short run, if equitable allocation of resources is a long-term goal. It is important to make clear and explicit the processes for identifying and allocating resources, to establish who is "legitimately" involved in such decisions, to understand how rapidly or slowly resources within the system can be reallocated and to determine who is accountable for increasing or reducing access to services.

There are many difficult decisions in resource allocation that we will need to confront as we attempt to rationalise delivery of services. If we have ten disorders to address, each with a different prevalence, different accompanying disability, different cost of treatment and different treatment effectiveness, should we aim at equal treatment outcomes even if one disorder is much more expensive to treat or maximise the reduction in total population impairment/disability? These are difficult decisions which need to be explicitly tackled in the context of public debate.

3.2
Quality Assurance

Governments need to think about putting systems in place to ensure quality of service inputs and processes and to measure the health and social outcomes achieved. Quality assurance of inputs and processes can be achieved by a combination of inspection systems, auditing and empowering users and carers to have certain specific expectations. In order to be successful, these functions need to be centralised at the administrative level responsible for resource allocation rather than shunted off to the treatment centres as tasks they are required to perform in order to maintain funding. Time and again, it has been shown that self-audits of the latter sort are perfunctory and aimed primarily at creating an administrative paper trail that justifies treatment decisions post hoc rather than providing accurate information on quality of care. Different models for using novel auditing procedures to monitor quality need to be studied, implemented, evaluated and revised if optimal treatment quality and cost-effectiveness are to be achieved (Farrar 1996; Miller et al. 1995; Commonwealth of Australia 1996; Ministry of Health 1997b; Rosen et al. 1995).

3.3
Evidence-Based Medicine

As far as possible, it is helpful for governments to set a policy that aims to be as evidence based as possible. We are now uniquely positioned to develop such an evidence-based approach, using reports from national morbidity surveys, capitalising on the advent of routine outcome measures and information on effective interventions. It is not always politically possible or indeed desirable to wait for all the possible evidence to be accumulated before government action is taken. Often, it is important to take some preliminary action and then to fine-tune the policy with specific research evidence at a later date (Conway et al. 1996).

4
Desiderata for National Policy

Mental health policy at a national level comprises the government's mission statement and strategic objectives for the mental health of the country.

4.1
Goals

Many countries are now starting to focus much more on the central goals of improving health and reducing morbidity, disability and mortality, with improved processes seen not as main goals in themselves but rather as intermediate steps towards real health gains (Jenkins 1990). England's Health of the Nation strategy set some specific targets to reduce morbidity, disability and mortality (Department of Health 1992, 1994a; Jenkins 1994a, 1996a,b), and Australia is another example of a country which has set national goals,

targets and strategies for better health outcomes into the next century (Commonwealth of Australia 1994). As a general principle, the overall framework can be summarised in five principal goals:

1. Promote good mental health
2. Prevent illness
3. Reduce morbidity (improve health and social functioning) of people with mental illness
4. Reduce disability associated with mental disorders and their ensuing discrimination
5. Reduce mortality (from suicide and physical illness) of people with mental illness

4.2 Needs Assessment

Mental health policy should be rooted in and supported by accurate information about the health and social care needs of the country, either by using national surveys of psychiatric morbidity or extrapolation from a combination of local surveys together with national surveys from other countries in order to obtain accurate estimates of the range, frequency, severity, chronicity and accompanying disability and mortality and their relationship to sociodemographic variables, including geographic variables. National surveys need to collect information on variation in local need to help decide which areas need more resource per capita. They cannot micro-manage local services, but they can provide some comparative information which may be helpful to local areas, and they can provide help for local needs assessment (Jenkins et al. 1997a,b; Meltzer et al. 1995a–c, 1996a–d; Wing 1992; Croghan et al. 1998). Mental health policy will also need to take into account information about the existing accessibility, quantity, quality, costs and outcomes of services (Jenkins and Knapp 1996; Glover and Gould 1996; Glover 1996; Goldberg and Gater 1991; Griffiths et al. 1992; Knudsen and Thornicroft 1996). Furthermore, it will need to take account of the existing resource infrastructure of the country in terms of the available capital, revenue, trained personnel and untrained staff in the context of managing any kind of transition to a new style service. Other inputs such as public beliefs and opinions, publicity about specific incidents and the political inclination of governing parties are crucial. Thus mental health is not just an issue for health and social care, but also for the general population, schools, workplaces, rural areas and cities.

4.3 Coherent Interdepartmental Policies

National policy needs to be coherent and complementary across the different government departments or ministries of health, social services, housing, social security, employment and finance. Ideally, any major policy move in any one sector needs to be preceded by the production of a mental health impact statement, analogous to an environmental impact statement. Many examples exist of national policies that are internally inconsistent and so lead to suboptimal resource allocation and treatment effectiveness due to conflicting intra-sector rationalities. The only hope of solving this problem is to empower an administrative team with an oversight of the departments in which coordination is sought. Interdepartmental coordination is unlikely in the absence of such a structural arrangement. National policy also needs to be coherent across the local geographic units where it is implemented, and these local planning units need to be coordinated and targeted as national priorities. Within each of these levels are multiple agencies and departments requiring coordination and common fiscal strategies (Department of Health 1995).

4.4 Support for the Voluntary Sector

The voluntary sector, encompassing as it does consumer and family organisations and charitable donations of time and money to the health service, needs to be encouraged, stimulated and supported. Governments play a vital role in pump-priming this system. The voluntary sector is also important because of its links with the wider community and hence its capacity as an agent for influencing public attitudes. In some countries, the voluntary sector is the largest part of the de facto mental and addictive services treatment system. This is the case, for example, in the United States, where there are more visits per year to self-help groups for psychiatric problems than to all professional speciality mental health services combined. Opportunities for increasing coordination between the voluntary sector and the professional service sectors need to be explored as one way of extending the limited professional resources available in even the most resource-rich society in treating psychiatric problems (Kessler et al. 1997a).

4.5
Strategies for Research and Development

It is vital to invest adequately in research and development to support policy and service developments and to ensure that we adequately research different methods of organising and financing care and different methods of service delivery. If money and activity are to be efficiently spent, it is important to have some degree of national coordination, oversight and stimulation.

4.6
Strategies for Manpower Training

Human resources are fundamental to the operation of psychiatric services, and it is essential to have national strategies for ensuring adequate flows of sufficient personnel for relevant disciplines, taking account of projected demographic trends; national strategies are also necessary to ensure appropriate professional training and continuing professional development. This includes not only specialist mental health personnel, but also generic staff in primary health care teams, social services, etc.

Manpower training is also a major opportunity for dealing with the issue of stigma. Training should include multi-disciplinary components so that there is increased awareness of the activities of other agencies and specialities (Mohit 1997).

4.7
Strategies for Mental Health Promotion

National government policy can affect the mental health of the population, e.g., through policies on poverty, unemployment, social exclusion, alcohol availability and restriction of firearms. Furthermore, there are some strategies for mental health promotion that need to be directed at the whole population, i.e., universal measures which, if they are to be implemented, will require some kind of national policy support (Jenkins and Üstün 1997), e.g., home visiting for the newborn; general measures to encourage health policies in the workplace that address both mental and physical health (Jenkins and Coney 1992; Jenkins and Warman 1993); teacher training as prevention intervention for school disruption, the encouragement of mental health promotion education in schools and physical health promotion, requiring liaison between the government, teacher-training bodies and schools (Durlak 1995).

4.8
Strategies for Suicide Prevention

Suicide is a major cause of avoidable death across the world (Murray and Lopez 1996). Those countries which have adopted a national plan for suicide prevention, e.g., Finland, Holland and the United Kingdom, are showing an encouraging decline in suicide rates, even in young men - thought to be the most difficult age-group to influence - and there are some aspects of suicide prevention that require action at a national level (Jenkins 1994a; Kingdon and Jenkins 1995). These measures include the following:

- Action to reduce access to means of suicide, e.g., measures on gun control, restricting the sale of certain prescribed medicines such as barbiturates, restricting pack size and availability of over-the-counter medicines such as paracetamol
- Action to support high-risk occupational groups by liaison between the government and the relevant occupational organisations
- Specific suicide prevention programmes in prisons and police cells, where suicide rates tend to be very high
- Research into the causes and prevention of suicide and action to ensure that all suicides and psychiatric homicides are investigated to learn the lessons for prevention and to disseminate them
- Action to ensure adequate continuing professional training programmes in the assessment and management of mental illness, especially depression, and of suicidal risk for both primary and secondary health care professionals, e.g., the United States' DART programme; the training programme for practitioners in Sweden for the prevention and therapy of depression (Rutz et al. 1996) and the World Psychiatric Association's educational programme on depression
- Working with the media to ensure that individual suicides are reported responsibly, without glamorising the victim and without reporting the specific method (Schmidtke and Hafner 1988).

Suicide prevention strategies should be a major component of the overall framework of government goals in order to reinforce the commitment of all the relevant sectors (Department of Health 1994a; Ministry of Health 1997a).

In the United States, there is an increasing focus on trying to reduce risk, which is done both through public policy as well as by formal coordination of professionals and by increasing the informal support systems.

5
Desiderata of Local Service Delivery

5.1
Goals

There is a need for individual service providers to establish their goals and targets, to monitor their success in achieving these outcomes and to make adjustments to reduce the negative outcomes. These local goals might include prevention of episodes, shortening duration of episodes, increasing duration between episodes, reducing residual impairment from mental disorders and reducing the amount of co-morbidity. Even less severe disorders can be disruptive to work and social relationships and entail considerable costs for society, employers and individuals.

5.2
Needs Assessment

Therefore, local geographic needs assessment has to be carried out for health and social care within which individual agencies can contribute to integrated assessment and management (Johnson et al. 1996). In order for overall mental health goals to be met, many other agencies besides traditional mental health providers need to be involved and managed in an integrated way.

When local health planners, commissioners or purchasers assess the health needs of their local population, they need to consider the epidemiological evidence, both national and local, on the prevalence of the different disorders, their severity and chronicity. The consequences of mental disorders – particularly disability and the degree of risk to self and others – also become more salient in determining the need for services and specific actions. Local planners also need to consider the policy context, again both national and local, and the current state of the specialty services, together with the current levels of in-patient and residential provision and the existing pressures on them. In particular, they need to examine the extent to which the full spectrum of health and social care is being provided to people with mental illness and the extent to which there are problems with the interfaces between the different agencies providing health and social care, the criminal justice system, etc. It is also important to examine local suicide rates (both those officially identified by the coroner and the "undetermined" deaths that are epidemiologically likely to be suicides). People with severe mental illness have highly standardised mortality rates from physical illness such as cardiovascular disease, malignancy and respiratory disease. Purchasers should therefore ensure that they also pay attention to assessing and meeting the physical health needs of their population with mental illness.

Variables which contribute to high rates of mental illness in local areas include social deprivation, urbanicity and the combination of deprived inner city areas, marital breakdown, family breakdown, single-person households, single-parent families, unemployment and levels of substance abuse. The presence of major rail termini within a service catchment area can aggravate the effects of the geographical drift of people with severe mental illness from more rural areas to inner city areas. Particular markers of high mental health needs are high proportions of homeless people, alienated, ethnic communities, refugees, children "at risk", children "in care", young males aged 15–45 and people who are single, divorced or widowed. Some of these variables may be influenced by public policy and general economic developments, while others may be influenced by clinical interventions.

Aspects of the specialist services where there are likely to be pre-existing heavy pressures include secure provision for mentally disordered offenders, acute beds, 24-h nursed care for "new long-stay" patients, supported housing, occupational rehabilitation, mother and baby units, eating disorder services and drug and alcohol services.

Local needs assessment should also include the need for primary health care, firstly to tackle the less severe disorders; secondly to deliver physical health promotion and physical health care to people with mental illness; thirdly to contribute as locally agreed to a shared care programme for people with severe mental illness; and fourthly to deliver primary prevention to high-risk groups such as the socially isolated, bereaved, physically disabled, unemployed and elderly.

Local planners/purchasers should ask themselves whether there is equity in their existing provision, with people from different geographic areas, cultures, social class and the homeless all having their needs met to an equal extent. They need to know what costs are associated with the current provision, what obstacles currently exist to planning coordinated and comprehensive services and how the services might be improved.

5.3
Education, Support and Deployment of Primary Care and Links with Specialists

Mental disorder is extremely common, and no country can afford sufficient specialists to cover everyone with

a mental disorder. The WHO multi-site study (Üstün and Sartorius 1995) has shown that around one third of general practice patients have a psychosocial problem and that the most common of all chronic disorders (physical or psychological) in general practice is depression. The essential goal of primary care is to tackle the majority of individuals with a mental disorder, and most people with a mental disorder will need to be seen by primary care teams, leaving highly trained specialists free to deal with those patients with disorders that are more severe and difficult to treat (Kingdon and Jenkins 1996).

Members of the primary care team, especially doctors and nurses, must be trained to assess, diagnose and manage common mental disorders and to know how to refer patients to specialists. This requires addressing undergraduate, postgraduate and continuing education of primary care doctors and continuing training of primary care nurses to ensure that adequate attention is given to mental disorders. It is relevant to mention here some innovative "cheap" strategies that there is reason to think might work. Computerised self-therapy programmes, for example, are very intriguing and have the potential to be an enormous boon in extending the number of people a single practitioner can "treat". The adjunctive use of self-help groups is another potentially useful strategy in which the amount of time the professional spends with each patient alone is reduced and the group is given the resources (e.g., educational tools) to assume some of the education and support and possibly even therapeutic functions that would otherwise be met by the professional. These kinds of initiative are important if we are to come close to having enough resources to meet the massive need for services that exists in the population.

The primary care team needs to consider the appropriate balance of doctors, nurses, counsellors and others and the appropriate division of labour, to ensure that the caseload is tackled within the resources available. A number of innovative models have evolved over the past two decades to coordinate primary care and specialty treatment of psychiatric disorders. These include the placement of mental health specialists (psychologists and social workers) within primary health care settings. It is important that these models be studied, evaluated and modified as necessary so as to increase the efficiency of coordination across service sectors. It should be the responsibility of the specialty sector to take the lead in developing these systems, as they are in a better position than their colleagues in primary care to evaluate options for innovative treatment along these lines (Jenkins and Field 1996).

Primary care teams need backup support from the specialist services (e.g., opportunities to agree on clinical guidelines and discuss cases; opportunities to learn additional psychological skills such as behaviour therapy techniques). This is a special challenge in the many parts of the world in which the population in need of treatment lives predominantly in rural areas. Problems of distance and communication and lack of availability of specialists in rural areas conspire here to create very serious problems, especially for patients with problems that are not sufficiently severe to warrant moving them to distant cities for in-patient treatment. Innovative models need to be developed here. New developments made possible by breakthroughs in communication technology, such as telepsychiatry and Internet consultation networks, need to be considered here. However, what is most crucial is that those involved in primary care should feel supported by the specialists and should feel they have adequate specialist backup. Failure to communicate effectively is often aggravated by stigma and leads to poor patient care and misunderstanding about the respective roles of primary and secondary care teams.

Primary health care teams need to have a certain amount of basic information about local mental health services, including an organisation chart of the local services (with the names of key clinical and management staff and maps of geographic or other boundaries) and an information booklet of therapies available together with named contacts for advice on appropriate referrals. Agreed criteria are required for referral to the specialist services, taking into account diagnosis, severity of symptoms, duration and risk of harm to self or others – and also evolving safe methods of shared care, whereby the primary care team can participate, e.g., with medication, physical health care and health promotion (Strathdee and Jenkins 1996; Lloyd and Jenkins 1995). Secondary care teams need specific information in the referral letters, including background family and social history, details of the presenting problem, interventions tried so far with outcomes, the reason for the present referral, the specific role that is being requested of the specialist team and the anticipated continuing role of the primary care team. Likewise, primary care teams need specific information in the reply from the specialist, which should include the following: a clear management plan with objectives and expected outcomes; an indication of the risk of suicide; what the patient has been told about his or her condition; the prognosis and likely continuing disabilities and influence of the patient's life style; the role that the primary care team is expected to play in the management plan; the role that the specialist staff will play; and who will be responsible for doing what and within what time schedule, including prescribing and monitoring roles and responsibilities.

5.4 Development of Comprehensive Health and Social Services

Countries vary in their capacity to afford the full scope of comprehensive specialist local health and social services, but even the richest will never be able to afford sufficient specialists to cater for the complete spectrum of mental disorders. There must therefore be explicit public policy and local availability of supportive and nurturing family and community environments as well as an appropriate balance of specialist services and primary care services which will need to be thought through in relation to the specific needs and resources available within each country (e.g., Wig and Murthy 1994; Murthy 1997). There is also a need to establish developmentally appropriate responses and services for children, adolescents, adults aged 16–65 and older people in terms of the opportunities for effective interventions and for reducing compounding factors.

Now that research has demonstrated the long-term damage to health and social functioning which prolonged institutionalisation can bring, most countries are attempting to transform services reliant on old, large asylums into services where people with severe mental illness are cared for in their own homes or in homely environments, as close to home as is compatible with the health and safety of the patient and the safety of the public. However, some people with severe mental illness will need to be admitted to acute beds in hospital for short periods of time for intensive assessment and management of acute episodes of severe illness; in addition, a proportion will need supported accommodation for long periods of time, and a small proportion will need very intensive skilled nursing care for long periods of time in order to enable regular supervision of medication and daily monitoring of their mental state. Thus every locality will need a range of services to supply the gradation of needs of people with mental illness for psychological treatments and medication, supported accommodation, work and occupational rehabilitation, leisure activities and day care (Ministry of Health 1997b; Commonwealth of Australia 1996; Department of Health 1994a). The use of these services needs to be carefully evaluated so that their predictors are well understood (e.g., Kessler et al. 1997a, 1998).

There is a risk that politicians, planners and health service managers may misuse de-institutionalisation programmes to divert resources into other areas of expenditure instead of re-investing the money from the institutions into local community mental health care. Table 1 gives the spectrum of care that needs to be available locally for people with mental illness. Multidisciplinary teams are needed to look after clients in both hospital, residential and home settings.

While many countries are making arrangements for the "old long-stay" patients in the old, large asylums to move in a planned and supported way into the community, most countries have not yet made adequate provision for the so-called new long-stay clients, i.e., that small number of very severely mentally ill people who, for a variety of reasons and despite excellent treatment, will nonetheless continue to need intensive, skilled, round-the-clock nursing for many years. In the absence of such provision, these clients either inappropriately "block" acute beds because there is nowhere else for them to go or they receive no or inadequate support in the community (either in inadequately supported hostels, unsupported in their own homes, homeless or in prison or secure units). The provision of 24-h nursed accommodation would release acute beds for those who require them and would provide a more homely and more socially rich and stimulating environment for those who need

Table 1. Range of care

	Acute/emergency	Rehabilitation/continuing care
Home-based	Sector teams Sustainable out-of-hours cover Intensive home support	Domiciliary services Key workers Care programme approach
Day care	Day hospitals	Drop-in centres Support groups Employment schemes Day care
Residential support	Crisis accommodation Acute units Local secure units	Ordinary housing Unstaffed group homes Placement schemes Care schemes Mental nursing homes 24-h nursed NHS accommodation Medium secure units

ongoing skilled nursing for long periods of time. This group includes those who need daily monitoring of their mental state, frequent monitoring of the risk of harm to self or violence towards others, storage, administration and supervision of medication which will usually be necessary on a daily basis, assistance with self-care and daily living skills, supervision and support to attend day care or rehabilitation activities, crisis intervention at night if required, skilled management of challenging behaviour and possibly skilled management of dual or even triple diagnosis.

There is an equivalent problem at every level in the service system, where those who are inappropriately at that level reduce the availability of services for those in need, thus resulting in diversion to more intensive or inadequately intensive services. In addition to attending to when a more intensive service is needed, systems are needed to determine when services can be discontinued and when less intensive services can be substituted for more intensive and usually more expensive services.

5.5
Development of Good Inter-agency Collaboration

There needs to be public policy and practice at each level that focuses on individual and family needs. Health professionals (doctors, nurses, psychologists, occupational therapists) will have to work closely with social workers, voluntary workers, probation officers and the police to ensure that care is properly coordinated (Kingdon 1994a). This can be very difficult, since all the separate agencies involved in the care of severely mentally ill people will have their own sets of principles and priorities which guide their work. However, it is essential that they work well together, and there are certain basic requirements without which working together can be very difficult. These include the principles of a commitment to working together at all levels of the agencies involved, including senior management; an agreed and jointly "owned" strategy for the care of severely mentally ill people; agreed and well understood principles for accessing services; appropriate and effective arrangements for inter-agency information exchange; joint commissioning, wherever possible, to maximise the use of available resources; a commitment to training, on a single and multi-agency basis, in order to encourage a better understanding of other agencies' roles and structures; and regular review and evaluation of arrangements for inter-agency working. One very appealing strategy that has been adopted in many systems is the use of a case manager to coordinate care across multiple service sectors – to interface with and optimise on a patient-by-patient basis the allocation and delivery of such diverse services as those involving public housing, income maintenance, social services, health and mental health. Although systems that have adopted one or more of the widely used case manager models have, perforce, an additional layer of administration than those without such an approach, experience has shown that the savings introduced by reduced inefficiency and coordination costs more than make up for the costs and complexity of this innovation.

5.6
Development of Mental Health Information Systems

Mental health information systems are valuable adjuncts to planning the delivery of specialist services to a defined population, assisting the coordination of care between a variety of professionals and helping to ensure that clients do not get lost to care. At their simplest, they comprise names, addresses, age, sex and date of referral to specialist services. However, many places are now developing data sets which include a common core of helpful clinical and sociodemographic data to assist health professionals in reviewing and coordinating care. These minimum data sets, which include measures of health and social functioning, allow health and social outcomes to be monitored.

It is essential that such information systems be developed in consultation with users and carers, to satisfy concerns about data protection and confidentiality and to develop sensible procedures for accessing information. Common information technology standards allow local information systems to "speak" to each other should a client move from one district to another with accompanying transfer of care. Interfacing information systems are also essential to examine the coordination of care at the population level. It is important to continue to evaluate these systems in terms of direct patient benefits against the costs of implementing and maintaining them (Wing et al. 1996).

5.7
Development of Good Practice Guidelines

As well as ensuring adequate quantity and quality of the inputs to the service in terms of beds, staff, housing, etc., it is also necessary to ensure adequate quantity and quality of the processes of care in both specialist and primary care services, to bring levels of care across the board up to the best standards. Good practice guidelines are a key method of implementing evidence-based services (Armstrong 1997; World Health Organisation 1996; NHS Executive 1996a,b).

Where possible, good practice guidelines should be developed in collaboration with the relevant professional bodies and should address the multi-axial elements of care, namely physical and psychological health care needs and social needs including housing, employment and benefits (Kingdon 1994b,c), as well as continuation of care (Tessler et al. 1986; Tessler 1987; Test 1979).

5.8 Prevention of Suicide

As well as the national measures to reduce access to means of suicide, research and audit of suicides and national measures to support high-risk occupational groups, it is also necessary for services to take a number of local steps to reduce suicide (Jenkins et al. 1994; Kingdon and Jenkins 1995). These include ensuring that primary and secondary health and social care professionals are regularly educated about the assessment and management of suicidal risk; that local arrangements are in place to support high-risk occupational groups, e.g., doctors, nurses, farmers, vets and pharmacists; and that local alliances are made with the media to ensure more responsible reporting of individual suicides whereby the specific method used is not reported and the event is not glamorised.

There are some important misconceptions that need to be dispelled about suicide. These include the belief that those who talk about it never do it; two thirds of people who commit suicide have mentioned their suicidal ideas and a third have expressed clear suicidal intent. Conversely, many also believe that asking about suicide might make it more likely. Good risk assessment and management are essential. The second misconception is that suicide is often a "rational choice". In fact, most people who kill themselves are mentally disordered at the time, usually with depression but also alcoholism and schizophrenia. Treatments are available for each and need to be utilised, but sometimes lack of knowledge of the skills of other professionals can mean patients are not referred. As Goldberg and Gater 1991 have cautioned, "those who work only in the environment of their own profession, (or in isolation) tend to develop the idea that if someone cannot be helped by their own brand of intervention then they cannot be helped at all and can therefore be discharged to suffer on their own." A further important message which is not clearly understood, even by general practitioners, is that mood can be significantly improved and suicidal risk reduced even in the face of seemingly overwhelming life events and circumstances such as severe physical illnesses.

The final major and particularly important misconception is that "if they want to do it, they will do it anyway." Unavailability of "acceptable" means, treatment of depression, life events and social support can all intervene and remove, or significantly reduce, risk. Patients who have been fortuitously intercepted during a determined suicide attempt and who have recovered and gone on to lead a meaningful existence are relatively common in clinical practice.

5.9 Mental Health Promotion and Prevention

Within primary prevention, there are three principal strategies (Mrazek and Haggerty 1994):

1. Preventing the occurrence of the risk factor
2. Improving the coping response
3. Altering environmental settings

The first strategy assumes that it is possible to control or prevent the occurrence of the causal agent and is called proactive primary prevention, while the second strategy assumes that the agent, if unavoidable, can be resisted and is termed reactive primary prevention. Thus reactive primary prevention can occur before or after the stressor, but it is aimed at preparing the individual to react effectively to the stressor. In contrast, proactive primary prevention attempts to avoid the stressor altogether.

Potentially targetable stressful life events include both predictable transition points or "normal" crises which are potentially periods of lengthened stress (e.g., starting school, moving school at 11, job entry, retirement) and which may be judiciously prepared for by parents, schools and workplaces if supported and educated to do so.

Unpredictable events or crises, such as unemployment, physical injury or war, cannot usually be anticipated or avoided, but social support can nonetheless be delivered in a variety of ways, including promoting existing natural support systems, creating new but natural support systems, educating carers, consulting specific organisations such as schools (Leaf et al. 1997; Sampaio Faria et al. 1997; Mubbashar 1997), police, the criminal justice system and social services, developing local community alliances and, last but not least, establishing a public education programme about mental health and illness.

Research indicates that local residents often have minimal knowledge about mental health services offered in their vicinity. It is particularly important to reduce stigma, and schools and the media have an important role here. Such mental health education aims to develop important competence within normal groups and groups at risk in order to improve the capacity to cope both with predictable life transitions and with less predictable stress factors.

5.10
Introduction of Mental Health Policies in the Workplace

While employers have long valued the physical health of their work force, and more recently have paid particular attention to reducing the health problems associated with alcohol, drugs and the human immunodeficiency virus (HIV), they have tended to ignore mental health problems. The reasons for this lack of attention to mental health are mixed – the stigma attached to mental illness, the fact that the common mental disorders of depression and anxiety are often hidden, the relative paucity of concrete research data on the full economic and social costs of mental health problems in the workplace – and employers have historically taken the view that, since they do not cause the majority of mental health problems, they do not need to address them. However, whether or not employers cause mental health problems, the economic burden of mental health problems in the workplace (reduced performance, labour turnover, sickness absence, accidents) is very great (Jenkins 1985a,b; Kessler and Frank 1997), and this burden is borne by both employers and society.

Health policies in the workplace are systematic statements of action on health agreed between staff, management and appropriate bodies such as unions. Mental health components may include a commitment to a healthy workforce; placing a high value on both physical and mental health; acknowledging that mental health problems have many causes, including stresses in the workplace and in the outside world; listing factors which may lead to increased stress in the organisation (customised based on discussion with staff and needs assessment); recognising that domestic factors (such as housing, family problems and bereavement) may add to levels of stress experienced by employees; and stating that the organisation is committed to an explicit course of action. This course of action might include increasing understanding of the causes of mental health problems in the workforce; action to combat workplace stresses and helping staff to manage stress; managing mental health problems that occur effectively through early recognition and appropriate management (including early access to counselling, providing advice on sources of help); and taking action to manage the return to work of those who have suffered mental health problems to ensure that their skills are not lost to the firm. Once employers recognise the cost to their businesses of inefficiency caused by mental disorders and employee turnover, they can became advocates for more effective services and greater access to services.

5.11
Development of Good Processes of Care

There needs to be a systematic approach to the care of mentally ill people, based on a systematic assessment of needs, both health and social; the production of a plan to meet those needs; the coordination of care by a nominated health or social professional; and regular review as appropriate until the person is better.

5.12
Risk Assessment and Management

A full assessment of risk should be part of the assessment process, covering both risk to the patients themselves and risk to others. Any risk identified will need to be managed appropriately. The key principle of risk assessment is to use all available sources of information – a proper assessment cannot be made in the absence of information about a patient's background, present mental state, social functioning, or past behaviour. In addition to the treatment team and the patient, sources may include relatives, carers, friends, the police, probation officers, housing departments, social workers, local press reports and concerns expressed by neighbours.

It is often possible to identify circumstances under which, based on past experience, it is likely that an individual will present an increased risk. An assessment can then go on to indicate what must change to reduce this risk, to propose how these changes might be brought about and to comment on the likelihood of interventions successfully reducing risk. An example might be when a patient stops medication or abuses alcohol or drugs. As well as undergoing training in these areas, team members will need to be aware of the underlying risk factors for suicide and be able to ask patients about possible suicidal intent. The period immediately after discharge from hospital is a time of particularly high risk of social withdrawal or violence and suicide, emphasising the need for proper assessment prior to discharge and effective follow-up afterwards.

5.13
Auditing

Even in the best-run service, the possibility remains that something may go wrong. It is essential that all those involved in planning, purchasing and providing mental health services investigate any serious assault involving a mentally ill person quickly and objectively and learn the lessons from any inquiry.

Optimal audits require separate and joint focus on intake (the decision of who to treat and who not to treat as well as decisions regarding allocation of patients across diverse service offerings), processes, outcome and follow-up. As noted earlier, it is important that these audits are genuine efforts to provide independent evaluation of quality at each of these critical phases of the treatment process. It is all too common that this is not the case and that ostensible auditing degenerates to a sleight-of-hand exercise that does more to justify current practice than to provide an independent review that might lead to necessary changes in the system.

Audit reviews should be comprehensive and multidisciplinary, involving everyone associated with the patient's care, including the primary care team and the patient's family. The purpose is not to find culprits, but to identify any lessons that can be learnt to prevent such incidents in the future. Auditing is of course only part of the picture, and all auditing should take place within the context of systematic monitoring of outcomes.

5.14
Use and Disclosure of Patient Information

The use of information is a sensitive issue for mental health services. There have been serious incidents where failure to pass on information, e.g., about a particular patient's previous acts of violence, has put a patient, member of staff or the public at risk. As a matter of good practice, it is vital to share information if multi-disciplinary and inter-agency care is to function effectively. Joint planning across agencies needs to specifically address this need for exchange of information. At the same time, mentally ill people are entitled to the same confidential handling of information about their health and care as any other patient or client.

Patient information is covered by a duty of confidentiality. As a general note, information given by the patient for one purpose should not be disclosed to a third party or used for a different purpose without the consent of the patient. However, this requirement need not be so rigidly employed that it disadvantages the patient or the public interest. Thus the patient should be made aware that, in order for health and social services (or others such as probation, housing or voluntary agencies) to plan and provide effective care, personal information may need to pass between them as part of the normal processes of assessing and meeting needs. However, any information passed on should be restricted to that in which the recipient has a legitimate interest. The recipient should not transmit it to a third party unless the latter is entitled to it or the patient has explicitly consented or is aware that information needs to be passed on to enable care to be coordinated properly. There may be particular circumstances in which disclosure of information is required by the court or, exceptionally in the absence of consent, can be justified in the public interest (e.g., in certain circumstances, this may be so if someone has a history of violence). Disclosures based on public interest involve weighing that interest against the duty of confidentiality in the particular set of circumstances. The balance can be delicate and it may be necessary to take legal advice. It is important to share the right amount of information with those who need to know. For example, telling staff that an individual is violent is likely to be less helpful than detailing in what circumstances he or she is likely to be violent and how such situations can be avoided.

5.15
Local Manpower Strategies

It is important to develop and retain sufficient manpower to deliver good mental health services. This prompted the German government, for example, to issue legislation in 1991 stipulating the minimum number of doctors, psychologists, nurses, social workers and occupational therapists in psychiatric hospitals according to the individual requirements of each hospital. Important aspects in retaining good staff include attention to induction processes, mentoring schemes and occupational health services. They also include sufficient flexibility in the job plan to allow the development of specific management expertise, involvement with professional bodies, development of research and academic skills, e.g., protected time for teaching and research, appropriate links to academic institutions and development of clinical expertise in special areas such as psychotherapy and forensic work.

6
Conclusions

Good services are those which show respect for individuals and their social, cultural, ethnic, religious and philosophical principles; where individuals' needs are taken fully into account; where treatment is provided in the least restrictive environment possible and is aimed at promoting each individual's self-determination and personal responsibility; and where care and treatment aim to achieve each individual's highest attainable level of health and well-being. Success depends on a critical analysis of where we are now, a vision of where we want to be, a strategy of how to get there and commitment from politicians, professionals, managers and the general public.

7 References

**Armstrong E (1997) The primary mental health care toolkit. Royal College of General Practitioners, London

Bierman KL (1986) Process of change during social skills training with preadolescents and its relation to treatment outcomes. Child Dev 57: 230–240

Bond LA, Compass BE (1989) Primary prevention and promotion in schools. Sage, Newbury Park

Commonwealth of Australia (1994) Better health outcomes for Australians – national goals, targets and strategies for better health outcomes into the next century. Australian Government Publishing Service, Canberra

Commonwealth of Australia (1996) National Mental Health Report 1995. Commonwealth Department of Health and Family Services, Canberra

Conway M, Shepherd G, Melzer D (1996) Effectiveness of intervention for mental illness – implications for commissioning. In: Thornicroft G, Strathdee G (eds) Commissioning mental health services. HMSO, London, pp 247–264

Croghan TW, Johnstone BM, Buesching DP, Gorospe KJ, Kessler RC (1998) Information needs for medication coverage decisions in a state medicaid program (in preparation)

Department of Health (1992) The health of the nation: a strategy for health in England. Cm 1986. HMSO, London

**Department of Health (1994a) Mental illness, (Health of the nation key area handbook) 2nd edn. HMSO, London

Department of Health (1994b) Report of the Working Group on High Security and Related Psychiatric Provision. Department of Health, London

Department of Health (1994c) Confidential enquiry into homicides and suicides by mentally ill people. Preliminary report on homicide (Reed Report). Department of Health, London

**Department of Health (1995) Building bridges. Department of Health, London

**Department of Health (1996) ABC mental health in the workplace. HMSO, London

Durlak JA (1995) School-based prevention programmes for children and adolescents. Sage, Thousand Oaks

Farrar M (1996) Monitor quality. In: Thornicroft G, Strathdee G (eds) Commissioning mental health services. HMSO, London, pp 293–308

Glover G (1996) Mental illness needs index (MINI). In: Thornicroft G, Strathdee G (eds) Commissioning mental health services. HMSO, London, pp 53–58

Glover G, Gould K (1996) Performance indicators in mental health services. In: Thornicroft G, Strathdee G (eds) Commissioning mental health services. HMSO, London, pp 265–272

Goldberg D, Gater R (1991) Estimates of need. Psychiatr Bull 15: 593–595

Griffiths S, Wiley I, Jenkins R (eds) (1992) Creating a common profile for mental health. HMSO, London

Hawkins J, Catalono RF, Morrison DM, O'Donnell J, Abbott RD, Day LE (1992) The Seattle Social Development Project: effects of the first four years on protective factors and problem behaviours. In: McCord J, Tremblay R (eds) Preventing antisocial behaviour: interventions from birth through adolescence. Guildford, New York

Jenkins R (1985a) Minor psychiatric morbidity in employed young men and women, and its contribution to sickness absence. Br J Indust Med 42: 147–154

Jenkins R (1985b) Minor psychiatric morbidity and labour turnover. Br J Indust Med 42: 534–539

**Jenkins R (1990) Towards a system of outcome indicators for mental health. Br J Psychiatry 157: 500–514

Jenkins R (1992) Developments in primary care of mental illness – a forward look. Int Rev Psychiatry 4: 237–242

Jenkins R (1994a) The health of the nation – recent government policy and legislation. Psychiatr Bull 18: 324–327

Jenkins R (1994b) Viewpoint – mental health at work – why is it so under-researched J Occup Med 43: 65–67

Jenkins R (1996a) England's policy on severe mental illness. Epidemiol Psychiatr Social 5: 31–37

Jenkins R (1996b) Psychiatry and the health of the nation: the view from the Department of Health. Br J Hosp Med 56: 155–158

Jenkins R, Coney N (eds) (1992) Prevention of mental ill health at work. HMSO, London

Jenkins R, Field V (1996) Primary care of schiozophrenia, 2nd edn. HMSO, London

Jenkins R, Griffiths (eds) (1991) Indicators for mental health in the population. HMSO, London

Jenkins R, Knapp M (1996) Use of health economic data by health administrators in national health system. In: Morscarelli M, Rupp A, Sartorius N (eds) Economics of schizophrenia. Wiley, New York

Jenkins R, Üstün TB (eds) (1997) Mental health promotion and prevention in primary care. Wiley, New York

Jenkins R, Warman D (eds) (1993) Promoting mental health policies in the workplace. HMSO, London

**Jenkins R, Griffiths S, Hawton K, Morgan G, Tylee A, Wylie I (eds) (1994) Prevention of suicide. HMSO, London

**Jenkins R, Bebbington P, Brugha T, Farrell M, Gill B, Lewis G, Meltzer H, Petticrew M (1997a) The national psychiatric morbidity surveys of Great Britain – strategy and methods. Psychol Med 27: 765–774

**Jenkins R, Bebbington P, Brugha T, Farrell M, Gill B, Lewis G, Meltzer H (1997b) The national psychiatric morbidity surveys of Great Britain – initial findings from the household survey. Psychol Med 27: 775–789

Johnson S, Thornicroft G, Strathdee G (1996) Assessing population needs. In: Thornicroft G, Strathdee G (eds) Commissioning mental health services. HMSO, London, pp 37–52

Kellam SA, Rebock GW (1992) Building development and aetiological theory through epidemiologically based preventive intervention trials. In: McCord J, Trembloy RE (eds) Preventing antisocial behaviour: interventions from birth through adolescence. Guildford, New York, pp 162–195

Kessler RC, Frank RG (1997) The impact of psychiatric disorders on work loss days. Psychol Med 27: 861–874

Kessler RC, Olfson M, Berglund M, Katz SJ, Lin E, Leaf P (1997a) Differences in the use of psychiatric outpatient services between the United States and Ontario. N Engl J Med 226: 551–557

Kessler RC, Mickelson KD, Zhao S (1997b) Patterns and correlates of self help group membership in the US. Social Policy 27: 27–46

Kessler RC, Olfson M, Berglund PA (1998) Patterns and predictors of treatment contact after first onset of a psychiatric disorder. Am J Psychiatry (in press)

Kingdon DG (1994a) Interprofessional collaboration in mental health. J Interprof Care 6: 141–148

Kingdon DG (1994b) The care programme approach. Psychiatr Bull 18: 68–70

Kingdon DG (1994c) Making care programming work. Adv Psychiatr Treat 2: 41–46

Kingdon D, Jenkins R (1995) Suicide prevention in England. Ital J Suicidol 5: 9–17

Kingdon D, Jenkins R (1996) Adult mental health policy. In: Thornicroft G, Strathdee G (eds) Commissioning mental health services. HMSO, London, pp 1–12

Knudsen HC, Thornicroft G (1996) Mental health service evaluation. Cambridge University Press, Cambridge

Leaf PJ, Bognor M, Webb MB (1997) The East Baltimore mental health partnership. In: Henggeler SE, Santos AB (eds) Innovative approaches for difficult to treat populations. American Psychiatric Press, Washington, pp 117–138

Lloyd K, Jenkins R (1995) The economics of depression in primary care – Department of Health initiatives. Br J Psychiatry 166 [Suppl 27]: 60–62

Lloyd K, Jenkins R, Mann A (1996) The longterm outcome of patients with neurotic illness in general practice. Br Med J 313: 26–28

Meltzer H, Gill B, Petticrew M, Hinds K (1995a) OPCS surveys of psychiatric morbidity in Great Britain, report no 1. The prevalence of psychiatric morbidity among adults living in private households. HMSO, London

Meltzer H, Gill B, Petticrew M, Hinds K (1995b) OPCS surveys of psychiatric morbidity in Great Britain, report no 2. Physical complaints, service use and treatment of adults with psychiatric disorders. HMSO, London

Meltzer H, Gill B, Petticrew M, Hinds K (1995c) OPCS surveys of psychiatric morbidity in Great Britain, report no 3. Economic activity and social functioning of adults with psychiatric disorders. HMSO, London

Meltzer H, Gill B, Petticrew M, Hinds K (1996a) OPCS surveys of psychiatric morbidity in Great Britain, report no 4. The prevalence of psychiatric morbidity among adults living in institutions. HMSO, London

Meltzer H, Gill B, Petticrew M, Hinds K (1996b) OPCS surveys of psychiatric morbidity in Great Britain, report no 5. Physical complaints, service use, treatment of residents with psychiatric disorder. HMSO, London

Meltzer H, Gill B, Petticrew M, Hinds K (1996c) OPCS surveys of psychiatric morbidity in Great Britain, report no 6. Economic activity and social functioning of residents with physical disorders. HMSO, London

Meltzer H, Gill B, Petticrew M, Hinds K (1996d) OPCS surveys of psychiatric morbidity in Great Britain, report no 7. Psychiatric morbidity among homeless people. HMSO, London

Miller V, Rosen A, Parker G (1995) A guide to standards of care and quality assurance for area integrated mental health services (AIMHS). AIMHS Standards Project, Chatswood, Australia

Ministry of Health (1997a) An approach for action – phase two in the development of a national strategy to help prevent youth suicide in New Zealand. Ministry of Health, Wellington

Ministry of Health (1997b) The national mental health standards. Ministry of Heath, Wellington

Mohit A (1997) Training packages in developing countries. In: Jenkins R, Üstün TB (eds) Preventing mental illness – mental health promotion in primary care. Wiley, Chichester, pp 253–259

Mrazek PJ, Haggerty RJ (1994) Reducing risks of mental disorders – frontiers for preventive intervention research. National Academy Press, Washington

Mubbashar M (1997) School mental health program in Pakistan. In: Jenkins R, Üstün TB (eds) Preventing mental illness – mental health promotion in primary care. Wiley, Chichester, pp 329–336

Murray JL, Lopez AD (1996) The global burden of disease. Harvard University Press/WHO, Boston

Murthy RS (1997) Applications of interventions in developing countries. In: Jenkins R, Üstün TB (eds) Preventing mental illness – mental health promotion in primary care. Wiley, Chichester, pp 117–130

NHS Executive (1996a) The spectrum of care: local services for people with mental health problems. Department of Health, Leeds

NHS Executive (1996b) 24-Hour nursed care for people with severe and enduring mental illness. Department of Health, Leeds

Paykel E, Jenkins R (1996) Prevention of mental disorder. Gaskell, London

Rosen A, Miller V, Parker G (1995) Area integrated mental health services standards (AIMHS). AIMHS Standards Project, Chatswood, Australia

Rotheram MJ (1982) Social skill training with underachievers, disruptive and exceptional children. Psychol Schools 19: 532–539

Royal College of Psychiatrists (1996) Report of the confidential inquiry into homicides and suicides by mentally ill people. Royal College of Psychiatrists, London

Rutz W (1996) Prevention of suicide and depression. Nordic J Psychiatry 50 [Suppl 37]: 61–67

Sampaio Faria J, Weare K, Gray G (1997) Mental health promotion in schools. In: Jenkins R, Üstün B (eds) Preventing mental illness. Wiley, Chichester

Schmidtke A, Hafner H (1988) The Werther effect after television films: new evidence for an old hypothesis. Psychol Med 18: 665–676

Strathdee G, Jenkins R (1996) Purchasing mental health care for primary care. In: Thornicroft G, Strathdee G (eds) Commissioning mental health services. HMSO, London, pp 71–84

Stevens A, Gabbay J (1991) Needs assessment. Health Trends 23: 20–23

Tessler RC (1987) Continuity of care and client outcome. Psychosocial Rehabil J 11: 39–53

Tessler RC, Willis G, Gubman GD (1986) Defining and measuring continuity of care. Psychosocial Rehabil J 10: 27–38

Test MA (1979) Continuity of care in community treatment. New Direction Mental Health Services 2: 15–23

Üstün TB, Sartorius N (1995) The background and rationale of the WHO Collaborative Study on Psychological Problems in General Health Care. In: Üstün TB, Sartorius N (eds) Mental illness in general health care. Wiley, Chichester

Wig NN, Murthy R (1994) From mental illness to mental health. Health Millions 20(4): 2–4

Wing JK (1992) Epidemiology based mental health needs assessment: a review on psychiatric disorders. ICD10 F2-F6. HMSO, London

Wing J, Curtis R, Beevor A (1996) Health of the nation outcome scales (HoNOS). Royal College of Psychiatrists Research Unit, London

World Health Organisation (1996) Diagnostic and management guidelines for mental disorders in primary care. ICD10. Chapter V: Primary care version. Hogrefe and Huber/WHO, Göttingen

CHAPTER

12

W. Gaebel

Quality Assurance in Psychiatry

Translator: S. Johnson

1 General Principles

Attempts to regulate the quality of clinical practice date back as far as medicine itself. The Hippocratic Oath includes as fundamental principles that doctors should use their ability and judgement to benefit patients as much as they can and that they should refrain from any harmful or mischievous course of action. Safeguarding and improving existing standards of practice is now accepted as a central priority. The introduction of measures aimed at quality assurance (QA) is in part a response to the growth of knowledge and to increasing specialisation. However, it also reflects increased public scepticism about medical practice and, last but not least, pressures to contain costs because of increasingly restricted resources in public sector health services (Eichhorn 1997).

Medical care encompasses not only diagnosis and treatment, but also prevention, rehabilitation and aftercare. Planning, organisation and co-ordination of this whole field requires an overall framework of health policies. However, the starting point for the safeguarding and, where possible, improvement of medical care needs to be the definition of quality standards. Following this, methods need to be developed for implementing these standards and for continuously monitoring how far they have been attained. Quality assurance in medicine should provide patients with a guarantee that they will receive treatment which is rational according to the current status of specialist knowledge. This applies in psychiatry as in all other areas of health care.

In psychiatry, a systematic approach to quality assurance is particularly necessary. This is because psychiatric disorders are common and their predisposing, precipitating, maintaining and modifying factors are complex, so that diagnosis, treatment and rehabilitation need to be multi-dimensional and to be provided in a diverse range of care settings. This chapter will set out the conceptual, methodological and organisational framework required for such a systematic quality assurance.

1.1 Concepts and Definitions

The starting point for a concept of quality assurance in medicine must be a definition of quality. Outside medicine, in manufacturing and service industries, quality control and quality assurance are familiar concepts. Here the Deutsche Institut für Normung (DIN; the German Standards Authority) defines quality as the "whole set of characteristics of an entity which determine its suitability to fulfil specified and predetermined requirements" (DIN ISO 8402; DIN 1992). Within the field of medicine, Fifer (1980) gives an operational definition of quality as "the degree of adherence to a standard", thus making it clear that one can only assess quality in relation to clearly defined standards.

The main ways of defining standards are on the basis of statistical/quantitative norms and/or of qualitative norms; these two should generally be combined so that they complement each other. Examples of quality standards in medicine are those published by the American Medical Association (1986):

- Emphasis on health promotion, prevention of illness and disability, early recognition and treatment
- Treatment provided in a timely manner, without inappropriate curtailment or unnecessary delay, discontinuity or prolongation
- Achieving the informed co-operation and participation of the patient in the care process and in decisions about that process
- Treatment based on accepted principles of medical science and proficient use of appropriate technologies and professional resources
- Sensitive care that takes into account the anxiety and tension which illness causes; concern for the welfare of the whole family
- Attainment of a good treatment outcome through efficient use of technologies and other health system resources
- Sufficient documentation of the patient's condition to allow continuity of care and peer evaluation

These standards are fully applicable to psychiatry.

Specific treatment standards should increasingly be based on the results of empirical evaluations of treatment, taking into account modifications needed to apply these experimental treatments in clinical practice.

"Evidence-based medicine" (Ellis et al. 1995; Naylor 1995) or in this case evidence-based psychiatry (Goldner and Bilsker 1995) starts from the assumption that medical practice should be based on the best possible information. This will be the results of studies which satisfy appropriate scientific standards, evaluated through meta-analysis. This requires a paradigm shift:

Evidence-based medicine places less weight on intuition, unsystematic clinical experience and pathophysiological basic principles than on having adequate foundations for making clinical decisions. The emphasis is on seeking empirical evidence by means of clinical research (Evidence-Based Medicine Working Group 1992, as quoted by Goldner and Blisker 1995).

With this aim, practising clinicians should make great efforts to achieve a better balance between the use of clinical experience and acumen (*phronesis* in the Aristotelian sense) and the application of scientific knowledge (Aristotle's *techne*) (Goldner and Bilsker 1995). Thus doctors must become well acquainted with scientific ways of thinking and must be in a position to inform themselves about currently available empirical knowledge. Study results, meta-analyses or guidelines available via modern information systems can provide the necessary foundation for this (e.g. see Antes et al. 1995; Cochrane Collaboration Working Group 1996).

The basic concepts in quality assurance are described – not always in a uniform way – by the expressions "norm", "criterion", "standard", "indicator", "threshold value" and "guideline" (Bertolote 1993; Donabedian 1982; Fauman 1989; Wilson and Phillips 1992c; see Table 1).

According to this terminology, the dosage specified in a psychopharmacological formulary is an example of a quality *norm*, established by empirical and statistical means. An example of a corresponding quantitative *criterion* is the range of doses across which a drug is effective for the average subject. A *standard* is the normative specification of a particular range of doses which indicates optimum quality. For various reasons, total unanimity cannot be expected (see below). A *threshold value* or reference range therefore needs to be specified, above or below which a review of quality (audit) should be triggered. Possible methods for evaluating quality of treatment based on a particular formulary include assessing the extent of deviations from this dosage standard. The goal for *guidelines* for treatment with a particular formulary would be to provide instructions on how to achieve optimal treatment. The guidelines would need to include practical explanations of indications, dosage and duration of treatment, investigations which must routinely be carried out, contra-indications and modifications required for particular individuals or situations.

Standards are thus the reference points for medical care based on the principle of high quality. They indicate the current status of medical knowledge and clinical practice and thus also provide foundations for basic and higher training and for continuing professional development. They establish the scientific foundations on which the quality assurance process should be based (Gaebel 1995a).

In applying standards, the risks, benefits and costs of trying to achieve particular goals need to be weighed, and standards need to be adapted to varying needs at different stages in treatment (see Linden 1994). Standards are not rigid prescriptions, and their applicability is limited by doctors' freedom to make decisions and practice according to their own clinical judgement (Buchborn 1993). Finally, an important principle is that, in certain individual cases, it will be reasonable and necessary to depart from the principles which have been established statistically as valid for groups. Unless this principle is recognized, the "art" of medicine will diminish into a merely mechanical process, and the development of innovative treatments will cease to be possible. Because of its lack of clarity – and the overtones of strict obligation and of a threat of being held legally accountable – it might be best to abandon the use of the expression standard (see Selbmann 1996).

In medico-legal terminology, the term "regulation" is applied to the binding rules governing medical

Table 1. Fundamental concepts in quality assurance

Concept	Definition
Norms	The general precepts on the basis of which quality is defined (Donabedian 1982)
Criterion	A quantifiable characteristic which can be used in assessing quality of care (Donabedian 1982)
Standard	The precisely defined level of quality which is deemed adequate, acceptable or optimal (Donabedian 1982)
Indicator	The variables measured in order to detect deviations from standards (Bertolote 1993)
Threshold value	The predetermined value of an indicator at which an intensive assessment is triggered (Fauman 1989)
Guideline	Systematically developed sets of precepts which provide health professionals and patients with a basis for provision of appropriate care in particular clinical situations (Agency for Health Care Policy and Research, as quoted in AWMF 1995)
Structure (structural quality)	The physical, organisational and other characteristics of the system of care and of its context (Donabedian 1966, 1986)
Process (process quality)	Encompasses all the activities which contribute to the care of the patient (Donabedian 1966, 1986)
Outcome (outcome quality)	What is achieved: generally an improvement in health, but also improvements in practice and knowledge which may benefit health in future (Donabedian 1966, 1986)

practice. Guidelines are formulated on the basis of diagnostic and therapeutic standards, while recommendations and statements of opinion simply provide information and suggestions for management (Klinkhammer 1995).

Guidelines (see below) are not absolute rules, but doctors do have some obligation to take account of them. They must correspond to the current state of knowledge. According to Sect. 70 of the Fifth Book of the German Social Welfare Code (see below), they must also be confined to the basic steps which are necessary to achieve the goal and must take account of economic viability.

1.2 Domains of Quality

More detailed discussions of quality and of quality assurance measures are usually based on Donabedian's (1966) model, which distinguishes between structure, process and outcome domains of quality (see Table 1).

Ideally, one might expect outcome quality to reflect the combined effects of structure and process quality in care, so that it would be the fundamental domain for evaluation of the whole process of care. However, even given state-of-the-art treatment, the relationship between intervention and outcome is not necessarily a straightforward linear one. A great many modifying variables, including variations in the natural history of disorders, co-morbidity and compliance, also influence outcomes, which may thus be suboptimal with ideal conditions and inputs. These modifying variables must therefore be taken into account in any evaluation of outcomes (Fauman 1989), e.g. when comparing institutions whose patient populations have markedly different initial characteristics. Essentially, the attainment of the best possible structure and process quality through implementation of valid standards can only increase the *likelihood* of an optimum outcome (Schybe and Prevost 1990). Few investigations of these relationships have so far been published.

1.3 The Quality Improvement Cycle

Quality control and quality assurance are elements in a cyclical process whose goal is quality improvement: the Plan–Do–Check–Act (PDCA) cycle (Fig. 1) (see Selbmann 1995). Thus planning related to quality (problem analysis), monitoring (problem recognition, evaluation) and implementation of solutions to problems and of safety measures make up a continually repeated series of stages.

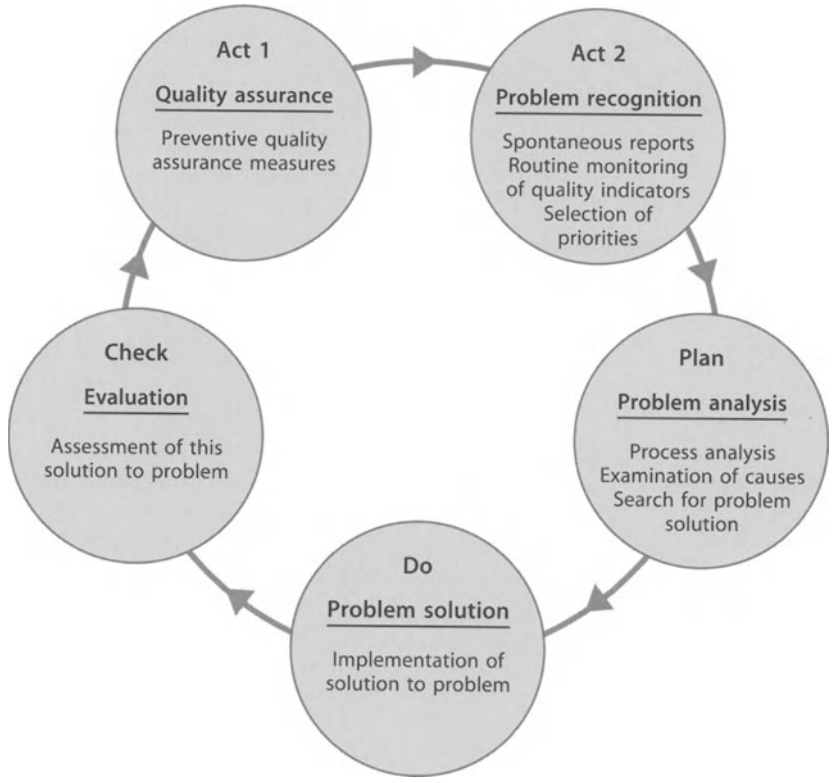


Fig. 1. Problem-oriented quality assurance

From a production engineering perspective, quality control can be understood as the whole procedure through which the actual value of a particular target variable, e.g. an actual treatment process, is compared with its *desired* value, as defined by a guideline. Applying the domains of quality outlined above, structural parameters, ways and levels of functioning and outcome criteria are all variables which can thus be used as starting points for such analyses.

Quality assurance consists of the problem analysis which takes place when a deviation is found from a fixed range of acceptable values and of the corresponding problem solution (Fifer 1980). The terminology of the German Standards Authority characterises quality assurance as "all the planned and systematic activities which are necessary to achieve reasonable confidence that a product or service will fulfil the quality requirements specified for it" (DIN ISO 8402; DIN 1992). "Confidence through quality" is also the motto of a quality assurance programme in a Munich hospital (Piwernetz et al. 1991).

In North America, a distinction is drawn between quality assurance and utilisation review (UR). Quality assurance is concerned with the question of whether care is in keeping with accepted medical standards and thus whether it fulfils specific quality criteria. In contrast, utilisation review investigates the cost-benefit relationships which may be arrived at within existing resources (Sederer 1991). Utilization review is carried out by specific organisations and involves exercises such as peer review of cases identified by specific criteria (e.g. length of stay). Together with the Health Maintenance Organizations (HMO), it is a component of managed care (see Arnold et al. 1997). Indirect cost control through examination and review of clinical practice (utilisation review), control of service budgets and, in the case of HMO, restricting the choice of doctor to certain practitioners with contracts with the organization are mechanisms used to switch from a fee-for-service system to a system of flat rates. The principle is insurance and service delivery from a single provider (Selbmann 1997), and the patient becomes a partner in the insurance company and must thus be involved in judging the quality of services. To guard against deterioration in quality, systems for accreditation and quality management need to be put in place. Advantages of HMO are the possibilities of better horizontal (between different treatment facilities) and vertical (between levels of care) integration in health care provision (Selbmann 1997).

The central idea behind the concept of "continuous improvement of care" is that of a continual process of quality improvement, which is decentralised and self-regulating as opposed to centrally controlled.

1.4 Current Status and Models

In planning and implementation, a distinction should be made between internal and external quality assurance measures. Internal quality assurance programmes are independent and specialty specific, have explicit criteria, are based on sampling frames which are as representative as possible and apply to specific illnesses or specific situations. They are to be preferred to purely external types of programmes (Eichhorn 1987).

However, fulfilling legal requirements (see below) largely involves external quality assurance measures. Without internal measures, external programmes cannot be effective; in practice they need to complement each other in a meaningful way.

In U.S. health services, external models of quality assurance have a longer tradition than in Germany (see Kaltenbach 1991; Sederer 1991). Methods of investigating quality such as utilisation review and quality assurance originated in the early 1960s. Following the introduction of the Medicare programme and the funding of the Community Mental Health Care movement by the U.S. Government, mechanisms needed to be developed for controlling rapidly escalating costs. In 1972, organisations called Professional Standard Review Organizations (PSRO) were established by law. These were local medical organisations within the public health service entrusted with control of resource allocation in inpatient and occasionally also outpatient services. Conflicts between cost control and quality improvement led in 1977 to the PSRO being placed under the scrutiny of the Health Care Financing Administration (HCFA), a division of the Social Security Administration, with a view to more vigorous cost control. Because costs continued to escalate in the public health service and medical organisations continued to oppose the PSRO, they were replaced in 1982 by bodies called Peer Review Organizations (PRO). These were organised at state level and have been associated with a system of flat rates per case, based on diagnosis. In psychiatry, flat rates per case have not so far been introduced. PRO are now generally set up and maintained by state medical associations. Their main task is to use utilisation review and quality assurance to control the costs of Medicare, but also of Medicaid and other funders. The goal is to reduce the use of expensive methods, e.g. outpatient treatments are favoured rather than hospital admissions. Over the years, their remit has developed to encompass quality assurance as well as cost control. Utilisation review has become an element in managed health care systems, in which HMO make use exclusively of selected doctors, their own protocols and accredited institutions. The

work of the Joint Commission on the Accreditation of Hospitals (now the Joint Commission on Accreditation of Health Care Organizations) has defined and operationalised quality assurance, making use of guidelines. This forms the basis on which these clinics are obliged to identify and resolve problems in care. Initially, use was mainly made of structural and process measures, but outcome-oriented quality criteria play an increasing role in a system of care which is now more tightly regulated. For some years, the American Psychiatric Association (APA) played a substantial role in development of quality assurance standards. However, it withdrew from this activity at the end of the 1980s because of the increasing economic interests of the bodies involved (Hamilton 1992). Its main activities now lie in the development of practice guidelines (see below).

In the Federal Republic of Germany, the development of quality assurance measures in medicine dates from the mid-1970s, when quality assurance programmes were initiated in neonatology, gynaecology and general surgery (Eichhorn 1987). These have subsequently been implemented nationally. Other examples of activities aimed at external quality assurance are found in laboratory medicine, radiology, nuclear medicine, pathology, cardiac surgery, neurosurgery and paediatric cardiology.

In these programmes, special multi-centre surveys take place during an annual study period examining characteristics relevant to quality. For example, for the surgical treatment of gallstones, patient characteristics, risk factors, treatment procedures, intraoperative diagnosis, results of operation and outcome at follow-up are recorded. Survey statistics are used to record trends and to compile a profile for each clinic. The position of each clinic can then be evaluated in relation to the average or to a fixed standard, and this is reported back to a central project office. Thus identification of weaknesses in particular places, initially on an anonymous basis, can serve as a starting point for the analysis of deficiencies in treatment and of how these may be remedied (Baur-Felsenstein 1994).

1.5 Range of Application

Quality assurance can be applied at a variety of levels of planning and of care (Bertolote 1993). Health policy and programmes are the highest level at which quality standards can policy be established. They are implemented within a health care system, which is made up of treatment facilities (inpatient/partial hospitalisation/outpatient and community/social care services), in which particular treatment procedures are available, which are applied for specific illnesses. The domains of quality discussed above can usefully be applied within each of these levels.

A variety of legal requirements also need to be observed, and many of these are specific to a region (*Land*) in Germany. German law regulating medical practice requires all doctors to implement measures for quality assurance in clinical practice. The law regulating doctors who treat patients under insurance schemes stipulates further measures aimed at quality assurance, and here groups known as quality circles (see below) have a particular significance. Finally, regional laws regulating local hospitals require internal and external quality assurance in the hospital. The fifth book of the Social Welfare Code is specifically concerned with quality assurance and includes regulations relating to care by doctors with health service contracts (Sect. 135), to prevention and rehabilitation in community care (Sect. 135a) and to hospital treatment (Sect. 137). For example, Sect. 137 states that "hospitals registered under Sect. 108 are obliged to implement quality assurance measures. These measures must encompass quality of treatment procedures, of the process of care and of treatment outcomes. They should be designed so to allow comparative investigations." Thus this section relates to external quality assurance. Section 70 states that the basis of service provision of appropriate quality is that "health insurers and care providers should guarantee to insured patients service provision which is appropriate to needs, equitable and reflects the generally agreed status of medical knowledge." Treatment must therefore be "adequate and goal-directed", "must not go beyond what is necessary" and must be "carried out economically".

The following section will discuss in greater detail specific applications of quality assurance in psychiatry.

2 Applications in Psychiatry

In psychiatry, the range of application for quality assurance measures includes all aspects of multidisciplinary mental health care, from diagnosis to identifying indications for and carrying out treatment, prevention, rehabilitation and aftercare. Institutional structures and the administration and organization of care also fall within their scope.

Concepts, methods and applications for quality assurance have been discussed in a series of reviews (Liptzin 1974, 1991; Mattson 1984; Fauman 1989; Bertolote 1993; Gaebel 1995a) and collections of papers (Gaebel 1995b; Haug and Stieglitz 1995; Berger and Gaebel 1996). Statements on this topic have been published by the Royal Australian and New Zealand

College of Psychiatrists (1982) and the Canadian Psychiatric Association (Cahn and Richman 1985). The Committee on Quality Assurance of the APA has issued a *Manual of Psychiatric Quality Assurance* (Mattson 1992a), which includes criteria for reviewing quality in different types of psychiatric treatment facility and methods of therapy (see below).

The specialist medical associations have played a particular role in the development and implementation of quality assurance measures (see below). Organizational elements of quality assurance programmes include the following (adapted from Eichhorn 1987):

1. Framework for quality assurance
   - Internal–external
   - Within specialty/across specialties
2. Extent of organisation of quality assurance
   - Improvised/institutionalised
3. Degree of autonomy in programme implementation
   - Autonomous/regulated by government
   - Optional/compulsory
4. Domains for assessing quality
   - Outcome/process/structure
5. Method of sampling for quality assessment
   - Total/representative group/based on selected examples
   - Random/non-random sampling
   - Selection based on illness type/based on situation
6. Degree to which indicators of quality and quality requirements are formally specified
   - Implicit/explicit
7. Timeframe for quality assessment
   - Prospective/concurrent/retrospective
8. Strategies for promoting quality
   - Organisational development
   - Training
   - System of incentives
9. Organisational and technical infrastructure for quality assurance
   - Analysis of single cases/statistical methods
   - Study of documentation/direct observation/ surveys and interviews

Particularly in psychiatry, quality standards are not the only basis for clinical judgement and care. As well as economic resources, local social and cultural norms have an important role at all levels of service planning and treatment (Bertolote 1993). As required by the World Health Organization (WHO), the views of those directly (patients) and indirectly (relatives) affected by psychiatric care are increasingly taken into account in the development and implementation of quality standards (the "community participation" and "consumerism" movements). This can also be seen as a significant step towards greater "consumer protection" in psychiatry.

The domains of quality discussed above are not fully separate and distinct in their range of application. However, for the sake of clarity, structural process and outcome quality will be described separately in the following sections.

2.1 Structural Quality

The term "structure" is used here to describe all qualitative and quantitative aspects of mental health policy and service organisation and of the buildings, rooms, staffing and equipment resources whose function is to enable needs for psychiatric care to be met. Levels of psychiatric morbidity and of those forms of social deprivation which are known to be associated with psychiatric illness can be used to assess need for psychiatric services in a region (Wing et al. 1992).

The sole basis for the humanitarian approach to care is assessment of the needs of suffering individuals. The realistic approach, on the other hand, also takes into account the degree to which a method of treatment is available and effective and can be funded. Recent expert opinion concludes that needs for care should be assessed on the basis of epidemiological data. Calculations should take into account the prevalence of psychological disorders and the evidence on their treatability as well as current levels of demand for services.

Needs-led delivery of health care is based on specific treatment goals and is provided within a structured care system. Goals are formulated in the context of health policies which provide overall guidelines for care, such as the following recommendations formulated by the *Psychiatrie-Enquete* (enquiry into German psychiatric services; BMJFG 1975) and by the 1988 Expert Commission which followed this (BMJFFG 1988; see Rössler and Salize 1995a):

- Needs-led, community-based systems of services should be established.
- There should be equity between the mentally and physically ill in terms of legal provisions and rights, financial entitlements and social status.
- There should be co-ordination and collaboration between the elements in the care system.

For the individual patient, what this should provide is a guarantee of diagnosis, treatment and rehabilitation which are tailored to his or her illness and disability, personality and living circumstances and which are the best possible on the basis of current knowledge. Reaching this goal requires the implementation of explicit guidelines within satisfactory care structures. Thus there is increasing emphasis on an approach to care which is goal-directed and tailored to individuals rather than centred on institutions.

The quality of a system of care (see Vol. 1, Part 2, Chap. 11) essentially depends on how far needs for treatment can be met in a way which is satisfactory according to clearly defined criteria (Rössler and Salize 1995b). In the Federal Republic of Germany, the 1975 *Psychiatrie-Enquete* provided a decisive impetus to improvement in the structural quality of the psychiatric care system. This was partly evaluated through model programmes and was modified and carried further in the recommendations of the 1988 Expert Commission.

Care takes place within a system with several levels, as summarised in Goldberg and Huxley's (1980) filters of care model. These levels encompass non-professional (self-help, neighbourhood, volunteer-based schemes) and professional types of help. The latter can further be divided into non-specialist, community-based primary care services (general practitioners, community nursing schemes, social services) and specialist services. Specialist services may be categorized as community and outpatient, inpatient, partial hospitalization or social care and include independent outpatient psychiatrists, social psychiatric services, day hospitals, outpatient clinics, specialist psychiatric hospitals, university units, departments in general hospitals, hostels, group homes, workshops and day centres. A variety of quantitative statistics can be used as indicators of the quality of a health care system. Examples include levels of provision of inpatient beds, the size (total number of beds) of psychiatric hospitals and the average length of stay in inpatient care.

Comparing European countries, the number of psychiatric beds per 1000 inhabitants in the period from 1987 to 1991 ranged from 0.55 in Italy to 2.6 in Luxembourg. In Germany, the average was 0.94 in 1994, with figures for different regions ranging from 0.8 to 1.5. When elderly patients were included, the figure was approximately 1.4 (Rössler et al. 1993; Rössler and Salize 1995b).

The deinstitutionalisation process has involved a restructuring of the care system in almost all Western countries. Psychiatric hospitals have diminished in size, while departments of psychiatry have been developed within general hospitals. In some areas, considerable deficits remain in the provision of day hospital places and in social care services such as community residential facilities, work schemes and day and drop-in centres.

Checklists and glossaries for evaluation of the quality of treatment facilities were published by the WHO following testing in 13 countries (Bertolote 1994). In the United States, "generic quality screens" have been developed in the course of the work of the HCFA (Mattson 1992b). These are used to examine aspects of quality of care in psychiatric hospitals on which accreditation is conditional. Likewise in Germany, criteria for structural and process quality have been published for psychiatric inpatient units (Bundesarbeitsgemeinschaft der Träger Psychiatrischer Krankenhäuser 1990; Landeswohlfahrtsverband Hessen 1991, unpublished; von Cranach 1994) and for child and adolescent psychiatric units (Bundesarbeitsgemeinschaft der Leitenden Ärzte kinder- und jugendpsychiatrischer Kliniken und Abteilungen 1993). These can be operationalised and applied in the quality assurance process.

The development of guidelines for the structural quality of other components of the care system, such as partial hospitalisation, outpatient and community programmes and social care services, has not so far been satisfactorily achieved (Spengler 1991; APA Task Force 1992; Wilson 1992; Wilson and Phillips 1992a; Bertolote 1994). This applies also to specialist units within adult psychiatry (Böhme et al. 1994; Gaebel 1995c), child and adolescent psychiatry (Bundesarbeitsgemeinschaft der Leitenden Ärzte kinder- und jugendpsychiatrischer Kliniken und Abteilungen 1993, unpublished), geriatric psychiatry (Moak 1990) and addictions (Miller and Phillips 1992). Here, the continuing discussions about the advantages and disadvantages of integration, specialisation and sectorisation are of particular relevance. Evaluative studies which have compared outcomes for hospital and community-based settings, including some which have involved health economic evaluation, have arrived at some controversial results (Häfner and an der Heiden 1994; Brenner 1995). There is also controversy about whether single or multiple levels are needed in inpatient psychiatric care systems and how tasks should be divided between general departments and specialist clinics (Rössler and Salize 1993). Currently, these ideologically charged controversies tend in any case to be resolved through a pragmatic "anything goes" approach, mainly based on local factors (Finzen 1994).

Finally, the area in which research is most markedly lacking concerns the effect on outcome quality of the architecture and design of treatment facilities, the equipment available and the organisational structures. This also applies to issues concerning appropriate planning and decision-making processes and appropriate legal frameworks for service provision (Kukla 1995).

Psychiatric treatment and care is based on a team of professionals. In outpatient care, the number of independent practitioners providing psychiatric care to outpatients under insurance schemes (*niedergelassene Nervenärzte*) rose continuously up to 1993, when the number in Germany was around 4145 (Rössler and Salize 1995a). Since then, the social welfare legislation has prescribed a level of provision of care of one such practitioner for every 17,348 inhabitants. The nature of

the reimbursement system is such that the extent of ancillary professionals working in these practices cannot be adequately assessed.

In inpatient and partial hospitalisation programmes, a significant step towards optimising process quality was made with the appearance of the 1991 decree on psychiatric staffing levels (the *Psychiatrie-Personalverordnung*), which set out a method for calculating multi-professional staffing requirements on the basis of service activity (Kunze and Kaltenbach 1994). So far, no evaluation is available of the effects on quality of this method of calculation, but Sect. 4 (4.2) of the decree (*Verordnung über Maßstäbe und Grundsätze für den Personalbedarf in der stationären Psychiatrie*. Bundesgesetzblatt 1990, part I, pp 2930–2939) does make this a fundamental requirement:

> The contracting partners have agreed on a framework of criteria which allow assessment of whether the staffing patterns stipulated in the decree result in satisfactory service provision.

Under Sect. 275 of the German social welfare code, the *Medizinischer Dienst der Krankenversicherer* (MDK; medical advisory staff of the health insurers) has an advisory and regulatory role in matters of quality assurance. This function applies to determining which treatment groups patients fall into for the purposes of calculating staffing requirements according to Sect. 4 of the *Psychiatrie-Personalverordnung* and also to ways of applying this decree in such a way that quality is improved (Banaski et al. 1993). Draft versions of a checklist for improvement of quality of care using the additional staff input for which the decree provides have been devised (Kunze et al. 1994).

There have also been discussions of the staffing structures needed in social care services, such as residential facilities and day centres, in order to be able to provide treatment and rehabilitation adapted to individual needs (Kruckenberg et al. 1994).

The qualifications of health care staff constitute a further aspect of structural quality. Appropriate higher specialist training is a requirement for good process quality (Klieser et al. 1995). A recent example is the improvement of the quality of medical training through the reorganisation of specialist training organisations, which has given doctors the option of training as specialists in psychiatry and psychotherapy (Berger 1993). Establishing new subspecialties besides those which already exist (psychotherapeutic medicine, psychotherapy, psychoanalysis) not only requires co-ordination of service provision (Deutsche Gesellschaft für Psychiatrie, Psychotherapie und Nervenheilkunde 1996, unpublished), but must also be subject to critical scrutiny from the point of view of the European harmonisation of specialist medical training which the Union of European Medical Specialists (UEMS), the European specialist medical body, is currently trying to bring about. A law regulating the involvement of psychologists in psychotherapeutic practice has not so far been passed.

Another important issue in quality assurance is the provision of psychiatric training to medical students, currently being discussed in the context of the review of basic medical education. Obviously, similar considerations apply to basic training in the other health professions as to the specialist training of members of the various mental health professions. There needs to be an investigation of ways in which continuing professional development for specialists could be made more intensive, following the example of continued medical education (CME) programmes in the United States and the recently introduced compulsory further training in Switzerland. Under medical professional regulations, continuing professional training is an obligation, and doctors must be able to demonstrate to medical associations that they are undertaking it. It includes attendance at appropriate meetings, further clinical training, study of specialist literature and use of audiovisual teaching and learning materials.

2.2 Process Quality

The term "process quality" refers to all diagnostic, therapeutic and rehabilitative activities and particularly to the extent to which the identification of indications for these and the way they are carried out conforms to guidelines. Treatment and care are increasingly needs-led; thus selective provision of services, which is determined by the types of services and staff available, is giving way to adaptive provision, which emphasizes fitting treatments to individual needs.

In considering indications for particular treatment settings within the health care system, the development of operational criteria, such as those for inpatient admission or for determining length of stay, has particular importance (Prunier and Buongiorno 1989). The availability of alternative forms of treatment is a particularly central issue for quality evaluation where compulsory inpatient admission is concerned. Community rather than inpatient care is not valid as a principle in all cases. Even where there is no immediate danger to self or others, inpatient treatment may in certain cases prevent further damage, e.g. if it protects a patient from a pathogenic environment or relieves relatives of burden.

According to Sect. 112 of the Fifth Book of the Social Welfare Code (see above), two-way contracts regulate admission and discharge of insured parties and determination of need for and duration of hospital

treatment. Section 301 sets out requirements for the contents of admission documents, including reason for admission and initial diagnosis. Section 115a regulates hospital care preceding and following hospitalisation. The purpose of care prior to admission is to establish whether admission is necessary or to prepare for it, while care following hospitalisation safeguards or consolidates improvements made during the hospital stay.

Inappropriate placement and multiple episodes of care are to be avoided and continuity of care needs to be safeguarded, extending, if necessary, over several treatment facilities. This requires a functional network between the various components of the care system. In response to the small-scale fragmentation of the regional system of care, with gaps in provision, inappropriate provision and overprovision, the concept of the community psychiatric association has developed. This includes a community outreach service, a drop-in service and a day service and has been realized in the form of social psychiatric service with a single funder or as associations in which functions are divided between facilities with a more complex funding structure. The concept of care co-ordination through case management has been applied in this context, and in retrospect it has not been clearly established that it is effective in preventing repeat admissions to inpatient services (Rössler et al. 1992). Looking to the future, there is likely to be increasing emphasis on developing differentiated services which have different main focuses but are linked in a functional network within the health care system.

The introduction and further development of operationalised diagnostic systems, such as DSM-IV (APA 1994) and ICD-10 (Dilling et al. 1992) and of clinical interview instruments has been an important advance in improving diagnostic process quality. Because of the multi-dimensional nature of causes, presentations and treatment strategies for psychiatric disorders, supplementary diagnostic methods generally also need to be applied, in keeping with advancing understanding of aetiological factors and the development of modern diagnostic methods. This means applying scales measuring symptoms and psychological tests, such as intelligence, personality and neuropsychological tests, and above all establishing a clear general medical and neuropsychiatric diagnosis, e.g. by use of imaging techniques (see Gaebel 1995c). Quality assurance in diagnosis involves supervision, specialist training meetings and case conferences as well as regular psychopathological and diagnostic rater training.

Psychiatric treatment takes a wide variety of different forms, conventionally classified into physical, psychological and sociological forms. Quality monitoring measures are needed to ensure that indications for each are properly identified and that they are combined in appropriate ways and carried out correctly (Gaebel 1995c). Psychiatric treatment here also includes prevention and rehabilitation as components included in the overall concept of care in a structured health care system (Wig 1993). As well as treatment methods whose high level of effectiveness has been empirically demonstrated, a variety of inadequately evaluated methods are still in use (Grawe et al. 1994). This applies particularly to evaluation of different service models, so far undertaken only to a limited extent within German university psychiatry, not least because of the methodological difficulties involved (Rössler and Meise 1993). In psychiatry, progress has so far been greatest in establishing treatment standards and methods of quality assurance for drug treatment. In order to develop evidence-based medicine, a task for the future will be to establish a rational basis for the use of other medical and non-medical interventions or else to abandon them if there is no clear evidence for their effectiveness.

Donabedian (1988) distinguishes between technical and interpersonal aspects of therapeutic activity. Technical aspects include scientific knowledge and specialist judgement and skills. These are a central element in practice guidelines (see below), which are based on empirical findings and on establishing an expert consensus. In psychiatry, there is a particular requirement for this technical competence to be applied in combination with interpersonal competence, which has its own kind of "technical" quality and needs to be learnt.

Most notably in psychotherapy, a series of treatment manuals have been developed. The APA has published brief guidelines on various forms of treatment, including psychotherapy (Gray 1992), psychopharmacotherapy (Kane et al. 1992) and electroconvulsive therapy (ECT) (Weiner et al. 1992). These allow a more clearly goal-directed choice of treatment, which is a basic requirement in quality assurance programmes.

In the management of individual disorders, all therapeutic measures need to be co-ordinated within an overall treatment plan (Munich 1990; Munich et al. 1990). This structures the content and timing of the therapeutic process, taking into account treatment prognosis. Practice guidelines enable the professional to carry out diagnosis and treatment according to the valid rules of the art, but allow medical freedom to make modifications in treatment where required in particular individual cases. Guidelines must themselves fulfil a number of quality requirements. The principle characteristics of effective and efficient practice guidelines are as follows (Field and Lohr 1990, cited in Selbmann 1996):

- Validity: demonstrated effectiveness and efficiency
- Reliability when applied
- Clinical flexibility: allowing deviations

- Clearly written
- Clinical applicability: indications clearly specified
- Monitoring of effectiveness and efficiency: planned deadlines for revision
- Reproducibility: process of development reliable
- Balance and acceptability: product of a multidisciplinary development process
- Adequate documentation of development process and of guidelines

Practice guidelines, which – unlike textbook recommendations – are the product of a special procedure for establishing consensus (see below) have been published for a variety of illness types. These include schizophrenia (Andrews et al. 1986; Kissling 1991; Frances et al. 1996), depression (Armstrong and Andrews 1986; Rush 1993), eating disorders (Wilson and Phillips 1992b) and personality disorders (Royal Australian and New Zealand College of Psychiatrists 1991a,b). The APA has published guidelines for major depression (APA 1993b), bipolar affective disorders (APA 1995b), eating disorders (APA 1993a) and substance dependence (APA 1995a), with further guidelines on dementia and schizophrenia in preparation.

Introduction of practice guidelines will obviously have a greater impact on outcomes when accompanied by the use of internal educational programmes which are directly relevant to patient management and to specific illness types (Grimshaw and Russell 1993). An example is the linking of a clinical quality review with measures for postgraduate training, e.g. in clinical psychopharmacology (Awad 1987).

In setting up guidelines, certain ground rules should be observed and should prevent groupings which are not clearly recognised within their specialties from producing their own sets of guidelines on which an adequate consensus has not been established (see Gaebel and Falkai 1996). The specialist medical associations should take a lead in formulating professional standards in collaboration with relevant specialist organisations, established experts and practitioners and in deriving from these standards regulations, guidelines and recommendations (Selbmann 1996). This process must itself satisfy methodological standards.

In consensus development, one can distinguish between *non-formalised* procedures, such as statements from single experts (e.g. in text books) or groups of experts, and *formalised* procedures (Ellis and Whittington 1993; Deutsche Gesellschaft für Chirurgie 1995 as quoted in AWMF 1995). Formalised procedures are the most effective way of developing consensus statements. There is now substantial literature on methods of establishing consensus in groups (for an overview, see Ellis and Whittington 1993). Where there is a need for rapid development of simple guidelines, an expert group is adequate (AWMF 1995). For longer-term development of more detailed guidelines, other techniques for consensus formation should be added, particularly the Delphi conference and the use of focus groups (Ellis and Whittington 1993; AWMF 1995; Deutsche Gesellschaft für Chirurgie 1995 as quoted in AWMF 1995).

Practice guidelines should be incorporated in basic and higher training, and corresponding specific indicators issued for quality assurance measures (Fauman 1989).

2.3 Outcome Quality

Outcome quality can be defined as the degree of congruence between the treatment goal (ideal) and treatment result (reality). Outcome quality reflects the interaction of structural and process quality in care and is thus undoubtedly the most important dimension in judging quality. As already explained, outcome quality does not have a simple, direct relationship to these input variables. If the outcome of treatment is suboptimal, an investigation should generally take place of whether the assumptions behind treatment and the way it has been carried out satisfy defined quality criteria. In practice, in contexts such as quality assurance programmes, such an investigation will generally be initiated if the treatment outcome falls outside a defined tolerance zone.

A thorough discussion of outcome quality in psychiatry needs to take into account the multidimensional nature of outcome in psychiatric illness. The various outcome characteristics (such as symptoms, rate of recidivism and social adaptation) constitute independently valid criteria for assessment, which are related to partially correlated, relatively stable longitudinal functional systems ("open linked systems"; Strauss and Carpenter 1977). Conceptualisations of quality of life (e.g. see Awad 1992; see also Vol. 1, Part 2, Chap. 7) principally emphasize the patient perspective on outcome quality. Patients' views play an increasing role in quality assurance (Leimkühler 1995). The relationships which may exist between patients' diagnoses and their views need to be taken into account and have not so far been elucidated (Kelstrup et al. 1993).

It is increasingly assumed that consumer expectations must be taken into account in establishing treatment goals. Ways of doing this include surveying satisfaction with treatment among patients and relatives, routine provision of patient advocates so that patients' interests are more effectively represented, setting up bodies such as complaints commissions and canvassing patients' and relatives' experiences when treatment structures are planned, set up or modified. In

this process, uncritical agreements should not be reached and role conflicts should be discussed; in addition, it needs to be accepted that illness-related restrictions on capacity to judge and decide mean that psychiatric treatment to which the patient does not consent will sometimes be required and can be legally enforced. There is a considerable need for research regarding the complicated relationships which exist in this area.

Research on prognosis and outcomes are important prerequisites for quality assurance measures (Möller et al. 1995).

Length of stay in psychiatric institutions is a global indicator of efficiency. Worldwide, length of stay in inpatient units has fallen, while admission rates have risen. For example, between 1991 and 1992, the average length of stay in Germany is reported to have fallen from 70.1 to 60.6 days (Rössler and Salize 1995b).

Obviously, this outcome marker has multiple determinants and cannot therefore be viewed as an unambiguous indicator of outcome quality (Gaebel 1995c). Short stays do not necessarily reflect a more successful treatment than longer stays. Patient population and peculiarities of the regional care system play a modifying role, and if this is not taken into account, false conclusions are inevitable (see Böhme et al. 1994). As a variable on which data can quite readily be collected, this indicator can however be used in an internal quality assurance process as a starting point for an "auditing" process. In such a process, certain situations can be examined more closely, such as very long stays within particular diagnostic groups.

Because of the complexity of outcome quality, a risk management approach is often preferred, which focuses on serious adverse incidents involving patients. Surveys are conducted of events such as violent incidents among inpatients, accidents and complications of treatment (Liptzin 1991; Way et al. 1985), with the aim of preventing further such events (Clements et al. 1985; DGS 1993; Kibbee 1988; see also Gaebel 1995c). Obviously, indicators of this type allow only a broad, global estimate of quality of care, and with retrospective analysis quality improvement is only possible after quite a long delay. The same applies to the drug monitoring process (Cole and Katz 1988; Helmchen et al. 1985; Molnar and Feeney 1985), unless it is carried out "online" with a capacity to trigger changes immediately.

3 Framework for Quality Assurance

This concluding section will describe the framework which needs to be in place for effective implementation of quality assurance measures.

3.1 Organizational Framework

Various institutions are involved in the development and organisation of quality assurance in Germany. In 1993, the 96th Deutscher Ärztetag (German doctors' conference) declared that assurance of the quality of medical work is an integral part of medical activity. The Bundesärtzekammer (national medical association), the Kassenärtzliche Bundesvereinigung (national association of doctors treating patients under insurance schemes), the Deutsche Krankenhausgesellschaft (German association of hospitals) and the Spitzenverbände der Gesetzlichen Krankenkassen (associations of providers of statutory health insurance) have increased their activities in quality assurance since the 1989 health reforms, including work in the field of psychiatry. In 1993, they founded a working party for the advancement of quality assurance in medicine.[1,2]

The specialist medical associations and their parent body, the Arbeitsgemeinschaft Wissenschaftlich-Medizinischer Fachgesellschaften (AWMF; federation of specialist medical associations), are taking a leading role in the development of specialty-specific quality assurance measures. In each region, associations of hospitals, doctors' associations and providers of health insurance collaborate in a working party, whose main project office is located within the regional doctors' association (Baur-Felsenstein 1994; Kolkmann 1995). External quality assurance programmes are carried out here.

Within institutions, "quality policy", "quality management", "quality assurance system" and "quality control" are key expressions in the organisation and implementation of quality assurance measures (DIN ISO 8402; DIN 1992). All those involved need to understand and accept the obligation to carry out internal quality assurance programmes. In literature on quality assurance in the hospital, quality assurance is presented as an element in a "total quality management" programme (Eichhorn 1987; Kaltenbach 1991). According to DIN ISO Standards 8402 and 9000–9004, quality management encompasses "all the activities by means of which quality philosophy, goals and responsibilities are established and are implemented through quality planning, quality control, quality assurance and quality improvement" (see Selbmann 1995). This

[1]Gesetz zur Strukturreform im Gesundheitswesen (Gesundheits-Reformgesetz - GRG) 1988, Bundesgesetzblatt, part 1, no 62, Bundesanzeiger, Bonn.

[2]Gesetz zur Sicherung und Strukturverbesserung der gesetzlichen Krankenversicherung (Gesundheitsstrukturgesetz, GSG) 1992, Bundesgesetzblatt, part 1, no 59, pp 2266–2334, Bundesanzeiger, Bonn.

process involves the use of specific forms of organization, such as the introduction of quality officers and commissions and of quality circles based on participatory group work (Antoni 1990). The development of a quality assurance system should progress from simple to more complex types of indicator. The concepts on which multi-organisational external programmes are based need to be quite simple, while internal programmes can be more complex.

In inpatient care, models of external quality assurance have been developed on the basis of Sect.137 of the Fifth Book of the Social Welfare Code and have followed other medical specialties in being based on particular tracer diagnoses (on depression, see Wolfersdorf et al. 1996; on schizophrenia, see Janssen et al. 1998). Specific screening criteria can then be applied to all patients or to a random sample who fall outside a defined tolerance zone, e.g. related to length of stay.

Models of internal quality assurance are currently being evaluated in outpatient care in the form of quality assurance circles made up of outpatient psychiatric practitioners (Berger et al. 1995; Härter and Berger 1996). In inpatient care, a "primer of quality assurance in psychiatric clinics" has been published as a basis for internal quality management programmes (BMG 1996; Gaebel 1996).

3.2 Documentation and Data Processing

A prerequisite for effective quality management is documentation that meets the recognized quality criteria (reliability, objectivity and validity), permitting analyses which are sufficiently precise and concurrent with treatment. In inpatient settings, the psychiatric case notes are the major instrument of documentation of all interviews and observations relating to the patient and of therapeutic measures and results of treatment (see Gaebel 1995c).

However, lack of standardisation limits the usefulness of case notes in quality analyses ("medical audit"). In charting the course of symptoms, it is valuable to incorporate the use of standardised instruments such as the AMDP (Arbeitsgemeinschaft für Methodik und Dokumentation in der Psychiatrie; Schaub 1994). A basic psychiatric form for recording basic psychiatric data has been developed so as to generate a minimum data set for analyses (*Basisdokumentation*, BADO; Dilling et al. 1983). However, the implementation of this in Germany has been patchy, and its suitability for quality assurance is limited as it does not include process or outcome measures (see Cording 1995).

The Deutsche Gesellschaft für Psychiatrie, Psychotherapie und Nervenheilkunde (DGPPN, German Psychiatric and Psychotherapeutic Society) has recently published an extended version of this form, aimed specifically at collecting data required for quality assessment (Cording et al. 1995). As well as information on structure (e.g. catchment area, type of hospital, structure of regional care system, characteristics of patient population), it incorporates process data (e.g. diagnosis, treatment) and outcome data (e.g. progress of treatment, outcome). Where initial conditions such as characteristics of patient populations are similar, an instrument of this type allows the external quality comparisons required by law to be made. The addition of such a set of instruments would greatly increase the usefulness of centralised health monitoring, which is essential from the point of view of service planning but remains weak in the area of data on process quality (Rössler and Salize 1995a).

Further requirements for both internal and external quality assurance are the availability of satisfactory data analysis techniques together with expert data management (Craig and Mehta 1994; Smith 1992; Schröder 1993). To this end, appropriate hardware and software need to be installed, allowing prompt and user-friendly processing of data. Only when these conditions are met will a satisfactory information base be available for prospective quality planning and control.

3.3 Needs for Evaluation and Research

Quality assurance should not be equated with research (Selbmann 1995). Without research, there is no possibility of progress in medicine, so that quality of care cannot be improved. On the one hand, research is the source of advances in scientific knowledge, while on the other hand there is a constant need for research evaluating the implementation in clinical practice of guidelines based on scientific knowledge (Fauman 1990). Research must itself satisfy specific quality criteria (Falkai et al. 1995).

There are particularly prominent gaps in research available in the area of service evaluation, especially regarding groups of patients who are cared for in non-university facilities (Klein 1994). This restricts not only the development of empirically based practice guidelines, but also their implementation in practice, as the generalisability of research results is uncertain. In principle, the introduction of systems of documentation and quality assurance structures should open up opportunities for research in non-university institutions. The quality assurance process may itself foster a more research-friendly culture, providing a basis for a more rational understanding of the treatment process and thus for the improvement of this process.

3.4 Ethical Aspects

Ethical principles such as those set out by the APA (1973) require medical practice to have a scientific basis. This is also the fundamental principle underlying evidence-based medicine. Practice guidelines developed to reflect the current status of scientific knowledge should enable doctors to follow this principle in their practice. Through its commitment to the principle of applying empirically developed standards to clinical practice, quality assurance thus has an ethical justification which protects it from misuse in the service of purely economic interests.

3.5 Economic Aspects

The fundamental purpose of quality assurance is the maintenance and improvement of quality of care. Economic considerations are secondary, but not necessarily irrelevant. Rational treatment involves a willingness to analyse costs and benefits. This process is best initiated by clinicians themselves through peer review rather than it being imposed on them from outside based on standards which are not acceptable to them. Where measures labelled as quality assurance involve examination of the treatment process with the sole aim of making it cheaper and more uniform, there will necessarily be negative consequences from the perspective of patient management which aims at the best possible treatment. Effective opposition to such developments requires clear identification of situations in which treatment standards that have been reached or could in principle be reached have had to be abandoned for economic reasons. Explicit and evaluated quality standards are needed so that a consensus can be reached on the standard of treatment which a society is able and willing to afford.

4 Future Directions

A central requirement for the further development and implementation of quality assurance measures is the co-ordination at the levels of content and methodology of the work of different committees and organisations. Formulation of guidelines which are illness specific and which take into account the modifications required for different settings is a priority. A practicable system of documentation which includes sections recording service activity of specific types needs to be introduced, and frameworks need to be established and evaluated for internal and external quality assurance. In the work on these tasks, counterproductive competition and unrealistically high goals need to be avoided.

Without quality assurance, accountable and responsible further development of the public health service cannot be envisaged (Deutscher Ärztetag 1994). This principle is entirely valid for the field of psychiatry.

5 References

American Medical Association, Council on Medical Service (1986) Quality of care. JAMA 256: 1032–1034

Andrews S, Vaughan K, Harvey R et al (1986) A survey of practising psychiatrists' views on the treatment of schizophrenia. Br J Psychiatry 149: 357–364

Antes G, Egger M, Zellweger T (1995) Randomised trials in German-language journals. Lancet 347(3): 1047–1048

Antoni CH (1990) Qualitätszirkel als Modell partizipativer Gruppenarbeit. Huber, Bern

APA (1973) The principles of medical ethics with annotations especially applicable to psychiatry. Am J Psychiatry 130(9): 1058–1064

APA Task Force (1992) Guidelines for psychiatric practice in community mental health centers. In: Mattson MR (ed) Manual of psychiatric quality assurance. A report of the American psychiatric association committee on quality assurance. American Psychiatric Association, Washington, DC, pp 215–218

APA (1993a) Practice guideline for major depressive disorder in adults. American Psychiatric Association, Washington, DC

APA (1993b) Practice guideline for eating disorders. American Psychiatric Association, Washington, DC

APA (1994) Diagnostic and statistical manual of mental disorders, 4th edn (DSM-IV). American Psychiatric Association, Washington, DC

APA (1995a) Practice guideline for treatment of patients with bipolar disorder. American Psychiatric Association, Washington, DC

APA (1995b) Practice guideline for the treatment of patients with substance use disorders. Alcohol, cocaine, opioids. American Psychiatric Association, Washington, DC

APA (1997a) Practice guideline for the treatment of patients with schizophrenia. APA, Washington, DC

APA (1997b) Practice guideline for the treatment of patients with Alzheimer's disease and other dementias of late life. APA, Washington, DC

Armstrong MS, Andrews G (1986) A survey of practising psychiatrists' views on treatment of the depressions. Br J Psychiatry 149: 742–750

Arnold M, Lauterbach KW, Preuß KJ (eds) (1997) Managed care. Ursachen, Prinzipien, Formen und Effekte. Schattauer, Stuttgart

Awad A (1987) Integrating a clinical review process with postgraduate training in clinical psychopharmacology. QRB 13: 279–282

Awad A (1992) Quality of life of schizophrenic patients on medications and implications for new drug trials. Hosp Community Psychiatry 43: 262–265

AWMF (Arbeitsgemeinschaft Wissenschaftlicher Medizinischer Fachgesellschaften) (1995) Protokoll der AWMF-Konferenz "Leitlinien", 4 October 1995, Hamburg

Banaski D, Flachsmeyer E, Grundig E, Henskes S, Leuffert U (1993) Der Einsatz des MDK bei der Umsetzung der Psychiatrie-Personalverordnung (Psych-PV) – Erfahrungsbericht. Gesundheitswesen 55: 493–499

Baur-Felsenstein M (1994) Qualitätssicherung aus der Sicht der Selbstverwaltung. Arzt Krankenhaus 1(94): 24–28

Berger M (1993) Der neue Facharzt für Psychiatrie und Psychotherapie. Spektrum 22: 4–9

*Berger M, Gaebel W (eds) (1996) Qualitätssicherung in der Psychiatrie. Springer, Berlin Heidelberg New York

Berger M, Barth-Stopik A, Gaebel W (1995) Qualitätszirkel in der ambulanten psychiatrisch-psychotherapeutischen Versorgung. Spektrum 5: 217–219

**Bertolote JM (1993) Quality assurance in mental health care. In: Sartorius N, De Girolamo G, Andrews G, German GA, Eisenberg L (eds) Treatment of mental disorders. A review of effectiveness. American Psychiatric Press, Washington, pp 443–461

**Bertolote JM (1994) Quality assurance in mental health care. Division of Mental Health, WHO, Geneva

BMG (ed) (1996) Leitfaden zur Qualitätsbeurteilung in Psychiatrischen Kliniken: Projekt 1994–1996 im Auftrag des Bundesministeriums für Gesundheit. Nomos, Baden-Baden

BMJFG (1975) Bericht über die Lage der Psychiatrie in der Bundesrepublik Deutschland – Zur psychiatrischen und psychotherapeutisch/psychosomatischen Versorgung der Bevölkerung. Bundesministerium für Jugend, Familie und Gesundheit, Bonn

BMJFFG (1988) Empfehlungen der Expertenkommission der Bundesregierung zur Reform der Versorgung im psychiatrischen und psychotherapeutisch-psychosomatischen Bereich. Bundesministerium für Jugend, Familie, Frauen und Gesundheit, Bonn

Böhme K, Cording C, Ritzel G, Spengler A, Trenckmann U (1994) Thesen zur Qualitätssicherung (QS). Spektrum Psychiatr Nervenheilkd 23: 58–62

Brenner HD (1995) Stand der Diskussion zur Kosten-Effektivitätsfrage in der Gemeindepsychiatrie und Klinikpsychiatrie. Schweiz Arch Neurol Psychiatr 1: 24–32

Buchborn E (1993) Der ärztliche Standard. Dtsch Ärztebl 90(28,29): B1446–1449

Bundesarbeitsgemeinschaft der Leitenden Ärzte kinder- und jugendpsychiatrischer Kliniken und Abteilungen (1993) Zielsetzungs- und Orientierungsdaten kinder- und jugendpsychiatrischer Kliniken und Abteilungen. Landschaftsverband Rheinland, Cologne

Bundesarbeitsgemeinschaft der Träger Psychiatrischer Krankenhäuser (1990) Zielsetzungs- und Orientierungsdaten psychiatrischer Krankenhäuser. Landschaftsverband Rheinland, Cologne

*Cahn C, Richman A (1985) Quality assurance in psychiatry. Can J Psychiatry 30: 148–152

Clements CD, Bonacci D, Yerevanian B et al (1985) Assessment of suicide risk in patients with personality disorder and major affective diagnosis. QRB 11: 150–154

Cochrane Collaboration Working Group (1996) Die Cochrane Collaboration. Schweiz Ärztez 77(3): 117–120

Cole JO, Katz DL (1988) Drug therapy monitoring in a private psychiatric hospital: a consideration of its risks and benefits. McLean Hosp J 13: 114–157

Cording C (1995) Basisdokumentation und Ergebnisqualität. In: Gaebel W (ed) Qualitätssicherung im psychiatrischen Krankenhaus. Springer, Berlin Heidelberg New York, pp 173–181

Cording C, Gaebel W, Spengler A et al (1995) Die neue psychiatrische Basisdokumentation. Eine Empfehlung der DGPPN zur Qualitätssicherung im (teil-)stationären Bereich. Spektrum Psychiatr Nervenheilkd 24: 3–41

Craig TJ, Mehta RM (1984) Clinician-computer interaction: automated review of psychotropic drugs. Am J Psychiatry 141: 267–270

Cranach M von (1994) Leitfaden für die Begehung des psychiatrischen Krankenhauses. In: Kunze H, Kaltenbach L (eds) Psychiatrie-Personalverordnung. Textausgabe mit Materialien und Erläuterungen für die Praxis. Kohlhammer, Stuttgart, pp 203–211

DGPPN (Deutsche Gesellschaft für Psychiatrie, Psychotherapie und Nervenheilkunde (1997) Die Behandlung psychischer Erkrankungen in Deutschland. Positionspapier zur aktuellen Lage und zukünftigen Entwicklung. Springer, Berlin Heidelberg New York

DGPPN (1998) Praxisleitlinien in Psychiatrie und Psychotherapie. Behandlungsleitlinie Schizophrenie. Redaktion: Gaebel W, Falkai P. Steinkopff, Darmstadt

DGS (Deutsche Gesellschaft für Suizidprävention) (1993) Leitlinien zur Organisation von Krisenintervention. Köhler, Harsum

Deutscher Ärztetag (1994) Gesundheitspolitisches Programm der deutschen Ärzteschaft (blue paper). Dtsch Ärztebl [Suppl] 24: 1–42

Dilling H, Balck F, Bosch G et al (1983) Zur psychiatrischen Basisdokumentation. Nervenarzt 54: 262–267

Dilling H, Mombour W, Schmidt MH (1992) Internationale Klassifikation psychischer Störungen (ICD-10). Huber, Bern Göttingen Toronto

DIN (Deutsches Institut für Normung e.V.) (ed) (1992) Qualitätssicherung und angewandte Statistik. Verfahren 3: Qualitätssicherungssysteme. Beuth, Berlin

Donabedian A (1966) Evaluating the quality of medical care. Milbank Q 44:166

Donabedian A (1982) Explorations in quality assessment and monitoring. II. The criteria and standards of quality. Health Administration Press, Ann Arbor

Donabedian A (1986) Criteria and standards for quality assessment and monitoring. QRB 12(3): 99–108

Donabedian A (1988) The quality of care: how can it be assessed? J Am Acad 260: 1743–1748

Eichhorn S (1987) Krankenhausbetriebslehre. Theorie und Praxis der Krankenhaus-Leistungsrechnung, vol III. Kohlhammer, Stuttgart

*Ellis R, Whittington D (1993) Quality assurance in health care. A handbook. Arnold, London

Ellis J, Mulligan I, Rowe J, Saccett DI (1995) Inpatient general medicine is evidence based. Lancet 346: 407–410

Falkai P, Gaebel W, Wölwer W (1995) Qualitätssicherung in der Psychiatrischen Forschung. Psycho 21: 236–240

*Fauman MA (1989) Quality assurance monitoring in psychiatry. Am J Psychiatry 146: 1121–1130

Fauman MA (1990) Monitoring the quality of psychiatric care. Psychiatr Clin North Am 13: 73–88

Fifer WR (1980) Quality assurance in health care. In: Awad AG, Durost HB, McCormick WO (eds) Evaluation of quality of care in psychiatry. Pergamon, Toronto Oxford New York Sydney Paris Frankfurt, pp 1–12

Finzen (1994) Zukünftige Strukturen psychiatrischer Versorgung – Zwischenbilanz und Perspektiven nach zwei Jahrzehnten Psychiatriereform. In: Reimer F (ed) Versorgungsstrukturen in der Psychiatrie. Springer, Berlin Heidelberg New York, pp 103–110

Frances A, Docherty JP, Kahn DA (1996) The expert consensus guideline series. Treatment of schizophrenia. J Clin Psychiatry 57[Suppl 12B]: 1–58

*Gaebel W (1995a) Qualitätssicherung in der Psychiatrie. Nervenarzt 66: 481–493

*Gaebel W (ed) (1995b) Qualitätssicherung im psychiatrischen Krankenhaus. Springer, Berlin Heidelberg New York

Gaebel W (1995c) Qualitätssicherung diagnostischer und therapeutischer Maßnahmen im psychiatrischen Krankenhaus. In: Gaebel W (ed) Qualitätssicherung im psychiatrischen Krankenhaus. Springer, Berlin Heidelberg New York, pp 87–108

Gaebel W (1996) Leitfaden zur Qualitätsbeurteilung in Psychiatrischen Kliniken. Nervenarzt 67: 968–970

Gaebel W, Falkai P (1996) Praxisleitlinien in der Psychiatrie. Zu Methodik und Stand von Leitlinienentwicklungen. Nervenarzt 67: 179–181

Goldberg D, Huxley P (1980) Mental illness in the community. The pathway to psychiatric care. Tavistock, New York

Goldner EM, Bilsker D (1995) Evidence-based psychiatry. Can J Psychiatry 40: 97–101

Grawe K, Donati R, Bernauer F (1994) Psychotherapie im Wandel – von der Konfession zur Profession. Hogrefe, Göttingen

Gray SH (1992) Quality assurance and utilization review of medical psychotherapies. In: Mattson MR (ed) Manual of psychiatric quality assurance. A report of the American Psychiatric Association Committee on quality assurance. American Psychiatric Association, Washington, DC, pp 153–159

*Grimshaw JM, Russell IT (1993) Effect of clinical guidelines on medical practice: a systematic review of rigorous evaluations. Lancet 342: 1317–1322

Häfner H, an der Heiden W (1994) The evaluation of mental health care systems. In: Mezzich JE, Jorge MR, Salloum IM (eds) Psychiatric epidemiology. Assessment concepts and methods. Johns Hopkins University Press, Baltimore, pp 494–504

Hamilton JM (1992) Introduction to the American Psychiatric Association's involvement in quality assurance and utilization review. In: Mattson MR (ed) Manual of psychiatric quality assurance. A report of the American Psychiatric Association Committee on quality assurance. American Psychiatric Association, Washington, DC, pp 7–9

Härter M, Berger M (1996) Qualitätszirkel – eine Maßnahme der Qualitätssicherung in der ambulanten psychiatrisch-psychotherapeutischen Versorgung. In: Berger M, Gaebel W (eds) Qualitätssicherung in der Psychiatrie. Springer, Berlin Heidelberg New York, pp 89–98

*Haug HJ, Stieglitz RD (1995) Qualitätssicherung in der Psychiatrie. Enke, Stuttgart

Helmchen H, Hippius H, Müller-Oerlinghausen B, Rüther E (1985) Arzneimittel-Überwachung in der Psychiatrie. Nervenarzt 56: 12–18

Janssen B, Burgmann C, Held T, Hoff P, Jänner M, Mecklenburg H, Prüter C, Ruth A, Saß H, Schneider F, Gaebel W (1998) Qualitätsindikatoren der stationären Behandlung schizophrener Patienten. Ergebnisse einer Pilotstudie zur externen Qualitätssicherung mit Hilfe einer Tracer-Diagnose. Psychiatr Prax 25: 303–309

Kaltenbach T (1991) Qualitätsmanagement im Krankenhaus. Bibliomed, Melsungen

Kane JM, Evans DL, Fiester SJ et al (1992) Psychopharmacological screening criteria. In: Mattson MR (ed) Manual of psychiatric quality assurance. A report of the American Psychiatric Association Committee on quality assurance. American Psychiatric Association, Washington, DC, pp 189–205

Kelstrup A, Lund K, Lauritsen B, Bech P (1993) Satisfaction with care reported by psychiatric inpatients. Acta Psychiatr Scand 87: 374–379

Kibbee P (1988) The suicidal patient – an issue for quality assurance and risk management. J Nurs Qual Assur 3: 63–71

Kissling W (ed) (1991) Guidelines for neuroleptic relapse prevention in schizophrenia. Springer, Berlin Heidelberg New York

Klein HE (1994) Probleme der experimentellen psychiatrischen Forschung. Spektrum 23: 20–24

Klieser E, Lehmann E, Strauß WH (1995) Ärztliche und psychiatrische Weiterbildung als Mittel und Aufgabe der Qualitätssicherung. In: Gaebel W (ed) Qualitätssicherung im psychiatrischen Krankenhaus. Springer, Berlin Heidelberg New York, pp 66–75

Klinkhammer G (1995) Leitlinien zur Qualitätssicherung diskutiert. Dtsch Ärztebl 92(14,7): B742–43

Kolkmann FW (1995) Qualitätssicherung aus der Sicht der Bundesärztekammer. In: Gaebel W (ed) Qualitätssicherung im psychiatrischen Krankenhaus. Springer, Berlin Heidelberg New York, pp 11–20

Kruckenberg P, Jagoda B, Aktion Psychisch Kranke (1994) Personalbemessung im komplementären Bereich – von der institutions- zur personenbezogenen Behandlung und Rehabilitation. Aktion Psychisch Kranke, Bonn

Kukla R (1995) Strukturqualität psychiatrischer Krankenhäuser aus Trägersicht. In: Gaebel W (ed) Qualitätssicherung im psychiatrischen Krankenhaus. Springer, Berlin Heidelberg New York, pp 52–57

Kunze H, Kaltenbach L (eds) (1994) Psychiatrie-Personalverordnung. Textausgabe mit Materialien und Erläuterungen für die Praxis. Kohlhammer, Stuttgart

Kunze H, Wienberg G, Vitt KD, Buss G (1994) Strukturierende Gesichtspunkte für die Auswertung von Unterlagen psychiatrischer Krankenhäuser/Abteilungen zur Umsetzung der Psych-PV in ein entsprechendes Behandlungsangebot. In: Kunze H, Kaltenbach L (ed) Psychiatrie-Personalverordnung. Textausgabe mit Materialien und Erläuterungen für die Praxis. Kohlhammer, Stuttgart, pp 194–211

Leimkühler AM (1995) Die Qualität klinischer Versorgung im Urteil der Patienten. In: Gaebel W (ed) Qualitätssicherung im psychiatrischen Krankenhaus. Springer, Berlin Heidelberg New York, pp 163–172

Linden M (1994) Therapeutic standards in psychopharmacology and medical decision-making. Pharmacopsychiatry 27 [Suppl]: 41–45

*Liptzin B (1974) Quality assurance and psychiatric practice – a review. Am J Psychiatry 131: 1374–1377

*Liptzin B (1991) Quality assurance and treatment outcome: a medical perspective. In: Mirin SM, Gossett JT, Grob MC (eds) Psychiatric treatment. Advances in outcome research. American Psychiatric Press, Washington, pp 265–278

Mattson MR (1984) Quality assurance: a literature review of a changing field. Hosp Community Psychiatry 35: 605–616

**Mattson MR (ed) (1992a) Manual of psychiatric quality assurance. A report of the American Psychiatric Association Committee on quality assurance. American Psychiatric Association, Washington, DC

Mattson MR (ed) (1992b) Generic quality screens – psychiatric. Developed by the Health Care Financing Administration for use by peer review organizations. In: Mattson MR (ed) Manual of psychiatric quality assurance. A report of the American Psychiatric Association Committee on quality assurance. American Psychiatric Association, Washington, DC, pp 207–213

Miller SI, Phillips KL (1992) Chemical dependency disorders: guidelines for review of inpatient therapy and rehabilitation. In: Mattson MR (ed) Manual of psychiatric quality assurance. A report of the American Psychiatric Association Committee on quality assurance. American Psychiatric Association, Washington, DC, pp 161–166

Moak GS (1990) Improving quality in psychogeriatric treatment. Psychiatr Clin North Am 13: 99–112

Möller HJ, Deister A, Laux G (1995) Outcome-Forschung als Mittel der Qualitätssicherung. In: Gaebel W (ed) Qualitätssicherung im psychiatrischen Krankenhaus. Springer, Berlin Heidelberg New York, pp 147–162

Molnar G, Feeney MG (1985) Computer-assisted review of antipsychotics on acute care units. QRB Qual Rev Bull 11: 271–274

Munich RL (1990) Quality assurance and quality of care. I. Finding the linkages. Psychiatr Hosp 21: 13–24

Munich RL, Hurley B, Delaney J (1990) Quality assurance and quality of care. II. Monitoring treatment. Psychiatr Hosp 21: 71–77

Naylor D (1995) Grey zones of clinical practice: some limits to evidence-based medicine. Lancet 345(4): 840–842

Piwernetz K, Selbmann HK, Vermeij DJB (1991) "Vertrauen durch Qualität": Das Münchner Modell der Qualitätssicherung im Krankenhaus. Krankenhaus 11: 557–560

Prunier P, Buongiorno PA (1989) Guidelines for acute inpatient psychiatric treatment review. Gen Hosp Psychiatry 11: 278–281

Rössler W, Meise U (1993) Neue Trends in der psychiatrischen Versorgung. Neuropsychiatrie 7(4): 171–175

Rössler W, Salize HJ (1993) Psychiatrische Versorgung: Leitlinien für die Reformpraxis. Dtsch Ärztebl 90(10,39): 2526–2528

Rössler W, Salize HJ (1996) Die psychiatrische Versorgung chronisch Kranker – Daten, Fakten, Analysen. Schlußbericht des Forschungsprojekts Strukturanalyse zur psychiatrischen Versorgung unter besonderer Berücksichtigung des Bereichs chronischer psychischer Störungen. Arbeitsgruppe Versorgungsforschung, Mannheim. Nomos, Baden-Baden

Rössler W, Salize HJ (1995b) Qualitätsindikatoren psychiatrischer Versorgungssysteme. In: Gaebel W (ed) Qualitätssicherung im psychiatrischen Krankenhaus. Springer, Berlin Heidelberg New York, pp 39–51

Rössler W, Löffler W, Fätkenheuer B, Riecher-Rössler A (1992) Does case management reduce the rehospitalization rate? Acta Psychiatr Scand 86: 445–449

Rössler W, Salize HJ, Häfner H (1993) Gemeindepsychiatrie. Grundlagen und Leitlinien. Planungsstudie Luxemburg. Integrative Psychiatrie, Innsbruck

Royal Australian and New Zealand College of Psychiatrists (1982) The Quality Assurance Project: a methodology for preparing 'ideal' treatment outlines in psychiatry. Aust NZ J Psychiatry 16: 153–158

Royal Australian and New Zealand College of Psychiatrists (1991a) Treatment outlines for borderline, narcissistic and histrionic personality disorders. The quality assurance project. Aust NZ J Psychiatry 25: 392–403

Royal Australian and New Zealand College of Psychiatrists (1991b) Treatment outlines for avoidant, dependent and passive-aggressive personality disorders. The quality assurance project. Aust NZ J Psychiatry 25: 404–411

Rush AJ (1993) Clinical practice guidelines. Good news, bad news, or no news? Arch Gen Psychiatry 50: 483–490

Schaub RT (1994) Quality assurance in psychiatric care – the example of routine use of the AMDP system. Pharmacopsychiatry 27[Suppl]: 46–50

Schröder M (1993) Auswirkungen des GSG auf das Informationsmanagement und die Krankenhausinformatik. Krankenhaus 10: 460–470

Schyve PM, Prevost JA (1990) From quality assurance to quality improvement. Psychiatr Clin North Am 13: 61–72

Sederer LI (1991) Quality, cost, and contracts: administrative aspects of inpatient care. In: Sederer LI (ed) Inpatient psychiatry. Diagnosis and treatment. Williams and Wilkins, Baltimore, pp 419–431

Selbmann HK (1995) Konzept und Definition medizinischer Qualitätssicherung. In: Gaebel W (ed) Qualitätssicherung im psychiatrischen Krankenhaus. Springer, Berlin Heidelberg New York, p 3

*Selbmann HK (1996) Entwicklung von Leitlinien in der Medizin – Kunst oder Können? Chirurg 35(3): 61–65

Selbmann HK (1997) Managed care: Ein Ansatz zur Verbesserung der Qualität der Krankenversorgung? In: Arnold M, Lauterbach KW, Preuß KJ (eds) Managed care. Ursachen, Prinzipien, Formen und Effekte. Schattauer, Stuttgart, pp 253–258

Smith AP (1992) Design a clinical information system. Br Med J 305: 415–417

Spengler A (1991) Institutsambulanzen in der psychiatrischen Versorgung. Vandenhoeck and Ruprecht, Göttingen

Strauss JS, Carpenter WT (1977) Prediction of outcome in schizophrenia. III. Five-year outcome and its predictors. Arch Gen Psychiatry 34: 159–163

Way BB, Braff J, Steadman HJ (1985) Constructing an efficient inpatient incident reporting system. Psychiatry Q 57: 147–152

Weiner RD, APA Task Force on ECT (1992) Electroconvulsive therapy guidelines and criteria. In: Mattson MR (ed) Manual of psychiatric quality assurance. A report of the American Psychiatric Association Committee on quality assurance. American Psychiatric Association, Washington, DC, pp 181–187

Wig N (1993) Rational treatment in psychiatry: perspectives on psychiatric treatment by level of care. In: Sartorius N, Girolamo de G, Andrews G, German GA, Eisenberg L (eds) Treatment of mental disorders. A review of effectiveness. American Psychiatric Press, Washington, pp 423–441

Wilson GF (1992) Issues in the review of adult outpatient therapy. In: Mattson MR (ed) Manual of psychiatric quality assurance. A report of the American Psychiatric Association Committee on quality assurance. American Psychiatric Association, Washington, DC, pp 149–152

Wilson GF, Phillips KL (1992a) Residential treatment centers: quality assurance and utilization review guidelines. In: Mattson MR (ed) Manual of psychiatric quality assurance.

A report of the American Psychiatric Association Committee on quality assurance. American Psychiatric Association, Washington, DC, pp 173–180

Wilson GF, Phillips KL (1992b) Eating disorders: quality assurance and utilization review guidelines. In: Mattson MR (ed) Manual of psychiatric quality assurance. A report of the American Psychiatric Association Committee on quality assurance. American Psychiatric Association, Washington, DC, pp 167–172

Wilson GF, Phillips KL (1992c) Concepts and definitions used in quality assurance and utilization review. In: Mattson MR (ed) Manual of psychiatric quality assurance. A report of the American Psychiatric Association Committee on quality assurance. American Psychiatric Association, Washington, DC, pp 23–30

Wing J, Brewin CR, Thornicroft G (1992) Defining mental health needs. In: Thornicroft G, Brewin CR, Wing J (eds) Measuring mental health needs. Gaskell, London, pp 1–17

Wolfersdorf M, Stieglitz RD, Metzger R et al (1996) Modellprojekt zur Qualitätssicherung der klinischen Depressionsbehandlung. In: Berger M, Gaebel W (eds) Qualitätssicherung in der Psychiatrie. Springer, Berlin Heidelberg New York, pp 67–87

CHAPTER

13

M. Linden

Psychiatric Disorders in Primary Care

Translator: S. Johnson

1 Epidemiology of Psychiatric Disorders

Epidemiological studies carried out in recent years both in Germany and internationally have shown that between 15% and 25% of the population are suffering from psychiatric illnesses for which treatment is currently indicated (Bebbington et al. 1981; Henderson et al. 1979; Dilling et al. 1984; Häfner 1978; Schepank 1987; Robins and Regier 1991). In industrial countries, around two thirds of all the people in the population regularly consult general practitioners because of a variety of illnesses and complaints or else for preventive health care measures (RCGP and OPCS 1979). Thus, on the basis of these facts alone, it is clear that many patients in general practices are suffering from psychiatric disorders. If we take into account the further observation that people with chronic illnesses, and especially those with psychiatric disorders, consult doctors about twice as often as other patients (Finlay-Jones and Burvill 1978; Williams et al. 1986; Goldberg and Huxley 1980), then it is not surprising to find that surveys on this topic demonstrate that about every fourth patient in general practices is suffering from a psychiatric disorder.

Following the early work of Shepherd and his colleagues (Shepherd et al. 1966), a series of studies have been published reporting investigations of the prevalence of psychiatric disorders among service users in primary care (Burvill 1990; Üstün and Sartorius 1995; Dilling et al. 1978; Zintl-Wiegand et al. 1978; Gastpar 1984; Schulberg and Burns 1988; El-Rufaie and Absood 1993; Leon et al. 1995). These studies indicate that the overall proportion of patients who have psychiatric illnesses lies between 11% and 36%. Around two thirds of these patients can be seen as having chronic disorders and a third as having new episodes of illness. The degree of variation between the studies cited may be explained in terms of differences in prevalence between different settings and regions, different definitions of the threshold for "caseness", different study instruments and variations in sample inclusion criteria. Even so, the basic order of magnitude is in keeping with the values which would be predicted on the basis of population epidemiological studies. There are further substantial areas of agreement between the various studies. Throughout this literature, women, divorced people, middle-aged individuals and those of low social status are found to have raised rates of psychiatric disturbance (Goldberg and Huxley 1980; Jenkins and Shepherd 1983).

Comparisons between industrialized and developing countries are also interesting. They indicate that in poorer countries the rate of psychiatric disorders among patients receiving primary-care treatment is of a similar order of magnitude or in some instances may be even higher than in richer countries (German 1987; Ndetei and Muhangi 1979; Gautam et al. 1980; Harding et al. 1980; Mari and Williams 1984). This is despite the greater difficulties which are sometimes encountered in obtaining access to medical care (German 1987; Ndetei and Muhangi 1979; Gautam et al. 1980; Harding et al. 1980; Mari and Williams 1984).

A more recent study of psychological problems among patients in primary health care was carried out under the leadership of the World Health Organization (WHO) in parallel in industrialized and developing countries, encompassing a range of different cultures and health care systems (Üstün and Sartorius 1995). Despite language barriers, a central aim of the study was to examine patients using an identical research instrument, which allowed problems to be diagnosed in relation to current psychiatric classification systems. The structured Composite International Diagnostic Interview, Primary Care Version (CIDI) was used and diagnoses made according to ICD-10. Self-assessment questionnaires such as the General Health Questionnaire (GHQ) were also used. Figure 1 (Goldberg and Lecrubier 1995) shows that, according to ICD-10 definitions, 24% of practice attenders were suffering from a psychiatric disorder. Among a further 9%, clinically significant symptoms of illness were found without the requirements for an ICD-10 diagnosis being met. Finally, a further 31% suffered from two or more psychological symptoms without reaching the threshold for an illness being diagnosed. The results of the GHQ self-assessment scale indicated a similar picture. A total of 23.3% of practice visitors had a score of 5 or higher, suggesting the presence of a psychiatric disorder of clinical significance.

Table 1 shows the spectrum and frequency of the illnesses found (Goldberg and Lecrubier 1995). The most important psychiatric disorders are depressive illnesses, occurring in 10.4%, and generalized anxiety disorders, occurring in 7.9%, followed by neurasthenia (5.4%), alcohol misuse (3.3%), alcohol dependence (2.7%), somatization disorder (2.7%), dysthymia (2.1%), panic disorders (1.1%) and agoraphobia (1.0%). A total of 9.5% of the patients met criteria for more than one ICD-10 diagnosis, indicating a psychological co-morbidity.

The probability of suffering from psychiatric illnesses rises with the level of severity of physical illnesses. Patients whose physical health was classified as poor had a 1.4-fold risk of also suffering from a psychiatric disorder. This corresponds to the findings of a series of similar studies, which in some cases have indicated that the risk is as much as a 3.6-fold (Weyerer 1990; Leon et al. 1995). Causal mechanisms behind these associations may be psychiatric illness

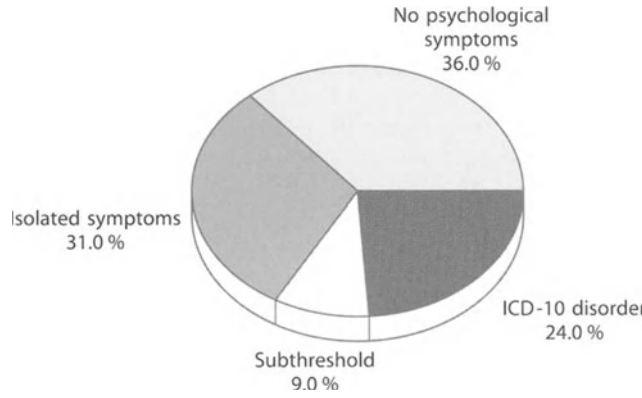


Fig. 1. Psychiatric disorders in primary care. (From Goldberg and Lecrubier 1995)

Table 1. Spectrum of psychiatric disorders in primary care (based on Goldberg and Lecrubier 1995)

Psychiatric disorder	Incidence (%)
Depressive episode	10.4
Generalized anxiety disorder	7.9
Neurasthenia	5.4
Alcohol misuse	3.3
Alcohol dependence	2.7
Somatization disorder	2.7
Dysthymia	2.1
Panic disorder	1.1
Agoraphobia with panic	1.0
Hypochondria	0.8
Agoraphobia without panic	0.5
At least one ICD-10 diagnosis	24.0

acting as a pathogenic factor in physical illnesses, organic illnesses manifesting themselves in partly psychiatric symptoms, or psychiatric disorder arising as a consequence of an organic illness such as a carcinoma (Brown and Parasekvas 1982; Massie and Holland 1984), a cardiovascular illness (Bass and Wade 1984), arthritis (Gardiner 1980) or a wide variety of other illnesses (Katon 1982).

The studies also provided some confirmation of a finding already mentioned above, which is that women suffer more frequently than men from psychiatric disorders and have different patterns of illness. Thus the prevalence of depression among women, for example, was found to be 12.5%, 1.9 times higher than the prevalence in men, which was 7.1%. However, if we take into account the full range of psychiatric disorders, including alcohol misuse, very little overall difference can be found. A total of 22.8% of men and 24.7% of women met criteria for an ICD-10 diagnosis. There was also an association with social class. If education is used as an indicator of social class, people who have had higher education have a lower risk of illness by a factor of 0.8.

So far, the results given in these international studies are summary results covering all the countries and cultures included. Depressive disorders and anxiety syndromes are the most frequent illnesses in the European countries, just as in Shanghai, Bangalore or Ibadan. Interestingly, the symptom profile of depression is also in essence the same across all centres. However, there are also significant differences between the various centres. For example, the overall prevalence of psychiatric disorders varies between 7.5% in Shanghai and 52.5% in Santiago. These differences provide further confirmation of the phenomenon of great inter-regional differences in diagnosis and treatment, a phenomenon which has already been frequently described in the literature. It is still an open question whether the explanation for these variations should be sought in differences between cultures, between health service structures or between the actual treatment facilities. An example of a difference related to culture is the finding that the rate of alcohol dependence among patients in Germany lies between 5% and 7%, while the corresponding rate in Ankara is only 1%.

2 Distinctive Aspects of Diagnosis and Classification

There is a long tradition of research on the extent to which general practitioners recognize and are thus in a position to treat their patients' psychiatric disorders. Studies which compare diagnoses of the general practitioners managing patients with psychiatrists'

diagnoses relating to those patients or else with the results of standardized surveys as a rule report a substantial rate of "underdiagnosis" (Blacker and Thomas 1988; Goldberg and Huxley 1980; Marks et al. 1979; Hankin and Oktay 1979; Skuse and Williams 1984; Casey et al. 1984; von Korff et al. 1987; Johnstone and Goldberg 1976; Shapiro et al. 1987; Zung et al. 1983; Hoeper et al. 1979). Goldberg and Blackwell (1970) applied the term "hidden psychiatric morbidity" to encapsulate this phenomenon. However, its origins do not lie, or at least do not principally lie, in a lack of specialist psychiatric knowledge among the general practitioners involved. Rather it is above all a reflection of the distinctive characteristics of the disorders concerned. The issues which give rise to problems relate to the specificity of presenting complaints, the particular ways in which patients present themselves and questions regarding the usefulness of making diagnoses in the usual standard ways.

Classification systems for psychiatric disorders have traditionally been developed in the context of psychiatric hospitals and have focused primarily on psychotic illnesses. In these very conspicuous illnesses, the principal symptoms guiding differential diagnosis, such as delusions of poverty or hallucinations, have a high level of validity and specificity. They are also not usually readily overlooked or misinterpreted. The situation is quite different for the disorders which predominate in primary care. They tend to be of mild to moderate severity and are not diagnosed on the basis of single characteristic symptoms so much as through a variable pattern of symptoms. Thus none of the symptoms to be observed has in itself a very marked pathological significance. Anhedonia, disturbances in concentration, lack of motivation, loss of appetite, sleep disorders or disturbances of energy may be manifestations of depression, but equally of many other psychiatric disorders.

Following the introduction of standardized assessment instruments and defined diagnostic algorithms, this marked lack of specificity of presenting symptoms has led to the phenomenon known as co-morbidity, with many patients fulfilling to similar extents the criteria for anxiety, depression, neurasthenia or somatization. Thus Stein et al. (1995), for example, found that the prevalence of mixed anxiety and depression was greater than that for pure cases of either, and as a result of this the mixed category has even been adopted as a distinct diagnosis in ICD-10. Moreover, the symptoms discussed here can also be manifestations of physical illnesses. In a study of older patients, Linden et al. (1995) were able to demonstrate that the average score on the Hamilton Depression Scale fell by half when they did not count as symptoms of depression those complaints which were seen by general physicians carrying out a simultaneous examination as symptoms of physical illnesses. Finally, the complaints discussed above can also be prodromal symptoms or indicators of relapse in a still wider range of disorders, so that it may only be possible to arrive at a valid differential diagnostic classification after more prolonged observation. This difficulty in establishing the boundaries of specific categories has resulted among other things in the WHO proposing a separate classification of psychiatric disorders for the field of primary care (Üstün et al. 1995a). Its principal characteristic is that diagnostic categories are broader.

Another problem which arises in assessing symptoms in general practice is how to establish thresholds for making diagnoses (Linden and Geiselmann 1996). The classical psychopathological symptoms already mentioned, such as hallucinations or delusions, are to a large extent categorical either/or phenomena, whose appearance always signifies illness and is always to be interpreted in psychopathological terms. In contrast, the symptoms of disorders which present in primary care are dimensional phenomena, which without any discontinuity make the transition from normal through to psychopathological, as with the fluid transition from excellent capacity for concentration via various gradations of limited concentration capacity through to a complete inability to concentrate. Epidemiological knowledge indicates that less severe manifestations of such symptoms are more common than severe ones. Minor alterations in the threshold at which a symptom is seen as being of clinical significance will therefore necessarily lead in primary care settings to substantial shifts in prevalence figures. The dimensional nature of the complaints encountered also has consequences for the recognition of disorders. Above all, it is the less severe disorders which do not yet impair social functioning that are not diagnosed (Coyne et al. 1995; Tiemens et al. 1996).

From such theoretical considerations, it follows that diagnostic conclusions regarding the disorders mainly seen in primary care settings must essentially carry a greater uncertainty than is the case in psychiatric settings and with more severe illnesses. Thus agreement rates between two observers or between observations based on different methodologies will necessarily be characterized by greater variability. We may even ask who is really coming closer to the "truth" when researchers and general practitioners arrive at different conclusions about diagnosis, since the lack of specificity of symptoms, for example, is not taken into account in using standardized survey instruments. In the WHO primary care study, the general practitioners arrived at a rate of psychiatric disorders among their patients of 23.4%, which is almost identical with the prevalence of 24.0% arrived at using the CIDI, but falls below the overall rate for the CIDI of 32.5%, arrived at when the subthreshold

CIDI patients were also included. However, agreement in making diagnosis was less satisfactory. Figure 2 (Üstün and von Korff 1995) indicates that only 13.3% of practice attenders were detected both by the doctor and the research interview (Üstün and von Korff 1995). Apart from the jointly detected patients, there are not only those who are ill according to the CIDI but are not recognized by the doctors, but also 10.9% who, according to the doctor's evaluation, are mentally ill without being detected by the CIDI. Interestingly, the correlation is greater between medical diagnoses and the self-rated instrument, the GHQ ($r = 0.83$) than between the CIDI diagnoses and the GHQ ($r = 0.73$).

Finally, another fundamental problem in the diagnosis of psychiatric disorders in primary care lies in the fact that, in less severe disorders, unlike for example in psychotic illnesses, patients themselves have a substantial influence on how a symptom is seen. It is up to them to decide which complaints they experience as burdensome and which they wish to report. Patients who go to a general practitioner will therefore as a rule be rather more inclined to report physical problems than those who seek help from a psychiatrist. Correspondingly, all studies in this area show that only few patients with psychiatric problems make direct reports of psychiatric complaints to general practitioners. Instead, they describe their somatic symptoms (Bridges and Goldberg 1985). Thus, in the international WHO study, 32.8% of the practice attenders presented with complaints relating to various physical conditions, and a further 29.3% complained about a variety of pains; 6.9% complained of tiredness and sleeping problems. Only 5.3% spontaneously talked about psychological problems and complaints (Üstün and von Korff 1995). Goldberg (1990) demonstrated clear variations between doctors in rates of recognition, which were independent of the patient's presenting complaint. However, the form of the presenting complaint also has a direct influence on the type of diagnosis made. Cremniter et al. (1995) found that, in depressive disorders, general practitioners did not see depressed mood as the primary problem, but sleep disturbance in 31.8%, lack of energy in 29.9% and anxiety in 24.6%. Finally, Tylee et al. (1995) also showed that the recognition rate was approximately eight times higher when psychological complaints were reported by the patient at the beginning of a consultation rather than first emerging later in its course. In explaining this phenomenon, we need to take into account the fact that the task of the general practitioner does not consist primarily of managing psychiatric disorders, but of considering the whole spectrum of illnesses. Thus it is understandable that, in patients who have known physical illnesses, the likelihood of psychological problems being recognized is lowest. In a study by Goldberg (1990), the recognition rate fell from 85% in the absence of other somatic illness to 33% with known physical illness. With tiredness in the presence of known diabetes, it is good medical practice to begin by considering whether a better insulin regime might not overcome the problem. Further, it needs to be borne in mind that, when patients present with particular complaints, they are also asking for treatment and are indicating what sort of treatment they expect. Someone who goes to the doctor because of headaches or muscle tension quite understandably expects that this complaint will be made the initial object of treatment. He or she will not expect to give information about personal problems. From this perspective, the role of doctors is to some extent to receive instructions. Doctors do not always have the right to go beyond the boundaries set and to try to see "behind the façade", as they may run the risk of infringing patients' rights to privacy and autonomy.

Alongside these basic problems in detecting and describing psychiatric disorders in patients in primary care settings and in transferring classification systems from psychiatry into primary care, there are also doctor-related factors which impede detection of such disorders. Studies of diagnostic techniques have repeatedly shown that general practitioners fail to collect significant pieces of information about current state and previous history and, for example, are too easily satisfied by explanations in terms of "problems in living" (Langwieler and Linden 1993) or focus too narrowly on the symptom initially reported, rather

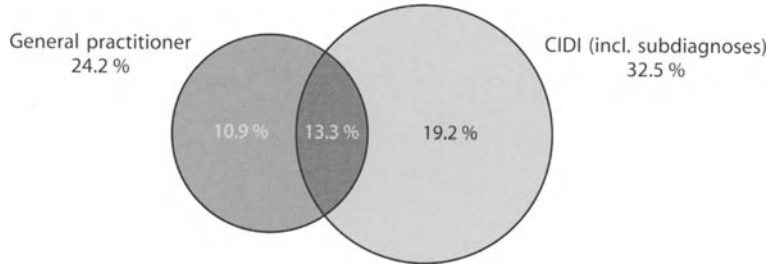


Fig. 2. Level of agreement between general practitioners' diagnoses and Composite International Diagnostic Interview, Primary Care Version (CIDI) research diagnoses. (From Üstün and von Korff 1995)

than enquiring further (Badger et al. 1994). A final fundamental problem which needs to be considered relates to the consequences which follow from making a diagnosis in a particular case. Given the uncertainty already discussed surrounding assessment of symptoms in this area, shifts in parameters such as the threshold at which a particular diagnosis is made will lead to substantial alterations in the sensitivity and specificity of whatever assessment criteria are being used to make this diagnosis. Thus a question arises about the clinical relevance of a false-positive diagnosis or of a failure to recognize a condition. While the spontaneous course of a severe psychiatric illness obviously has serious consequences for the person concerned and his or her environment and, in such a case, treatment can lead to improvements of obvious significance, this may not be true of the disorders encountered in general practice. At least on the basis of cross-sectional observation, it is an open question how far neurasthenic or somatoform complaints have really "serious" consequences. On the whole, we are only beginning to have scientifically evaluated and effective treatment methods for such disorders. Making a diagnosis which does not have treatment consequences may be of scientific interest, but from the general practitioner's perspective, it is on the whole contraindicated. This is not only because of the time and energy involved, but also because psychiatric diagnoses continue to be experienced as stigmatizing, so that if such diagnoses are to be made, there needs to be some clear prospect that the patient will benefit.

Goldberg (1979) was one of the first to show that an improved diagnosis which also influenced treatment plans could shorten the course of illness. This effect became more marked with increasing severity of disorder. Hoeper et al. (1984), who carried out an analogous study which was based only on giving doctors feedback about diagnosis, without making sure that some therapeutic action followed from the diagnosis, found that there was no alteration in the course of illness. Similarly, Tiemens et al. (1996) found that a higher recognition rate did not by itself lead to a better course of illness. On the other hand, Ormel and Giel (1990) reported that, in comparison to undiagnosed patients, diagnosed patients received psychotropic medication 4.5 times more often, psychotherapy 12.2 times more often and a referral to a specialist 3.3 times more often. This was accompanied by a sixfold better psychopathological state and a fivefold better level of social functioning.

3 Public Health Significance

With regard to the significance of psychiatric disorders in general and those encountered in general practice in particular, it is well established that to a greater extent than physical illnesses they are directly linked to a reduction in quality of life. Further, a series of findings have been reported showing that even the less severe psychiatric disorders which we have discussed have substantial social costs associated with them, above all because of the disturbances they cause in social adjustment and in the performance of work and family roles (Broadhead et al. 1990; von Korff et al. 1992; Wohlfarth et al. 1993; Hecht et al. 1990; Ormel et al. 1993).

Figure 3 shows data from the WHO study cited above regarding the association between severity of psychiatric disorder and the average number of days of sick leave. It shows that patients in general practice who have been diagnosed according to ICD-10 as having a psychiatric disorder have had around 6 days of sick leave in the previous month compared with 2 for the rest of the patients. As well as this tripling of days missed which occurs with clear-cut illnesses, it is of interest that even marginal disorders lead to a doubling of time off sick (Üstün et al. 1995b). This demonstrates the relevance from a public health

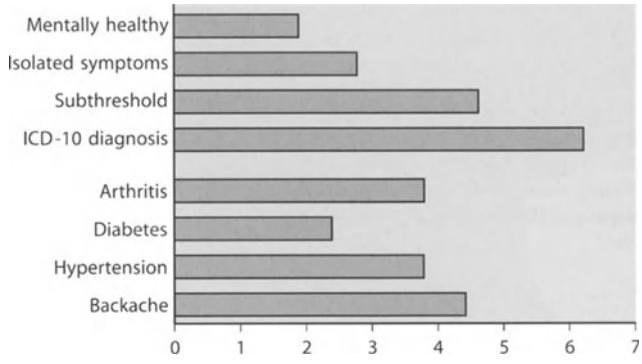


Fig. 3. Average days of sick leave in the last month in psychiatric disorders and selected physical illnesses. (From Ormel and Costa e Silva 1995; Üstün and Sartorius 1995)

perspective of disorders which may not seem especially impressive on cross-sectional study. The same result was also found when the Groningen Social Disability Schedule (SDS) was used as an indicator of illness-related disability or when global ratings made by the treating doctor were used as the basis of analysis. In all cases, a significant correlation was found between the degree of psychiatric disorder and limitations in the fulfilment of social roles.

It is of interest to consider to what extent physical co-morbidity plays a role in producing this effect, since, as already discussed, a correlation has been established between psychiatric and physical disorders. A logistic regression analysis was used to examine the relative contributions of physical and psychiatric disorders. As Table 2 shows and as would be expected, both physical and psychiatric disorders contribute to unfitness to work, but the relative influence of psychiatric disorders is certainly greater (Ormel and Costa e Silva 1995). To clarify this point, days of unfitness for work are twice as frequent in depressive as in comparable chronic physical illnesses. These unambiguously negative consequences of psychiatric disorders which on cross-sectional study might seem relatively mild may seem surprising at first glance. However, this finding is wholly in keeping with everyday clinical experience. Only a very small number of physical illnesses are clearly and directly linked to incapacity for activity, so that even where physical illness is present, interaction with psychological factors has a substantial mediating role in the extent of absence from work and sick leave.

Psychiatric co-morbidity also has a significant role as a determinant of the course and therefore the costs of treatment of physical illnesses. Sherbourne et al. (1996) found that patients with hypertension or diabetes who at the same time had a psychiatric disorder, particularly anxiety and depression, showed significantly worse health status on a health-related quality of life questionnaire for physical illnesses. Psychiatric disorders also lead to raised health care costs as a result of greater demands from these patients for general medical care. Wilkinson et al. (1988) observed over a 20-year period that every year 85%–90% of patients with psychiatric disorders consulted the general practitioner because of physical illnesses, compared with 60% of a control group. Similarly, Henk et al. (1996) used health insurance company data to show that, when a simultaneous questionnaire survey showed a raised depression score, an additional expenditure of on average US$ 1498 was incurred per year for patients receiving treatment from general practitioners for acute health problems. Simon et al. (1995) used a somewhat different methodology to calculate a corresponding additional expenditure of US$ 718, and McCombs et al. (1990) found additional costs of US$ 1043 for depressive patients who had not been successfully treated. These and other comparable studies (Manning and Wells 1992; Levenson et al. 1992; Widmer and Cadoret 1979; Greenberg et al. 1993) demonstrate that psychiatric co-morbidity is thus associated with substantial costs in health services.

The disabilities associated with psychiatric disorders are made more significant by the fact that in many cases the limitations imposed in various aspects of life are enduring. K.R. Lloyd et al. (1996) studied clinical histories in a follow-up study carried out 11 years after initial examination and found that 54% of patients were currently ill at the point of the follow-up examination and 37% could be classified as having chronic psychiatric illnesses.

Robins and Regier (1991) and Üstün and Sartorius (1993) have suggested describing the public health significance of illness according to criteria such as frequency, consequences (e.g. unfitness for work or treatment costs), public perception and possibilities for treatment. When this is done, psychiatric disorders of the type seen in general practice are of great significance and warrant increased awareness from the points of view of both treatment provision and research.

Table 2. Influence of psychiatric and physical disorders on unfitness to work: results of logistic regression analysis (Ormel and Costa e Silva 1995)

	Beta	SD	*p* value	Odds
Severity of somatic illness	0.39	0.08	0.001	1.47
Number of physical complaints	0.47	0.08	0.001	1.6
Number of psychological complaints	0.91	0.09	0.001	2.6
Number of ICD-10 diagnoses	0.92	0.09	0.001	2.52
Centre			0.001	

4 Selected Psychiatric Problems and Their Treatment

Some psychiatric illnesses are seen by the general practitioner and the specialist in psychiatry alike. However, there are also many disorders which are to be found almost exclusively in the general practitioners' practice and almost never in that of the specialist. Important examples in this latter category are generalized anxiety disorders, neurasthenia or somatoform disorders.

4.1 Psychotic Disorders

As already discussed, psychotic illnesses occur rather infrequently among patients in primary care. As they occur in the general population only with relatively low prevalences of around 1%–2%, they could not constitute a large patient group in general practice even if there were no specialist treatment. Severely mentally ill patients are also as a rule referred directly to specialists or hospitals, if these are available. However, general practitioners often remain involved in the physical management of these patients after referral, and in many cases they even play an important role in the management of patients with illnesses such as schizophrenia. Thus, in a study by Nazareth et al. (1995), 88% of the general practitioners interviewed felt that they had some share in responsibility for the treatment of their patients with schizophrenia. In 64% of patients, this took place in co-operation with a psychiatrist. On the other hand, 37% of these patients had no contact with a psychiatrist, so that their treatment remained exclusively in the hands of a general practitioner.

4.2 Depressive Illnesses

The most frequent psychiatric disorders in general practice are depressive episodes. They have the same symptomatic picture as is seen in psychiatric practice and described under the definition of "depressive episode" in ICD-10. Despite this prominent phenomenological similarity, there is evidence of some substantial differences between the depressive patients of psychiatrists and those of general practitioners. In a study of the distribution of patients between general practitioners and psychiatrists, it was found that in primary care the illnesses were less established as chronic and responded better to treatment with a serotonin-specific re-uptake inhibitor (SSRI) antidepressant (Dittmann et al. 1997). In addition, Blackburn (1984) reported that patients who were recruited for a treatment study from a psychiatric specialist clinic responded better to amitriptyline and less well to cognitive behavioural treatment, whereas the reverse applied for patients in general practices, which was attributed to differences in patient characteristics that exist between the two settings despite patients having the same diagnosis.

An impediment to the recognition of depressive illnesses is that they are typically linked with complaints about burdens experienced in life or experiences of loss. Depressive illnesses are thus seen as not requiring treatment because they can be understood. Such causal assumptions are, however, not scientifically as well-founded as they appear at first sight. Despite these links with difficulties in living, there is evidence that, when symptoms reach a moderate level of severity, depressive episodes should be treated regardless of their aetiology. Thus Zung et al. (1983) used a randomized design to show that patients in primary care who were diagnosed by means of a questionnaire as having depression showed an improvement in 64% of cases when an antidepressant was prescribed, whereas only 28% of untreated patients improved. This corresponds with the results of the many controlled trials on the effectiveness of antidepressants. As an alternative, specific psychotherapy, e.g. cognitive behavioural or interpersonal psychotherapy, can also be indicated. However, this is beyond the range of treatments generally available at a primary care level, and it will therefore require referral.

A problem which remains in part unresolved is how to treat mild depressive episodes or dysthymia, the label given to mild but chronic depressive mood. Paykel et al. (1988) found in a treatment study that antidepressants led to clinically significant benefits in more severe depressions, but could not bring about convincing improvements in milder ones. This finding fits with those of Blackburn (1984), cited above. Since the introduction of the newer antidepressants, there is however an increasing body of evidence that a course of treatment also makes sense for less severe but chronic depressive illnesses and that good results can be expected (Kocsis et al. 1995). Scientific evidence about the effectiveness in depressive illnesses of general counselling, addressing the psychological conflicts patients are experiencing, is lacking. However, this is not a reason not to give these patients the opportunity to talk about their feelings and living situation and to examine their depressive cognitions.

4.3 Generalized Anxiety Disorder

Generalized anxiety disorder is characterized by an increased level of tension, increased autonomic arousability and a tendency to constantly worry (Zubrägel and Linden 1997). Neither the doctor nor the patient may recognize these disorders for what they are, as the patient is of the view that his or her worries are justified. In psychological terms, worrying about something means always taking care and taking control of the situation (Butler et al. 1991). Thus patients do not identify the exaggerated worry as the problem, but rather the object of anxiety, whether it is that something could happen to a child on the way to school, that the washing machine might break down,

that the patient's husband might have a road accident etc. Patients therefore present because of the physical symptoms accompanying the anxiety, or because of worries about their own health or their children's, or with an "overwork" syndrome. Only by asking questions about the core symptoms can a diagnosis be made, while premature "understanding", as with depression, leads to a misdiagnosis.

Almost all psychotropic substances, prescribed and unprescribed, have effects on anxiety (Linden et al. 1988). This applies to alcohol, barbiturates and other sedatives, benzodiazepine tranquillizers or hypnotics, high- and low-potency neuroleptics and sedating and non-sedating antidepressants. However, none of the options mentioned can be seen as ideal, as each has significant limitations. Alcohol, barbiturates or benzodiazepines are ruled out because of the development of tolerance and dependence, as the nature of generalized anxiety disorder is such that treatment needs to be medium to long term. High-potency neuroleptics in low doses are in widespread use in Germany, for example (Lehmann 1989). However, empirical evidence on the effectiveness and risks of this treatment is largely unsatisfactory, so that it should currently be seen as an experimental treatment, which the prescriber must have particular reasons for employing it in an individual case. Further, even with a low dose, the risk of a tardive dyskinesia cannot be excluded. The sedating tricyclic antidepressants are also in wide use and can be used long term, lead to no problems of dependence, have positive effects on the depressive symptoms which sometimes occur with anxiety and also can have a positive effect on sleep disturbance. A disadvantage is that the anticholinergic and α-adrenergic side-effects can be experienced by patients with anxiety as an increase in the primary symptoms. Currently, the most promising pharmacotherapeutic alternatives are the 5HT-1A agonists (e.g. buspirone). These do not have an immediate effect, but, like antidepressants, only work after a treatment introduction phase. They are largely free from subjectively experienced side-effects and produce clinically significant reductions both in the primary psychopathology and in the dysfunctional cognitions associated with it.

In psychotherapeutic terms, clinically significant improvements can be achieved through behavioural therapy focusing on worrying as the problematic behaviour, in contrast to treatment measures which aim only at supporting patients in coping with problems (Durham et al. 1994; Butler et al. 1991). Even though the basis has not been scientifically worked out in detail, enough evidence is available to propose some treatment guidelines for psychological treatment by general practitioners. A central principle must be that not the problem, but the worry itself is the object of treatment. One must therefore work with the patient on developing more and more alternative ways of thinking, e.g. about their child returning from school a little later than usual. Patients must also be encouraged to take good care of their psychological health. This includes having rest periods which are free from any commitments, sufficient physical activity, care in the use of alcohol or coffee and regular sleeping times.

4.4 Chronic Fatigue Syndrome and Neurasthenia

Patients with chronic fatigue syndrome (Holmes et al. 1988; A. Lloyd et al. 1988) are seen almost exclusively by general practitioners and, because of the particular characteristics of their disorder, almost never by psychiatrists. There are a variety of very different views on the classification of this syndrome, reflected in synonyms such as effort syndrome (Wood 1941), exhaustion syndrome (Dowden and Johnson 1929; Macy and Allen 1934), fibromyalgia (Yunus 1989), neurocirculatory asthenia (Wheeler et al. 1950; Mantysaari et al. 1988), environmental hypersensivity disorder (Stewart 1987), post-viral or post-infectious fatigue syndrome (Behan et al. 1985; Bannister 1988) and more specifically chronic Epstein-Barr infection (Tobi and Straus 1985). The descriptions mirror the professional backgrounds of the authors concerned rather than really having a scientific foundation (Holmes et al. 1988; David et al. 1988; Wessely and Powell 1989). In any case, fatigue as a primary complaint produces substantial problems of differential diagnosis and, from a practical point of view, there are limits to the extent to which logical explanations can be provided under general practice conditions (Ward et al. 1996). Common characteristics are a physical and mental susceptibility to exhaustion not accompanied by the affective and cognitive indicators of a depressive illness, typically together with myalgic complaints in the absence of other somatic findings. The greatest diagnostic problem with this disorder is that tiredness is an exclusively subjective experience, so that in the study carried out by Jenkins et al. (1988) observers achieved a reliability score of only 0.02.

In around three quarters of cases, patients who present in specialist outpatient clinics with chronic fatigue syndrome suffer from clear-cut psychiatric disorders, and nearly half of these are depressive illnesses (Allan 1944; Manu et al. 1988; Morrison 1980; Sugarman and Berg 1984; Taerk et al. 1987; Wessely and Powell 1989; Wessely et al. 1996). This is true even when the sample is drawn exclusively from patients in whom blood tests have led to an Epstein-Barr infection being suspected (Katon et al. 1988; Kruesi et al. 1989).

Neurasthenia has a largely similar symptomatic picture as chronic fatigue. In contrast to the chronic

fatigue syndrome, which through nomenclature is conceptualized as a somatic disorder, neurasthenia is understood as a psychiatric disorder and included in ICD-10 among the psychiatric illnesses (Sartorius 1997). Nonetheless, in the first formulation by Beard (1869, 1880), it was described as an overtaxing of the nervous system which was to be understood in organic terms. Only in the writing of later authors such as Freud (1895) was it understood as a psychoneurosis (Linden 1991). Neurasthenic symptoms may also be an element in a depressive syndrome, so that, as with cases of chronic fatigue syndrome, the majority of cases of neurasthenia also meet the criteria for other psychiatric disorders, particularly depression (Kleinman 1982). For this reason, this disorder is no longer listed separately in DSM-IV, but is subsumed under dysthymia.

Despite this high co-morbidity or hierarchical incorporation into the categories of depressive and anxiety disorders, there still remains in all studies a group of patients who complain exclusively about exhaustion and easy fatigability and who therefore must be viewed as a distinct group. For this reason, the ICD-10 criteria allow neurasthenia (fatigue syndrome) to be diagnosed only when no depression or anxiety is present (Dilling et al. 1994).

Neurasthenia is by definition a chronic illness. According to a study by Kroenke et al. (1988), affected patients in general practice had on average already been suffering from their complaints for over 3 years, and around 70% still had them after a further year. This has been confirmed by the observations made by other authors (Nelson et al. 1987; Valdini et al. 1987).

These disorders are associated with substantial reductions in subjective well-being and are also chronic and not amenable to any other forms of treatment; thus, whatever the aetiological debates, the therapeutic challenge of trying to improve the well-being of the individual remains. Pharmacotherapeutic strategies have not so far been clearly established for this. Mild stimulants, e.g. nootropic drugs, such as piracetam or centrophenoxin, are probably worth a try. However, decisive factors are changes in patients' attitudes towards their own well-being, a greater tolerance of physical discomfort and above all a shift from a focus on their own internal state to an orientation towards action (Kuhl 1992). The significant pathogenic factor is the development of avoidance of unpleasant subjective states. This needs to be counteracted by establishing activity as a therapeutic measure. One treatment approach involves so-called paradoxical interventions. This means that the patient can withdraw and have a rest when he or she is feeling good, but that feelings of tiredness must be counteracted through activity. The therapeutic procedure thus has a similar form as for somatization disorders.

4.5 Somatoform and Functional Disorders

Somatoform disorders are defined as physical complaints for which there is no recognizable organic explanation or which can be described as exclusively functional disorders (Lipowski 1988). They include somatosensory misperceptions (somatization disorder), an autonomic dysregulation of organ functioning (somatoform autonomic dysfunctioning) or an anxious interpretation of normal or misregulated somatosensory perceptions connected with anxiety as indicating the presence of a particular physical illness (hypochondriacal disorder). Finally, combinations of all of these are also found.

In somatization disorders, patients complain of multiple non-specific physical symptoms, such as headaches, limb and joint complaints or heart complaints. If the patient is asked for further complaints, this generally results in a substantial increase in the number of complaints. Patients are firmly convinced that they are suffering from one or more physical illnesses, for which further investigative and treatment procedures are required (Samuels 1995). Negative test results do not reassure the patient.

In hypochondriacal disorder, the focus of complaints is not so much on a general and direct physical experience as on a particular organ, with a lack of confidence in this organ's functioning or fears that there is something major wrong with the organ. In autonomic functional disorders, there is a general autonomic instability or autonomic dysregulation, with an increase in dermatographism, sweating, palpitations and rise in blood pressure, stomach and bowel complaints, shaking and so on. These functional disturbances generally increase under stress.

The somatization disorders also present particular problems of differential diagnosis. Thus psychosomatic complaints are an integral part of the depressive syndrome, and various studies report that around 48%–94% of patients examined are suffering from a depressive disorder (Smith 1992). Other important differential diagnoses are anxiety and dependence disorders (Starcevic et al. 1992; Katon and Russo 1989).

In aetiological terms, both psychological and somatic factors are significant in somatization disorders. In psychological terms, somatosensory perception preferences (Barsky 1979), behaviour learnt from models in the family (Benjamin and Eminson 1992), attitudes to health (Tyrer et al. 1990), negative affectivity (Watson and Pennebaker 1989), societal sensitization (David and Wessely 1995) or attention from other parties, e.g. relatives and doctors (Mayou 1976, 1993), all play a role. From a biological point of view, there are indications of genetically determined

vulnerability (Bohman et al. 1984; Cloninger et al. 1984) and of a constitutional or acquired susceptibility to increased autonomic or hormonal arousal (Sharpe and Bass 1992).

The development of proven treatment strategies for these disorders is as yet only in its initial stages. Bass and Benjamin (1993) give general treatment recommendations, which can also be applied in general practice. One important principle is to recognize the disorder for what it is, so as to avoid mistaken physical management. A consistent relationship then needs to be built up, which also prevents patients from constantly changing their doctors. Patients should be given clear medical information about negative physical findings and about positive results regarding reactivity to stress. The suffering of the patient should be met with understanding and should never be devalued. To this end, the individual illness theories of the patient should not be dismissed too hastily, but seen as the expression of the experience and cognitions of the patient and taken seriously. The psychological factors in the patient's current situation and personal history which promote the disorder should be discussed with the patient and, where applicable, other ways of problem-solving should be examined.

In one of the few treatment studies, Real-Perez et al. (1996) report a brief family therapy, carried out by the general practitioner him- or herself. The patients had been suffering for at least 1 year from a somatization disorder. In 61% of patients, the results indicated some success which still persisted after 6 months.

4.6 Alcohol Dependence

In a survey by Roche et al. (1995), the overwhelming majority of general practitioners saw the treatment of alcohol dependence as one of their important tasks. Dependence disorders are treated by general practitioners and specialists as well. However, it is largely reserved for the general practitioner to see people with addiction illnesses which are in their early stages and not yet associated with grave damage. This gives the general practitioner opportunities for early recognition and possibly targeted early intervention to prevent problems becoming chronic.

The means by which this can be achieved are almost exclusively psychological. The first step is careful history taking. The second is to present the patient with a clear formulation of the problem without covert overtones of reproach. As a third step, self-monitoring and limits to the volume of alcohol consumed should be agreed with the patient. As patients with dependence problems on the whole cannot stick to these, the next step is to undertake an analysis of the circumstances which promote drinking and those which reinforce self-control. It is important to adjust to the fact that treatment success is not likely to be achieved by giving information on a one-off basis, but requires consistent, longer-term support (Schmidt 1997).

4.7 Dementia

With the increasing numbers of old people, illnesses associated with ageing, i.e. primarily dementing illnesses, represent a major care task. From around the age of 70, the prevalence of dementia rises in an approximately linear relationship with age, with the result that approximately half of those in their nineties are suffering from a dementing illness (Jorm et al. 1987; Reischies et al. 1997). Old people generally also have high levels of need for medical help because of other forms of morbidity, explaining the finding of Eefsting et al. (1996) that 5.2% of the patients of general practitioners are suffering from dementing illnesses.

Studies of the use of medical services indicate that old people have a tendency only to consult one doctor with their mostly chronic and multiple complaints and that they are less likely to be referred on to specialists (Ryynanen et al. 1997; Linden et al. 1996a). This means that general practitioners are responsible for diagnosis and treatment of psychiatric disorders in the elderly population (Cooper et al. 1992). There is also a danger that psychiatric disorders among old people are not recognized as illness (Eefsting et al. 1996; Pond et al. 1994), but seen as by-products of the accompanying physical multiple morbidity, as consequences of the social situation or simply as an unavoidable manifestation of old age. In around half of cases, the patients themselves, according to Newens et al. (1994), are not aware that they have a psychiatric disorder.

The focuses of general practitioners' treatment encompass not only the actual dementia syndrome, but also the other psychological and above all physical co-morbidity. Finally, support for relatives acting as carers is an important therapeutic task.

Currently, the possibilities for direct intervention in dementing illnesses are limited. However, the so-called anti-dementia drugs or nootropics represent important treatment possibilities, whose effectiveness can be seen as essentially established (Moeller 1991). The question is rather about when the success which can be expected is great enough to justify the treatment costs. Epidemiological treatment studies suggest that the current situation is that only a few per cent of those affected are treated with nootropics, and there are thus reasons for suspecting undermedication (Helmchen et al. 1996). After all, given the severity of the illness and the considerable costs incurred for care and nursing,

even small improvements or halting the progression of the illness may have great significance for those affected, their relatives and the wider network involved in supporting them. The same applies to the treatment of co-occurring physical illnesses. Here, too, epidemiological treatment studies indicate that, with increasing severity of dementia, less medical treatment is provided (Helmchen et al. 1996), even though overall physical state is of considerable significance for the development of the illness. Of all groups of doctors, general practitioners have the best basis for managing overall morbidity within an integrated treatment plan, and they can do this with a competence similar to that of specialists (Colenda et al. 1996).

General practitioners have a particular role in supporting the relatives of those with dementia. By their nature, these illnesses demand that patients are supervised and, at a later stage, also intensively nursed. Even among those with the greatest needs for nursing, the majority of patients are not cared for in nursing homes, but at home by relatives, which makes great demands both of them and of the treating doctors (Linden et al. 1996b; Tyler and Bourguet 1997). The ability of relatives to assume care depends among other things on the network of professional support (Vernooij-Dassen et al. 1997). General practitioners have a good knowledge of the social and family environment of the patient, and the work of the caring relative can therefore be facilitated not only through good provision of medical care, but also through advice and support on nursing aspects of care and in relation to social needs. Doctors' basic training gives them only limited education in areas which concern social needs. According to the studies carried out by Cheok et al. (1997) and Shah and Harris (1997), they feel less confident about advising on these areas than on other aspects of diagnosis and management, and relatives often expect more support than they obtain (Brodaty et al. 1994; Commissaris et al. 1995).

5 Primary Medical Care in a Complex Health Care System

5.1 Structures and Tasks of Primary Medical Care

The care provided for a patient is not only a function of medical knowledge, but depends just as much on the setting in which it is delivered and also on the treatment strategies which are permitted or available within a particular treatment setting. Economic considerations now have a greater influence on medical practice, it is not the doctor's task to do what is medically conceivable or optimal, but, as dictated by the law governing medical insurance, for example, only what is deemed to be necessary and sufficient, and sometimes merely what is practicable. In this respect, the general practitioner takes on a key position in many health care systems. He or she is responsible not only for basic care, but also for either mobilizing or avoiding the involvement of specialists and sometimes for co-ordinating specialist input. The significant facets of primary or general practitioner treatment thus include both the initial contact and first treatment and also continuing care following specialist medical intervention.

There are substantial international differences in the ways in which primary medical care is organized, depending both on the available material resources of a society and on political views. Thus the organization of the general practitioner system differs between Holland and Germany, for example, despite comparable standards of living. While in Germany a patient with any complaint is free to go to whatever doctor he or she chooses, in Holland the patient first has to go to a general practitioner and may then also see a specialist on referral from a general practitioner. This is the so-called gatekeeper function of the primary care physician and has direct consequences for service use pathways. In Holland, over 95% of patients who present anew in general practice have never been in treatment elsewhere for the problem concerned, whereas the proportion in Germany is only around 75% (Üstün and von Korff 1995).

Another aspect of the organization of primary care which is important in making international comparisons is the distinction between hospital and family doctor systems. These forms of care also exist alongside each other in a health care system in some cases. According to Starfield (1992), collective or hospital care can be differentiated from family doctor care in that patients consult the hospital and not an individual doctor, that the caring doctor in some circumstances may change from visit to visit and does not necessarily have prior knowledge of the patient, that a multiprofessional team and not a single doctor takes on the care of the patient and that the treating doctor as a rule does not receive direct reimbursement for this work. This way of organizing work has direct consequences for the diagnosis and treatment of the disorders treated. Thus the rate of diagnosed psychiatric disorders in hospital facilities is only around half as high as in family doctor practices, for example. This effect can be demonstrated even when corresponding facilities in the same city are compared (Üstün and von Korff 1995). The most plausible explanation is that, in the family doctor system, the doctor's longer acquaintance with both the patient and the relatives makes much more information available to him or her than in a hospital setting.

These structural characteristics of primary care medicine are paralleled by therapeutic tasks. These can be summarized as detection of the onset of illness, initial treatment and basic care, as well as the integration of specialist diagnostic procedures and treatment and referral for specialist treatment interventions.

The problems already outlined of non-recognition of psychiatric disorders have a particular relevance in relation to this description of the general practitioner's role in care. It is with just those patients who are least ready to accept the idea that they have a psychiatric illness that there is a particular need for the diagnostic and therapeutic skills of the general practitioner. He or she is the only person who can detect such cases and work to prevent adverse outcomes from illness such as suicide. In many cases, he or she is the only person in a position to treat those patients who find going to a psychiatrist an unacceptable idea. Thus in Great Britain, for example, the Royal College of General Practitioners has initiated a campaign together with the Royal College of Psychiatrists to fight the stigma associated with psychiatric illnesses and above all to improve awareness of such disorders in primary care settings (Sims 1993).

5.2 Co-operation Between General Practitioners and Psychiatrists

The next question to be considered in relation to the treatment of mental illnesses is the division of roles between general practitioners and psychiatrists, psychotherapists, psychiatric hospitals, social care services such as day centres, hostels and workshops and other providers of services for the mentally ill.

Williams and Clare (1986) have described three models of co-operation between general practitioners and specialists: the referral, the substitution and the consultation model. The referral model envisages the general practitioner as the primary treatment provider, who, depending on the type of disorder, may send the patient on to a specialist for additional and further treatment. In the substitution model, the specialist is seen as the primary doctor for illnesses which fall into his or her area of competence and is also approached directly by patients if they suffer from disorders in this category. The consultation model sees the treatment of a patient as almost entirely in the hands of the general practitioner, who in difficult cases presents the patient to a specialist for explanation and advice with regard to further treatment or else makes use of other groups of professionals such as social workers and psychologists for particular types of help. Which model is implemented depends partly, as already discussed, on the organization of the health care system. In many countries, all three models may be encountered at the same time. Which form of collaboration is chosen also depends not at least on the patients or general practitioners themselves.

Surveys in practices indicate that the current rate of onward referrals and co-operation is around 1%–2% of practice clients, with a great deal of variation not only between different general practitioners but also between regions and health care systems. This variation cannot be attributed exclusively to differences in local prevalence of psychiatric disorders. Around 10% of patients have had at least one contact with a specialist in the past (Zintl-Wiegand et al. 1978; Gastpar 1984; Dilling et al. 1978; Carey et al. 1994; Arreghini et al. 1991; Geiselmann and Linden 1989). Patient-related factors which make referral more likely are a known psychiatric history, psychiatric primary complaints, social problems and the family doctor making a relevant diagnosis. The latter, however, is of less significance than the primary complaints of the patient. The presence of clearly defined physical illnesses make referral less likely (Arreghini et al. 1991; Strathdee et al. 1990; Chithhiramohan et al. 1993; Verhaak 1993). Older patients also have a tendency to remain with their general practitioner and not to seek out the help of an additional doctor such as a psychiatrist (Lingg et al. 1995; Helmchen et al. 1996). Variables associated with the doctor are another contributing factor, as referral is less likely if the definition of the physician's role is narrowly confined to the management of physical disorders and is influenced by his or her view about whether he or she has the appropriate competence (Verhaak 1993; Fritzsche et al. 1993). Finally, referral also depends on organizational factors such as the structure, proximity and availability of specialist services (Dilling et al. 1984; Verhaak 1993; Arreghini et al. 1991).

Studies on needs for referral have as a rule tended to start out from the premise that the rate of referrals is markedly low (Dilling et al. 1978). Indications for referral identified by Helmchen (1991) are as follows: (a) doubts about psychiatric diagnosis and problems of differential diagnosis in multi- and co-morbidity, (b) severity of illness and threat of complications such as suicidality, (c) length of illness, degree to which it has become chronic and treatment resistance, (d) requirement for long-term medication, (e) unwanted side-effects of medication and (f) need for specific psychological treatments. In studies in which psychiatrists directly examined patients in family practices, half of the patients seen were in need of psychiatric treatment (Dilling et al. 1978; Zintl-Wiegand et al. 1978). In a newer study by Schulberg et al. (1995) on depressive illnesses in primary care, it was found that, among 283 patients with depression, 70% could be appropriately treated by the general practitioner, 13%

needed referral to a psychiatrist and 17% had other disorders in need of other forms of treatment.

Empirical studies on when general practitioners make referrals show that by then, as a rule, some treatment has already taken place (Maguire et al. 1995). In 61%, drug treatment had already been given, while 67% had received some form of advice or counselling. In 30%, the reasons for referral were for treatments not available in the primary care setting, in 20% because of treatment resistance and in 14% for a sharing of the burden of care in chronic disorders. This corresponds on the whole to guidelines for the management of depressive illnesses such as those issued by the Agency of Health Care Policy and Research of the USA (AHCPR 1993), which recommend that general practitioners should generally make a first attempt at treatment, including drug treatment. However, some psychiatrists are critical of such recommendations and advocate earlier referral (Munoz et al. 1994).

Empirical evidence is required as to when referral results not only in additional costs, but also in a better clinical outcome. According to the available studies, the types of disorders which are mainly managed in general practice are not necessarily managed better by a specialist than by general practitioners themselves. Jenkins and MacDonald (1994) randomly allocated 65 older depressive patients who had been detected in the context of a screening programme in general practices either to management over 9 months by a multiprofessional psychogeriatric team or to continued management as usual by the general practitioner. Their results indicated no significant differences between the two groups. In a study by Katon et al. (1992), half of a group of patients of 18 general practitioners who had in the past been heavy users of medical services were randomized to receive a psychiatric consultation. After 6 months, there was a significant increase in antidepressant prescriptions for the referred patients. However, there were no significant differences with regard to psychopathological status, degree of disability or the continuing use of medical services. In a similar study by the same authors (Katon et al. 1995), patients with depression were randomly allocated to joint management with a psychiatrist. At follow-up, the intervention patients were receiving significantly more antidepressants. Overall, there was a significantly better course of illness only for those with major depression, while the improvement rate for minor depression was the same in the intervention and the control groups.

In a study by Scott and Freeman (1992), 121 patients were randomized either to a psychiatrist, a psychologist behaviour therapist, a social worker or treatment by the general practitioner alone. After 4 and 16 weeks, independent raters detected clear improvements in all patient groups, but no clinically significant differences between the treatment groups. Meanwhile, treatment by a specialist rather than the general practitioners was associated with four times as much expenditure of time and twice as great a cost. Patients were most satisfied with the psychological treatment and above all management by the social workers (Scott and Freeman 1992; J. Scott et al. 1994). Thus undiscriminating referral of all those with psychiatric disorders from the general practitioner to the specialist cannot necessarily be expected to lead to a better course of illness. The question to be resolved is which patients it is appropriate to refer. Helmchen's (1991) proposed criteria, listed above, can be seen as practicable guiding principles which are supported by research evidence.

6
Specific Methods of Diagnosis and Treatment

When patients go to the general practitioner as the first port of call, the doctor's task is to clarify what the current problems are, to decide which diagnostic procedures and treatment measures need to be introduced now and, where feasible, to make an initial attempt at treatment. The diagnostic and treatment strategies which are typically available in primary care and the ways these are used will now be outlined.

6.1
Diagnostic Procedures

The particular difficulties of diagnosing psychiatric disorders among patients in general practice and the resulting failure to detect and thus to treat a significant sector of this patient group have already been discussed. The following are recommendations for improving the detection rate.

A factor which has a considerable impact on clinical detection of psychiatric disorders is the doctor's interviewing technique. Goldberg (1990) used video observation of doctors with better and worse recognition rates and identified a set of criteria which allow these groups to be distinguished:

1. Beginning of interview
   a) Eye contact
   b) Clarification of complaints
2. Interview style
   a) Reacts to verbal cues
   b) Reacts to non-verbal signals
   c) Can steer patient's flow of talk
   d) Can cope with interruptions
3. Style of questioning

a) Precise psychiatric questions
b) Supportive comments
c) Asks about domestic relationships

The successful doctors were also on the whole more experienced, had a greater interest in psychiatric questions and were happier in their practices. Interviewing technique is thus certainly dependent to a degree on personality and experience. However, as Goldberg et al. (1980a,b) were also able to show, it was possible through video feedback and targeted training to improve interviewer behaviour.

An option for improving the recognition of psychiatric disorders is routine use of diagnostic scales as screening instruments. The problems in implementing such a method are however substantial. According to Wittchen and Essau (1990), a screening instrument which could be used for all patients in general practice would need to fulfil a set of essential requirements. These include the criteria that it should be short and not time-consuming, simple to fill out, easily rated, useable in very different forms of practice organization, applicable with very different patient groups, e.g., young and old, and sensitive for the whole spectrum of possible psychiatric disorders. It also needs to be able to be interpreted without much training, to produce results which are sufficiently concrete to allow treatment to be initiated on the basis of it, to avoid false-positives results due to co-occurring physical morbidity, to have adequate psychometric reliability, specificity and sensitivity, to provide information as far as possible not only on the symptomatic but also the diagnostic level, to pick up disorders in their early stages and ideally also to encompass additional aspects, such as level of psychosocial functioning. It is clear from this list that such an instrument is not likely to ever exist.

The currently available screening instruments tend to cover a group of core symptoms which are very non-specific and can therefore occur in almost all psychiatric disorders. They do not allow diagnostic classification. An example is the GHQ (Goldberg and Williams 1988), which asks questions about general physical and psychological health, about a series of somatic complaints including sleep and about psychological complaints including suicidality. The questionnaire is available in a longer form of 60 items and shorter forms with 28 and 12 items. From the point of view of practicability, there would be no major problems in giving every patient who attends the practice such a questionnaire in the waiting room and having it rated by the clinic receptionist. From our own experience, we know that this is accepted by patients and in some cases even positively welcomed. With repeated administration, the questionnaire can also be used as an instrument for the documentation of course. Data about this are not available, but despite this basic practicability, it is the exception rather than the rule for such scales to be applied in general practices. One can only speculate that the problems of underdiagnosis may not be taken as seriously by doctors as the literature suggests, so that measures to address this are not seen as necessary.

Once it is established that a condition meeting criteria for a psychiatric diagnosis is present, starting properly targeted treatment will often require a precise diagnosis to be established. On the basis of the modern psychiatric classification systems ICD-10 and DSM-IV, it is now possible to distinguish between important psychiatric diagnoses with limited psychopathological knowledge, through structured questioning and use of the algorithms provided. In research projects, these structured interviews, e.g., CIDI (Wittchen et al. 1989), are sometimes also carried out by lay people after relatively short training. However, one cannot envisage this working in practices, as the interviews are time-consuming and commit the doctor or his or her colleagues to using up expensive time.

A possible alternative is the use of computer-supported interviews, whose suitability for application in general practice has to some extent already been demonstrated (Weissman et al. 1995; Olfsson et al. 1995a). The CIDI is also available in a computer version. The patient is seated directly at the computer and uses screen and keyboard to answer all the questions. At the end of the interview, diagnostic scores can immediately be generated using the computer. The patient needs only to be observed by a staff member from a distance, in case there are any queries or problems. From our own experience, such a procedure is feasible under practice conditions and is also accepted by patients. In each case, a doctor will need to check over the results of this questioning, but on the basis of the information available this can be done in a much more focused way.

6.2 Drug Treatments

For the disorders which are most significant for primary care, the most important pharmacological treatment strategies according to currently available literature have already been outlined. While in the initial years after the introduction of psychotropic drugs the treatment of psychotic illnesses was in the foreground, there is now increasing interest in the further development of psychopharmacotherapy for non-psychotic and milder disorders, such as anxiety disorders, dysthymia, somatization disorders or even personality disorders. The assumption that psychoses should be treated pharmacologically and never

psychotherapeutically, while the so-called neuroses should be treated psychotherapeutically and never pharmacologically, is no longer valid. Advanced understanding of the psychological and biological bases of these disorders permits understanding of why biological and psychological treatment may be effective in the same disorder and why in some cases combination treatments are also indicated. Thus drug treatment for those disorders which are on average milder is not merely an emergency solution or an inadequate substitute treatment, but is one of the core treatment strategies and, compared with alternatives, is relatively well evaluated. It should therefore play an important role in the armoury of the general practitioner. In the following, individual disorders and their management will not be discussed again, but rather we will look at the epidemiology of prescribing practices among general practitioners.

Pharmaco-epidemiological studies indicate that general practitioners rather than psychiatrists prescribe the largest total quantities of psychotropic drugs (Schäfer 1990).This can be explained in terms of the fact that there are around 20 times as many general practitioners as psychiatrists (Thust 1997), and per doctor they treat around twice as many patients. Considering the quantity of prescriptions per individual doctor or even per treated patient, psychiatrists clearly prescribe more psychotropic drugs than general practitioners, as would be expected. Psychiatrists also have a higher rate of prescribing a combination of psychotropic drugs (Meredith et al. 1994; Dittmann et al. 1997).

Pharmaco-epidemiological studies have repeatedly shown substantial differences both between regions and between professional groups. These are essentially related to differences in the prescribing behaviour among general practitioners (Wessling et al. 1991; Bellantuono et al. 1988; Friebel 1989). Such data show that choice of treatment and prescribing behaviour are basically not only dependent on medical factors in the narrower sense, but also on gender, attitudes of patients and doctors, organizational and political guidelines, and ethical and cultural context (Raynes 1979; Williams 1983; Hohmann 1989; Lloyd and Moodley 1992; Maslowski 1987; Linden 1994; Morabia et al. 1992).

In the WHO primary care study (Üstün and von Korff 1995; Linden et al. 1999), it was found across all international sites that 11.5% of practice attenders were treated pharmacotherapeutically for psychiatric disorders, with considerable variations from 2% in Shanghai to 29.6% in Santiago. Surveys by other authors have arrived at comparable or even higher rates (K.R. Lloyd et al. 1996; Joukamaa et al. 1995; Linden et al. 1996b). Among the patients who were mentally ill in the view of doctors, an average of 51.3% were treated with psychotropic drugs in the widest sense. Looking only at the patients meeting criteria according to the CIDI for an ICD-10 diagnosis, 27.6% were given appropriate treatment. These data firstly indicate that psychotropic drugs are a first-line method of treatment by general practitioners for psychiatric disorders. Secondly, they form a further basis for the premise which has already been discussed, and which is in fact obvious, that the recognition of such disorders is an indispensable requirement for treatment. Of the range of psychotropic drugs which are in use, tranquillizers and sedatives play the greatest role in 26.3% of patients recognized as mentally ill, followed by antidepressants in 15.0% and herbal and tonic substances in 13.2%. These data already give an indication that to some extent the treatments concerned are symptomatic and non-specific. This is confirmed by observation of patients with acute depressive episodes according to ICD-10 who were also diagnosed by the treating physician. Only 22.2% received an antidepressant. A further 27.6% received tranquillizers and another 23.2% other psychotropic drugs. On theoretical grounds, it could be concluded that treatment is largely non-specific in two thirds of cases.

6.3 Psychological and Social Treatments

In the conventional use of the term, psychotherapy is a specific treatment method which generally involves a 50-min interaction once a week with a specifically trained psychotherapist. Even if some general practitioners have a psychotherapeutic training and treat some individuals in a psychotherapeutic framework, the time requirements alone mean that this is not a form of primary care treatment. However, general practitioners often play an important part in directing people to such psychotherapy. They are also often asked to be joint participants in treatment in cases where drug treatment or another physical treatment is indicated in addition to psychotherapy in order to optimize the treatment outcome.

In addition to the traditional form of psychotherapy already discussed, there are various further forms of psychiatric and general practitioner psychotherapy. Basic supportive psychotherapy, as practised in many psychiatric contexts, is defined by the doctor having specific training, spending at least 20 min with the patient and, among other things, making some systematic use of the doctor–patient relationship in treatment. Thus it is a focused psychotherapeutic procedure for the treatment of different psychological disorders. "Syndrome-related verbal interventions" are brief psychotherapeutic interventions related to psychiatric disorders in which particular account is taken

of effects of psychopathology in limiting communication and functioning. Finally, basic care also involves the provision of general counselling and of specific counselling on particular problems arising from illness and treatment.

According to information provided by the doctors in the WHO primary care study (Üstün and von Korff 1995), 52.4% of patients recognized as mentally ill received relevant advice or counselling. Patients raise a great variety of topics in consultations with general practitioners. Zwernemann (1997) studied the types and frequencies of themes discussed during 100 general practitioner contacts. Table 3 summarizes these data, comparing patients with and without psychiatric disorders. In a further study, Olfsson et al. (1995b) studied the forms of psychological intervention used by general practitioners. They carried out some form of psychological intervention with 24.1% of all patients. Above all, these consisted of listening to problems (22.4%) and giving advice (19.0%). Such an intervention was used with 66.7% of patients with psychiatric disorders. According to the results of this study, advice and support with problems are an important aspect of the work of general practitioners, an aspect on which adequate scientific research has not yet been carried out.

A question of particular interest is the extent to which, in the treatment of psychiatric disorders by general practitioners, there is an inverse relationship between drug treatment and psychological intervention. An assumption sometimes expressed is that drug treatment is a substitute for personal attention and talk. In relation to this question, Fletcher et al. (1995) carried out a study in which they worked together with general practitioners to make targeted counselling available for patients to varying extents. Their results were unexpected and contradicted their initial hypothesis. The more counselling related to personal problems and other psychological complaints was offered, the higher the rate of prescription of antidepressants and tranquillizers. Apparently, detection of such disorders was associated with a greater overall intensity of treatment. These results were confirmed by a study carried out by Katon et al. (1996). This evidence thus indicates that conversations and counselling do not replace psychopharmacological treatments. Under general practice conditions, they are not alternatives, but complementary aspects of treatment.

Table 3. Themes of consultation for patients in general practice with and without psychiatric disorders (Zwernemann 1997)

Themes	Psychiatric diagnoses (%)	Physical diagnoses (%)
General topics	32	62
Brief pieces of advice given	18	2
Discussion of problems	20	6
Advice about problems	4	2
Psychotherapeutic intervention	6	0

A point that has already come up several times is that each course of treatment is necessarily influenced by the organizational context. This applies especially to the provision of counselling, which depends on the availability of time and relevant experience. In a study by Meredith et al. (1996), the type and range of counselling provision and psychological intervention for depression through general practitioners, psychiatrists and psychologists was investigated and examined in relation to the payment system. As one might expect, there were clear differences between the professional groups. While with psychiatrists and psychologists more than 90% of patients were counselled in some way, the corresponding rate for general practitioners was only around 40%–50%. Psychiatrists and psychologists more often made use of the methods of psychoanalytic psychology and of behavioural therapy, whereas doctors mainly gave advice. Across specialities, they found that doctors who received a global payment per patient rather than payment for each contact, as well as doctors who had greater numbers of patients to handle, were less likely to give advice and spent less time on this. The structuring of the remuneration system must therefore not only be considered in financial terms, but should receive as much scientific attention as the pharmacokinetics of drugs, as its influence on treatment is at least as great.

However, treatment in primary care does not only encompass the doctors' own activities, but also help for patients which is provided by other workers in the practice or polyclinic or by various other collaborators. The studies by Fletcher et al. (1995) and Katon et al. (1996) already cited involved increasing the availability of counselling and support by placing psychologists, social workers or nurses in the practices. A further example is a study by Blanchard et al. (1995) in which it was shown that, for older depressed patients, additional intensive care from a community nurse led to better depression scores than among a group of patients cared for solely by general practitioners.

6.4 Effects of Intensified Diagnostics and Therapy

In view of the particular characteristics of general practice already discussed, including the many factors apart from the nature of the disorder itself, empirical evidence is needed on the results of more targeted and

specific diagnosis and treatment. Goldberg (1979) was one of the first to show that an improved diagnosis which also influenced treatment plans could shorten the course of illness. This effect became more marked with increasing severity of disorder. In any case, there is no necessary connection between diagnosis and treatment. Hoeper et al. (1984), who carried out an analogous study based only on giving doctors feedback about the diagnosis, without making sure that some therapeutic action followed, found that there was no alteration in the course of illness. Similarly, Tiemens et al. (1996) found that a higher recognition rate did not by itself lead to a better course of illness. On the other hand, Ormel and Giel (1990) reported that, in comparison to undiagnosed patients, diagnosed patients received psychotropic medication 4.5 times more often, psychotherapy 12.2 times more often and a referral to a specialist 3.3 times more often. This was accompanied by a sixfold better psychopathological state and a fivefold better level of social functioning. Schulberg et al. (1996) carried out a comparative study on the treatment of depressive patients in primary care. A total of 92 patients remained in routine treatment (RT) as usual from the general practitioner, 91 were treated following the research protocol with nortriptyline (NT) at a daily dose of 190–270 mg over 8 months, and a further 93 patients received interpersonal psychotherapy (IPT). The results give clear support to a targeted course of treatment. After 8 months, 48% of the NT and 46% of the IPT active treatment groups were symptom free compared with 18% of the RT group. The differences are even clearer when only those who remained in treatment rather than all initially included are considered. The percentage of symptom-free patients was 67% in the NT group, 72% in the IPT group and 20% in the RT group, while the proportion of patients whose score on the Hamilton Depression Scale was just as high after 8 months as at the beginning of treatment (>12) was 13% in the NT group, 15% in the IPT group and 48% in the RT group. These data confirm that the specificity of the treatment, the adequacy of its implementation and the consistency with which it is carried out are important requirements for treatment success.

However, it should be borne in mind that, while these data refer to patients recruited in general practice, they were specifically selected for the research project and the treatment was also carried out by specialists. A rather more realistic study was carried out by Callahan et al. (1994), who randomized 103 general practitioners and 175 depressive patients aged over 60 in equal numbers between a routine therapy and an intervention group. The intervention consisted of the doctor receiving detailed information about their patients' depression and hints on treatment. In the intervention group, significantly more antidepressants were prescribed, but psychiatrists were not consulted about treatment more often. In contrast to Schulberg et al. (1996), they did not find a difference between groups with regard to mental status on longitudinal follow-up. With respect to the management of depressive illnesses in general practices, basic treatment was supplemented by Katon et al. (1996) by an offer of targeted counselling for patients. In comparison with the routine therapy group, the patients in the intervention group were significantly happier with treatment, were more likely to be receiving a prescription for psychotropic drugs and above all were significantly more compliant with this medication. Further, their status improved to a significantly greater degree in the course of treatment, above all in the more serious cases.

7 Conclusion

The primary care setting is an important setting for the treatment of patients with psychiatric disorders and is in fact where the greatest number of such patients are seen. Both theoretical considerations and evidence in practice indicate that knowledge from the narrower realm of psychiatry cannot be automatically transferred without further consideration to the primary care setting. Instead, there is a need for setting-specific knowledge about diagnosis and treatment. It also cannot just be assumed that all affected patients can or should be automatically referred to specialists. Instead, co-operation between specialists and general practitioners should be intensified and to this end new models of co-operation should be sought. Finally, this area needs to become a more prominent field for scientific research endeavours in psychiatry, so as to allow the best setting and type of care to be determined for different groups of patients.

8 References

**AHCPR (Agency for Health Care Policy and Research) (1993) Depression in primary care, vol 2. Treatment of major depression. U.S. Department of Health and Human Services, Public Health Service, Agency for Health Care Policy and Research, Rockville/MD (publ no 93–0551)

Allan F (1944) The differential diagnosis of weakness and fatigue. N Engl J Med 231: 414–418

Arreghini E, Agoostini C, Wilkinson G (1991) General practitioner referral to specialist psychiatric services: a comparison of practices in North- and South-Verona. Psychol Med 212: 485–494

Badger LW, DeGruy FV, Hartman J et al (1994) Patient presentation, interview content, and the detection of depression by primary care physicians. Psychosom Med 56: 128–135

Bannister B (1988) Post-infectious disease syndrome. Postgrad Med J 64: 559–567

Barsky AJ (1979) Patients who amplify bodily sensation. Ann Intern Med 91: 63–70

Bass C, Benjamin S (1993) The management of chronic somatization. Br J Psychiatry 162: 472–480

Bass C, Wade C (1984) Chest pain with normal coronary arteries: a comparative study of psychiatric and social morbidity. Psychol Med 14: 51–61

Beard G (1869) Neurasthenia or nervous exhaustion. Boston Med Surg J 3: 217–220

Beard G (1880) A practical treatise on nervous exhaustion (neurasthenia). Wood, New York

Bebbington PE, Hurry J, Tennant C et al (1981) Epidemiology of mental disorders in Camberwell. Psychol Med 11: 561–579

Behan P, Behan W, Bell E (1985) The post-viral fatigue syndrome: an analysis of the findings in 50 cases. J Infect 10: 211–222

Bellantuono C, Fiorio R, Williams P, Arreghini E, Carson G (1988) Urban-rural differences in psychotropic drug prescribing in northern Italy. Eur Arch Psychiatr Neurol Sci 237: 347–350

Benjamin S, Eminson DM (1992) Abnormal illness behaviour: childhood experiences and long-term consequences. Int Rev Psychiatry 4: 55–70

Blackburn IM (1984) Setting relevant patient differences: a problem in phase IV research. Pharmacopsychiatry 17: 143–147

Blacker CVR, Thomas JM (1988) Treatment of psychiatric disorder in primary care. Presented at the NIMH conference on psychiatry and primary care, Pittsburgh

Blanchard MR, Waterreus A, Mann AH (1995) The effect of primary care nurse intervention upon older people screened as depressed. Int J Geriatr Psychiatry 10: 289–298

Bohman M, Cloninger CR, von Knorring AL et al (1984) An adoption study of somatoform disorders III. Cross-fostering analysis and genetic relationship to alcoholism and criminality. Arch Gen Psychiat 14: 872–878

Bridges DN, Goldberg DP (1985) Somatic presentation of DSM-III psychiatric disorders. J Psychosom Res 29: 563–569

Broadhead WE, Blazer DG, George LK, Tse CK (1990) Depression disability days, and days lost from work in a prospective epidemiologic survey. JAMA 264: 2524–2528

Brodaty H, Howarth GC, Mant A, Kurrle SE (1994) General practice and dementia. A national survey of Australian GPs. Med J Aust 160: 10–14

Brown JH, Parasekvas F (1982) Cancer and depression. Cancer presenting with depressive illness: an autonomic disease? Br J Psychiatry 141: 227–232

Burvill PW (1990) The epidemiology of psychological disorders in general medical settings. In: Sartorius N, Goldberg D, De Girolamo G, Costa e Silva J, Lecrubier Y, Wittchen U (eds) Psychological disorders in general medical settings. Hogrefe and Huber, Toronto, pp 9–20

Butler G, Fennell M, Robson P, Gelder M (1991) Comparison of behavior therapy and cognitive behavior therapy in the treatment of generalized anxiety. J Consult Clin Psychol 59: 167–175

Callahan CM, Hendrie HC, Dittus RS, Brater DC, Hui SL, Tierney WM (1994) Improving treatment of late life depression in primary care: a randomized clinical trial. J Am Geriatr Soc 42: 839–846

Carey T, Owens J, Mulligan P, Moran D (1994) An analysis of general practice referral behaviour to psychiatric out-patient clinics. Ir J Psychol Med 11: 177–179

Casey PR, Dillon S, Tyrer PJ (1984) The diagnostic status of patients with conspicuous psychiatric morbidity in primary care. Psychol Med 14: 673–683

Cheok AS, Cohen CA, Zucchero CA (1997) Diagnosing and managing dementia patients. Practice patterns of family physicians. Can Fam Physician 43: 477–482

Chithhiramohan RN, Ballard CG, Baxter MA et al (1993) Factors influencing general practitioner referral to a child psychiatric service. Ir J Psychol Med 10: 144–147

Cloninger CR, Sigvardsson S, von Knorring AL et al (1984) An adoption study of somatoform disorders. II. Identification of two discrete somatoform disorders. Arch Gen Psychiatry 41: 863–871

Colenda CC, Rapp SR, Leist JC, Poses RM (1996) Clinical variables influencing treatment decisions for agitated dementia patients: survey of physician judgements. J Am Geriatr Soc 44: 1375–1379

Cooper B, Bickel H, Schäufele M (1992) The ability of general practitioners to detect dementia and cognitive impairment in their elderly patients: a study in Mannheim. Int J Geriatr Psychiatry 7: 591–598

Commissaris CJ, Jolles J, Verhey FR, Kok GJ (1995) Problems of caregiving spouses of patients with dementia. Patient Educ Couns 25: 143–149

Coyne JC, Schwenk TL, Fechner Bates S (1995) Nondetection of depression by primary care physicians reconsidered. Gen Hosp Psychiatry 17: 3–12

Cremniter D, Guelfi JD, Fourestie V, Fermanian J (1995) Analysis of the terms used by general practitioners to characterize patients considered by them as depressed: a prospective study on 682 patients. J Affect Disord 34: 311–318

David AS, Wessely SC (1995) The legend of Cammelford. J Psychosom Res 39: 1–10

David AS, Wessely SC, Pelosi AJ (1988) Postviral fatigue syndrome: time for a new approach. Br Med J Clin Res 296: 696–699

Dilling H, Weyerer S, Enders I (1978) Patienten mit psychischen Störungen in der Allgemeinarztpraxis und ihre psychische Überweisungsbedürftigkeit. In: Häfner H (ed) Psychiatrische Epidemiologie. Springer, Berlin Heidelberg New York, pp 135–160

*Dilling H, Weyerer S, Castel R (1984) Psychische Erkrankungen in der Bevölkerung. Enke, Stuttgart

Dilling H, Mombour W, Schmidt MH, Schulte-Markwort E (1994) Internationale Klassifikation psychischer Störungen. ICD-10 Kapitel V (F): Forschungskriterien. Huber, Bern

Dittmann RW, Linden M, Osterheider M, Schaaf B, Ohnmacht U, Weber HJ (1997) Antidepressant drug use: differences between psychiatrists and general practitioners. Pharmacopsychiatry 30: 28–34

Dowden C, Johnson W (1929) Exhaustion states. JAMA 93: 1702–1706

Durham RC, Murphy T, Allan T, Richard K, Treliving LR, Fenton GW (1994) Cognitive therapy, analytic psychotherapy and anxiety management training for generalized anxiety disorder. Br J Psychiatry 165: 315–323

Eefsting JA, Boersma F, van den Brink W, van Tilburg W (1996) Differences in prevalence of dementia based on community

survey and general practitioner recognition. Psychol Med 26: 1223–1230

El-Rufaie OE, Absood GH (1993) Minor psychiatric morbidity in primary health care: prevalence, nature and severity. Int J Soc Psychiatry 39: 159–166

Finlay-Jones RA, Burvill P (1978) Contrasting demographic patterns of minor psychiatric morbidity in general practice and the community. Psychol Med 8: 455–466

Fletcher J, Fahey T, McWilliam J (1995) Relationship between the provision of counseling and the prescribing of antidepressants, hypnotics and anxiolytics in general practice. Br J Gen Pract 45: 467–469

Freud S (1895) On the grounds for detaching a particular syndrome from neurasthenia under the description "anxiety neurosis", standard edn, vol 3. Hogarth, London

Friebel HH (1989) Psychopharmakaverbrauch im internationalen Vergleich. In: Heinrich H, Linden M, Müller-Oerlinghausen B (eds) Werden zu viele Psychopharmaka verbraucht? Methoden und Ergebnisse der Pharmakoepidemiologie und Phase-IV-Forschung. Stuttgart, Thieme, pp 7–41

Fritzsche V, Haasen C, Stark FM (1993) Betreuung depressiver Patienten durch Allgemeinärzte: Eine Fragebogen-Studie in Hamburg. Fortschr Med 111: 35–38

Gardiner BM (1980) Psychological aspects of rheumatoid arthritis. Psychol Med 10: 159–163

Gastpar M (1984) Studies in general practice: interpractice differences. Pharmacopsychiatry 17: 148–151

Gautam S, Kapur Rl, Shamasundar C (1980) Psychiatric morbidity and referral in general practice: a survey of general practitioners in Bangalore. Ind J Psychiatry 22: 295–297

Geiselmann B, Linden M (1989) Überweisung psychisch kranker Patienten vom Allgemeinarzt zum Nervenarzt. Münch Med Wochenschr 131: 50–52

German GAG (1987) Mental health in Africa. I. The extent of mental health problems in Africa today: an update of epidemiological knowledge. Br J Psychiatry 151: 435–439

Goldberg D (1979) Detection and assessment of emotional disorders in a primary care setting. Int J Ment Health 8: 30–48

**Goldberg D (1990) Reasons for misdiagnosis. In: Sartorius N, Goldberg D, De Girolamo G, Costa e Silva J, Lecrubier Y, Wittchen U (eds) Psychological disorders in general medical settings. Hogrefe and Huber, Toronto, pp 139–145

Goldberg DP, Blackwell B (1970) Psychiatric illness in general practice: a detailed study using a new method of case identification. Br Med J 2: 439–443

*Goldberg DP, Huxley P (1980) Mental illness in the community. The pathway to psychiatric care. Tavistock, London

Goldberg DP, Lecrubier Y (1995) Form and frequency of mental disorders across centers. In: Üstün TB, Sartorius N (eds) Mental illness in general health care: an international study. Wiley, Chichester, pp 323–334

Goldberg DP, Williams P (1988) A user's guide to the general health questionnaire. NFER-Nelson, London

Goldberg D, Steele J, Smith C (1980a) Teaching psychiatric interview techniques to family doctors. Acta Psychiatr Scand 62: 41–47

Goldberg D, Steele J, Smith C (1980b) Training family doctors to recognize psychiatric illness with increased accuracy. Lancet 3: 521–524

Greenberg PE, Stiglin LE, Finkelstein SN, Berndt ER (1993) The economic burden of depression in 1990. J Clin Psychiatry 54: 405–418

*Häfner H (ed) (1978) Psychiatrische Epidemiologie. Springer, Berlin Heidelberg New York

Hankin J, Oktay JS (1979) Mental disorder and primary medical care: an analytical review of the literature (series D, no 5). National Institute of Mental Health, Washington, DC

Harding TW, De Arango MV, Baltazar J et al (1980) Mental disorders in primary care: a study of their frequency and diagnosis in four developing countries. Psychol Med 10: 231–241

Hecht H, von Zerrsen D, Wittchen HU (1990) Anxiety and depression in a community sample: the influence of comorbidity on social functioning. J Affect Disord 18: 137–144

Helmchen H (1991) Allgemein- und Nervenarzt: Im Wechselspiel von Kompetenz und Selbstkritik. Therapiewoche 41: 115–122

Helmchen H, Baltes MM, Geiselmann B et al (1996) Psychische Erkrankungen im Alter. In: Mayer KU, Baltes PB (eds) Die Berliner Altersstudie. Akademie, Berlin, pp 185–219

Henderson S, Duncan-Jones P, Byrne DG et al (1979) Psychiatric disorders in Canberra: a standardized study of prevalence. Acta Psychiatr Scand 60: 355–374

Henk HJ, Katzelnick DJ, Kobak KA, Greist JH, Jefferson JW (1996) Medical costs attributed to depression among patients with a history of high medical expenses in a health maintenance organization. Arch Gen Psychiatry 53: 899–904

Hoeper EW, Nycz GR, Cleary PD, Regier DA, Goldberg ID (1979) Estimated prevalence of RDC mental disorder in primary medical care. Int J Ment Health 8: 6–15

Hoeper EW, Nycz GR, Kessler L, Burke J, Pierce W (1984) The usefulness of screening for mental illness. Lancet 1: 33–35

Hohmann AA (1989) Gender bias in psychotropic drug prescribing in primary care. Med Care 27: 478–490

Holmes G, Kaplan J, Gantz N et al (1988) Chronic fatigue syndrome: a working case definition. Ann Intern Med 108: 387–389

Jenkins D, MacDonald A (1994) Should general practitioners refer more of their elderly patients to psychiatric services? Int J Geriatr Psychiatry 9: 461–465

Jenkins R, Shepherd M (1983) Mental illness and general practice. In: Bean P (ed) Mental illness: changes and trends. Wiley, Chichester, pp 16–21

*Jenkins R, Smeeton N, Shepherd M (1988) Classification of mental disorder in primary care. Psychol Med Monogr Suppl 12: 1–59

Johnstone A, Goldberg D (1976) Psychiatric screening in general practice: controlled trial. Lancet 20: 605–609

Joukamaa M, Sohlman B, Lehtinen V (1995) The prescription of psychotropic drugs in primary health care. Acta Psychiatr Scand 92: 359–364

Jorm AF, Korten AE, Henderson AS (1987) The prevalence of dementia: a quantitative integration of the literature. Acta Psychiatr Scand 76: 465–479

Katon W (1982) Depression: somatic symptoms and medical disorders in primary care. Compr Psychiatry 23: 274–287

Katon W, Russo J (1989) Somatic symptoms and depression. J Fam Pract 29: 65–69

Katon W, Riggs R, Gold D, Corey L (1988) Chronic fatigue syndrome: a collaborative virologic, immunologic and psychiatric study. Presented at the American Psychiatric Association, Montreal, Canada

Katon W, von Korff M, Lin E, Bush T, Russo J, Lipscomb P, Wagner E (1992) A randomized trial of psychiatric

consultation with distressed high utilizers. Gen Hosp Psychiatry 14: 83–85

Katon W, von Korff M, Lin E et al (1995) Collaborative management to achieve treatment guidelines: impact on depression in primary care. JAMA 273: 1026–1031

Katon W, Robinson P, von Korff M et al (1996) A multifaceted intervention to improve treatment of depression in primary care. Arch Gen Psychiatry 53: 924–932

Kleinman A (1982) Neurasthenia and depression: a study of somatization and culture in China. Cult Med Psychiatry 6: 117–190

Kocsis JH, Friedman RA, Markowitz JC, Miller N, Gniweschl L, Bram J (1995) Stability of remission during tricyclic antidepressant continuation therapy for dysthymia. Psychopharmacol Bull 31: 213–216

Kroenke K, Wood D, Mangelsdorff D, Meier N, Powell J (1988) Chronic fatigue in primary care: prevalence, patient characteristics and outcome. JAMA 260: 929–934

Kruesi M, Dale J, Straus S (1989) Psychiatric diagnosis in patients who have chronic fatigue syndrome. J Clin Psychiatry 50: 53–56

Kuhl J (1992) A theory of self-regulation: action versus state orientation, self-discrimination, and some applications. Appl Psychol 41: 97–129

Langwieler G, Linden M (1993) Therapist individuality in the diagnosis and treatment of depression. J Affect Disord 27: 1–12

Lehmann E (1989) The dose effect relationship of 0.5, 1.0 and 1.5 mg fluspirilene and anxious patients. Neuropsychobiology 21: 197–204

Leon AC, Olfson M, Broadhead et al (1995) Prevalence of mental disorders in primary care: implications for screening. Arch Fam Med 4: 857–861

Levenson JL, Hamer RM, Rossiter LF (1992) Psychopathology and pain in medical in-patients predict resource use during hospitalization but not rehospitalization. J Psychosom Res 36: 585–592

Linden M (1991) Somatogenic neurasthenia. In: Gastpar M, Kielholz P (eds) Problems of psychiatry in general practice. Hogrefe and Huber, Lewiston, pp 71–78

Linden M (1994) Therapeutic standards in psychopharmacology and medical decision-making. Pharmacopsychiatry 27: 41–45

Linden M, Geiselmann B (1996) Subdiagnostische psychiatrische Morbidität: Beschwerdeprofil und Konsequenzen am Beispiel depressiver Störungen. In: Saß H (ed) Psychopathologische Methoden und psychiatrische Forschung. Fischer, Jena, pp 106–116

Linden M, Geiselmann B, Helmchen H (1988) Anxiolytika und Sedativa: Aktueller Stand und neuere Entwicklungen. Münch Med Wochenschr 130: 571–574

Linden M, Borchelt M, Barnow S, Geiselmann B (1995) The impact of somatic morbidity on the Hamilton depression rating scale in the very old. Acta Psychiatr Scand 92: 150–154

Linden M, Gilberg R, Horgas AL, Steinhagen-Thiessen E (1996a) Die Inanspruchnahme medizinischer und pflegerischer Hilfe im hohen Alter. In: Mayer KU, Baltes PB (eds) Die Berliner Altersstudie. Akademie, Berlin, pp 475–495

**Linden M, Maier W, Achberger M, Herr R, Helmchen H, Benkert O (1996b) Psychische Erkrankungen und ihre Behandlung in Allgemeinarztpraxen in Deutschland. Nervenarzt 67: 205–215

Linden M, Lecrubier Y, Bellantuono C, Benkert O, Kisely S, Simon G (1999) Psychotropic drug prescribing by primary care physicians: an international collaborative study. J Clin Psychopharmacol (in press)

Lingg A, Jugl P, Bacher R (1995) Betreuung und Behandlung alterspsychiatrischer Patienten durch den Praktiker. Wien Med Wochenschr 145: 541–544

Lipowski ZJ (1988) Somatization: the concept and its clinical application. Am J Psychiatry 145: 1358–1368

Lloyd A, Wakefield D, Boughton C, Dwyer J (1988) What is myalgic encephalomyelitis? Lancet 1: 1286–1287

Lloyd KR, Moodley P (1992) Psychotropic medication and ethnicity: an inpatient survey. Soc Psychiatr Epidemiol 27: 95–101

Lloyd KR, Jenkins R, Mann A (1996) Long-term outcome of patients with neurotic illness in general practice. Br Med J 313: 26–28

Macy J, Allen E (1934) Justification of the diagnosis of chronic nervous exhaustion. Ann Intern Med 7: 861–867

Maguire N, Cullen C, O'Sullivan M, O'Grady Walshe A (1995) What do Dublin GPs expect from a psychiatric referral? Ir Med J 88: 215–216

Manning WG, Wells KB (1992) The effects of psychosocial distress and psychological well-being on use of medical services. Med Care 30: 541–553

Mantysaari M, Antila K, Peltonen T (1988) Blood pressure reactivity in patients with neurocirculatory asthenia. Am J Hypertens 1: 132–139

Manu P, Matthews D, Lane T (1988) The mental health of patients with a chief complaint of chronic fatigue: a prospective evaluation and follow-up. Arch Intern Med 148: 2213–2217

Mari JJ, Williams P (1984) Minor psychiatric disorder in primary care in Brazil: a pilot study. Psychol Med 14: 223–227

Marks JN, Goldberg DP, Hillier VF (1979) Determinants of the ability of general practitioners to detect psychiatric illness. Psychol Med 9: 337–353

Maslowski J (1987) Characteristic differences in the use of psychotropic drugs as a stigma of cultural influences. Bull Inst Marit Trop Med Gdynia 38: 133–138

Massie MJ, Holland JC (1984) Diagnosis and treatment in the cancer patient. J Clin Psychiatry 42: 25–28

Mayou R (1976) The nature of bodily symptoms. Br J Psychiatry 129: 55–60

Mayou R (1993) Somatization. Psychother Psychosom 59: 69–83

McCombs JS, Nichol MB, Stimmel GL, Sclar DA, Beasley CM, Gross LS (1990) The cost of antidepressant drug therapy failure: a study of antidepressant use patterns in a Medicaid population. J Clin Psychiatry 51: 60–69

Meredith LS, Wells KB, Camp P (1994) Clinician specialty and treatment style for depressed outpatients with and without medical comorbidities. Arch Fam Med 3: 1065–1072

Meredith LS, Wells KB, Kaplan SH, Mazel RM (1996) Counseling typically provided for depression: role of clinician specialty and payment system. Arch Gen Psychiatry 53: 905–912

Moeller HJ (1991) Die Rolle der Nootropika in der medikamentösen Therapie dementieller Erkrankungen. In: Moeller HJ (ed) Hirnleistungsstörungen im Alter. Springer, Berlin Heidelberg New York, pp 51–69

Morabia A, Fabre J, Dunand JF (1992) The influence of patient and physician gender on prescription of psychotropic drugs. J Clin Epidemiol 45: 111–116

Morrison J (1980) Fatigue as a presenting complaint in family practice. J Fam Pract 10: 795–801

Munoz RF, Hollon SD, McGrath E, Rehm LP, VandenBos GR (1994) On the AHCPR depression in primary care guidelines: further considerations for practitioners. Am Psychol 49: 42–61

Nazareth I, King M, Davies S (1995) Care of schizophrenia in general practice: the general practitioner and the patient. Br J Gen Pract 45: 343–347

Ndetei DM, Muhangi J (1979) The prevalence and clinical presentation of psychiatric illness in a rural setting in Kenya. Br J Psychiatry 135: 269–272

Nelson E, Kirk J, McHugo G, Douglass R, Ohler J, Wasson J, Zubkoff (1987) Chief complaint fatigue: a longitudinal study from the patient's perspective. Fam Pract Res J 6: 175–188

Newens AJ, Forster DP, Kay DW (1994) Referral patterns and diagnosis in presenile Alzheimer's disease: implications for general practice. Br J Gen Pract 44: 405–407

Olfsson M, Leon AC, Broadhead et al (1995a) The SDDS-PC: a diagnostic aid for multiple mental disorders in primary care. Psychopharmacol Bull 31: 415–420

Olfsson M, Weissman MM, Leon AC, Higgins ES, Barrett JE, Blacklow RS (1995b) Psychological management by family physicians. J Fam Pract 41: 543–550

Ormel J, Costa e Silva JA (1995) The impact of psychopathology on disability and health perceptions. In: Üstün TB, Sartorius N (eds) Mental illness in general health care: an international study. Wiley, Chichester, pp 335–346

Ormel J, Giel R (1990) Medical effects of nonrecognition of affective disorders in primary care. In: Sartorius N, Goldberg D, De Girolamo G, Costa e Silva J, Lecrubier Y, Wittchen U (eds) Psychological disorders in general medical settings. Hogrefe and Huber, Toronto, pp 146–158

Ormel J, von Korff M, van den Brink W, Katon WJ, Brilman EM, Oldehinkel T (1993) Depression, anxiety, and social disability show synchrony of change in primary care patients. Am J Public Health 83: 385–390

Paykel ES, Hollyman JA, Freeling P, Sedgewick P (1988) Predictors of therapeutic benefit from amitriptyline in mild depression: a general practice placebo-controlled trail. J Affect Disord 14: 83–95

Pond CD, Mant A, Kehoe L, Hewitt H, Brodaty H (1994) General practitioner diagnosis of depression and dementia in the elderly: can academic detailing make a difference? Fam Pract 11: 141–147

Raynes NV (1979) Factors affecting the prescribing of psychotropic drugs in general practice consultations. Psychol Med 9: 671–679

RCGP (Royal College of General Practitioners), OPCS (Office of Population Censuses and Surveys) (1979) Morbidity statistics from general practice 1971–1972, 2nd national study. HMSO, London (Studies on medical and population subjects, no 36)

Real Perez M, Rodriguez Arias Palomo JL, Cagigas Viadero J, Aparicio Sanz MM, Real Perez MA (1996) Terapia familiar breve: una opcion para el tratamiento de los trastornos somatoformes en atencion primaria. Aten Primaria 17: 241–246

Reischies FM, Geiselmann B, Geßner R, Kanowski S, Wagner M, Wernicke F, Helmchen H (1997) Demenz bei Hochbetagten. Ergebnisse der Berliner Altersstudie. Nervenarzt 68: 719–729

Robins LN, Regier DA (eds) (1991) Psychiatric disorders in America: the epidemiologic catchment area study. Free Press, New York

Roche AM, Parle MD, Stubbs JM, Hall W, Saunders JB (1995) Management and treatment efficacy of drug and alcohol problems: what do doctors believe? Addiction 90: 1357–1366

Ryynanen OP, Myllykangas M, Kinnunen J, Takala J (1997) Doctors' willingness to refer elderly patients for elective surgery. Fam Pract 14: 216–219

Samuels AH (1995) Somatisation disorder: a major public health issue. Med J Aust 163: 147–149

Sartorius N (1997) Diagnosis and classification of neurasthenia. In: Judd LL, Saletu B, Filip V (eds) Basic and clinical science of mental and addictive disorders. Karger, Basel, pp 1–5

Schäfer T (1990) Arzneimitteltransparenz und Arzneimittelberatung am Beispiel der Region Dortmund. Bundesministerium für Arbeit und Sozialordnung, Bonn (Forschungsbericht Gesundheitsforschung no 198)

*Schepank H (1987) Psychogene Erkrankungen in der Stadtbevölkerung. Springer, Berlin Heidelberg New York Tokyo

Schmidt LG (1997) Früherkennung und -intervention bei Alkoholismus. Dtsch Ärztebl 94: 2905–2908

*Schulberg HC, Burns JP (1988) Mental disorders in primary care: epidemiologic, diagnostic and treatment research directions. Gen Hosp Psychiatry 10: 79–87

Schulberg HC, Madonia MJ, Block MR et al (1995) Major depression in primary care practice: clinical characteristics and treatment implications. Psychosomatics 36: 129–137

*Schulberg HC, Block MR, Madonia et al (1996) Treating major depression in primary care practice: eight-month clinical outcomes. Arch Gen Psychiatry 53: 913–919

Scott AIF, Freeman CPL (1992) Edinburgh primary care depression study: treatment outcome, patient satisfaction, and cost after 16 weeks. Br Med J 304: 883–887

Scott J, Moon CA, Blacker CV, Thomas JM (1994) A. I. F. Scott & C. P. L. Freeman's "Edinburgh Primary Care Depression Study". Br J Psychiatry 164: 410–415

Shah S, Harris M (1997) A survey of general practitioner's confidence in their management of elderly patients. Aust Fam Phys 26[Suppl 1]: 12–17

Shapiro S, German PS, Skinner EA et al (1987) An experiment to change detection and management of mental morbidity in primary care. Med Care 25: 327–339

Sharpe M, Bass C (1992) Pathophysiological mechanisms in somatization. Int Rev Psychiatry 4: 81–97

Shepherd M, Cooper B, Brown AC, Kalton GW (1966) Psychiatric illness in general practice. Oxford University Press, London

Sherbourne CD, Wells KB, Meredith LS, Jackson CA, Camp P (1996) Comorbid anxiety disorders and the functioning and well-being of chronically ill patients of general medical providers. Arch Gen Psychiatry 53: 889–895

Simon GE, von Korff M, Barlow W (1995) Health care costs of primary care patients with recognized depression. Arch Gen Psychiatry 52: 850–856

Sims A (1993) The scar that is more than skin deep: the stigma of depression. Br J Gen Practice 43: 30–31

Skuse D, Williams P (1984) Screening for psychiatric disorder in general practice. Psychol Med 14: 365–377

Smith GR (1992) The epidemiology and treatment of depression when it coexists with somatoform disorders, somatization, or pain. Gen Hosp Psychiatry 14: 265–272

Starcevic V, Kellner R, Uhlenhuth EH et al (1992) Panic disorder and hypochondriacal fears and beliefs. J Affect Disord 24: 73–85

Starfield B (1992) Primary care: concept, evaluation, and policy. Oxford University Press, New York

Stein MB, Kirk P, Prabhu V, Grott M, Terepa M (1995) Mixed anxiety-depression in a primary-care clinic. J Affect Disord 34: 79–84

Stewart D (1987) Environmental hypersensitivity disorder, total allergy and 20th century disease: a critical review. Can Fam Phys 33: 405–410

Strathdee G, Brown RMA, Doig RJ (1990) Psychiatric clinics in primary care: the effect on general practitioner referral patterns. Soc Psychiatry Psychiatr Epidemiol 25: 95–100

Sugarman J, Berg A (1984) Evaluation of fatigue in a family practice. J Fam Pract 19: 643–647

Taerk K, Toner B, Salit I, Garfinkel P, Ozersky S (1987) Depression in patients with neuromyasthenia (benign myalgic encephalomyelitis). Int J Psychiatry Med 17: 49–56

Thust W (1997) Ärztliche Versorgung in Deutschland. Dtsch Ärztebl 94(19)

*Tiemens BG, Ormel J, Simon GE (1996) Occurrence, recognition, and outcome of psychological disorders in primary care. Am J Psychiatry 153: 636–644

Tobi M, Straus S (1985) Chronic Epstein-Barr disease: a workshop held by the National Institute of Allergy and Infectious Diseases. Ann Intern Med 103: 951–952

Tylee A, Freeling P, Kerry S, Burns T (1995) How does the content of consultations affect the recognition by general practitioners of major depression in women? Br J Gen Pract 45: 575–578

Tyler CV, Bourguet C (1997) Primary care of adults with mental retardation. J Fam Pract 44: 487–494

Tyrer P, Fowler-Dixon R, Ferguson B, Keleman A (1990) A plea for the diagnosis of hypochondriacal patients. Psychol Med 23: 167–173

Üstün TB, Sartorius N (1993) Public health aspects of anxiety and depressive disorders. Int Clin Psychopharmacol 8: 15–20

**Üstün TB, Sartorius N (eds) (1995) Mental illness in general health care: an international study. Wiley, Chichester

Üstün TB, von Korff M (1995) Primary mental health services: access and provision of care. In: Üstün TB, Sartorius N (eds) Mental illness in general health care: an international study. Wiley, Chichester, pp 347–360

Üstün TB, Goldberg D, Cooper J, Simon GE, Sartorius N (1995a) New classification for mental disorders with management guidelines for use in primary care: ICD-10 PHC chapter five. Br J Gen Pract 45: 211–215

Üstün TB, Simon G, Sartorius N (1995b) Discussion. In: Üstün TB, Sartorius N (eds) Mental illness in general health care: an international study. Wiley, Chichester, pp 361–370

Valdini A, Steinhardt S, Jaffe A (1987) Demographic correlates of fatigue in a university family health center. Fam Pract 4: 103–107

Verhaak PFM (1993) Analysis of referrals of mental health problems by general practitioners. Br J Gen Pract 43: 203–208

Vernooij-Dassen M, Felling A, Persoon J (1997) Predictors of change and continuity in home care for dementia patients. Int J Geriatr Psychiatry 12: 671–677

von Korff M, Shapiro S, Burke JD et al (1987) Anxiety and depression in a primary care clinic: comparison of Diagnostic Interview Schedule, General Health Questionnaire and practitioner assessment. Arch Gen Psychiatry 44: 152–156

von Korff M, Ormel J, Katon W, Lin E (1992) Disability and depression among high utilizers of health care. Arch Gen Psychiatry 49: 91–100

Ward M H, DeLisle H, Shores JH, Slocum PC, Foresman B (1996) Chronic fatigue complaints in primary care: incidence and diagnostic patterns. J Am Osteopath Assoc 96: 34–46

Watson D, Pennebaker JW (1989) Health complaints, stress, and distress: exploring the central role of negative affectivity. Psychol Rev 96: 234–254

Weissman MM, Olfson M, Leon AC et al (1995) Brief diagnostic interviews (SDDS-PC) for multiple mental disorders in primary care: a pilot study. Arch Fam Med 4: 220–227

Wessely S, Powell R (1989) Fatigue syndromes: a comparison of chronic "postviral" fatigue with neuromuscular and affective disorders. J Neurol Neurosurg Psychiatry 52: 940–948

Wessely S, Chalder T, Hirsch S, Wallace P, Wright D (1996) Psychological symptoms, somatic symptoms, and psychiatric disorder in chronic fatigue and chronic fatigue syndrome: a prospective study in the primary care setting. Am J Psychiatry 153: 1050–1059

Wessling A, Bergman U, Westerholm B (1991) On the differences in psychotropic drug use between the three major urban areas in Sweden. Eur J Clin Pharmacol 40: 495–500

Weyerer S (1990) Relationship between physical and psychological disorders. In: Sartorius N, Goldberg D, De Girolamo G, Costa e Silva J, Lecrubier Y, Wittchen U (eds) Psychological disorders in general medical settings. Hogrefe and Huber, Toronto, pp 34–46

Wheeler E, White P, Reed E, Cohen M (1950) Neurocirculatory asthenia (anxiety neurosis, effort syndrome, neurasthenia). JAMA 142: 878–889

Widmer RB, Cadoret RJ (1979) Depression in family practice: changes in pattern of patient visits and complaints during subsequent developing depression. J Fam Pract 9: 1017–1021

Wilkinson G, Smeeton N, Skuse D, Fry J (1988) Consultation for physical illnesses by patients diagnosed and treated for psychiatric disorders by a general practitioner: 20 year follow up study. Br Med J 297: 776–778

Williams P (1983) Factors influencing the duration of treatment with psychotropic drugs in general practice: a survival analysis approach. Psychol Med 13: 45–55

Williams P, Clare A (1986) Psychiatric general practice. In: Hill P, Murray RH, Thorley A (eds) Essentials of postgraduate psychiatry. Academic Press, London, pp 112–118

Williams P, Tarnopolsky A, Hand D, Shepherd M (1986) Minor psychiatric morbidity and general practice consultation: the West London Survey. Psychol Med Monogr Suppl 9

Wittchen HU, Essau CA (1990) Assessment of symptoms and psychosocial disabilities in primary care. In: Sartorius N, Goldberg D, De Girolamo G, Costa e Silva J, Lecrubier Y, Wittchen U (eds) Psychological disorders in general medical settings. Hogrefe and Huber, Toronto, pp 111–136

Wittchen HU, Burke JD, Semler G, Pfister H, von Cranach M, Zaudig M (1989) Recall and dating reliability of psychiatric symptoms. Test-retest reliability of time related symptom questions in a standardized psychiatric interview (CIDI/DIS). Arch Gen Psychiatry 46: 437–443

*Wohlfarth TD, van den Brink W, Ormel J, Koeter MWJ, Oldehinkel AJ (1993) The relationship between social dysfunctioning and psychopathology among primary care attenders. Br J Psychiatry 163: 37–44

Wood P (1941) Da Costa's syndrome (or effort syndrome). Br Med J 1: 767–772, 805–811, 845–851

Yunus M (1989) Fibromyalgia syndrome: new research on an old malady. Br Med J 298: 474–475

Zintl-Wiegand A, Schmidt-Maushard R, Leisner R, Cooper B (1978) Psychische Erkrankungen in Mannheimer Allgemeinpraxen: Eine klinische und epidemiologische Untersuchung. In: Häfner H (ed) Psychiatrische Epidemiologie. Springer, Berlin Heidelberg New York, pp 111–133

Zubrägel D, Linden M (1997) Generalisierte Angsterkrankung. Münch Med Wochenschr 139: 168–170

Zung WWK, Magill M, Moore JT, George DT (1983) Recognition and treatment of depression in a family medical practice. J Clin Psychiatry 44: 3–6

Zwernemann B (1997) Untersuchung zu psychischen Störungen in der Allgemeinarztpraxis. Dissertation, Freie Universität, Berlin

CHAPTER 14

A. Diefenbacher

Consultation and Liaison Psychiatry

Translator: S. Davison

1 Definitions

Increasing specialisation within modern medicine has meant that individual doctors are no longer able to have expertise in all areas and apply this to the benefit of their patients. When a doctor requires advice and help with a problem in a patient's management, he or she seeks the advice of a specialist who has particular experience and qualifications in the specialty in question (Schliack 1992). The discussion between several doctors, in this way, in the assessment and management of a single case has long been known as consultation.

In consultation psychiatry, therefore, a doctor who is not specialised in psychiatry enlists the help of a psychiatrist, who gives advice to his or her referring colleagues but is not actively responsible for managing the patient.

The term liaison psychiatry (from the French *liaison*, meaning relationship or connection), as currently used, was coined by Billings in the 1930s in the USA (Billings 1941). It describes a broader approach. It was originally used to describe the activity of a multidisciplinary team of psychiatrists and social workers who worked in a general hospital in the absence of an autonomous psychiatry department. However, the concept broadened in the following years, particularly in English-speaking countries. Liaison psychiatry differs from consultation psychiatry because of its more comprehensive and time-intensive integration within a particular area of somatic medicine. For example, a liaison psychiatrist may be present on a regular basis in a general medical or oncology department if he or she has not received a specific referral. Additional tasks of the liaison psychiatrist over and above direct patient contact include providing advice to the whole medical team on psychosocial issues, training and research (Strain 1996a; Lipowski 1992; Levy 1989).

2 Epidemiology and Effects of Psychiatric Co-morbidity in Physically Ill Patients in the General Hospital

It is known from community studies that members of the general population with a physical illness are at increased risk of suffering from a psychiatric disorder. The proportion of hospital patients with co-morbid psychiatric disorders is even higher. A series of methodologically sound studies from around the world have shown that between 30% and 50% of physically ill general hospital inpatients have co-morbid psychiatric disorders (Creed 1996). Three groups of diagnosis are significantly over-represented among general hospital in-patients in comparison to physically ill out-patients (Arolt 1997; Weyerer 1990). These are organic psychoses, alcohol and other addictions and acute stress reactions.

A German study of general hospital in-patients in the town of Lübeck showed the following prevalence rates of psychiatric disorder in surgical and medical in-patients: organic disorders 17.5%, depression 16.3% and addictions 11%. These figures are even higher in particular patient groups, such as patients on liver transplant programmes (Arolt 1997; Surman 1992).

Studies now show a significant association between psychiatric co-morbidity and more complicated illness courses, with longer lengths of stay in hospital and more frequent re-admissions. The presence of delirious states and depressive and dementing syndromes present particular risk factors in this respect (Saravay and Lavin 1994).

In a U.S. study, Levenson et al. (1990) diagnosed 27.5% of a population of general medical in-patients as very depressed, 27.5% as very anxious and 20.2% as having cognitive dysfunction. The median length of stay of patients with high levels of psychopathology was 40% longer, and the mean hospital costs 30% higher, than that of a comparison group with low levels of psychopathology.

These high rates of psychiatric co-morbidity leading to more complicated illness courses in physically ill general hospital in-patients are the "bread and butter" of consultation–liaison psychiatry.

It is important to know how many patients actually need either joint psychiatric care during their in-patient medical admission or out-patient psychiatric follow-up in order to plan the requirements of a consultation psychiatric service. There have, however, been very few studies in this regard.

In the Lübeck general hospital study, it was estimated using ratings by experts that 16% of acute hospital in-patients required a psychiatric intervention, whereas only a quarter of these patients had actually been referred to the psychiatrist (Arolt 1997). Wancata et al. (1998) reported similar findings. Swigar et al. (1992) found even more remarkable findings in older patients. They considered that 36% of older medical patients needed a psychiatric opinion, whereas the actual referral rate was 3%.

The only representative study at a national level was carried out in the USA and found that only about 0.9% of all in-patients admitted to general hospitals were referred to a consultation psychiatrist (Wallen et al. 1987). Higher referral rates of up to 10% of in-patient admissions (e.g. at Massachusetts General Hospital in

Boston) were noted and were attributed to particular local circumstances.

The discrepancy between estimated need and actual referral rates can be explained by the combination of multiple factors. Co-morbid psychiatric disorders are under-recognised by non-psychiatrists (Margolis 1994). This applies particularly to cognitive dysfunction and to depressive and anxiety disorders (e.g. Katon 1991). In addition, many doctors have negative preconceptions about what a consultation psychiatrist can achieve. Finally, in contrast to these individual factors, the organisational/sociological problems that may either facilitate or hinder the implementation of a consultation service have to be considered (Koch and Siegrist 1988).

3 Psychiatry in the General Hospital

The comparatively long tradition and well-established nature of consultation psychiatry in the USA can be traced back to the early establishment of psychiatric departments within general hospitals, a process that had already begun in the 1920s and 1930s (Greenhill 1977; Levy 1989; Panse 1964). The rapid growth of consultation psychiatry, with several journals and specialist associations, as well as its proposed status as a recognised sub-specialty within psychiatric training by the American Psychiatric Association would have been unthinkable without this development (Lipowski 1992).

In Germany, the issue of psychiatric departments in general hospitals was a "hot potato" for a long time (Panse 1964). It was only from the 1970s onwards that psychiatric departments were increasingly established in general hospitals under the influence of the *Psychiatrie-Enquete* (national inquiry into mental health services in Germany). The aim was to stop the discriminating segregation and poor treatment of psychiatric patients. The inquiry suggested the discharge of the chronically ill patients, predominantly suffering with schizophrenia, from institutions (Häfner 1991) and aimed overall to integrate psychiatry into the whole of medicine (Müller 1981). It was recognised that consultation psychiatry care could provide better treatment opportunities in general hospitals for physically ill patients with co-morbid psychiatric disorders (Winkler 1975). In addition, the inquiry called for the establishment of a consultation psychiatry service in every large hospital for the treatment of patients who had attempted suicide. It also suggested that it was important for consultation psychiatry services to be involved with primary prevention in high-risk groups within physical medicine, such as hospitalised accident victims, dialysis patients or transplantation patients (Deutscher Bundestag 1975, pp. 279–281, 392). However, these factors receded into the background compared with the improved medical management of psychiatric patients with co-existing physical illnesses through consultation with the medical and surgical specialties.

Bönisch and colleagues (Bönisch and Mayer 1975; Bönisch et al. 1986) had written at an early stage about extreme situations in medical management and had predicted the growing importance of liaison work. However, the impression prevailed that a consultation psychiatrist who was interested in psychosomatic medicine would be most likely to have to justify this interest (Blankenburg 1988). In Germany, consultation psychiatry remained an activity that individual psychiatrists had to take on in addition to their other work, leaving little space for scientific interests. In addition, German psychiatrists did not look back to their medical roots to establish their identity in the same way that their colleagues in the USA did: "There is no better way to establish medical identity than through consultation psychiatry" (Hackett and Cassem 1987).

The opportunity to use consultation psychiatry actively to demonstrate the usefulness of psychiatry in non-psychiatric departments in general hospitals, and in this way to drive forward the overall integration of psychiatry, was hardly taken. A survey of consultation psychiatry in Germany at the end of the 1980s thus pointed out that only the very few departments of psychosomatic medicine, principally within university settings, saw themselves as providing a consultation service; in contrast to departments of psychiatry (the two are separate specialties in Germany) (Herzog and Hartmann 1990). It should, however, be noted that this overview did not include consultation services provided by office-based psychiatrists and neurologists in private practice. The departments of psychiatry reported relevant consultation activities much less often, even though, in actual fact, they took on by far the major share of consultation patient care (Bender et al. 1983; Böker 1973). The results of this study were rather disappointing, if not altogether surprising: the extent and intensity of co-operation work was low among both general psychiatrists and, interestingly, psychosomatic specialists, and less than half of the psychosomatic departments and even fewer of the psychiatric departments held joint case conferences. Predominantly, the psychiatrists co-operated with all other departments (except neurology and psychosomatic departments), whereas only half of psychosomatic departments did so. On the other hand, psychosomatic departments were better staffed and saw their patients more frequently and for longer. Psychiatrists were more often available during the night and at weekends (24-h cover) and responded

more promptly to referrals than their counterparts in psychosomatic departments (Herzog and Hartmann 1990). At the time of this survey, there were 78 psychiatric departments situated in general hospitals in Germany. Since then, their number has risen sharply (125 departments in 1993; Rössler et al. 1996). In comparison, the USA had 1358 such departments in 1984 (the population of the USA is roughly just over three times that of Germany).

In Germany, although non-psychiatrists were increasingly recognising the importance of psychiatric consultation services in everyday clinical practice, it was not until the early 1990s (with a few notable exceptions; e.g., Möller and Lauter 1986) that psychiatrists increasingly began to devote their research to consultation psychiatry topics and to develop expertise in psychotherapeutic and psychopharmacological treatments for physically ill patients (e.g. Kapfhammer 1993; Möller and Scriba 1994; Saupe and Diefenbacher 1996a; Arolt 1997; Diefenbacher 1999).

4 Special Features of the Psychiatric Consultation

The classic psychiatric consultation on a medical ward is a complicated undertaking. In contrast to a surgeon providing an opinion to a physician or vice versa, the consultation psychiatrist is frequently regarded by his or her colleagues in physical medicine as an "unknown being". Conversely, psychiatrists carrying out such a consultation may well develop anxieties as to whether they can hold their own in the real or perceived hostile territory of physical medicine. This leads, not uncommonly, to reciprocal misunderstandings which jeopardise the success of the consultation (Spiess 1996; Greenhill 1977).

Meyer and Mendelson (1961) described three phases in a psychiatric consultation which are useful for consultation psychiatrists to be aware of. The initiation of the consultation may be the expression of non-specific uncertainty on the part of the referrer as to how to deal with a patient. This may not be explicitly verbalised. This requires the psychiatrist to work out independently what the actual question is. In order to arrive at his or her own definition of the problem, the psychiatrist's assessment should take into account the psychiatric examination of the patient; the inspection of the available findings and a corroborative history taken from third parties (such as the treatment team and relatives, if appropriate). The conclusion may be distorted if second-hand information is accepted uncritically. The ensuing intervention should take into consideration not only the patient, but also the whole system (e.g. the ward doctor, nurses, relatives). Consultation psychiatrists must see themselves as a catalyst who need the individual members of the system to mediate their work. This work will be unsuccessful if conducted in isolation and/or out of context (Saupe and Diefenbacher 1996a).

Although the consultation must take into consideration a series of environmental/situational factors in order to be successful, it is not usually the intention to bring about a change in the pattern of referrals of individual units, beyond the advice given in an individual case.

5 The Liaison Model

In contrast to the classic model of consultation psychiatry, the liaison model draws its foundations from the discrepancy between the high need for joint psychiatric care of general hospital patients and the low rates of referral for consultation.

According to the liaison model, the tip of the iceberg of psychiatric co-morbidity which is recognised by the general hospital doctor is merely the "numerator" of existing psychiatric disorder, whereas the unrecognised broad base is the "denominator" (Strain 1996a). Consultation psychiatry has to restrict itself of necessity to what is "counted" by the referring hospital doctor, such that the consultation psychiatrist is often only called at a late stage and is reduced to a "fire-fighting" function. In contrast to this, liaison psychiatry aims to have a substantial influence on the "denominator". To this end, two ways of accessing patients are employed over and above the classic consultation model: (1) referrals from nursing staff dealing with patients as well as doctors and (2) early screening of high-risk patients. This process is made easier by the more or less constant presence of the liaison psychiatrist (in contrast to the consultation psychiatrist) on the physical medicine ward.

As a result, emergencies and crises along the continuum from organic (e.g. acute states of confusion) to reactive disorders (e.g. preoperative anxiety states) are avoided as patients are more efficiently managed by the ward team. This approach is supported by empirical studies which show that not only does liaison psychiatry lead to an increase in the number of referrals, as can also be the case with the consultation model, but it also leads to changes in the pattern of referrals. Thus a drop in the number of referrals of acutely confused patients was observed following the establishment of a psychiatric liaison service as wards gradually began to manage them independently. On the other hand, there was an increased number of referrals of depressed patients, precisely the patient

group that non-psychiatrists frequently either misdiagnose or mismanage (Anderson and Philpott 1991).

Advice to medical and nursing staff and also to social workers with regard to diagnosis, therapy and guidance of patients is usually at the forefront of liaison work. The use of short rating scales for the identification of cognitive, depressive or anxiety disorders is recommended where relevant (Bass 1995). The early screening of high-risk patients by liaison psychiatrists has hitherto occurred predominantly in the context of research projects (e.g. Strain et al. 1991; John et al. 1996) or in specialist programmes such as transplant surgery (Sperling and Kalb 1995; Freeman et al. 1992). However, a European study is currently underway to develop a screening instrument for the detection of psychosocial risk factors in patients admitted to general hospital wards (Huyse et al. 1997).

6
Consultation and Liaison Psychiatry in Practice

Some of the different areas of practice in consultation psychiatry are presented below. The care of residents of old people's and nursing homes with psychiatric disorders is discussed in particular detail, as this area of consultation psychiatry is growing in size and importance.

6.1
Consultation Psychiatry in the General Hospital

Although Bönisch et al. (1986) found that the idea of consultation psychiatry as a psychiatric subspecialty in its own right had been only poorly taken up in the German-speaking world, and that the opportunity to integrate psychiatry fully in clinical medicine by interdisciplinary work had therefore been missed, there have been significant changes since that time.

Important papers describing the practice of consultation psychiatry in general hospitals that have since been published in the German-speaking literature are summarised in Table 1. A literature review by Hengeveld et al. (1984) is included to provide an international comparison. In addition, a consultation service provided by a university department of psychosomatic medicine is cited (Knorr et al. 1996). On the whole, German psychiatric consultation services lie within the scope of international psychiatric liaison activities, as described by Hengeveld et al. (1984).

In contrast to the German and international psychiatric consultation services described in the table, the psychosomatic medicine consultation service had some special features (Knorr et al. 1996). There were rarely referrals for psychosis, substance misuse problems or suicidal behaviour; approximately half of referrals were as a result of somatic symptoms in the absence of an organic aetiology, and psychopharmacological treatment methods were very rarely used. Unexplained physical symptoms occupied between second and fourth place in the order of frequency of referrals in the psychiatric consultation services and accounted for a considerable proportion of referrals in some services (Kapfhammer et al. 1992).

Psychiatric consultation services employ a wide range of treatment modalities. For example, Arolt (1997) reported the following recommendations for cases of alcohol-related problems: psychotherapeutic treatments in 44.8% of cases; psychosocial help in 34.5%; and psychopharmacological treatment in 25.9%. Bender et al. (1983) employed advice, behavioural psychotherapy, relaxation exercises and antidepressants most frequently. However, general doubts are expressed by psychiatrists as to whether the colleagues they advise act on that advice to a sufficient extent (see Sect. 9 for the problem of compliance).

Further details relevant to psychiatric work in medical specialties such as psycho-oncology, transplantation surgery, intensive care medicine and infectious diseases can be found in Vol. 2, Part 1 (see also Bönisch et al. 1986; Cravin and Rodin 1992; Breitbart and Holland 1993; Kopp et al. 1994; Rundell and Wise 1996).

6.2
Psychiatrists and Neurologists in Private Practice

A significant proportion of consultation psychiatry practice in German general hospitals is undertaken by office-based neurologists and psychiatrists in private practice (Fegers 1999). There is, however, inadequate data available describing the extent of this consultation work, particularly in nursing homes and homes for the elderly. This led the German Association for Psychiatry, Psychotherapy and Neurology (Deutsche Gesellschaft fur Psychiatrie, Psychotherapie und Nervenheilkunde, DGPPN) to order an overview of this area of practice within the framework of their section for health care delivery, which the German Task Force for Consultation Psychiatry and Psychotherapy (Arbeitsgemeinschaft Konsiliarpsychiatrie und -psychotherapie) also belongs to. A previous German study on the delivery of mental health care by office-based psychiatrists and neurologists, the so-called Psychiatrist-Neurologist Study (Nervenarzt-Studie; Bochnik and Koch 1990, pp. 54, 151), had not gone into this in detail.

In the USA, a survey of members of the American Psychiatric Association, conducted in 1991, revealed that 7.5% of the 36,000 or so members spent at least

Table 1. Consultation psychiatry in general hospitals in Germany and Austria

	Arolt et al. (1995)	Deister (1994)	Kapfhammer (1992)	Saupe u. Diefenbacher (1996b)[d]	Fiebiger et al. (1997)	Fleischhacker et al. (1986)	G. Herzog et al. (1993)	Knorr et al. (1996)[g]	Hengeveld et al. (1984)
Location of study	Lübeck	Bonn	Munich	Berlin (two hospitals)	Görlitz	Innsbruck	Graz	Berlin	Literature review[i]
Rate of referral to consultation psychiatric services (% of all admissions)	3.6	About 3–5[a]	2	About 1[e]	1.52	2	0.8	About 1	0.5–9.1
Proportion of referrals from general medicine (%)	58.8	23.5	29	–	47	58.7	About 10[f]	–	47.7–90
Proportion of referrals from surgical specialties (%)	29.3	31	11.6	–	12.5	25.5	About 2.4[f]	–	7–34.7
Proportion of referrals from other specialties (%)	5.8 (gynaecology)	25.5 (neurology, epilepsy, neurosurgery)	20.3[b] (neurology)	–	17	4.3 (neurology)	About 2.5[f] (dermatology)	–	1–26.9 (gynaecology)
Diagnoses (%)									
Organic psychoses	20.7	40.7	11.8	30.2	23.3	49	20.1	About 1	0.6–57
Neurotic, adjustment and somatoform disorders, personality disorders	34.8	13.2	63.6[c]	19.5	32.3	44.7	–	About 72[h]	2–48
Affective disorders	7.7	23.9	15	13.9	16.9	9.9	29.3	About 18	4–62
Substance-related disorders	24.6	28.3	5.6	17.4	23.3	–	7.9	About 2	0.6–28
Treatment recommendations (%)									
Psychotherapeutic measures	19.1	–	27.4	–	–	–	9		–
Psychotropic medication	41.3	–	41.8	57[e]	69	–	72.4	3	14–74.5
Transfer to a psychiatric ward	28.7	–	5.6	–	14	15.7	26		5–31
Reasons for referral (%)									
Suicide attempt/suicidal ideation	25.1	–	14.4	10	7	24.7	–	2	5.1–47
Addiction	19.8	–	–	11.5	3	28.3	–	0	–
Acute psychiatric symptoms	–	–	–	42.2	57	–	–	31	–
Physical symptoms with no organic explanation	–	–	–	10.3	18	–	–	55	12–22

[a]General medical, surgical and neurological patients.
[b]Out-patient pain clinic and physical medicine 14.3% (liaison work).
[c]Of these, 21.7% were somatoform disorders.
[d]Data from two hospitals: Urban-Krankenhaus and the Universitätsklinikum Rudolf Virchow of the Freie Universität (FU) in Berlin.
[e]Data only for the Universitätsklinikum of the FU in Berlin.
[f]Percentages each refer to all hospital admissions.
[g]Psychosomatic medicine consultation service of the Universitätsklinikum Benjamin Franklin of the FU in Berlin.
[h]Somatoform disorders about 16%, organic mental disorders (ICD-10F5) about 22%.
[i]Each entry gives the range of percentages in the reviewed publications.

25% of their working time engaged in liaison-consultation work, and 3.2% spent more than half of their time on such work. About 10.5% of members had a contractual arrangement with a consultation-liaison service. A greater proportion of general hospital psychiatrists than psychiatrists in private practice had these sorts of arrangements (Noyes et al. 1992).

A survey of neurologists and psychiatrists in private practice in Austria in the 1980s found that 41% and 21%, respectively, were engaged in consultation work in non-psychiatric hospital departments. A total of 87% of all hospitals, general and specialist, worked in collaboration with a consultation psychiatrist or neurologist; about 13% were served by specialists in private practice according to demand, and about 4% by a nearby psychiatric hospital. About 70% had at their disposal consultation psychiatrists employed by the hospital. However, the vast majority were part-time appointments for only a few hours each week. It was concluded on this basis that, overall, specialists in private practice undertook the greatest proportion of consultation work (Wancata and Gossler 1999).

6.3 Provision of Services to Homes for the Elderly and Nursing Homes

Since the 1970s, there has been a fundamental structural change in the composition of old people's and nursing homes. While 30% of residents were dependent on nursing care in 1969, this number rose to 70% in 1988 according to a German article (Vollhardt 1993). A large proportion of residents suffer from a mental disorder, usually in addition to at least one physical illness. Dementia is the most common psychiatric disorder.

Kim and Rovner (1996) found that 67.4% of the residents of a nursing home in Baltimore, USA, were suffering from dementia and a further 12.8% from a depressive illness. A total of 40% of the demented residents experienced additional psychiatric symptoms such as delusions, hallucination or depression which led to behaviour problems making their care more difficult (Kim and Rovner 1996; Burns et al. 1988). Similar figures have been found in German studies (e.g., 50% prevalence of dementia and 25% of depression in elderly nursing home residents in a Hamburg study; Wörle et al. 1992). The prevalence (57.5%) of psychiatric disorders in nursing homes has been found to be even higher than in residential homes for the elderly. In contrast, a prevalence rate of 24.2% was found in a German community sample of older people (Cooper and Sosna 1983).

During the 1980s, an expert commission set up by the German government confirmed once again that the provision of psychiatric services to mentally ill residents of homes was inadequate (Deutscher Bundestag 1975, p. 13). This had already been described by the German national inquiry into mental health services (*Psychiatrie-Enquete*). The inadequate provision of psychiatric and neurological care was condemned and the systematic introduction of consultation psychiatry services to homes was called for, among other measures (Bundesministerium für Jugend, Familie, Frauen und Gesundheit 1988; the BMJFG is the German Ministry for Youth, the Family, Women and Health). However, consultation psychiatry work in nursing homes and homes for the elderly remains rare, despite the pressing need.

A study of all nursing and old people's homes in the district of Gutersloh in Lower Saxony (Germany) described the psychiatric care of mentally ill residents as wholly insufficient; less than a half of the homes had nursing staff with psychiatric expertise at their disposal, and treatment by a specialist could only be guaranteed in about 25% of the homes. In contrast, the medical services from general practitioners were described as satisfactory (Steinkamp 1993). Similar findings have been published in the USA, where only between 2% and 14% of the mentally ill residents of homes are jointly managed with psychiatrists (Burns et al. 1988; Reichmann and Katz 1996, p. vii).

It is particularly those patients who are described by nursing staff as uncooperative (e.g., those who are described as manipulative or apathetic and withdrawn) who are not adequately assessed by psychiatric specialists with regards to psychiatric co-morbidity and therefore miss out on the possibility of more specific psychiatric treatments.

Rovner et al. (1992) compared "co-operative" and "unco-operative" patients and found a significantly higher rate of psychiatric co-morbidity in the latter group. They pointed out that the expert assessment of and differentiation between such behaviours facilitates more specific management; thus a demented patient whose delusional experiences lead to disturbed behaviour should be managed differently from a demented patient who is in pain or a dysphasic patient who does not understand instructions given by the nursing staff.

Non-specific management can lead to therapeutic nihilism and the development of a custodial milieu in which a pro-active nursing approach is lacking. Furthermore, such custodial management is associated with a high usage of sedative medication, in particular neuroleptics, despite the well-known associated risks such as falls and hip fractures. In contrast, psychosocial interventions are under-valued (e.g. modulation of the therapeutic milieu or supportive psychotherapy), although preliminary reports are promising (Wojnar and Bruder 1995).

Both the implementation of therapeutic milieu strategies in the management of dementias (Struwe 1995) and the offer of consultation from an out-patient old-age psychiatry service led to a significantly higher number of psychiatric consultations (Nißle 1994) and were associated with reduced use of psychotropic medication in nursing homes and residential homes for the elderly. In addition, such interventions were associated with increased use of anti-depressants instead of neuroleptics or benzodiazepines in the treatment of depressed residents, leading to significant clinical improvement in 60% of those managed in this way (Kim and Rovner 1996). Guidelines for such integrated care of demented residents of nursing homes have been drawn up in the German state of Nordrhein-Westfalen. Psychiatrists with experience of old-age psychiatry have an important role in the implementation of these guidelines (Höft and Paulus 1996; Vollhardt 1993).

The proposed models of co-operation all suggest a liaison psychiatry service which involves the regular attendance of a psychiatric specialist or neurologist in the homes for a specified number of hours per week, e.g., with regular weekly or fortnightly visits. The models require close co-operation with the doctor responsible for the general medical care of the residents. Emphasis is placed on the importance of giving advice about patients to nursing staff and regular contact with the management of the homes (Struwe 1995; Bienenfeld and Wheeler 1989; Sakauye and Camp 1992).

7 International Perspectives

Internationally, most consultation-liaison psychiatric services comprise elements from along a spectrum between classical consultation services and classical liaison services, the latter being provided to specialist areas (such as care of parasuicides, psycho-oncology, transplantation services). In most cases, consultation services account for the major share of the services provided. The exact balance depends on local conditions and the individual interests of the consulting psychiatrists. Where all services are provided, this is usually taken into account and the term consultation-liaison psychiatry is used. The British usually use the term liaison psychiatry (Guthrie and Creed 1996; Huyse et al. 1996a).

In English-speaking countries, the combined term "psychiatry and psychotherapy", which is used in German, is not in common usage because psychiatry includes psychotherapy (psychodynamic and/or cognitive/behavioural). The most important consultation-psychiatric organisation in the USA, the Academy of Psychosomatic Medicine, gave itself the additional name of Organization for Consultation-Liaison Psychiatry a few years ago to make clear its move away from classical psychosomatic concepts. There has since been discussion as to whether a further renaming to "medical-surgical psychiatry", for instance, might be more appropriate to reflect the increasing complexity of the field of consultation-liaison psychiatry (Thompson 1993). Although there are varying approaches in different countries (Greenhill 1977), the so-called German special way (A.E. Meyer 1992) must be mentioned separately because of its peculiarities.

Psychosomatic-psychotherapeutic theories were very influential in the USA in the 1930s and were very important for the development of consultation and liaison services within psychiatry. In contrast, in Germany in the 1920s, psychosomatic medicine developed as a discipline within general medicine, largely separate from classical German psychiatry (Lipowski 1992; A.E. Meyer 1992; Schwab 1989). Most other countries, such as the USA, have a unified model of consultation-liaison psychiatry. In contrast, the German "special way", with its two distinct medical specialties each with their own training, has led to a two-stranded system (psychiatric *and* separate psychosomatic consultation services), particularly within universities. In practice, the majority of consultation work is undertaken by the psychiatrists.

The dichotomous nature of consultation-liaison services in Germany is most evident where psychiatric and psychosomatic services co-exist in the same hospital (see Sect. 6.1). There is, for example, certainly a wide overlap in the care of patients with depressive symptoms. This would appear to be problematic given the differing therapeutic approaches, especially as there is no universally accepted way of allocating patients to one service or the other (Arolt et al. 1995; Knorr et al. 1996). In addition, an established trend has developed in Germany towards clinical psychology with the advent of behaviour therapy models in line with international developments. Behavioural medicine takes a genuinely psychological approach in the care of general hospital patients with psychiatric co-morbidity. The effectiveness of behavioural approaches can be proved in relation to a number of outcomes, including shorter hospital stays (Lupke et al. 1995; Friedman et al. 1995).

8 Benefits of Consultation Psychiatry Interventions

Consultation psychiatry care is seen overwhelmingly as positive by patients (Windgassen et al. 1997). It leads

to improved management of physically ill patients with co-morbid psychiatric disorder (Hall and Frankel 1996). Möller and Lauter (1986) reported that a psychiatric liaison service which provided out-patient follow-up of patients who had attempted suicide, which was provided by the psychiatrist who had first seen them, led to significantly higher compliance with follow-up appointments than in a control group.

The shortage of resources within health services has led increasingly to consideration of the issue of the costs of consultation psychiatry services (Hall and Frankel 1996). The argument in favour of the necessity of consultation psychiatry services on the grounds of cost-effectiveness was put forward early in the USA, as a result of the observation that patients with co-morbid psychiatric disorders had longer lengths of stay in hospital (Billings 1941). It was postulated that a psychiatric intervention might shorten the length of stay of these patients. Predominantly American research has tried to establish that consultation–liaison psychiatric services paid for themselves and furthermore led to cost savings (Schwab 1989; Strain et al. 1994).

Hip fractures are common in the elderly, so it is no wonder that the cost-effectiveness of consultation psychiatry treatment in this patient group was already being researched at the beginning of the 1980s. Strain et al. (1991) found that liaison psychiatry interventions in elderly patients with hip fractures were associated with an average length of stay that was 2 days shorter.

In a follow-up, 3 months after initial admission, patients who had been admitted during the study period had had fewer in-patient re-admissions than the control group. This suggested that the liaison intervention had not merely led to a transfer of costs from the in-patient to the out-patient sector. Lengths of in-patient stay are not conclusive as measures of the success of consultation psychiatry interventions as it is difficult to control adequately for the many confounding variables. Particularly in the field of hip fractures, other interventions have been described which are also associated with shorter admissions, e.g. where the most important factor has been found to be better co-operation with social services without any psychiatric involvement (Seyfarth-Metzger 1997). Whereas it has been convincingly established that patients with co-morbid psychiatric disorders have a longer length of stays on medical wards, study results of the effect of consultation psychiatry interventions with regard to shortening length of stay on medical–surgical wards must be considered as inconclusive, largely on methodological grounds.

9 Research

Some of the tasks and problems facing research in consultation psychiatry are outlined below (Guthrie and Creed 1996; Öhman et al. 1989). In recent years, there has been increasing discussion of the methodological problems involved in the diagnosis of mental illness in patients who are physically ill. For example, problems in the ascertainment of cases have been identified (problems with two-stage screening processes; see Clarke et al. 1993), as have problems in the diagnosis of depressive disorders. In the latter case, evaluation of the somatic symptoms (e.g. sleep disturbance, loss of appetite and weight) is very difficult in the presence of co-existing physical illness, such as malignancy (see Lagrhissi-Thode et al. 1996).

Despite the approaches that are available to understanding psychiatric problems in physically ill patients as a whole group, there has hitherto been no precise description of the phenomenology of psychiatric illnesses co-morbid with particular physical illnesses. No classification system has been developed to meet the specific needs of consultation–liaison psychiatry (Guthrie and Creed 1996; Diefenbacher and Saupe 1994).

Consultation psychiatry research in general hospitals should concentrate on common conditions, such as acute states of confusion in older patients, alcohol-related problems, depressive and adjustment disorders and post-traumatic stress reactions after accidents, with a view to improving diagnosis and treatment. In this respect, some relevant research findings already exist (e.g. Frommberger et al. 1996). Examples of further areas worthy of mention include the effects of modern high-tech medicine on patients in particularly stressful settings such as intensive care or during isolation following bone marrow transplantation (Schmidt-Degenhardt 1986; Andrykowski 1994), interactions of psychotropic medication with non-psychotropic drugs, as well as the psychiatric side-effects of non-psychotropic drugs, especially in older patients (Katz et al. 1994; Kasper and Jung 1995), and the prescription of psychotropic drugs for transplant recipients (Shapiro 1991).

The extent to which psychiatric co-morbidity during an admission for a physical illness is directly caused by the underlying physical illness and can therefore be expected to resolve with successful treatment of the physical illness is an important factor in determining the need for consultation services. This has been described in patients with depressive symptoms, for example (Pomerantz et al. 1992; Popkin et al. 1991).

The identification of risk factors is also important in planning consultation psychiatry interventions, as not

all psychiatric symptoms which develop in association with a physical illness resolve spontaneously once the physical illness improves. This would provide information about the necessity of providing psychiatric treatment to physically ill patients treated on medical-surgical wards (defining needs and planning services, Benjamin et al. 1994).

Treatment studies should aim to identify high-risk groups who benefit from consultation psychiatry interventions. This is important in view of the increasing restrictions on resources within the health service and the introduction of limited budgets and "carve-outs" that are likely to reduce the funds that third-party payers are willing to pay for particular conditions or particular treatment groups, especially if related to psychiatric co-morbidity in physically ill patients. If, for example, a psychiatric intervention was shown to contribute towards shorter lengths of stay in hospital, this might lead to savings.

The following points should be taken into account, among others, when planning relevant studies. Consultation psychiatry interventions can have differing effects on length of stay. They may be associated with shorter admissions (e.g. owing to joint triage by social workers and psychiatrists; Diefenbacher and Strain, in preparation), but in other cases they may be associated with longer admissions (e.g. due to the identification of depression and commencement of anti-depressant treatment; Callies and Popkin 1987). Poor compliance of the referring doctors in implementing the advice of the consultation psychiatrist may impair the success of the intervention. It has been found that only 70% of pharmacological treatment recommendations and 56% of recommendations for further investigations are followed (Popkin et al. 1991). Finally, the timing of the intervention should be considered. Early interventions are more likely to be associated with shorter hospital stays (Ormont et al. 1997).

Lyons and Larson (1989) have argued convincingly that it is not meaningful to measure such a complex phenomenon as the outcome of psychiatric consultations using a single parameter (e.g. length of stay). Instead, they propose a value matrix which takes multiple perspectives into account (e.g. the patient, the family, the psychiatrist, the medical team, the hospital, the medical insurance company, society), both clinical and economic.

Using this framework Strain et al. (1991, 1994) described economic outcomes (shorter in-patient stays, lower rates of re-admission), clinical outcomes (improvement in cognitive function and reduction of depressive symptoms), family/career outcomes (fewer relatives having to cut back on their work to care for the patient at home) and service-related outcomes (generation of funds from savings made). This demonstrates how important it is for every consultation-liaison service to consider how it wishes to affect each of the different parameters, all of which are relevant, more or less obviously, in every consultation.

While attempting to prove the economic benefits of consultation psychiatry, it should never be forgotten that the main goal of psychiatric consultation work is the alleviation of the patients' suffering in the hope that this will contribute towards an improvement in their quality of life (see Schmeling-Kludas 1995).

10 Training

The increasing integration of psychiatric departments within general hospitals will lead to a growing demand for medical care in collaboration with consultation psychiatrists (Creed et al. 1993). It is therefore necessary to develop a consultation-liaison curriculum within the framework of specialist psychiatric training in order to cover comprehensively the diagnosis and treatment of physically ill patients with co-morbid psychiatric symptoms. In addition to the acquisition of theoretical knowledge, practical experience should be provided during specialist psychiatric training, either in the form of rotation through a placement in a consultation psychiatry service or supervised consultation work during training.

Gitlin et al. (1996) suggest a tiered model for training in consultation-liaison psychiatry. They propose a basic compulsory core of knowledge that all psychiatrists should be taught, such as psychotherapy and psychopharmacology with physically ill patients. In addition, specialist centres would provide the opportunity to obtain further qualifications (e.g. transplantation psychiatry, psycho-oncology; e.g. see Craven and Rodin 1992; Breitbart and Holland 1993; Kissane 1993).

Furthermore, there is an urgent need to broaden and deepen non-psychiatric doctors' knowledge and experience of psychiatric issues, including psychosomatic ideas (Gask 1994). Consultation psychiatrists can therefore play an important part not only in training psychiatrists in the treatment of patients with physical illness, but also training general practitioners, physicians and surgeons in the care of patients with co-morbid psychiatric symptoms (Kathol et al. 1994).

There are interesting developments in the USA which provide joint training in somatic and psychiatric medicine. Combined medical-psychiatric specialist training has begun to be established in which a considerable portion of the programme (known as a

medical-psychiatric residency) takes place on so-called medical-psychiatric units.

Overall, there has been an increasing interest internationally in collaborative models in the outpatient and primary care sectors. Attempts are being made to integrate primary and secondary care of the general population better, with inclusion of other professional groups such as community nurses and social workers. Consultation-liaison models may have a useful role to play in this process (Gask et al. 1997; Creed 1996; Goldberg 1997; Hendriksche and Kroger 1997).

It should not be forgotten that the general hospital plays an important role in bringing patients whose psychiatric symptoms have hitherto gone unnoticed to the notice of specialist psychiatric services for the first time. Up to two thirds of patients seen by consultation-psychiatric services have never previously received psychiatric treatment (Saupe and Diefenbacher 1996b; Bass 1995; Creed et al. 1993). In this context, academic consultation psychiatry should strive to collaborate closely with academic departments of primary care and general medicine.

11 Further Liaison Models (Medical-Psychiatric Units)

Combined so-called medical-psychiatric units should be mentioned as an example of possible models of collaboration within the general hospital. They have experienced an increase in popularity over the last 10 years in the USA.

Kathol et al. (1992) have proposed a categorisation of in-patient units for the treatment of patients with physical and psychiatric problems. They suggest four types of unit depending on the level of what they describe as medical and psychiatric "acuity". Type I refers to the usual psychiatric in-patient units which treat primarily psychiatric problems, varying from very acute to less acute, but which can treat only a few non-severe medical problems. Type II units are their medical counterparts. These include general medical units or medical sub-specialty units that are associated with a psychiatric liaison service and provide low levels of psychiatric care to those admitted to the general medical setting. Medical-psychiatric units in the narrow sense are type III and IV units, as can be found in tertiary-care hospitals such as at the University of Iowa and the Emory University Hospital in the USA. They care for patients with concurrent and more severe psychiatric and medical problems in a unified setting. These patients would otherwise be frequently transferred back and forth between medical or surgical wards and psychiatric wards. Such medical-psychiatric units are run jointly by psychiatrists and physicians.

The following patients are given priority on a type III unit: delirious patients who cannot be managed on a medical ward, suicidal patients in need of acute medical treatment, patients whose physical state is unstable following intoxication etc. A type IV unit is a stage beyond that. It would be able to carry out more complicated medical procedures such as haemodialysis, peritoneal dialysis or parenteral treatment via central venous access (Stoudemire 1996).

In Germany, patients identified as having psychosomatic illnesses are referred to psychosomatic wards, mainly located in rehabilitation centres, and not in acute-care hospitals. Joint internal medicine-psychosomatic medicine units, corresponding to the American type II units, have seldom been tried (see Köhle et al. 1996, pp. 528–540). One should therefore avoid directly equating American medical-psychiatric units with German psychosomatic medicine units. Units similar to the American type III and type IV units can be found in Germany in such places as the Zentralinstitut für Seelische Gesundheit (the Central Institute for Mental Health) in Mannheim or the Wilhelm-Griesinger-Krankenhaus (Wilhelm-Griesinger Hospital) in Berlin. The latter is a neuropsychiatric intensive care unit.

12 Quality Assurance, Documentation and Further Reading

The following standards have been suggested for the organisation of consultation services in hospitals: consultation rates of 3%–5% of patients admitted to a general hospital; average time spent per patient about 2.5 h (including assessment and three follow-up appointments, advising the team and associated paperwork). If these standards were met, a psychiatrist employed by a consultation service would manage about 300 patients each year. Wherever possible, the psychiatrist should work with a multi-disciplinary team including social workers and specially trained liaison nursing staff (House and Hodgson 1994; Herzog and Hartmann 1990; Bönisch et al. 1986).

It remains to be seen to what extent the findings of the European Consultation Liaison Workgroup (ECLW) study can be usefully implemented (Huyse et al. 1996b). In this study, several hundred consultation psychiatrists throughout Europe documented their activities to provide empirical support for their suggestions.

Adequate documentation of consultation work is necessary for administrative purposes, but can also be

used for training, quality assurance and research. The American clinical electronic MICROCARES database provides access to a series of rating scales in addition to the individual service's consultation-psychiatric clinical data entry form. It exists in an optical pen entry form and can therefore be used at the bedside with a laptop computer without any extra time commitment (Hammer et al. 1995).

There is a specialist English-language consultation psychiatry literature with several journals (e.g. *General Hospital Psychiatry*, *Psychosomatics*, *International Journal of Psychiatry in Medicine*) and textbooks (e.g. Cassem 1997; Rundell and Wise 1996; Guthrie and Creed 1996; the only French textbook on consultation-liaison psychiatry is by Zumbrunnen 1991), covering the full range of psychiatric disorders and treatment possibilities. *Nursing Home Medicine - The Annals of Long-term Care* is a multi-disciplinary journal of great practical use for psychiatric consultation work in nursing and old people's residential homes. Strain (1996b) has produced a list of over 2000 consultation psychiatry references selected by experts. It is also available in electronic form on a 3.5″ disk with abstracts and, in some places, a commentary. It provides a comprehensive synthesis of relevant publications from recent years. It is helpful to both the consultation expert and the non-expert who wish to inform themselves about general or specific aspects of everyday clinical practice or about research in consultation-liaison psychiatry. The Academy of Psychosomatic Medicine has produced a collection of slides that are well suited to teaching and which include important data relevant to consultation-liaison psychiatry (Academy of Psychosomatic Medicine 1997).

13 References

Academy of Psychosomatic Medicine (1997) Mental disorders in general medical practice - adding value to healthcare through consultation-liaison psychiatry. Kendall/Hunt, Dubuque, IA

Anderson DN, Philpott RM (1991) The changing pattern of referrals for psychogeriatric consultation in the general hospital: an eight-year study. Int J Geriatr Psychiatry 6: 801-807

Andrykowski MA (1994) Psychiatric and psychosocial aspects of bone marrow transplantation. Psychosomatics 35: 13-14

Arolt V (1997) Psychische Störungen bei Krankenhauspatienten - eine epidemiologische Untersuchung zu Diagnostik, Prävalenz und Behandlungsbedarf psychiatrischer Morbidität bei internistischen und chirurgischen Patienten. Springer, Berlin Heidelberg New York

**Arolt V, Gehrmann A, John U, Dilling H (1995) Psychiatrischer Konsiliardienst an einem Universitätsklinikum - eine empirische Untersuchung zur Leistungscharakteristik. Nervenarzt 66: 347-354

Bass CM (1995) The role of liaison psychiatry. In: House A, Mayou R, Mallinson C (eds) Psychiatric aspects of physical disease. Royal College of Physicians and Royal College of Psychiatrists, London, pp 91-99

Bender W, Greil W, Meyer G (1983) Psychiatrischer Konsiliardienst an einem medizinischen Großklinikum: Evaluation dreier Jahrgänge. Psychiatr Clin 16: 324-339

Benjamin S, House A, Jenkins P (eds) (1994) Liaison psychiatry - defining needs and planning services. Gaskell, Glasgow

Bienenfeld D, Wheeler BG (1989) Psychiatric services to nursing homes: a liaison model. Hosp Community Psychiatry 40: 793-794

Billings EG (1941) The value of psychiatry to the general hospital. Hospitals 15: 30-34

Blankenburg W (1988) Der Leib - das gemeinsame Thema von somatischer und psychosomatischer Medizin. In: Bräutigam W (ed) Kooperationsfragen somatischer und psychosomatischer Medizin. Springer, Berlin Heidelberg New York, pp 61-71

Bundesministerium für Jugend, Familie, Frauen und Gesundheit (BMJFG) (1988) Empfehlungen der Expertenkommission der Bundesregierung zur Reform der Versorgung im psychiatrischen und psychotherapeutisch/psychosomatischen Bereich auf der Grundlage des Modellprogramms Psychiatrie der Bundesregierung. BMJFG, Bonn

Bochnik HJ, Koch H (1990) Die Nervenarzt-Studie. Deutscher Ärzte-Verlag, Cologne

Böker W (1973) Sozialpsychiatrische Konsultationstätigkeit im Allgemeinen Krankenhaus. Fortschr Med 91: 683-684, 714

Bönisch E, Meyer JE (1975) Medizinische Extremsituationen und der sterbende Patient. In: Kisker KP, Meyer JE, Müller C, Strömgren E (eds) Psychiatrie der Gegenwart, vol 3, 2nd edn. Springer, Berlin Heidelberg New York, pp 519-555

Bönisch E, Götze P, Meyer JE (1986) Zur Psychologie und Psychopathologie bei schweren und unheilbaren Organerkrankungen. In: Kisker KP, Lauter H, Meyer JE, Müller C, Strömgren E (eds) Psychiatrie der Gegenwart, vol 2, 3rd edn. Springer, Berlin Heidelberg New York, pp 178-227

Breitbart W, Holland JC (1993) Psychiatric aspects of symptom management in cancer patients. American Psychiatric Press, Washington, DC

Burns BJ, Larson DB, Goldstrom IG, Johnson WE, Taube CA, Miller NE, Mathis ES (1988) Mental disorder among nursing home patients: preliminary findings from the National Nursing Home Survey Pretest. Int J Geriatr Psychiatry 3: 27-35

Callies AL, Popkin MK (1987) Antidepressant treatment of medical-surgical inpatients by nonpsychiatric physicians. Arch Gen Psychiatry 44: 157-160

**Cassem NH (1997) (ed) The Massachusetts General Hospital handbook of general hospital psychiatry, 4th edn. Mosby, New York

Clarke DM, Smith GC, Herrman HE (1993) A comparative study of screening instruments for mental disorders in general hospital patients. Int J Psychiatry Med 23: 323-337

Cooper B, Sosna U (1983) Psychische Erkrankung in der Altenbevölkerung - eine epidemiologische Feldstudie in Mannheim. Nervenarzt 54: 239-249

Craven J, Rodin GM (1992) Psychiatric aspects of organ transplantation. Oxford University Press, Oxford

Creed F (1996) Developments in liaison psychiatry. Curr Opin Psychiatry 9: 433-438

Creed F, Guthrie E, Black D et al (1993) Psychiatric referrals within the general hospital: comparison with referrals to general practitioners. Br J Psychiatry 162: 204–211

Deister A (1994) Häufige Fragestellungen im Rahmen des psychiatrischen Konsiliardienstes. Internist 35: 807–813

Deutscher Bundestag (1975) Bericht über die Lage der Psychiatrie in der Bundesrepublik Deutschland – zur psychiatrischen und psychotherapeutisch/psychosomatischen Versorgung der Bevölkerung. (Drucksache 7/4200) Heger, Bonn

Diefenbacher A, Saupe R (1994) Psychopathologie beim chronischen subduralen Hämatom – Bemerkungen zur Arbeit des Konsilpsychiaters. Schweiz Arch Neurol Psychiatr 145: 7–10

Diefenbacher A (ed) (1999) Aktuelle Konsiliarpsychiatrie und Psychotherapie, vol 1. Enke, Stuttgart

Diefenbacher A, Saupe R (eds) (in preparation) Handbuch der Konsiliarpsychiatrie und Psychotherapie. Steinkopff, Darmstadt

Fegers S (1999) Konsiliarpsychiatrie aus der Sicht des niedergelassenen Nervenarztes. In: Diefenbacher A (ed) Aktuelle Konsiliarpsychiatrie und Psychotherapie, vol 1. Enke, Stuttgart

Fiebiger D, Ficker F, Winiecki P, Stein B, Herzog T, the European Consultation Liaison Workgroup (ECLW) (1997) Der Psychiatrische Konsiliardienst der Klinik für Psychiatrie an der Städtisches Klinikum Görlitz GmbH. Psychiatr Prax 24: 129–133

Fleischhacker WW, Barnas C, Haring C, Stuppäck C, Unterweger B, Wagner R (1986) Der psychiatrische Konsiliardienst – eine Analyse von Bedarf und Inanspruchnahme im a. ö. Landeskrankenhaus (Universitätsklinik) Innsbruck. Nervenarzt 57: 589–592

Freeman A, Davis L, Libb JW, Craven J (1992) Assessment of transplant candidates and prediction of outcome. In: Craven J, Rodin GM (eds) Psychiatric aspects of organ transplantation. Oxford University Press, Oxford, pp 9–21

Friedman R, Sobel D, Myers P et al (1995) Behavioral medicine, clinical health psychology, and cost offset. Health Psychol 14: 509–518

Frommberger U, Käppler C, Stieglitz RD, Schlickewei W, Kuner E, Berger M (1996) Die Entwicklung von posttraumatischen Belastungsstörungen nach Verkehrsunfällen. Erste Ergebnisse einer prospektiven Studie. In: Möller HJ, Engel RR, Hoff P (eds) Befunderhebung in der Psychiatrie: Lebensqualität, Negativsymptomatik und andere aktuelle Entwicklungen. Springer, Berlin Heidelberg New York, pp 309–312

Gask L (1994) Training for general practitioners in psychiatry. In: Pullen I, Wilkinson G, Wright A, Gray DP (eds) Psychiatry and general practice today. The Royal College of Psychiatrists and The Royal College of General Practitioners, London, pp 337–349

Gask L, Sibbald B, Creed F (1997) Evaluating models of working at the interface between mental health services and primary care. Br J Psychiatry 170: 6–11

Gitlin DF, Schindler BA, Stern TA et al (1996) Recommended guidelines for consultation-liaison psychiatric training in psychiatry residency programs. Psychosomatics 37: 3–11

Goldberg D (1997) Implications of epidemiological findings for the management of mental disorders encountered in primary care settings. Eur Psychiatry 12[Suppl 2]: 56–62

Greenhill MH (1977) The development of liaison programs. In: Usidin G (ed) Psychiatric medicine. Brunner and Mazel, New York, pp 115–191

*Guthrie E, Creed F (1996) Seminars in liaison psychiatry. Gaskell, London

Hackett TP, Cassem NH (1987) The Massachusetts General Hospital handbook of general hospital psychiatry, 2nd edn. PSG, Littleton, MA, pp xi, 1–13

Haag A, Stuhr U (1992) Über den Nutzen integrierter Psychosomatik im Allgemeinen Krankenhaus. In: Uexküll T von, Adler R (eds) Integrierte Psychosomatische Medizin in Praxis und Klinik. Schattauer, Stuttgart, pp 43–52

Häfner H (1991) Die Reform der Versorgung psychisch Kranker in der Bundesrepublik. In: Häfner H (ed) Psychiatrie: ein Lesebuch für Fortgeschrittene. Fischer, Stuttgart, pp 256–282

Hall RCW, Frankel BL (1996) The value of consultation-liaison interventions to the general hospital. Psychiatr Serv 47: 418–420

Hammer JS, Strain JJ, Friedberg A et al (1995) Operationalizing a bedside pen entry notebook clinical database system in consultation-liaison psychiatry. Gen Hosp Psychiatry 17: 165–172

Hendrischke A, Kröger F (1997) Systematische Familienmedizin – ein Modell für Kooperation im Gesundheitswesen. Dtsch Ärztebl 94/63: A294–296

Hengeveld MW, Rooymans HGM, Vecht-van den Bergh R (1984) Psychiatric consultations in a Dutch university hospital: a report on 1814 referrals, compared with a literature review. Gen Hosp Psychiatry 6: 271–279

Herzog G, Wieselmann G, Marguc K, Zapotoczky HG (1993) Psychiatrische Konsiliartätigkeit an einem allgemeinen österreichischen Krankenhaus und Universitätsklinikum (LKH-Graz). Psycho 19: 181–194

Herzog T, Hartmann A (1990) Psychiatrische, psychosomatische und medizinpsychologische Konsiliar- und Liaisontätigkeit in der Bundesrepublik Deutschland. Ergebnisse einer Umfrage. Nervenarzt 61: 281–293

Höft B, Paulus HJ (1996) Leitlinien für die integrative Betreuung dementer Bewohner in Altenpflegeeinrichtungen. Z Gerontol Geriatr 29: 150–158

House A, Hodgson G (1994) Estimating needs and meeting demands. In: Benjamin S, Hosue A, Jenkins P (eds) Liaison psychiatry – defining needs and planning services. Gaskell, Glasgow, pp 3–15

Huyse FJ, Herzog T, Malt UF (1996a) International perspectives on consultation-liaison psychiatry. In: Rundell JR, Wise MG (eds) Textbook of consultation-liaison psychiatry. American Psychiatric Press, Washington, DC, pp 228–255

Huyse FJ, Herzog T, Malt UF, Lobo A (1996b) The European Consultation Liaison Workgroup (ECLW) Collaborative Study. I. General outline. Gen Hosp Psychiatry 18: 44–55

Huyse FJ, Herzog T, Lobo A, Lyons JS, Slaets JPJ, Fink P, Stiefel F, de Jonge P (1997) Detection and treatment of mental disorders in general health care. Eur Psychiatry 12[Suppl 2]: 70s–78s

**John U, Hapke U, Rumpf HJ, Hill A, Dilling H (1996) Prävalenz und Sekundärprävention von Alkoholmißbrauch und -abhängigkeit in der medizinischen Versorgung. Nomos, Baden-Baden (Schriftenreihe des Bundesministeriums für Gesundheit, vol 71, pp 56–61)

Kapfhammer HP (1992) Psychische Störungen bei körperlichen Erkrankungen: Erfahrungen im psychiatrischen Konsiliardienst. In: Hippius H, Lauter H, Greil W (eds) Psychische Störungen bei körperlichen Erkrankungen. MMW, Munich, pp 11–30 (Psychiatrie für die Praxis, vol 16)

Kapfhammer HP, Buchheim P, Bove D, Wagner A (1992) Konversionssymptome bei Patienten im psychiatrischen Konsiliardienst. Nervenarzt 63: 527–538

Kapfhammer HP (1993) Die psychopharmakologische Behandlung von ängstlich-depressiven Syndromen im Kontext somatischer Erkrankungen. In: Möller HJ (ed) Therapie psychiatrischer Erkrankungen. Enke, Stuttgart, pp 801–818

Kasper S, Jung B (1995) Psychiatrisch relevante Nebenwirkungen der nichtpsychopharmakologischen Pharmakotherapie. Nervenarzt 66: 649–661

Kathol RG, Harsch HH, Hall RCW et al (1992) Categorization of types of medical/psychiatry units based on level of acuity. Psychosomatics 33: 376–386

Kathol RG, Katon W, Smith RG et al (1994) Guidelines for the diagnosis and treatment of depression for primary care physicians – implications for consultation-liaison psychiatrists. Psychosomatics 35: 1–12

Katon WJ (1991) Panic disorder in the medical setting. American Psychiatric Press, Washington, DC

Katz IR, Streim J, Parmelee P (1994) Psychiatric-medical comorbidity: implications for health services delivery and for research on depression. Biol Psychiatry 36: 141–145

Kim E, Rovner B (1996) The nursing home as a psychiatric hospital. In: Reichman WE, Katz PR (eds) Psychiatric care in the nursing home. Oxford University Press, Oxford, pp 3–9

Kissane DW (1993) Psychotherapy for physical disorders. Curr Opin Psychiatry 6: 332–336

Knorr C, Diefenbacher A, Paetzmann S, the European Consultation Liaison Workgroup (ECLW) (1996) Vergleich eines psychosomatischen und eines psychiatrischen Konsiliardienstes zweier Universitätsklinika in Berlin. In: Peters UH, Schifferdecker M, Krahl A (eds) 150 Jahre Psychiatrie, vol 1. Martini, Cologne, pp 634–638

Koch U, Siegrist B (1988) Psychosomatische Dienste in medizinischen Kliniken – die Kooperationsfrage unter forscherischer Perspektive. In: Bräutigam W (ed) Kooperationsfragen somatischer und psychosomatischer Medizin. Springer, Berlin Heidelberg New York, pp 81–97

Köhle K, Joraschky P, Reisinger E (1996) Die Institutionalisierung im klinischen Bereich. In: Uexküll T von (ed) Psychosomatische Medizin, 5th edn. Urban and Schwarzenberg, Munich, pp 516–540

Kopp M, Schweigkofler H, Fleischhacker W et al (1994) Psychoonkologischer Liaisondienst zur Versorgung von Krebspatienten im Rahmen einer Knochenmarktransplantation. Psychotherapeut 39: 380–385

Laghrissi-Thode F, Pollock BG, Szanto K, Reynolds CF (1996) Depression and suicide in medically ill patients. Curr Opin Psychiatry 9: 137–140

Levenson Jl, Hamer RM, Rossiter LF (1990) Relation of psychopathology in general medical inpatients to use and cost of services. Am J Psychiatry 147: 1498–1503

Levy NB (1989) Psychosomatik und Konsultations-/Liaison-Psychiatrie: ein Überblick. Nervenarzt 60: 724–731

Lipowski ZB (1992) Consultation-liaison psychiatry at century's end. Psychosomatics 33: 128–133

Lupke U, Ehlert U, Hellhammer D (1995) Effekte psychologischer Behandlung im Allgemeinkrankenhaus: Verlaufsuntersuchung an Patienten mit Somatisierungsverhalten. Psychother Psychosom Med Psychol 45: 358–365

Lyons JS, Larson DB (1989) A proposed value matrix for the evaluation of psychiatric consultations in the general hospital. Gen Hosp Psychiatry 11: 345–351

Margolis RL (1994) Nonpsychiatric house staff frequently misdiagnose psychiatric disorders in general hospital inpatients. Psychosomatics 35: 485–491

Meyer AE (1992) Eine kurze Geschichte der Psychosomatik – der Sonderweg der ehemaligen Bundesrepublik. In: Uexküll T von, Adler R (eds) Integrierte Psychosomatische Medizin in Praxis und Klinik. Schattauer, Stuttgart, pp 35–42

*Meyer E, Mendelson M (1961) Psychiatric consultations with patients on medical and surgical wards: patterns and process. Psychiatry 24: 197–220

Möller HJ, Lauter H (1986) Der psychiatrische Liaisondienst – neue Gesichtspunkte bei der stationären und poststationären Versorgung nach Suizidversuch. In: Helmchen H, Hippius H (eds) Psychiatrie für die Praxis, vol 3. MMW, München, pp 116–123

Möller HJ, Scriba PC (eds) (1994) Innere Medizin und Psychiatrie. Internist 35: 805–862

Müller C (1981) Psychiatrische Institutionen – ihre Möglichkeiten und Grenzen. Springer, Berlin Heidelberg New York

Nißle K (1994) Evaluation eines gerontopsychiatrischen ambulanten Behandlungskonzeptes. Psychiatr Prax 21: 143–146

Noyes R, Wise TN, Hayes JR (1992) Consultation-liaison psychiatrists – how many are there and how are they funded? Psychosomatics 33: 123–127

Öhman R, Free HL, Homkvist AF et al (1989) Interaction between mental and physical illness – needed areas of research. Springer, Berlin Heidelberg New York Tokyo

Ormont MA, Weisman HW, Heller SS et al (1997) The timing of psychiatric consultation requests: utilization, liaison, and diagnostic considerations. Psychosomatics 38: 38–44

Panse F (1964) Das psychiatrische Krankenhauswesen – Entwicklung, Stand, Reichweite und Zukunft. Thieme, Stuttgart

Pomerantz AS, Nesnera A, West AN (1992) Resolution of depressive symptoms in medical inpatients after discharge. Int J Psychiatry Med 22: 281–289

Popkin MK, Colon EA, Callies AL, Mackenzie TB (1991) The shift from outcome studies to epidemiological studies of specific medical illnesses in consultation–liaison psychiatry. Psychiatr Med 9/4: 607–621

Reichman WE, Katz PR (eds) (1996) Psychiatric care in the nursing home. Oxford University Press, Oxford

Rössler W, Salize HJ, Bauer M (1996) Psychiatrische Abteilungen an Allgemeinkrankenhäusern – Stand der Entwicklung in Deutschland. Psychiatr Prax 23: 4–9

Rovner BW, Steele CD, German P, Clark R, Folstein MF (1992) Psychiatric diagnosis and uncooperative behavior in nursing homes. J Geriatr Psychiatry Neurol 5: 102–105

*Rundell JR, Wise MG (1996) Textbook of consultation–liaison psychiatry. American Psychiatric Press, Washington, DC

Sakauye KM, Camp CJ (1992) Introducing psychiatric care into nursing homes. Gerontologist 32: 849–852

*Saravay SM, Lavin M (1994) Psychiatric comorbidity and length of stay in the general hospital: a critical review of outcome studies. Psychosomatics 35: 233–252

*Saupe R, Diefenbacher A (1996a) Praktische Konsiliarpsychiatrie und -psychotherapie. Enke, Stuttgart

Saupe R, Diefenbacher A, for the European Consultation Liaison Workgroup (ECLW) (1996b) Konsilpsychiatrie: Sozial- und angewandte Neuropsychiatrie. In: Peters UH, Schifferdecker M, Krahl A (eds) 150 Jahre Psychiatrie, vol 1. Martini, Cologne, pp 639–643

Schliack H (1992) Das ärztliche Konsilium – Gedanken eines Neurologen. Dtsch Ärztebl 89: B374–375

Schmeling-Kludas C (1995) Psychosomatik im Allgemeinen Krankenhaus – Belastungsspektrum, Bewältigung und Therapiemöglichkeiten bei internistischen Patienten. VAS, Frankfurt

Schmidt-Degenhardt M (1986) Oneiroides Erleben bei intensivbehandelten panplegischen Polyradikulitis-Patienten. Nervenarzt 57: 712–718

Schwab JJ (1989) Consultation-liaison psychiatry: a historical overview. Psychosomatics 30: 245–254

Seyfarth-Metzger I (1997) Vertrauen durch Qualität. Managem Krankenhaus 7: 9

Shapiro PA (1991) Nortriptyline treatment of depressed cardiac transplant recipients. Am J Psychiatry 148: 371–373

Sperling W, Kalb R (1995) Das psychiatrische Konsil vor Lebertransplantationen. Fortschr Med 113: 175–177

Spiess K (1996) Das subkulturelle Randphänomen in der konsiliarpsychosomatischen Begegnung. Gruppenpsychother Gruppendyn 32: 150–170

Steinkamp G, Tropberger F, Werner B (1993) Heimliche Gerontopsychiatrie oder Wer hilft den Heimen bei der Versorgung psychisch kranker alter Menschen? Eine Untersuchung aller Alten- und Altenpflegeheime des Kreises Gütersloh 1991. Z Gerontol 26: 494–500

Stoudemire A (1996) Medical-psychiatric units. In: Rundell JR, Wise MG (eds) Textbook of consultation-liaison psychiatry. American Psychiatric Press, Washington, DC, pp 900–913

*Strain JJ (1996a) Liaison psychiatry. In: Rundell JR, Wise MG (eds) Textbook of consultation-liaison psychiatry. American Psychiatric Press, Washington, DC, pp 38–51

Strain JJ (ed) (1996b) Consultation-liaison psychiatry database (1996 update). Gen Hosp Psychiatry 18/5 (special issue): 293–375

*Strain JJ, Lyons JS, Hammer JS et al (1991) Cost offset from a psychiatric consultation-liaison intervention with elderly hip fracture patients. Am J Psychiatry 148: 1044–1049

*Strain JJ, Hammer JS, Fulop G (1994) APM Task force on psychosocial interventions in the general hospital inpatient setting: a review of cost-offset studies. Psychosomatics 35: 253–262

Struwe B (1995) Das dementielle Syndrom: psychosoziale Behandlungsmöglichkeiten. Krankenhauspsychiatrie 6: 175–179

Surman O (1992) Liver transplantation. In: Craven J, Rodin GM (eds) Psychiatric aspects of organ transplantation. Oxford University Press, Oxford, pp 177–188

Swigar ME, Sanguineti VR, Piscatelli RL (1992) A retrospective study on the perceived need for and actual use of psychiatric consultation in older medical patients. Int J Psychiatr Med 22: 239–249

Thompson TL (1993) Some advantages of consultation-liaison (medical-surgical) psychiatry becoming an added qualification subspeciality. Psychosomatics 34: 343–349

Vollhardt BR (1993) Landesärztliche Tätigkeit in Altenheimen in Nordrhein-Westfalen. Psycho 19: 369–375

Wallen J, Pincus HA, Goldman HH et al (1987) Psychiatric consultations in short-term general hospitals. Arch Gen Psychiatry 44: 163–168

Wancata J, Benda N, Hajji M et al (1998) Psychische Erkrankungen in internen, chirurgischen und gynäkologischen Abteilungen: Prävalenz und Versorgungsbedarf. In: Meise U, Hafner F, Hinterhuber H (eds) Gemeindepsychiatrie in Österreich. VIP, Innsbruck Wien

Wancata J, Gössler R (1999) Die konsiliarpsychiatrische Versorgung in Österreich. In: Diefenbacher A (eds) Aktuelle Konsiliarpsychiatrie und Psychotherapie, vol 1. Enke, Stuttgart

Weyerer S (1990) Relationships between physical and psychological disorders. In: Sartorius N, Goldberg D, Girolamo G et al (eds) Psychological disorders in general medical settings. Hogrefe and Huber, Toronto, pp 34–46

Windgassen K, Weißen PH, Schmidt K (1997) Vorurteile und Urteile: Die psychiatrische Konsiliaruntersuchung aus der Sicht des Patienten. Psychiatr Prax 24: 134–137

Winkler WT (1975) Das psychiatrische Krankenhaus; organisatorische und bauliche Planung. In: Kisker KP, Meyer JE, Müller C, Strömgren E (eds) Psychiatrie der Gegenwart, vol 3, 2nd edn. Springer, Berlin Heidelberg New York, pp 221–254

Wörle J, Klingenfeld H, Bruder J, Dahme B (1992) Demenz und Depression bei Altenpflegeheim-Bewohnern. Z Gerontopsychol Psychiatr 5: 179–190

Wojnar J, Bruder J (1995) Psychotherapeutische Unterstützung pflegebedürftiger alter Menschen in Heimen. Z Gerontopsychol Psychiatr 8: 163–168

Zumbrunnen R (1991) Psychiatrie de liaison-consultation psychiatrique à l'hôpital général. Masson, Paris

CHAPTER

15

S. Poitras, J.M. Bertolote

Mental Health Legislation: International Trends

1 Introduction

1.1 Formal and Informal Norms

Communities worldwide have long developed norms, formal or informal, to deal with persons with mental disorders. In England, for example, formal legal instruments designed to govern mental health matters are documented as early as in the thirteenth century (Neugebauer 1978). It has become customary to refer to the ensemble formed by those binding norms as "mental health law", here employed in the sense of formal, binding, public and specific norms related to persons with mental disorders. However, due weight ought to be given to the legal tradition of a jurisdiction to appreciate the actual implications of mental health law provisions. Informal practices or usages concerning ways to deal with persons with mental disorders have also developed; they may have appeared in lieu of, in addition to or in opposition to formal laws.

1.2 Law and Mental Health: The Tradition

Perhaps to a significantly larger degree than elsewhere, mental health care is especially dependent on and affected by law. Social interactions of persons with mental disorders have traditionally raised two levels of concerns which appear to be the foundations for the development of early mental health norms.

The first level is directed at the individual affected in the first place, i.e., the person with a mental disorder. The very nature of a mental disorder will often make persons affected by them vulnerable in their dealings with society. This vulnerability typically affects persons with mental disorders in terms of decision-making and behaviour regarding their own health and safety and in terms of management of their property. Accordingly, measures have been designed to protect persons with mental disorders against themselves by removing a portion of their decision-making and management authority and by conferring it on someone else, to act as a "best friend." Typically, these measures have included involuntary hospitalization procedures, substituted consent to treatment procedures and the appointment of a legal guardian to manage property.

The second level of concern has been directed to family, neighbours, friends and other third parties in society at large who interact with persons with mental disorders. In a significant number of instances, actions or omissions by persons with mental disorders may affect others to the point that their health and safety may be jeopardized. This concern has justified lawmakers to adopt measures allowing designated authorities to limit the autonomy of persons with mental disorders felt to present a danger to the health and safety of others. Traditionally, this has been achieved through mandatory (involuntary) hospitalization.

Until recently, the purposes and patterns of legal instruments governing mental health have remained similar. They tended to be conceived as tools allowing societies to *react* to disturbing or unusual behaviours of persons with mental disorders (Curran and Harding 1978). Both levels of concern outlined above have traditionally rationalized and justified involuntary hospitalization. It is no surprise then that mental health laws have traditionally focused almost exclusively on this topic in the vast majority of countries in which formal enactments exist.

1.3 Modern Concerns and Trends

With the development of modern treatments for mental health disorders, the last two decades have witnessed a substantial shift in the pattern of some mental health laws. Many jurisdictions have adopted proactive legal provisions allowing or providing in some details for a new approach to mental health care, in particular through community-based mental health care.

In virtually all societies, persons with mental disorders are vulnerable in three ways in particular which now appear as modern justifications for mental health law:

1. They are ill and likely in need of treatment.
2. Their ability to understand, reason and decide may (or may not) be affected.
3. They tend to be stigmatized.

The need for their protection as a vulnerable group increases in contexts of civil disturbances and armed conflicts.

Justifications are also found with other groups. The families of persons with mental disorders typically bear an important portion of the burden associated with mental illness. Mental health professionals (e.g., psychiatrists, psychologists, nurses, occupational therapists, social workers), health care institutions and law enforcement authorities (e.g., the judiciary, the police) are all bound to face delicate situations involving persons with mental disorders. Ministries of health face the challenge of coordinating the above. The general public is entitled to be protected against threatening behaviour by persons affected by a mental disorder.

Formal rules governing these dimensions are warranted so that the rights and duties of all are known, rational, clear and predictable. Care for persons with mental disorders governed only by informal practices runs the risk of resulting in arbitrary decision-making and unbridled use of discretion.

1.4 International Norms

The international human rights movement set by the ratification of the Declaration of Human Rights in 1946 has had an impact on the development of specific international norms affecting the field of mental health. Three independent axes, explored below, have been of particular influence on national laws:

1. The adoption of reference principles by the United Nations
2. Rulings handed down under the European Convention for the Protection of Human Rights and Fundamental Freedoms
3. Norms put forward by non-governmental organizations (NGOs)

The first axis relates to UN norms. A number of international instruments affecting persons with mental disorders have been developed, including the following (United Nations 1994):

- *Declaration on the Rights of Mentally Retarded Persons*, United Nations General Assembly (UNGA) resolution 2856 (XXVI) of 20 December 1971
- *Declaration on the Rights of Disabled Persons*, UNGA resolution 3447 (XXX) of 9 December 1975
- *Principles for the Protection of Persons with Mental Illness and for the Improvement of Mental Health Care*, UNGA resolution 46/119 of 17 December 1991
- *Standard Rules on the Equalization of Opportunities for Persons with Disabilities*, UNGA resolution 48/96 of 20 December 1993

Three of these, described below, are indicative of current trends.

First, the UNGA unanimously adopted, in 1991, a set of 25 *Principles for the Protection of Persons with Mental Illness and for the Improvement of Mental Health Care*. Although, as mere principles, they are not formally binding on UN Member States, they are nevertheless seen as the main (and indeed the only) truly international set of modern principles on the rights of persons with mental disorders and the provisions of mental health care. Criticisms have been directed at these principles, essentially claiming that they are insufficient in their scope and poorly framed (Gendreau 1996; Rosenthal and Rubenstein 1993).

As part of its mandate to direct international health matters, the World Health Organization (WHO) has developed two documents to help disseminate the UN principles: a detailed set of questions aimed at delineating the potential scope of the UN principles (WHO 1995), and a brief working document, entitled *Mental Health Care Law: Ten Basic Principles* (WHO 1996), aimed to define, break down and help in the practical implementation of a core of ten very fundamental items drawn from the UN principles.

Secondly, stemming from the UN Decade of Disabled Persons (1983–1992) are the *Standard Rules on the Equalization of Opportunities for Persons with Disabilities*, a set of 22 proactive rules with annotations aimed at ensuring that persons with disabilities (including mental and intellectual) may exercise the same rights and obligations as others. Its stress on implementation in practice is reflected in Rule 15, entitled "legislation", which reads as follows: "States have the responsibility to create the legal bases for measures to achieve the objectives of full participation and equality for persons with disabilities."

Thirdly, the *Declaration of Caracas*[1] was adopted by a variety of legislators, associations, NGOs, mental health professionals, public health authorities and jurists at a conference on the Restructuring of Psychiatric Care in Latin America held in Caracas (Venezuela), 11–14 November 1990 under the auspices of the Pan American Health Organization. Innovatively, it calls upon various public institutions to support the restructuring of psychiatric care in accordance with six fundamental principles resulting from the premise that "[hospital-based] conventional psychiatric services do not allow for the attainment of objectives contained in a community-based care, that is decentralized, participatory, integrated, continuing and preventive." Importantly, in addition to declaring that the restructuring of psychiatric care should be performed on the basis of primary health care, this Declaration calls for any required redrafting of national legislation to safeguard human and civil rights of mental patients and to organize mental health services so as to guarantee the enforcement of these rights.

The second axis concerns case-law under the European Convention for the Protection of Human Rights and Fundamental Freedoms, which entered into force in 1953. Through this Convention, the European Court of Human Rights (ECHR) is given the authority to render rulings binding on States which have ratified it. The court has rendered a considerable number of opinions affecting persons with mental disorders,

[1] International Digest of Health Legislation 42: 336–338.

thereby contributing to influencing the trends of mental health law in Europe and beyond.

Revealingly, in its first ruling on the rights of persons with mental disorders, the *Winterwerp* case,[2] the Court opened the way to its numerous subsequent judgements on mental health, holding that confinement ("detention", under the meaning of the Convention) requires a government to show evidence of three items, known as the "Winterwerp requirements":

(1) "a true mental disorder has been established by objective medical expertise; and
(2) the mental disorder [is] of a kind and degree warranting compulsory confinement; and for any prolongation of detention the government must also be able to show that:
(3) the continued confinement is based on the persistence of the disorder" (Wachenfeld 1992).

As a result, confinement on account of a mental disorder in 40 European countries must be rationally established and documented, must be of a sufficient degree and must be persistent to justify deprivation of liberty.

The third axis consists of guidelines proposed by international NGOs. As an international consensus of interest groups, these guidelines provide useful references for lawmakers. Notable among the norms proposed are two documents: the Declaration of Madrid,[3] adopted by the World Psychiatric Association (WPA) in 1996, as an illustration of ethical guidelines adopted by a professional association and the Declaration of Caracas, to which several organizations contributed.

The Declaration of Madrid, adopted by the General Assembly of the WPA on 25 August 1996, reflects the perspective of the psychiatrists that the WPA represents. It supersedes the Declaration of Hawaii adopted in 1977 and updated in Vienna in 1983. It consists of seven main principles, is preceded by a preamble introducing the principles and is supplemented by five so-called Guidelines Concerning Specific Situations. These seven principles essentially do the following:

1. Delineate the purpose of psychiatry
2. Create a duty to pursue continuing education
3. Insist that patients should be made partners in the therapeutic process
4. Set out the exceptional circumstances under which treatment can be imposed on a patient against his or her will
5. Make it a duty to inform a patient if an intervention is conducted at the request of a third party
6. Delineate the duty to keep information disclosed in confidence
7. Spell out standards for research

The five guidelines governing specific situations concern euthanasia, torture, death penalty, selection of sex and organ transplantation.

2 Trends

2.1 Data and Methodology

The results presented here are largely drawn from a study conducted by the WHO from 1992 to 1995, the methodology of which is presented in the study report (WHO 1999). The study was based on a selected sample of 45 countries, several of which were selected for the intrinsic interest of their legal instruments. The study focused on specific legislation related to mental health care, as opposed to mental health in general. Topics falling outside this scope include, inter alia, criminal law and guardianship law.

Despite the bulk of international reference principles in the field, it appeared that, although not formally documented, a large proportion of competent jurisdictions worldwide operate without formal and binding norms governing mental health matters. Customary practices have likely developed in most of these jurisdictions, although little is known on these.

2.2 Terminology

The terminology found in national mental health legislation is an important variable to consider to assess world trends. In particular, it is indicative of the generation of laws from which a legal instrument was inspired and of the level of concern of lawmakers for contemporary concepts and schools of thoughts. One illustration is the "politically correct" movement which has recently emanated from the United States. Proponents of this movement advocate that expressions which, they feel, unduly label members of groups (e.g., handicapped, mentally ill) should be replaced by neutral and person-centred expressions (e.g., person with disability, person with mental disorder).

The consideration of terminology is inherently limited by an important dimension: the language. For the study, the consideration of terms was based on an English translation of terms found in legislation. Some

[2]Winterwerp Judgement, ECHR, Ser. A, No. 33, 24 October 1979.
[3]International Digest of Health Legislation 48: 240–241.

translations were official or formal translations, while other translations were unofficial private translations; in some cases, translations were done specifically for the study. While translators are presumed to have selected the English terms most suitable to translate the original expression, one cannot discount the risk of approximative or inaccurate translations. Moreover, in some cases, a concept in the language of origin may have no exact English equivalent. The present findings on terminology are therefore to be considered with the language caveat.

Consideration of the terminology used in mental health laws has revealed that the majority of countries considered have incorporated updated terms to refer to "persons with mental disorders" and to refer to "mental health care setting." With regard to the first expression, one finds "person with mental disorder", "person suffering from mental disorders", "mentally ill person", "mentally disordered", "patient" and "mental patient"; also found are "lunatic" and "person of unsound mind", but these are isolated examples. With regard to "mental health care settings", the following expressions were found: "psychiatric hospital", "hospital", "mental health establishment", "mental hospital", "establishment", "psychiatric facility", "facility" and "institutions"; in one isolated case, "asylum" was found.

The use of terms largely appeared to be in keeping with the age of legal instruments, i.e., antiquated terms were found mostly in older legislation. In conclusion, lawmakers seem, overall, to have given weight to the terminology aspect when updating legislation.

2.3 Age of Legislation

At a global level, the year of enactment of mental health law instruments is indicative of the magnitude of change. Although this indicator does not relate to the content of legal instruments, it provides some insight into the overall level of preoccupation of a jurisdiction with persons with mental disorders. Table 1 classifies the jurisdictions considered by the study according to the year in which the most recent significant legal instrument in the field of mental health that could be identified was enacted. It shows that, of the 45 countries reviewed, 21 have adopted significant legislation since 1978 (year of publication of previous WHO study), 12 countries apply pre-1978 legislation and 12 appear to apply informal systems.

Table 1 allows us to visualize three important features. The first is that a large proportion of jurisdictions included in the study (about 70%) have mental health legislation adopted after 1975, a year appears to be a cut-off point. Although of interest, this proportion may not, however, be representative of world trends. In fact, the data gathered for the review seem to suggest, on the contrary, that the majority of countries have either no formal mental health law or old (i.e., pre-1975) mental health laws.

The second feature is the association between the length of time a legal instrument has been in force and the level of resources available in the country in

Table 1. Period of significant mental health enactments

Period	Jurisdiction
1990–	Australia (Victoria), Brazil (Rio Grande do Sul), Canada (Quebec), Canada (Ontario), China[a], Costa Rica, Finland, France, New Zealand[b], The Netherlands[c], Tunisia, UK (England and Wales)
1985–1989	Australia (Queensland), Australia (Victoria), Finland, India (federal), Japan, Kenya, USA (IN)
1980–1985	Argentina (Capital), Barbados, Denmark, Germany (Bavaria), Ireland, Lebanon, Norway, Sweden, UK (England and Wales)
1975–1979	Italy, Norway, Senegal, Swaziland, Trinidad and Tobago, USA (MA), Zimbabwe
1970–1974	Ghana
1965–1969	Romania, Syrian Arab Republic, Fiji
1960–1964	–
1955–1959	Nigeria (Lagos), Tanzania
1950–1954	Malaysia
1930–1949	Brazil (federal)
1900–1929	Pakistan
Pre-1900	Benin[d], Sri Lanka
No enactment identified	Argentina (federal), Australia (federal), Canada (federal), Colombia[e], Ecuador[e], Ethiopia[e], Germany (federal), India (Karnataka), Jordan[e], Mozambique[e], Papua New Guinea[e], People's Democratic Republic of Korea[e], Thailand[e], USA (federal), Vietnam[e], Yemen[e]

[a] *Draft Mental Health Act* of September 1990 (Ninth Revision). In view of the concrete influence of this draft in several Chinese provinces, the principles of this draft have been considered in this publication.

[b] *Mental Health (Compulsory Assessment and Treatment) Act 1992*, entered into force on 1 November 1992. Review is of the *Mental Health (Compulsory Assessment and Treatment) Act 1989*.

[c] 1992 act was at draft stage at time of review.

[d] Law of 30 June 1838 adopted under French rule and reportedly inoperative; presumably Benin operates under an informal system.

[e] Countries presumably operating under informal systems.

question. Not surprisingly, older legal instruments are largely found in low-income countries.

Thirdly, an important proportion of the countries for which no formal legal instrument was identified seem to share socio-economic difficulties, suggesting an association between the absence of legal instruments dealing with mental health and the socio-economic context. If the adoption of a legal instrument or provisions dealing with mental health is indicative of the level of social priority assigned to mental health, it appears that countries in difficult socio-economic contexts give lower priority to mental health than other countries.

2.4 Scope of Law

Table 2 classifies countries in light of the scope embraced by their mental health law scheme. Considered for the purpose of this classification are the number and relative importance of the themes covered by policy statements and legal instruments, the substantive rights granted and the procedural safeguards enshrined in legislation. Among the relevant themes considered are rules governing hospitalization, rules governing treatment, rules governing guardianship (considered for this purpose only), community-based and integration principles, social support measures and other relevant themes.

Table 2 illustrates that the number of areas of concern addressed in legal instruments is also associated with the relative level of available resources in a country. The classification reveals that detailed legislative schemes are found almost exclusively in high-income countries. Moderately detailed legislative schemes are found in a mix of high-income and low-income countries, while basic legislative schemes are found in low-income countries.

Table 2. Scope of mental health law schemes

Scope	Jurisdiction
Detailed legal schemes	Australia (Victoria), Brazil (Rio Grande do Sul), China, Finland, France, Ireland, Japan, Norway, Senegal, UK (England and Wales), USA (MA), USA (IN)
Moderately detailed legal schemes	Argentina (capital), Australia (Queensland), Barbados, Canada (Ontario), Canada (Quebec), Costa Rica, Germany (Bavaria), Ghana, India, Italy, Kenya, Lebanon, New Zealand, Romania, The Netherlands, Tunisia
Basic legal schemes	Argentina (federal), Brazil (federal), Fiji, Nigeria (Lagos), Pakistan, Sri Lanka, Swaziland, Syrian Arab Republic, Trinidad and Tobago, Zimbabwe

Table 3. Decentralization of services reflected by mental health laws

Levels	Jurisdiction
High	Brazil (Rio Grande do Sul), Finland, Italy, UK (England and Wales)
Moderate	Australia (Queensland), Australia (Victoria), Canada (Ontario), Canada (Quebec), China, Costa Rica, France, Germany (Bavaria), India, Ireland, Japan, Lebanon, The Netherlands, New Zealand, Norway, Romania, Tunisia, USA (MA)
Low	Argentina (capital), Argentina (federal), Barbados, Brazil (federal), Ghana, Kenya, Nigeria (Lagos), Pakistan, Senegal, Sri Lanka, Swaziland, Syrian Arab Republic, Trinidad and Tobago, USA (IN), Zimbabwe
Insufficient data	Brazil (federal), Denmark, Fiji, Malaysia, Sweden, Tanzania

2.5 Decentralization of Mental Health Services

Decentralization has been identified as a leading theme of policy and social change over the last two decades. As a result, it was found relevant to assess the overall level of decentralization evidenced by legislation. Table 3 classifies countries in one of three categories representing the level of decentralization in administration and mental health care delivery that is reflected by the mental health law provisions. Variables considered for the purpose of this classification include: decentralization of powers, decentralization of implementation, composition and authority of decision-making authority, authority granted to personal representatives and other relevant variables.

Table 3 shows that overall decentralization efforts, if any, are not reflected very significantly in legal instruments. Only four countries (representing 12% of the jurisdictions reviewed) were found to have a mental health law incorporating a high level of decentralization.

Some level of decentralization can be achieved through private initiatives. For instance, in Quebec (Canada), a law passed in 1992 reforming the health care system stipulates conditions for private "foster family" initiatives eligible to some government funding.

2.6 Levels of Legislative Intervention

The legislative level of the lawmaker (i.e., federal, provincial, state, *Land*, canton, etc.) in charge of

adopting legal instruments in the field of mental health varies from country to country. In some countries, more than one level of government has the authority to adopt norms affecting mental health. In others, full jurisdiction to adopt legal norms in this field is vested in only one level.

Table 4 classifies countries according to the legislative level of the main mental health lawmaker(s). This account is limited to the legal instruments on which reporters were asked to concentrate their work, i.e., leading legal instruments governing mental health, to the exclusion of instruments only incidental to mental health. In a significant proportion of countries, mental health care is regulated by lawmakers of two levels of governments.

The type of powers granted to the various orders of government to adopt legal instruments is also indicative of the level of decentralization of mental health services. Some policy and legislative aspects may be best handled at federal or confederal levels, such as interstate cooperative efforts, the provision of financial support contingent upon compliance with specified basic standards, financial support to research and training, rules governing interstate issues and others. Such may also be the case for the redistribution of funds or the provision of services to regions most in need. However, norms adopted at local levels, which will tend to reflect the specific needs of the population, carry the best potential of being actually operationalized and enforced. This is true, in particular, of norms governing the delivery of mental health care or setting legal guarantees concerning involuntary hospitalization and treatment.

Interestingly, in a few jurisdictions, local authorities (as opposed to federal or federated authorities) are reportedly granted significant powers to adopt rules in the field of mental health care; these include China, Costa Rica, Finland, Italy and the United Kingdom (England and Wales).

Table 4. Level of main mental health lawmakers

Level	Jurisdiction
Unique, federal or confederal	Argentina (federal), Barbados, Brazil (federal), China, Costa Rica, Fiji, Finland, France, Ghana, Germany (federal), India, Italy, Japan, Kenya, Lebanon, Malaysia, Norway, The Netherlands, New Zealand, Norway, Pakistan, Romania, Senegal, Sri Lanka, Swaziland, Syrian Arab Republic, Trinidad and Tobago, Tunisia, Zimbabwe
State, provincial, *Land* or regional	Argentina (capital), Australia (Queensland), Australia (Victoria), Brazil (Rio Grande do Sul), Canada (Ontario), Canada (Quebec), China (Central), China (Shanxi), Germany (Bavaria), Ireland, Italy, Nigeria (Lagos), UK (England and Wales), USA (MA), USA (IN)

2.7 Increasing Differentiation Between Treatment and Hospitalization

Traditionally, provisions governing involuntary hospitalization were typically interpreted as including the authority to treat a person with a mental disorder even against the patient's will. It went without saying that a hospital stay was meant to be therapeutic. In most instances, no specific provision made this explicit.

The single most important legal aspect to notice with regard to treatment is the increasing distinction made by lawmakers and judges between involuntary hospitalization (deprivation of liberty) and involuntary treatment (interference with bodily/mental integrity). In this context, "treatment" encompasses any type of interference with the patient's bodily or mental integrity, such as drug treatment, electroconvulsive therapy and other types of interference. Two movements are noticeable which, although in opposition, both illustrate this differentiation.

In the first, found in a number of jurisdictions, the authority for the involuntarily hospitalization of persons with mental disorders does not extend to the authority to treat that person. Measures carried out to limit the autonomy of an aggressive patient in a crisis (emergency) situation, such as the use of physical or chemical constraints, are typically excluded from or constitute an exception to this definition of treatment.

An increased focus on human rights enforcement in the field of mental health care in recent years may explain the emergence of this new dichotomy. It no doubt challenges the traditional assertion that all persons labelled as mentally disordered are unable to understand, appreciate and decide for their own good. It seems to reflect an increasing acceptance of the principle that, while persons with mental disorders may occasionally be in a situation where they will be unable to exercise their right to consent or refuse treatment, it will not systematically be so, hence the conclusion that persons with mental disorders are not automatically incompetent for the purpose of deciding about their treatment.

The law in force in the United Kingdom illustrates this movement. Another illustration is the law in Canada (Quebec) where, under the Civil Code, a patient's wish not to receive treatment must be respected if care is categorically refused, except if purely hygienic or emergency care is involved.

The second manifestation of the increasing differentiation between hospitalization and treatment, quite

Table 5. Distinction between treatment and hospitalization

Distinction	Jurisdiction
Specific distinction made	Australia (Victoria), Canada (Ontario), Canada (Quebec), China, Germany (Bavaria), Italy, Japan, New Zealand, Norway, Romania, UK (England and Wales), USA (MA), USA (IN)
No specific distinction made	Australia (Queensland), Argentina (federal, capital), Barbados, Brazil (federal, Rio Grande do Sul), Costa Rica, Fiji, Finland, France, Ghana, India, Ireland, Kenya, Lebanon, Malaysia, The Netherlands, Nigeria (Lagos), Pakistan, Senegal, Sri Lanka, Swaziland, Trinidad and Tobago, Tunisia, Zimbabwe

in contrast to the previous one, is best illustrated by the legal scheme in force in Italy. Starting with the reform brought about by the 1978 Italian act, treatment, as opposed to hospitalization, is the only mandatory measure that is to be imposed on a patient found to fit legal criteria. In short, apart from exceptional cases, no hospitalization is to take place.

The extent of the growing treatment/hospitalization dichotomy is made explicit in Table 5, which highlights countries where this distinction has been specified in mental health legislation. Overall, nearly 30% of the formal mental health laws reviewed were found to carry this dichotomy.

2.8 Integration of Individuals and Legal Provisions

The social integration of persons with mental disorders is widely advocated. The idea is to allow persons with mental disorders to be integrated into the mainstream of life in society to the extent allowed by their health status, as opposed to having them taken care of *intra muros*, in closed institutions. Similarly, it is advocated that one way to remove the stigma traditionally attached to mental health is to incorporate the legal provisions addressing mental health issues into the mainstream of the legal system, i.e., by having legal provisions on mental health not singled out in a "mental health act" (Campbell 1994). It is useful to examine how these two facets of integration are addressed in current legal instruments on mental health.

Relatively few countries seem that have adopted legal instruments or provisions aimed strictly at integrating persons with mental disorders into the mainstream. Integration efforts in the form of community care laws were found in three of the seven countries visited during the study. Traces of efforts made towards the integration of mental health care into general health care are found in at least some legal instruments. In some instances, broad-based legal instruments have been adopted to foster the integration both of persons with physical disabilities and of those with mental disabilities. In at least a few cases, efforts to provide community-based care have been undertaken through national policy papers.

Efforts to integrate or reintegrate persons with mental disorders into society are also evidenced by the compatibility of standard mental health legislation with an integration strategy. In order to evaluate the impact of current mental health laws, an integration compatibility indicator was developed. The assessment takes into consideration features of mental health laws found to either foster or hinder the shift towards social integration of persons with mental disorders, such as whether the discharge procedure is burdensome, whether efforts or openings towards community care have been incorporated into the law and whether the least restrictive alternative principle has found its way into legislation. The results of this compatibility assessment are reported in Table 6.

A total of 11 jurisdictions were found to incorporate legal features fostering the integration of persons with mental disorders in society. The largest group of countries is found in a category described as including countries with mental health laws "compatible" with

Table 6. Compatibility of mental health laws with an integration policy

Compatibility level	Jurisdiction
Fostering (incorporates features to foster integration)	Brazil (Rio Grande do Sul), China, Costa Rica, Finland, France, Italy, Norway, Senegal, UK (England and Wales), USA (MA), USA (IN)
Compatible (incorporates features which allow integration)	Argentina (capital), Argentina (federal), Australia (Queensland), Australia (Victoria), Brazil (Federal), Canada (Ontario), Canada (Quebec), France, Ghana, India, Japan, Lebanon, New Zealand, The Netherlands, Trinidad and Tobago, Tunisia
Neutral (incorporates no features fostering or hindering integration)	Barbados, Germany (Bavaria), Ireland, Romania, Syrian Arab Republic
Incompatible (incorporates features hindering integration)	Fiji, Kenya, Nigeria (Lagos), Pakistan, Sri Lanka, Swaziland
Others (insufficient data)	Zimbabwe

reintegration. These first two categories account for about 71% of formal mental health laws reviewed, which is indicative of a significant policy trend. Few countries were found to have neutral legislation. Six countries have laws which incorporate features found to be incompatible with integration. The latter group is composed of countries with older laws.

The emergence of treatment arrangements allowing patients to live outside of mental hospital settings is another notable manifestation of the integration policy. Different modi operandi exist.

Firstly, it was found that the legislation of some countries authorizes the responsible authority to allow a patient to be discharged from hospital contingent on continued compliance with treatment requirements. Discharge may be subject to a time limit and renewals (e.g., in France, the law stipulates a 3-month maximum). In some countries, this is known as "supervised discharge". Typically, the patient's discharged status will be maintained as long the patient complies with the prescribed treatment, such as a monthly injection of long-acting neuroleptic medication.

Secondly, so-called mandatory community treatment procedures have been incorporated into legislation in a number of countries. Under this type of scheme, persons with mental disorders are to be treated in the community in which they live and maintain no link with standard hospital settings (e.g., New Zealand).

Obviously, arrangements for involuntary treatment in the community carries tremendous potential for least restrictive treatments. In contrast, they also carry some risks of human rights abuse. Special attention should be devoted to monitoring their operation. Their duration and a patient's right to regain full autonomy are critical issues for consideration. Table 7 lists jurisdictions (about 18% of the formal health laws reviewed) which have incorporated one type or another of these community-life arrangements into their laws.

It seems that little effort has been made to integrate mental health provisions in the mainstream of the body of law. This may be due to the fact that, in accordance with the protocol and research constraints, the current sample focused predominantly on mental health law provisions which are part of distinct "mental health" legal instruments, easier to identify and more accessible.

Lawmakers in the majority of countries selected for the study have adopted distinct mental health acts. In some cases, general constitutional provisions on health or liberty have been identified as the foundation of rules derived from it. Although not a country formally included in the present study, Spain is reported to have incorporated a large portion of the legal provisions addressing mental health issues into its Civil Code. In Quebec (Canada), key provisions governing the treatment and confinement of persons with mental disorders have been incorporated into the Civil Code in sections dealing with the integrity of these individuals, while a mental health act stipulates procedural guarantees in more details.

Table 7. Specific integration mechanism

Mechanism	Jurisdiction
Legal integration mechanism	Australia (Queensland)[a], Australia (Victoria), Canada[b], Canada (Ontario)[a], France, Italy, Japan[a], Lebanon, New Zealand, Norway, UK (England and Wales)
No legal integration mechanism documented	Argentina (federal, capital), Brazil (federal, Rio Grande do Sul), Canada (Quebec), China, Costa Rica, Fiji, Finland, Germany (Bavaria), Ghana, Ireland, Kenya, Malaysia, The Netherlands, Nigeria (Lagos), Norway, Pakistan, Romania, Senegal, Sri Lanka, Syrian Arab Republic, Trinidad and Tobago, Tunisia, USA (MA[a], IN), Zimbabwe

[a] A mechanism is reported but not documented.
[b] *1987 Model Act* of the Uniform Law Conference of Canada.

2.9 Legislative Features Encouraging Voluntary Care

It is widely acknowledged that health care of any nature yields the best results in individuals who endorse and support their treatment. This assertion has enhanced significance in the field of mental health care which, by nature, aims at treating disorders which may affect one's capacity to adopt a course of action. There is therefore a compelling interest to examine provisions dealing with care provided on a voluntary basis.

Mental health laws which govern voluntary hospitalization in an explicit manner are likely to entail an increased use of this mode of admission. In addition, this carries a potential of decreasing the use of involuntary hospitalization procedure in cases which do not warrant this mode of admission and hence do away with a potential barrier for accessing care. As a result, the level of incorporation of explicit voluntary admission procedures in mental health legislation is indicative of the tendency to prefer voluntary over involuntary admission whenever feasible.

The level of availability of voluntary care is an indication of accessibility to care and treatment in the least restrictive context. It also provides an additional instrument to determine whether integration with

Table 8. Explicit voluntary admission procedures

	Jurisdiction
Explicit voluntary admission procedure documented	Argentina (capital), Australia (Queensland), Australia (Victoria), Barbados, Brazil (federal), Canada (Quebec), Canada (Ontario), China, Costa Rica, Fiji, Finland, France, Germany (Bavaria), Ghana, India, Ireland, Italy, Japan, Kenya, Lebanon, Norway, Senegal, Swaziland, Romania, Sri Lanka, The Netherlands, Trinidad and Tobago, Tunisia, UK (England and Wales), USA (MA), USA (IN), Zimbabwe
No explicit voluntary admission procedure documented	Argentina (federal), New Zealand, Nigeria (Lagos), Pakistan, Syrian Arab Republic
Others (insufficient data)	Brazil (Rio Grande do Sul), Malaysia

Table 9. Grounds for involuntary hospitalizations

Criteria	Jurisdiction
Very strict	Australia (Victoria), Canada (Ontario), Finland, Ireland, Italy, The Netherlands, Norway, Tunisia, UK (England and Wales), USA (MA)
Strict	Canada (Quebec), France, Germany (Bavaria), India, Japan, New Zealand, Syrian Arab Republic, USA (IN)
Broad or none	Argentina (federal), Australia (Queensland), Barbados, Brazil (federal), China, Costa Rica, Fiji, Ghana, Kenya, Lebanon, Nigeria (Lagos), Pakistan, Romania, Senegal, Sri Lanka, Swaziland, Trinidad and Tobago, Zimbabwe

general health care is impinged by the lack of procedures allowing voluntary admission and treatment, as for other types of disorders.

A classification of countries according to the availability of an explicit voluntary admission procedure under the documented national legislation appears in Table 8. It shows that most of the legislative schemes (over 85% of the laws reviewed) now provides for the voluntary admission of mental patients, an issue which, in past decades, has been a key preoccupation (Curran and Harding 1978). However, the meaning attached to the expression "voluntary admission" or its equivalent was found to be heterogeneous, ranging from rules requiring explicit expressions of consent to others allowing for voluntary non-protesting individuals and cases of surrogate consent by third parties.

2.10
Grounds for Involuntary Hospitalization

Involuntary hospitalization is a common reaction to manifestations of mental disorders. Grounds specified in legislation as conditions for involuntary hospitalization tend to be related to the foreseeable dangerousness of individuals for self or others, although there is no consensus on a reliable method for foreseeing dangerousness with accuracy. Incidentally, several countries allow involuntary hospitalization regardless of an individual's potential dangerousness.

Table 9 classifies countries in one of three categories reflecting grounds for commitments, if any is specified in legislation. "Very strict" hospitalization criteria refer to the most demanding type of criteria, such as "expectation of imminent and serious danger", "prevent the chance of appreciable improvement in health status being lost" and other comparable conditions. "Strict" hospitalization criteria include intermediate types of conditions, such as "danger to self or other". "Broad" hospitalization criteria include conditions such as "disturbs the peace", "suffers from mental health disorder", "protection of others" and "patient's interest".

Table 9 shows that a large number of jurisdictions (about 50% of the formal mental health laws reviewed) rely on broad or on no criteria for involuntary hospitalization. This finding is disturbing in view of some of the 1991 UN mental health principles. The stricter criteria are found in systems derived from European law traditions, perhaps as a reflection of the influence of legal cultures.

2.11
Limited Periods of Hospitalization for Involuntary Care

It is advocated that the duration of involuntary mental health care (hospitalization or treatment) should be set to the shortest possible period to fulfil its raison d'être, i.e., should be as short as clinically required. It is further advocated that the length of involuntary hospitalization or treatment should be specified from the outset so that a reassessment of the appropriateness to maintain involuntary measures is performed upon termination of this period.

Typically, a time limitation of 90 days has been included in mental health legislation and has been found to provide a certain level of protection to persons with mental disorders involuntarily hospitalized. Hence, current laws have been examined from this perspective.

Table 10. Time limits for involuntary hospitalizations[a] and treatments[b]

Time limit	Involuntary hospitalizations	Involuntary treatments
90 days or less	Canada (Ontario), Finland, France, Lebanon, Swaziland, Tunisia, Zimbabwe	–
More than 90 days	Australia (Queensland), Australia (Victoria), Barbados, China, Germany (Bavaria), Ghana, Ireland, Kenya, New Zealand, Sri Lanka[c], The Netherlands, Trinidad and Tobago, UK (England and Wales), USA (MA)	New Zealand, Romania
No statutory time limit specified	Argentina (capital), Argentina (federal), Brazil (Federal), Canada (Quebec), Costa Rica, Fiji, India, Italy, Japan, Nigeria (Lagos), Norway, Pakistan, Romania, Senegal, Sri Lanka[d], Syrian Arab Republic, USA (IN)	Canada (Ontario), Canada (Quebec)[e], China, Germany (Bavaria), Italy, Japan, Norway[f], Pakistan

[a]Presumably extends to treatment in the absence of indications with specific regard to treatment.
[b]If addressed distinctively from hospitalization in legislation.
[c]"Hospitalization" status.
[d]"Detention" status.
[e]Courts tend to authorize treatment for a period of 1 year.
[f]One year.

Table 10 classifies countries in one of three categories to reflect the type of limitations, if any, imposed by legislation on the duration of long-term involuntary hospitalization and involuntary treatment. It indicates that nearly 45% of the formal mental health laws assessed carry no statutory time limit on involuntary hospitalization. Importantly, it also indicates that about 65% of countries whose laws make a distinction between hospitalization and treatment carry no statutory time limit to involuntary treatment.

2.12
Role of Health Care Providers

The involvement, at appropriate levels, of an enlarged range of mental health care providers – typically nurses, but also traditional healers, social workers and others – is an essential component of the primary health care approach to mental health. This is due to the fact that the delegation of some level of clinical authority to non-physicians carries the potential of increasing the availability of health care output at primary level and, as a result, of improving public health. In particular, low-income countries are likely to benefit greatly from this approach. In this context, the level of empowerment of non-physicians caregivers in mental health legislation may be indicative of a country's commitment to a primary health care approach. Further, the delegation of some managerial and decision-making authorities to non-physicians, such as career managers, jurists, clergymen and others, is indicative of an effort to balance the multiple dimensions of decision-making processes in the field of public health.

Table 11 classifies countries according to the level of authority conferred upon non-physicians in mental health legislation; this overall assessment considers both the types of powers and the number of different non-clinicians empowered. It shows that only a handful of jurisdictions (about 16% of the formal mental health laws reviewed) provide for the significant involvement of non-physician health care providers.

2.13
Emergency Care Procedures

Emergency procedures are essential to allow timely action in crisis situations. However, they carry the

Table 11. Involvement of non-physicians reflected by mental health laws

Level of authority	Jurisdiction
Significant	China, Italy, The Netherlands, Trinidad and Tobago, UK (England and Wales), USA (MA)
Moderate	Barbados, Brazil (Rio Grande do Sul), Canada (Ontario), Canada (Quebec), France, India, Ireland, Japan, Kenya, Lebanon, New Zealand, Norway, Senegal, Syrian Arab Republic, Tunisia
Low or none	Argentina (capital), Argentina (federal), Australia (Queensland), Australia (Victoria), Brazil (federal), Costa Rica, Germany (Bavaria), Fiji, Ghana, Nigeria (Lagos), Pakistan, Romania, Sri Lanka, Swaziland, Zimbabwe, USA (IN)

potential of serious human rights infringements. One risk lies in the systematic use of the exceptional rules applicable to emergency situations, which are typically less demanding. Another lies in repeated extensions of the emergency status so as to bypass the standard procedure. A third type of risk exists if no health care provider is involved in the emergency decision-making process.

Table 12 classifies countries according to the time limits imposed on the use of emergency procedures, whereas Table 13 classifies countries in terms of whether a medical assessment is explicitly made part of the emergency procedure process. These tables reveal that a large proportion of jurisdictions have incorporated time limit and medical assessment safeguards into their emergency care procedure.

Table 12. Time limits for emergency hospitalizations

Time limit	Jurisdiction
3 days or less	Australia (Queensland), Barbados[a], Canada (Ontario), Canada (Quebec), Fiji[b], France, Germany (Bavaria), India, Japan, New Zealand, Norway, Trinidad and Tobago, Tunisia, UK (England and Wales), USA (IN)
More than 3 days	Fiji[c], Ghana, Kenya, Nigeria (Lagos), Romania, Senegal, Sri Lanka, Swaziland, USA (MA)
No statutory time limit	Argentina (capital), Argentina (federal), Brazil (federal)

[a]Observational hospitalization.
[b]If determined by hospital.
[c]On magistrate's order.

Table 13. Medical assessment for emergency procedures

Assessment	Jurisdiction
Medical assessment required	Argentina (capital), Brazil (federal), Canada (Ontario), Canada (Quebec), Germany (Bavaria), Ghana, Japan, Nigeria (Lagos), Norway, Romania, Senegal, Swaziland, Trinidad and Tobago, Tunisia, UK (England and Wales)[a], USA (IN), USA (MA)
No medical assessment required	Australia (Queensland), Barbados[b], Fiji, France, India, Kenya, New Zealand, Sri Lanka

[a]In some but not in all cases.
[b]Observational hospitalization.

2.14 Overseers and Review Bodies

The need for a specialized authority in charge of monitoring the implementation and enforcement of legislative provisions is also largely advocated. Enforcement of norms specific to the mental health field is better supervised by a specialized body, as opposed to a court of law in charge of a host of other matters.

In this respect, three indicators have been considered:

1. Whether they are provided for in legislation (Table 14)
2. Whether they are composed of representatives of a variety of disciplines (Table 14)
3. What type of authority is conferred upon them and how the involvement of this authority can be triggered (Table 15)

Table 14 illustrates how widely in use overseers and review bodies have become; a significant proportion of jurisdictions rely on this safeguard (nearly 55% of the formal mental health laws reviewed). Further, it indicates that a large proportion of countries with overseers and review bodies (about 80% of the formal mental health laws with overseers or review body) require that they be composed of individuals with some diversity in backgrounds. These countries are mostly countries with recent enactments.

Table 15 classifies jurisdictions according to the types of authority conferred upon specialized overseers or review bodies and the mechanisms triggering their involvement. With regard to the type of authority, mandates conferred upon bodies have been classified into two categories: adjudicative or reporting in nature. Adjudicative mandates include the power to order an enforceable legal redress. For instance, it may include the power to order that a given patient be discharged or to order clinical or logistical changes. Although the nature of the legal redresses that adjudicative bodies can impose may vary significantly among jurisdictions, all are indicative of some level of decentralization of power and of independent human rights monitoring. Turning to reporting mandates, although they carry no authority to order a legal redress, they include the power to report and recommend courses of actions to a higher authority and, as such, may constitute an important safeguard.

With regard to the types of involvement, jurisdictions were classified into four categories: involvement upon the independent discretion of the body (power to investigate on their own); involvement at set periodical intervals; involvement upon the request of an interested party (such as a patient, a patient's parent or

Table 14. Overseers and review bodies

Pluridisciplinary	Not pluridisciplinary	Not specified
Australia (Queensland), Australia (Victoria), Brazil (Rio Grande do Sul), Canada (Ontario), Canada (Quebec), France, Ghana, India, Ireland, Japan, Kenya, Lebanon[a], New Zealand, Norway, Swaziland, The Netherlands, Trinidad and Tobago, Tunisia (bidisciplinary), UK (England and Wales)	Barbados, Finland, Nigeria (Lagos), Pakistan, Senegal, Sri Lanka	Argentina (federal, capital), Brazil (federal), China, Costa Rica, Fiji, Germany (Bavaria), Ireland, Italy, Malaysia, Romania, Swaziland, USA (IN, MA), Syrian Arab Republic, Zimbabwe

[a]Reportedly made explicit by special legislation unavailable to the study.

guardian or others); and involvement at the discretion of a state authority. This classification is indicative of the potential for compliance with legislative and human rights principles. Table 15 shows that about 27% of the formal mental health laws reviewed have granted an adjudicative/binding mandate upon an overseer or a review board, arguably the human rights monitoring mechanism which offers the highest protection.

2.15 Operationalization of Norms

The translation of the norms enshrined in mental health legislation from theory into practice has been identified as a prime area for concern. It was found that mental health laws are widely unimplemented and misinterpreted. Implementation failures may be the result of several different causes, two of which found to be of particular importance are explored below: that of the understanding of the norms by the participants in the health care system and that of the planning of the implementation of theory in practice.

Legislative Understanding

The unprecedented amount of new dilemmas which have appeared as a result of the evolution of clinical practice and of legal principles over the last decades have had an impact on the bulk and pattern of legal instruments in the field of mental health. As a result of advances in psychiatry and of increasingly demanding and complex legal standards, mental health laws have noticeably grown in size. It has become a challenge for law drafters to embody the substantive principles and procedural requirements involved in a comprehensive, yet intelligible format. Although most countries have been affected to some degree by this growth, mental health legislations have been designed in significantly different formats and size around the world. Legal traditions have of course also influenced this.

Table 15. Involvement and authority of overseers and review bodies

	Involvement by independent decision of body	Mandated periodical assessments	Involvement upon request of interested party	At discretion of State authority
Adjudicative/ binding mandate	Lebanon[a], The Netherlands	Australia (Victoria), Canada (Quebec), Norway, UK (England and Wales)	Australia (Queensland), Australia (Victoria), Canada (Ontario), Canada (Quebec), Ireland, New Zealand, Norway, Trinidad and Tobago, UK (England and Wales)	Brazil (Rio Grande do Sul)
Reporting/ recommendation mandate	Australia (Victoria), Canada (Ontario), France, Japan, New Zealand, Nigeria (Lagos), Sri Lanka, Swaziland	Australia (Victoria), France, India, Ireland, New Zealand, Pakistan, Senegal[b], Sri Lanka, Trinidad and Tobago, Tunisia	Australia (Victoria), France, Japan	Australia (Queensland), Australia (Victoria), Sri Lanka, Tunisia

[a]Reportedly made explicit by special legislation unavailable to the study.
[b]Available to patients admitted voluntarily only.

Two opposite trends seem to have emerged. In one of them, legal instruments have become very detailed as they are conceived to systematically address and regulate the multiple facets through various substantive provisions and procedural safeguards. This trend is the untamed result of the above-noted evolution. It is found mostly in countries influenced by the common law tradition, but also in other legal traditions. Good examples are Japan and in New Zealand. Legal instruments found in this category are reportedly widely perceived as complex sets of rules rather inaccessible to the lay person. They typically embody extensive legal definitions of terms and expressions used therein. Cross-references between sections are frequent. As a disadvantage, users' guides and visual representations are generally found to be required if the norms specified in this type of legislation are to be understood and followed by health care providers.

The second trend includes brief descriptive legal instruments which are only a few articles long. Legal instruments found in Italy and Tunisia exemplify this trend. In some cases (e.g., Spain), the provisions involved are no longer to be found in distinct acts, but have been integrated into the general body of health law or into a code, thus contributing to remove the stigma attached to a distinct group. This trend is found mostly in countries influenced by the civil law tradition. As a decentralization measure, operationalization efforts and regulation may have been delegated to local authorities, although this generally does not lead to overly detailed sets of legal instruments at a local level. Legal instruments found in this group tend to be less legalistic in their format and hence more accessible to non-jurists. This fosters operationalization and accurate enforcement. One criticism directed at the legal instruments in this group is their very variable level of precision, resulting in higher levels of discretion for the authorities in charge of their operationalization. Another is that the high level of abstraction may require complementary and incidental legal instruments. Some confusion as to the applicable norms may ensue.

As detailed and comprehensive as they may be, the substantive principles, procedural requirements and structures of authority specified in legislation have to be abided by if they are to mean anything. Unlike many other areas where non-jurists have the option and the time to defer to jurists to interpret legislative provisions, mental health legislation will often require immediate action and will seldom allow mental health care givers to obtain legal advice on the spot. They often have to be able to understand the legal mechanisms involved by themselves or they will resort to exceptional procedures (e.g., emergency procedures) or, more simply, to their common sense. Of course, training sessions and internal manuals specifying legal requirements, if they can be made available, will help. However, they may not cover the growing number of exceptional procedures. In the end, too complex or detailed mental health laws carry the potential of remaining significantly non-operationalized and unenforced, with entire portions systematically disregarded or ignored.

In order to depict world trends with regard to the issue of understanding, an indicator which evaluates the relative level to which legal instruments are accessible to non-jurists was developed. This indicator takes into consideration both the format of the instrument and its content. Although inevitably subjective in nature, the results of this assessment were found to be useful to identify tendencies. These results are reported in Table 16, which shows that a very significant number of jurisdictions (50% of the formal mental health laws reviewed) now apply legislative schemes in the field of mental health assessed as "complex"; these are mostly derived from the common law tradition.

Strategic Planning

It has been a recurring comment of national informants and experts that a leading cause explaining enforcement failures in mental health legislation lies in the way the entry into force of new legislative features is prepared and the way its continued enforcement is ensured. Typically, the conception of modern legal instruments in the field of mental health draws much energy and takes place over a process which may – and usually does – last several years. Task forces and/or commissions are set up. Attention is devoted, inter alia, to substantive rules, procedural guarantees, constitutionality and human rights as well as a variety of other aspects. In several countries, public participation also forms an

Table 16. Complexity of mental health laws

Complexity level	Jurisdiction
Complex	Australia (Queensland), Australia (Victoria), Barbados, Canada (Ontario), France, India, Ireland, Japan, Kenya, New Zealand, Pakistan, Sri Lanka, Trinidad and Tobago, UK (England and Wales), USA (IN)
Moderately complex	Canada (Quebec), China, Germany (Bavaria), Ghana, Nigeria (Lagos), Norway, Swaziland, The Netherlands
Simple	Brazil (Rio Grande do Sul), Italy, Senegal, Italy, Lebanon, Syrian Arab Republic, Tunisia
Insufficient data	Argentina (federal, capital), Brazil (federal), Costa Rica, Fiji, Finland, Malaysia, Norway, Romania, USA (MA), Zimbabwe

integral part of the process, contributing to enriching and prolonging it. Debates take place and consensus is reached or not. The results are increasingly detailed laws embodied in increasingly complex formats.

Yet, in comparison, a relatively insignificant amount of time and effort seems to be typically devoted to preparing the scene for the entry into force of a new legislative scheme or of amendments to any existing scheme. Although in some jurisdictions some level of preparation may be provided for the individuals involved in the operation of mental health laws prior to or at their inception, this appears to be exceptional and/or limited in scope. The few instances of preparatory efforts that are reported are typically limited to clarifying the more obscure or more problematic features of the legislative schemes. Little energy appears to be devoted to a number of facets, including informing people at various levels in the mental health care field about philosophical and practical changes in the law that should have an impact on their usual practices, giving due weight to the time variable, itemizing funds to prepare the operationalization (e.g., transition) as opposed to the application of the law and explaining the meaning of "the rule of law" (i.e., legal instruments are binding). In short, no global, coherent and rational strategic plan for the launching of mental health legislation appears to be generally followed despite increasingly detailed legislative features.

3 The Italian Experience

The Italian Public Law 180 on Mental Health, enacted in 1978, has become known worldwide for the radical shift in mental health care it envisaged: from institutional segregation and control to rehabilitation and reintegration into normal social life in the community. Influenced by leading psychiatrist Franco Basaglia and embodied in a brief 11-article format, it was essentially designed to do the following:

1. Phase out public mental hospitals by admitting no new patients
2. Move the treatment of persons with mental disorders to community-based mental health centres
3. Move any needed psychiatric hospitalization to general hospitals
4. Limit involuntary hospitalizations to cases where "(a) urgent intervention is required, (b) the patient refuses necessary treatment and (c) alternative community-based treatment is unavailable or not feasible"[4]

The implementation of Law 180 was delegated to regional governments.

Legal principles and mechanisms put forward in Law 180 towards these ends include the following:

- Setting out the principle that medical examinations and treatment shall, as far as possible, be voluntary (article 1 para. 1)
- Granting decision-making authority to the local mayor (article 1 para. 6)
- Setting out the principle that prevention, treatment and rehabilitation of patients shall normally be carried out by outpatient psychiatric departments and centres (article 6)

After two decades of operation, Italian Public Law 180 provides a good precedent for consideration by international observers. Four remarks on the wide-ranging reforms it has striven to bring about can be made:

1. Variations in the pace, extent and even existence of reforms are reportedly very high from one Italian region to another, with regions in the north generally more compliant and successful than regions in the south (Burti and Benson 1996). Explanations reported for this include difficult economic and political circumstances in the late 1970s and the early 1980s, difficulty in reallocating funds from mental hospitals to community services and severe cuts in social programmes during the 1980s. Clearly, the availability (or shortage) of funds and the level of commitment of local health authorities appear to be leading determinants of successes and failures in the regional implementation of Law 180.
2. Documented successes in several Italian provinces strongly support the theory advocated by Basaglia and his followers that a key to recovery and reintegration of mental patients lies in their empowerment.
3. Reported data suggests that the deinstitutionalization brought about by the reform in successful regions has not, as feared by some, entailed corresponding increases in the criminal justice system or in private mental health institutions (Barbato 1998).
4. On the finding by a Special Commission in 1989 that there were significant implementation failures a decade after its adoption, the Italian Senate recommended that an adequate budget be set aside for the *implementation* of Law 180, resulting in a national

[4]For an English translation of the Act, see International Digest of Health Legislation 30(1): 75–79.

plan passed by the Italian Parliament in 1994. Clearly, the Italian experience in translating a commendable yet ambitious mental health care reform into practice attests the need for early planning and budgeting in anticipation of the entry into force of a new mental health law.

4 The Situation in China

China has no nationwide legislation governing mental health matters. However, it cannot be said that the country operates through informal practices. There are two explanations for this:

1. Successive draft mental health acts have been considered by both the Ministry of Public Health (MPH) and the People's Congress for many years. The most recent version is the Tenth Draft Revision introduced in July 1991; although not formally adopted, informants report that it is widely observed.
2. Several provinces have adopted this Tenth Draft Revision (as submitted to national authorities as a report from a MPH/WHO National Workshop on Mental Health and Law, which took place in Chengdu, Sichuan, in October 1990) as their provincial law.

In addition to this draft, a number of laws relevant to the field of mental health have been adopted at both national and provincial levels. They include, inter alia, a *National Law on Experts' Assessment (Competence) of Mental Disorders* (1989), *Rules for the Assessment of Psychiatric Offenders* (1989), *Programme for Special Education for the Mentally Disordered* (includes mental retardation and other mental disorders) (1989), *Law on the Protection of the Disabled* (physical, sensorial and mental disability) (1990), *Programme for the Protection of the Disabled* (includes e.g., education, employment) (1991) and *Law on Maternal and Infant Health Care* 1994.

Since most of the Chinese still live in rural areas, the majority of mental health care is provided at primary health care level; psychiatric hospitals are available only in medium-sized and large cities, and psychiatric wards in general hospitals usually only exist in teaching hospitals. In view of the size of the population, the relative number of mental health workers is rather limited, making nursing aides a key player in the provision of mental health care.

Institutional mental health care in China does not seem to differ much from that found in other countries with a similar level of development. At community level, however, the persistence of the extended family, particularly in rural areas, favours the integration of people with chronic mental disorders into community and social life.

China's peculiar situation is further accentuated by two factors. The first is the inclusion of disabilities and handicaps due to mental disorders among all other disabilities (i.e., originating from physical and sensory problems). For historical reasons, persons with disability benefit from a series of laws and other measures for their protection.

The second factor is the importance given, in China, to the so-called second-generation human rights. Certain Western perceptions of China tend to be of generalized human rights violations, referring mostly to so-called first-generation human rights (i.e., civil and political rights). However, second-generation rights (i.e., social, economic and cultural rights) are largely abided by. These rights are important to persons with mental disorders. Chinese legislation referred to above is largely in pursuit of these rights, and site visits by the second author to three Chinese provinces (Yantai, Shanxi and Sichuan) have left little doubt that efforts are made to implement these provisions in practice. No signs of discrimination against people with mental disorders were found in the sites visited by the second author. In fact, people with mental disorders were generally treated with a respect similar to that demonstrated in relation to people with disabilities of other nature and the elderly.

It should be kept in mind that the predominance of the rural over urban population and of the extended family are changing quickly in China. This may have a heavy impact on the integration of people with mental disorders into community life.

5 Concluding Remarks

The reliance on binding norms protecting persons with mental disorders and facilitating their full participation in society as normal citizens has become the reference standard worldwide. Persons with mental disorders deserve the protection of the rule of law: "decisions should be made by application of known principles [. . .] without the intervention of discretion in their application" (Black et al. 1991).

Like other legislation, mental health legislation is a tool. To achieve the desired results, it must be handled with skill and care, especially when a vulnerable group is targeted. Most importantly, it must be used, i.e., its spirit and features must be implemented and specific provisions enforced in practice. It should not be left

lying in a statute book. Mental health legislation reforms which remain unimplemented turn into a formidable disservice to persons with mental disorders by creating a false sense of achievement.

Beyond legal provisions, the dignity of persons with mental disorders is also ours.

6 References[5]

Barbato A (1998). Psychiatry in transition: outcomes of mental health policy reform in Italy. Aust NZ J Psychiatry 32: 673–679

Black HC, Alibrandi MN, Connolly MJ (1991) Black's law dictionary. West, Wadsworth

Burti L, Benson PR (1996). Psychiatric reform in Italy: developments since 1978. Int J Law Psychiatry 19: 373–390

Campbell TD (1994) Mental health law: institutionalized discrimination. Aust NZ J Psychiatry 28: 554–559

**Curran WJ, Harding T (1978) The law and mental health: harmonizing objectives. WHO, Geneva

*Gendreau C (1996) Le droit du patient psychiatrique à consentir à un traitement: élaboration d'une norme internationale. University of Montreal, Montreal

Neugebauer R (1978) Treatment of the mentally ill in medieval and early modern England: a reappraisal. J Hist Behav Sci 14: 158–169

*Rosenthal E, Rubenstein LS (1993) International human rights advocacy under the Principle for the Protection of Persons with Mental Illness. Int J Law Psychiatry 16: 257–300

United Nations (1994) A compilation of international instruments, vol I (second part): universal Instruments. United Nations, New York

Wachenfeld M (1992) The human rights of the mentally ill in Europe under the European Convention on Human Rights. Nord J Int Law (special issue: Acta scandinavia juris gentium): 107–292

**WHO (1995) Guidelines for the promotion of human rights of persons with mental disorders. WHO/MNH/MND/95.4. WHO, Geneva

**WHO (1996) Mental health care law: ten basic principles. WHO/MNH/MND/96.9. WHO, Geneva

WHO (1999) Mental health care legislation: international assessment and basic principles. WHO/MNH/MBD/99.1. WHO, Geneva

[5]The reader is referred to two journals as general reference sources: the *International Digest of Health Legislation*, a periodical published by the WHO four times a year, and the *International Journal of Law and Psychiatry*, published by Elsevier/Pergamon several times a year.

CHAPTER

16

A.R. Felthous, H.L. Kröber, H. Saß

Forensic Evaluations for Civil and Criminal Competencies and Criminal Responsibility in German and Anglo-American Legal Systems

1 General Remarks

1.1 Two-Step Evaluation Process

A forensic evaluation to examine a person's mental state, where his or her mental state is of legal significance, should follow a two-step process. The first step determines whether a psychiatric illness or a serious mental disorder exists at all. Kurt Schneider (1948) formulated an approach in which the following two issues must be determined:

1. Whether a mental disorder exits and how the disorder is to be denoted according to psychiatric nomenclature
2. Proceeding from the clinical findings and diagnosis, how this disorder affects the mental capacity in question

This two-step process is basically equally valid for the assessment of competency to work, to earn a living, to practice a profession, to do business, to make a will, to consent, to be interrogated, to stand trial, to commit a crime (i.e., criminal responsibility), to drive a vehicle, and to serve as a witness.

1.2 Psychiatric Diagnosis

The first step consists of a comprehensive psychiatric evaluation that includes the medical history and the interpretation of available relevant medical, psychiatric, and legal documents to consider whether such materials produce evidence for mental illness. Based on the collected data, a psychiatric diagnosis is established that is consistent with the criteria and terminology of a currently valid classification system such as ICD-10 (World Health Organization 1992) or DSM-IV (American Psychiatric Association 1994). Forensic experts and lawyers must clearly understand that these diagnostic and classification manuals correspond to changing conventions and are therefore not final, permanent diagnostic authorities. Expert witnesses and lawyers should also appreciate that a psychiatric diagnosis per se is insufficient for direct testimony about impairment of competency.

1.3 Competency Assessment

The second step, i.e., application of the collected data in a particular case to the legal criteria for the competency to be addressed, is a separate and equally important task of the forensic psychiatrist as the collection of data. As a rule, unless there is reason to question the issue, any adult is presumed to be competent unless and until proven to be incompetent. Criteria for specific competencies are usually defined in concise legal wording, which in Germany is presented in the opinions of the highest court and in Anglo-American law can be articulated in either state or federal statute or appellate court opinion. The forensic psychiatrist must know how the absence of the relevant competency is legally defined. The psychiatrist's knowledge of the definition of competency will guide him or her in the specific, purposeful exploration and application of clinical knowledge about typical limitations in such pathological conditions. The nature and extent of such limitations should be clearly presented in the forensic report.

1.4 Principle of the Psychopathological Reference System

The principle of the psychopathological reference system (Saß 1985) is applied by comparing the evaluee's presenting symptomatology with the psychiatrist's extensive knowledge of psychopathological conditions. Assessment of mental phenomena is conducted with the background of biographical knowledge about the healthy and pathologically disturbed developmental stages of life, about the mechanisms of reaction to stressors, and about the influences of specific therapeutic efforts and the natural phases of maturation and aging. In contrast to a common misunderstanding in forensic discussions, the psychopathological focus should not limit the assessment to pathologically disturbed psychological phenomena alone, but rather the understanding of the person's psychopathology is clearly based on the psychiatrist's total knowledge of both pathological and atypical, but not abnormal mental phenomena in contrast to more or less pure mental health.

This chapter cannot replace any textbook of forensic psychiatry (Gunn and Taylor 1993; Nedopil 1996; Rosner 1994; Sadoff 1988; Venzlaff and Foerster 1994) or mental health law (Brakel et al. 1985; Perlin 1989; Slovenko 1973; Weiner and Wettstein 1993); it simply summarizes current guidelines for the forensic assessment of individuals whose competency in a specific area has come into question. The examining psychiatrist must be careful to order supplemental examination procedures when indicated to address specific questions and especially psychological testing to help clarify the diagnosis.

The following sections will discuss first civil competencies – legal competency, competency to sue, competency to make a will, competency to consent to

treatment, and related competencies – and then competencies in criminal law – competency to be interrogated/to confess, to stand trial, to commit a crime (i.e., criminal responsibility), and to be sentenced. Each competency will be explained from the perspective of the law and forensic practices in the Federal Republic of Germany followed by commentary of special relevance to forensic psychiatry in North America and England.

2 Civil Competencies

2.1 Germany

2.1.1 Legal Competency

In Germany, legal competency is the ability to autonomously bring about desired legal results through legal transactions or, more specifically, through legal documents and contracts. Legal competency is required in order to effectively conduct and conclude business actions. Children aged 6 or younger are not considered to be competent to enter contracts. Children aged 7 or older and adolescents up to the age of 18 have *limited legal competency*. To conduct legal acts, older children require the consent of their parents as legal representatives, provided that their parents do not themselves profit from the transaction. As will be discussed, testamentary capacity and competency to consent are basically special subtypes of legal competency.

2.1.2 General Legal Competency

Legal competency is presumed at majority. In Germany, legal competency is governed by the Sec. 104 *Bürgerliches Gesetzbuch* (BGB) (German Civil Code). According to the relevant statute, a person is legally incompetent if he or she (a) has not reached the age of 7 or (b) has a condition involving a pathological disturbance of mental processes that destroys free decisional capacity, provided the condition is not temporary in nature.

2.1.3 Temporary Conditions

For purely temporary conditions, e.g., temporary severe clouding of consciousness or loss of consciousness such as could occur with an epileptic attack, high fever, or an intoxicated state, the German Civil Code states the following: "A legal document that is proffered during a condition of unconsciousness or temporary disturbance in mental processes is therefore void" (Sec. 105(2) BGB). Once the condition has passed, the impairment in legal competency naturally no longer exists.

Individuals who are incompetent according to Sec. 104(2) BGB cannot participate in the business transaction themselves, but their legal representatives can act on their behalf. If it can be shown in retrospect that, at the time of the legal transaction, the individual concerned was legally incompetent due to a temporary pathological disturbance of mental functioning which eliminates voluntary decision-making capacity, any contracts concluded (e.g., rent, sale, loan contracts) can be voided. Competency to do business is presumed to exist until incompetency is proven.

2.1.4 Forensic Questions

Apart from appropriate treatment considerations, forensic issues concerning legal competency refer to two possible questions:

1. At the time of the legal transaction, did a temporary disturbance of mental functioning exist which eliminated the individual's voluntary decision-making capacity (Sec. 105(2) BGB)? Examples of such temporary disturbances include a partial complex seizure or a state of intoxication.
2. At the time of the legal transaction, did legal incompetency exist that was more than a brief pathological disturbance of mental functioning eliminating the person's voluntary decision-making capacity (i.e., according to Sec. 104(2) BGB)? This possibility usually arises as a result of a dementing process or a psychotic illness such as schizophrenia or manic-depressive illness, but not from an acute organic brain disorder. In principle, it could also apply to disorders such as personality disorders, abnormal reactions, and mental retardation, provided that these disorders are comparable in severity to a major mental illness.

2.1.5 Retrospective Assessment

Retrospective assessments to address legal competency should follow the same two-step process described at the beginning of this chapter. If a disorder did exist, the second step is then to determine whether it resulted in the elimination of the subject's ability to make autonomous decisions, i.e., the elimination of the "normal decision-making ability resulting from normal motivations" (Aschaffenburg 1906) or "normal decision-making ability through rational deliberation" (Federal Court decisions).

According to an opinion given by the *Bundesgerichtshof* (BGH, German Supreme Court of Justice) in 1918, individuals are considered incompetent to do business (according to Sec. 104(2) BGB) if their "deliberations and voluntary decisions are no longer based on a general opinion corresponding to an appreciation of external affairs and living circumstances, but rather are influenced through pathological feelings, pathological ideas and thoughts, or through influences by a third person of a lasting nature, (such) that they are in fact no longer free but follow the so-called contrary-to-rule actions without inhibition or restriction" (Diederichsen 1994). To address the question of whether the individual's ability to make autonomous decision has been impaired, the psychiatrist must retrospectively apply clinical knowledge about the person's volitional processes, decision-making, and behavior at the time of the business transaction.

Organic brain disorders which involve substantial impairment of the abilities to register and/or remember information (e.g., multinfarct or Alzheimer's dementia) can abolish legal competency by impairing requisite cognitive decisional foundations for the development of free choices or by pathologically altering the individual's emotional control. More or less purely affective disorders and organic brain disorders (e.g., with depressive or excitable moods) can impair volitional capacity and decision-making. This clearly also applies to abulia and conditions which can render the individual extraordinarily susceptible to influences such as can occur with some organic brain disorders, mental retardation, and schizophrenic illnesses.

2.2 United States

What is usually meant by civil competency in the United States is *competency to handle one's own affairs*. Competency to do business, though less frequent, occasionally has to be investigated by a forensic psychiatrist; basically, the criteria for competency to do business is the ability to understand the proceedings in which the individual is to become involved (Sadoff 1988). Much of what has already been discussed about forensic assessments would apply independent of political jurisdiction.

Criteria for competency to handle one's own affairs includes the capacity to know the extent of one's possessions and to handle one's affairs prudently (Sadoff 1988, p. 149). Depending on jurisdictional law, the person may be declared incompetent to handle some functions but not others. Related concepts are competency as to the estate and competency as to the person. A person who is found by the court to be *incompetent as to the estate* is barred from entering into business contracts (Weiner and Wettstein 1993, p. 285). Such transactions can be handled only by an estate guardian, appointed by the individual or by the court itself. Estate incompetency generally includes lack of competency to do business, to make a will, to sue, and to enter into contracts. In contrast, an individual who is declared *incompetent as to the person* also loses the right to make personal decisions which may include treatment decisions (Weiner and Wettstein 1993, pp. 285–286). The guardian then typically makes such decisions on behalf of this person.

2.3 Competency to Sue

A subcategory of competency to do business is competency to sue, which is in effect competency to do business in court. This refers to the capacity to conduct a lawsuit oneself or through a self-designated representative. Competency to sue is, however, quite different from competency to stand trial (see below).

2.4 Testamentary Capacity

Testamentary capacity is another special subtype of legal competency. Those who lack testamentary capacity are not capable of making a will or concluding an inheritance contract. According to the relevant section of the German Civil Code (Sec. 2249(4) BGB), a will cannot be made by someone who, due to a pathological disturbance of mental processes, mental retardation, or a disturbance in consciousness, is not in a position to recognize the meaning of the testament submitted by the individual and to act in accordance with such recognition.

2.4.1 Presumptions

According to an opinion given by the German Supreme Court, testamentary capacity requires that testators be capable of knowing that they are making a will and of knowing the content of the will. Testators must be in a condition to form a clear judgment about their will and the extent of their instructions, especially about the effects they will have on their personal and economic relationships. Testators must be able to act according to such formed judgment, free from the influences applied by third persons, although this does not preclude testators from strongly considering suggestions made by third persons in arriving at some of their decisions. In considering the suggestions of others, testators must retain the capacity to form

concepts on their own in determining and expressing the contents of their will.

2.4.2 Assessment Procedure

Some wills are contested after the death of the testator with the claim that the testator lacked testamentary capacity at the time he or she made out the will. Since the deceased person is not available for direct examination, the forensic psychiatrist must examine the contents of relevant documents and records and must methodically interview witnesses and acquaintances of the deceased to determine whether the testator had a relevant mental disorder at the time in question. The most frequent clinical basis for testamentary incapacity is progressive dementia. In such cases, the incompetency of the testator is usually not disputed; however, counterclaims of testamentary capacity can be asserted in cases where the testator was thought to have been in a "lucid interval" when the will was made.

2.4.3 Lucid Intervals

Intervals of Long Duration

The concept of the so-called lucid interval originated from the view that mental illness is an enduring process of mental decompensation (corresponding to Sec. 104(2) BGB). In the German Civil Code, the concept of a protracted lucid interval is based on the well-known clinical observation that some mental illnesses such as bipolar disorder occur in phases and are interrupted by intervals devoid of obvious psychopathology, intervals that can last for weeks or months or sometimes even years. During these lucid intervals, the individual is competent to do business and to make out a will (see also Diederichsen 1994).

Intervals of Short Duration

Invariably, however, forensic mistakes will occasion a challenge in court regarding a testator's competency to do business and to make a will, especially in those cases wherein the will was signed or a contract was finalized during a lucid interval of only a few hours or even a shorter period of time. From a psychiatric standpoint, this inquiry is meaningless, and forensic psychiatrists in Germany have expressed their concerns in recent decades (Rasch and Bayert 1985; Rose 1986; Foerster 1994). With irreversible progressive organic brain disorders, the lucid interval is no different from a relative improvement in the person's condition; this does not change the person's fundamental loss of memory and consequent loss of biographical and personal information for making appraisals. This also applies to schizophrenic and psychotic mood-disordered individuals, whose disorders may be authentically interspersed with brief disorder-free intervals.

2.4.4 United States and Great Britain

In Anglo-American law, three criteria are universally required for testamentary capacity. The testator must know and understand the following:

1. The purpose of the will
2. The nature and extent of his or her property
3. The recipients of his or her bequests, i.e., the natural beneficiaries of his property, usually family members (Sadoff 1988, p. 150; Weiner and Wettstein 1993, p. 285) or the claims others have on his property (Briscoe et al. 1993, p. 109)

A will can be invalidated based upon an "insane delusion" if the challenging party can prove that the will resulted from the testator's delusion (Ciccone 1994, p. 253). In the United States, a will can be challenged on the basis of a claim that a beneficiary had an "undue influence" over the testator. Once a claim of undue influence is made, the person who was left a bequest then has the burden of disproving that he or she exercised undue influence upon the testator (Slovenko 1973).

2.5 Competency to Consent

Competency to consent refers to the ability to give truly effectively consent for medical treatments, surgical interventions, participation in scientific studies, or invasive diagnostic procedures. The legal definition of competency to consent has only recently emerged from more general definitions of competency, and the rate of this development has varied even within Europe (Koch et al. 1996). Some overlap exists between the criteria for competency to consent and those for testamentary capacity and competency to do business. As for these other competencies, the two-step assessment procedure is required for the evaluation of competency to consent. After determining whether a mental disorder exists, the forensic psychiatrist has to assess the person's ability to do the following:

1. *Understand* certain facts
2. *Process* the information in a rational manner
3. *Assess* the information appropriately
4. *Determine* and *express* his or her own intentions on the basis of understanding, processing, and assessing

The assessment of these abilities demands that the grammar of psychological criteria are translated into a system of operationalized questions. This is usually

done by means of an informal psychiatric interview, but can also be carried out by using more elaborate and standardized techniques (Grisso and Appelbaum 1995). The capacity to consent has to be regarded as a specific and not as a global concept. This implies that the issue cannot be whether a person is competent to consent at all but whether he or she is competent to make a specific decision and to agree to or to reject a particular medical intervention; he or she may be competent to make certain decisions while at the same time being unable to accept or to reject a different medical proposal.

The concept of competency to consent is not only a relative concept, but also a dimensional one. Many patients will be rated as neither fully competent nor completely incapacitated. They may be unlikely to function perfectly in one or in all of the above-mentioned psychological areas and fall somewhere in between. Thus, whether or not a patient is evaluated as competent ultimately depends not only on the standard of assessment and the number of psychological dimensions that have actually been evaluated, but also on the borderline where the threshold between competency and incompetency is drawn. In general, this threshold will be determined by two variables: the risk–benefit ratio of the intended intervention and the direction of a patient's decision. In other words, if the risk–benefit ratio of the medical intervention is questionable or unfavorable, the standard of assessment and the threshold of competency must be high, while even the decision of a marginally competent patient may be honored if the individual consents to the application of a well-established procedure which is known to have a favorable risk–benefit ratio.

2.5.1 United States

In the United States, competency to consent to medical and surgical treatments requires that the patient knows the nature and consequences of the procedure and the consequences of accepting no treatment whatsoever (Sadoff 1988). Particularly with respect to psychotropic medication, this capacity is sometimes expressed in the negative, i.e., competency to *refuse* treatment (Sadoff 1988). A delusion concerning the procedure could impair a patient's understanding and ability to consent. With the psychiatric patient's right to refuse treatment being increasingly acknowledged and formalized over the last four decades, the law governing consent to psychiatric hospitalization, psychotropic medication, and electroconvulsive treatment (ECT) and a qualified right to refuse each of these procedures has become more extensive and more complex and certainly not uniform across state jurisdictions. Thus both treating and forensic psychiatrists have an obligation to familiarize themselves with the relevant mental health law governing competency. Some of these competencies are defined in the state health laws and other legal regulations (Weiner and Wettstein 1993). Some state mental health laws require a separate court hearing even if a guardian for an otherwise incompetent patient authorizes treatment with psychotropic medicine or ECT where the patient has not consented.

In *Zinermon v. Burch*,[1] the United States Supreme Court held that a Florida case could be litigated where a patient claimed he was mistakenly allowed to sign himself into the hospital when he was psychotic and incompetent to do so. Some states have a mental health law that requires patients to have competency to decide to enter the hospital before being allowed to sign themselves in voluntarily (Weiner and Wettstein 1993).

2.5.2 England and Wales

Competency to consent to medical treatment for a mental disorder in England and Wales is defined in the Code of Practice for the Mental Health Act 1983: "In order to have capacity an individual must be able to understand what the medical treatment is and why he needs it, its principal benefits and risks, and the consequences of not receiving the treatment, and he must have the capacity to make a choice."

3 Criminal Competencies

3.1 Competency to Confess or be Interrogated

3.1.1 Competency to be Interrogated in Germany

Competency to be interrogated in Germany is the capacity of a witness or defendant to undergo an interrogation by the police or legal authorities and to make a declaration. It presumes the capacity to follow the inquiry, to understand the intended meaning of the questions, and to offer clarifications with voluntary decision-making.

3.1.2 Competency to Confess in the United States

The concept in United States jurisprudence that is most akin to Germany's competency to be interrogated is competency to confess. However, unlike other

[1] 494 U.S. 113, 110 S.Ct. 975 (1990).

competencies in criminal law, competency to confess in the United States is essentially devoid of cognitive or functional capacities. In its 1986 *Colorado v. Connelly* decision,[2] the United States Supreme Court considered a confession to be voluntary and therefore constitutional if it was given without intimidation, coercion, or deception by others (i.e., the police). In that specific case, since the police did not obtain the confession by force or trickery, the suspect's cognitive competency was not considered relevant, even though he was quite psychotic at the time.

3.2 Competency to Stand Trial

As defined by commentaries on German criminal procedure law, competency to stand trial is the capacity of an accused, in or out of the court hearing, to perceive his or her interests rationally, to conduct his or her defense sensibly and in an intelligible way, and to give and receive declarations during the lawsuit.

3.2.1 Presumptions

Competency to stand trial covers the capacity to mentally follow court procedures concerning the court session, to conduct the litigation, and to fulfill procedural duties. The defendant should be capable of understanding the explanations of other participants in the court and of proffering his or her own requests in an understandable manner. Competency requires more than just passive participation; it requires the ability to actively look out for one's legal interests. Competency to stand trial is presumed for court hearings except for commitment procedures, as stated in Sec. 63 *Strafgesetzbuch* (StGB) (German Criminal Code). Anyone intentionally causing their own state of incompetency to stand trial (e.g., by voluntary intoxication) can be prosecuted in absentia.

3.2.2 Determination

Competency to stand trial can be eliminated by a psychiatric or physical illness. Whether or not an accused is incompetent to stand trial is a legal question to be decided by the court alone. The forensic psychiatrist merely provides potential reasons for the decision. According to Sec. 81 of the German Code of Criminal Procedure (StPO), the court can have a defendant admitted to a public psychiatric hospital for a maximum of 6 weeks in order to undergo a forensic evaluation to assess the defendant's competency to stand trial.

3.2.3 Expert Medical Reports

The forensic report, whether based on outpatient or inpatient assessment, must address the diagnosis, cause of the illness, foreseeable duration of the illness taking into account all possible therapies, effects of the disorder on the participation of the defendant in the main trial, and the danger this poses to the defendant's health.

Within the psychiatric realm, dementia and severe organic psychoses constitute possible reasons for lasting incompetency, while acute florid psychoses usually account for only a temporary state of incompetency to stand trial. There are no grounds for the assumption that the participation of a defendant or witness in a courtroom procedure generally presents a health risk if the person is already mentally ill or unstable; the clarification of a stressful situation through court participation can actually affect the defendant favorably.

3.2.4 Great Britain

In England and Wales, the concept of fitness to plead can include abilities to plead, comprehend the evidence, follow the court proceedings, advise one's lawyers, and know about challenging jurors (Briscoe et al. 1993, p. 44). In actuality, it is unusual for someone to be determined unfit to plead, affecting only about 20 defendants each year (Grubin 1991).

3.2.5 Canada

Until Canada's Criminal Code was revised in 1992, the Canadian Criminal Code did not specify criteria for fitness to stand trial, although criteria applied by courts were similar to the new standard (Davis 1993). The revised code now states that a defendant can be found unfit if he or she is unable to (a) understand the nature or object of the proceedings, (b) understand the possible consequences of the proceedings, or (c) communicate with counsel.

3.2.6 United States

In the United States, competency or fitness to stand trial is frequently assessed, affecting a large number of defendants. The criteria used by most jurisdictions in

[2] 107 S.Ct. 515 (1986).

the United States are known as the *Dusky* standard after the 1960 United States Supreme Court decision which established the minimal constitutional requirements for legal standards for competency to stand trial. According to the *Dusky* decision, the test of competency is "whether [the defendant] has sufficient present ability to consult with his lawyer with a reasonable degree of rational understanding and whether he has a rational as well as a factual understanding of the proceedings against him."[3]

3.3 Competency to Waive Counsel in the United States

In the United States, competency to plead and competency to waive counsel are separate but related capacities pertaining to criminal defendants. Competency to waive counsel should logically include competency to represent oneself pro se, because any concern for fairness involves *functional* capacity at trial, not just *decisional* capacity when the defendant asks to release his or her attorney (see Felthous 1979, 1994; Perlin 1996). Although legislators and forensic psychiatrists are free to address such relevant concerns, in the *Godinez v. Moran* case,[4] the U.S. Supreme Court held that if the decision to waive counsel is knowing and voluntary, the *Dusky* criteria for competency to stand trial are constitutionally sufficient for competency to plead and competency to waive counsel.

3.4 Criminal Responsibility

3.4.1 Germany

Anyone who, when committing an act, is incapable of understanding that this act is wrong or of acting in accordance with this understanding due to a disorder of mental illness, a profound disturbance in consciousness, mental retardation or due to any other serious mental abnormality acts without guilt.

Since 1975, Sec. 20 StGB has been the standard for "lack of criminal responsibility due to mental disorders"; since then, possible reasons to presume mental disturbance have explicitly included psychiatric illness, mental retardation, and disturbance of consciousness. The concept "other serious mental abnormality" pertains to severe personality disorders, neuroses, and sexual deviations.

[3] *Dusky v. United States*, 80 S.Ct. 789 (782) (1960).

[4] 61 U.S. LW 4749; 113 S.Ct. 2680 (1993).

3.4.2 Consequences of Lack of Criminal Responsibility

Section 20 StGB provides the criteria for finding a defendant to be without criminal responsibility, which in turn results in *exculpation*, acquittal from criminal responsibility. The prerequisites for "*deculpation*," diminished responsibility, and corresponding diminished "punishability" are presented in Sec. 21 StGB: "[If], when committing the act, the capacity of the actor to understand that the act is wrong or to act according to this understanding, (is) substantially diminished by one of the reasons described in Sec. 20, the punishment can be reduced pursuant to Sec. 49(1)." Pursuant to Sec. 63 StGB, if the psychiatric disorder which led to the criteria of Sec. 20 or 21 being fulfilled persists and is expected to persist in the future, measures can be taken for hospitalization in a psychiatric inpatient facility.

3.4.3 Procedures for Forensic Examination

Examinations in forensic psychiatry follow the two-step process: First, through psychiatric exploration and examination, the evaluator must address whether the defendant has a disorder that corresponds to one of the four named legal concepts in Sec. 20 StGB. If a qualifying mental condition is diagnosed, the second step is to examine whether, at the time of the act, the condition resulted in a substantial impairment or elimination of the "capacity to understand" and/or the "capacity to control one's behavior." A general discussion about the capacity to control is not permitted without first determining whether one of the four prerequisite conditions was present.

In the context of retrospective examinations in forensic psychiatry, we shall discuss only persistent mental disorders, i.e., psychotic illnesses, severe personality disorders, and sexual paraphilias. Other qualifying conditions will be addressed briefly.

3.4.4 Profound Disturbance in Consciousness

The category of "profound disturbance in consciousness" is reserved for exceptional normal psychological conditions that are affectively intensified (Janzarik 1995) in which so-called expressive or affective crimes have been committed. A wide consensus has now been reached within the administration of justice concerning the criteria for assessing the existence and the effects of such a profound disturbance in consciousness (see Saß 1993).

3.4.5 Imbecility

Since the reformed German Penal Code of 1975, the legal concept of "imbecility" has played a more limited role than it used to in criminal forensic psychiatry evaluations. Imbecility is understood as a congenital intellectual deficiency with an intelligence quotient of below about 70, which is generally associated with a substantially defective ability to read, write, and calculate and with a clear impairment of social competence.

3.4.6 Brief Drug-Related Mental Disorders

Brief mental disorders, namely conditions of intoxication from alcohol or other drugs, frequently occur among otherwise psychiatrically healthy offenders and are thus of substantial practical relevance. The German Supreme Federal Court has attempted to simplify judgments by providing blood alcohol limits for which diminished (blood alcohol content of 2 per 1000th) or no (3 per 1000th) criminal responsibility are probable. However, these limits fail to take into account the wide range of reactions by different individual reactions to even comparable blood alcohol concentrations through habituation and the development of tolerance to steady drinking in contrast to episodic, social consumption (Kröber 1996). The German Supreme Court has since been modifying its policy on this issue, basing its decision more on psychological factors than on laboratory data.

The German Supreme Court has now corrected its guideline and formulated the following new basic principle (judgment of April 29, 1997):

There is no verified medical/statistical principle derived from experience according to which, without taking psychodiagnostic judgment criteria into account and only on the basis of a certain blood alcohol concentration at the time of the criminal act, considerably reduced ability of control as a function of alcohol can be assumed as a rule.

3.4.7 Evaluation of Mentally Ill Offenders

The evaluation of mentally ill offenders who suffer from a persistent "pathological mental disorder" was long substantially influenced by the agnostic position of Kurt Schneider (1948), who called for exculpation in the sense of the current Sec. 20 StGB in any case involving a defendant with a psychotic illness, since it is not always possible to distinguish the extent to which the illness alone influenced otherwise apparently normal premeditation and the decision to act. In previous decades, Janzarik developed and advanced a unified position that a graduation of impairment leading to the act is possible and that – particularly in moderately impaired schizophrenic patients – diminished or absence of criminal responsibility should be considered. On the other hand, if a delusional offender proceeds purposefully in a planned manner taking the situation into account, this does not preclude the absence of criminal responsibility; it is in fact characteristic for a particularly dangerous group of violent schizophrenic offenders (see Böker and Häfner 1973) whose external behavior is well organized and who are driven by a delusional system that guides their behavior.

Many crimes committed by individuals with schizophrenia are misdemeanor property offenses committed in the context of social marginalization. The crimes are suggested to these individuals because of their tendency to be easily influenced. In the psychotic context, the crimes may appear to be delusionally motivated calls for help or self-defense in a presumed emergency situation taken against imaginary threateners or persecutors. As a rule, the issue of criminal responsibility for misdemeanors is less compelling that the issue of whether serious offenses can be expected in the future such that hospitalization would be justified pursuant to the German Penal Code (Sec. 63 StGB), the state mental health code, or the law concerning people in need of care.

3.4.8 Crimes with Psychotic Mood Disorders

Patients with psychotic mood disorders such as unipolar depressive or bipolar manic depressive disorders are occasionally charged with criminal offenses. Crimes committed by manic patients, such as embezzlement, falsification of a legal document, or traffic violations, usually result from the diminished ability of the individuals to control their behavior, even though they appear to have been well organized with regard to the questionable transaction. The illness is manifested less by disorganized thought processes and more by pathologically altered mood with resultant abolition of the person's appraisal capacity.

3.4.9 Criminal Forensic Evaluations of Persons with Organic Brain Disorders

In forensic examinations of defendants with organic brain syndromes, the significance of the disorder essentially depends on its severity. These defendants can generally be classified according to a continuum of severity from very mild to extremely profound disturbances. This type of assessment is best performed by

someone with expertise in psychiatry or the treatment of nervous disorders; psychological test data alone, which portray only the defendant's intellectual functioning but not any severe changes in affect and critical capacity, are not sufficient.

3.4.10 Determining Severity

Since the current German insanity law became effective in 1975, determining the severity of any "other serious mental abnormality" within the context of the pathological reference system (Saß 1991) has been difficult. Not all personality disorders or character traits can be subsumed under this concept. Kröber (1995) reviewed the various interpretations of "other serious mental disorder" and proposed a corresponding classification whereby the regulation of self-esteem is constantly threatened by decompensation and persistently points to dangerous psychodynamic mechanisms, especially with flagrant borderline personality disorders and severe sexual paraphilias in which the sexual realm is distorted.

3.4.11 Dissocial Behavioral Pattern

With regard to dissocial behavioral patterns, Saß (1987) differentiated between dissociality as a behavioral pattern, psychopathy as a psychopathological personality construct, and sociopathy as a combination of personality disorder and dissocial behavior. In contrast, the Anglo-American psychopathological concept, at least as presented in DSM-III, DSM-III-R, and DSM-IV, is limited to a personality disorder that leads to criminal or socially destructive behavioral patterns. Of course dissociality only corresponds to the legal concept of "other serious mental abnormality" if it is accompanied by a profound disturbance of the personality and psychodynamic processes that are demonstrated to exist beyond the criminal behaviors alone. The possibility has to be considered that dissocial patterns of behavior can also represent precursors or accompanying phenomena of a psychotic illness. In addition to establishing the presence of other, similar disorders, it is the forensic psychiatrist's responsibility to carefully illuminate the entire spectrum of possible mental disorders from a psychopathological perspective.

3.4.12 Criminal Responsibility in England

The M'Naghten rule (also spelled "McNaughten") or "right-wrong test" derived from a case in England in 1843 by the same name. Still in use over a century and a half later, this standard criterion for the defense of insanity states the following:

> . . . to establish a defense on the grounds of insanity, it must be conclusively proved that, at the time of committing the act, the party accused was labouring under such a defect of reason, from disease of the mind, as not to know the nature and quality of the act he was doing; or if he did know it, that he did not know what he was doing was wrong.[5]

In order to find a middle option between full responsibility and full punishment, which used to result in death by hanging, and not guilty by reason of insanity, England borrowed from Scotland the principle of diminished responsibility (Briscoe et al. 1993). Diminished responsibility was incorporated into England's Homicide Act of 1957:[6]

> Where a person is party to the killing of another, he shall not be convicted of murder if he was suffering from such abnormality of mind (whether arising from a condition of arrested or retarded development of mind or any inherent causes or induced by disease or injury) as substantially impaired his mental responsibility for his acts and omissions in doing or being a party to the killing.

The definition of "abnormality of mind" was given a broad interpretation by the Court of Appeal in 1960:[7]

> . . . a state of mind so different from that of ordinary human beings that the reasonable man would term it abnormal. It appears to us to be wide enough to cover the mind's activities in all its aspects, not only the perception of physical acts and matters and the ability to form a rational judgement whether an act is right or wrong, but also the ability to exercise will-power to control physical acts in accordance with that rational judgement.

The *Byrne* court opinion allowed less serious mental disorders such as neurotic and character disturbances to be considered for diminished responsibility and opened the possibility of arguing that an irresistible impulse can be used as a partial defense against the charge of murder (Briscoe et al. 1993). If the defendant can prove diminished responsibility "on the balance of probabilities" (equivalent to the U.S. standard proof of "preponderance of the evidence" or over 50%

[5] 8 Eng. Rep 718, 8 Eng Rep 722 (1843).

[6] 5&6 Eliz. 2, Ch. 2, Sec.

[7] *Regina v. Byrne* 2 QB 396 (1960); 3 All ER 1 (1960); 44 Cr App Rep 246.

certainty), the defendant can be convicted of manslaughter rather than the more serious offense of murder. Diminished responsibility is now more the rule and not guilty by reason of insanity the exception as a defense and verdict in England.

3.4.13 Criminal Responsibility in Canada

Unlike the United States, where states and federal jurisdictions have developed their own rather diverse insanity laws, Canada has a single insanity law that is followed in all provinces. In 1992, the aforementioned revised Criminal Code (see above) produced substantial changes. Whereas under the previous code, either the prosecution or the defense could raise the defense of insanity before the trial, to prevent potential abuses, the revised code restricts the prosecutor from raising the issue until after the defendant has been judged guilty of having committed the offense. The term "insanity" was replaced by "mental disorder" and "not guilty" by "not criminally responsible." The Canadian insanity verdict is now termed "not criminally responsible on account of mental disorder" (Davis 1993).

The criteria for not criminally responsible on account of mental disorder are not specified in the Criminal Code, because they have been established through case law which generally follows the *M'Naghten* right-wrong test. Until the Supreme Court of Canada's *Regina v. Chaulk* decision[8] in 1990, "wrong" was interpreted to mean legally wrong; however, in this decision, the High Court broadened the definition to include morally wrong as well.

3.4.14 Criminal Responsibility in the United States

Insanity jurisprudence in the United States is complicated by jurisdictional differences. Each of the 50 states, the District of Columbia, and the Federal Government has its own separate insanity law. To simplify, basically four standards or variants thereof have been codified or established by court opinion: the American Law Institute (ALI) insanity test, the M'Naghten test, the product test, and the *mens rea* approach.

Mentioned first because of its remarkable similarity to the two-pronged test of insanity in Germany, the ALI test was proposed in 1955 by the American Law Institute in its Model Penal Code:[9] "A person is not responsible for criminal conduct if at the time of such conduct as a result of mental disease or defect he lacks substantial capacity either to appreciate the criminality (wrongfulness) of his conduct or to conform his conduct to the requirements of law." The second paragraph of the ALI test was intended to exclude antisocial personality disorder as a mental condition qualifying for the insanity defense (Elliott 1996): "As used in this article, the terms 'mental illness or defect' do not include an abnormality manifested only by repeated criminal or otherwise antisocial conduct" (American Law Institute Model Penal Code). Twenty states and the District of Columbia use the ALI test or a modified ALI test (Steadman et al. 1993).

For the first half of the 20th century, the M'Naghten rule was essentially the law of the land in the United States, having been adopted by all but one jurisdiction, the state of New Hampshire (see below). Besides the United States, where 26 states now have a M'Naghten or modified M'Naghten test, England, Canada, Australia, and New Zealand also use the M'Naghten test for insanity (Elliott 1996).

In 1869, the New Hampshire Supreme Court adopted a product test of insanity in contrast to the more widespread acceptance of the M'Naghten test in other states. From the beginning of this century, when most states followed the M'Naghten test, up to the present, with a mosaic of different tests, New Hampshire has used the *product* test. According to this test, a murder defendant is to be acquitted if the homicidal act was the offspring or product of mental disease.[10] Since 1989,[11] however, the insanity test in New Hampshire is "whether insanity negated criminal intent";[12] and this insanity test leaves to the jury to formulate *how* mental illness negated *mens rea* in a particular case.

Finally, four states have now abolished a special defense of insanity: Idaho, Montana, Nevada, and Utah. In these states, at least theoretically, the defendant can still offer a mental illness-related defense by attempting to prove the absence of the requisite *mens rea*. Curiously, research by Steadman et al. (1993) showed that, at least in Montana, mental illness dispositions and diversions from criminal punishment continue unabated, even though a special defense of insanity no longer exists in that state, i.e., diversions are accomplished by a finding of incompetency to stand trial followed by dismissing the charge and civil commitment (Steadman et al. 1993).

It should be noted that the German insanity law would permit a severe personality disorder or a severe paraphilia as a condition which can satisfy the diagnostic criterion for the defense of insanity. In the United States, the second paragraph of the ALI test is

[8] 62 C.C.C. (3d) (1990).

[9] Sec. 401.1 (1), test draft no. 4, 1955.

[10] *State v. Pike*, 49 N.H. (1869).

[11] *State v. Shackford*, 127 N.H. 695; 506 A 2d 315 (1986).

[12] New Hampshire Statutes Annotated 628.2 (11).

intended to exclude personality disorders, and some state insanity laws explicitly exclude personality disorders. Likewise, some state insanity laws clearly bar paraphilias for consideration. It is the forensic psychiatrist's task to examine both the diagnostic and functional elements of the insanity test that pertains to the case in question and then to follow the two-step evaluation process.

3.4.15 Diminished Capacity

About one third of the states have adopted the diminished capacity principle (Slovenko 1995). Unlike diminished responsibility in England, Germany, and Scotland, which have a special diminished responsibility defense, diminished capacity is established by showing the defendant lacked the requisite *mens rea*. This approach is typically used as a defense against a murder charge, which requires specific intent to commit murder. If psychiatric testimony demonstrates that the defendant did not have the mental capacity to commit murder, i.e., the capacity to form the specific intent to commit murder, he or she may be convicted of the less serious offense of manslaughter instead. In California, the court-formulated diminished capacity doctrine fell into general disfavor and was replaced by the California legislature with the principle of diminished actuality. For a successful plea, the defendant must show that he or she actually did not form the specific intent. Because of the variation in rules governing diminished capacity and actuality in different jurisdictions, forensic psychiatrists need to familiarize themselves with the relevant statutory and case law (Miller 1994).

3.4.16 Heat of Passion and Extreme Emotional Disturbance Defense in the United States

The legal concept of "heat of passion" has long been used in the United States to reduce murder, a capital offense, to manslaughter or a similar, less serious offense. Defendants must establish that they were under the influence of heat of passion when they committed the act, that a normal person would respond similarly to the sequence of events that led to the heat of passion, and that there was no time from the provocation to the act in which to "cool off" (Goldstein 1989). It should be noted that no mental disorder is presumed because the emotional and behavioral reaction are considered to be "normal."

From this "heat of passion" concept, however, several states have adopted a broader and more formalized defense that allows for some degree of psychological disturbance, the "extreme emotional disturbance defense." New York's extreme emotional disturbance defense is often cited as an example. The state's revised penal code of 1967[13] allowed this defense where the defendant committed the act while under an extreme emotional disturbance, and "there must be a reasonable explanation or excuse, the reasonableness of which is to be determined from the viewpoint of a person in the defendant's situation under the circumstances as the defendant believed them to be" (p. 107). Pursuant to an opinion of the New York Court of Appeals (New York's highest court), a defendant could advance this defense if he or she suffered a "significant mental trauma" that can last for a while and then erupt unexplainably, resulting in the homicidal act.[14]

3.4.17 Guilty but Mentally Ill

About one quarter of the states (Slovenko 1995) have adopted a "guilty but mentally ill" verdict to address gray areas and ensure that defendants who were mentally ill at the time of the offense but did not qualify for the defense of insanity would receive both treatment for their mental disorder and punishment for committing the crime. Michigan was the first state to adopt this verdict, and its statute is the prototypical model, although significant variations are seen in the laws of other states on the verdict of guilty but mentally ill. Criteria for this finding of guilty but mentally ill in Michigan are as follows:

1. The defendant is guilty of an offense
2. The defendant was mentally ill at the time the offense was committed
3. The defendant was not legally insane at the time the offense was committed[15]

All three criteria must be proven beyond reasonable doubt. Even though one of the purposes of legislation on the verdict of guilty but mentally ill was to limit the number of insanity acquittals, most states with an option of a guilty but mentally ill verdict have retained the defense of insanity as well. Typically, a test of mental illness to establish whether a defendant is guilty but mentally ill is specified that is less stringent than the insanity test. For example, if the state uses the M'Naghten rule for its insanity test, the test used to establish presence of mental illness at the time of the act for a verdict of guilty but mentally ill is essentially the ALI test.

[13]New York Penal Law §125.25, subd. L. par. [a]; §125.27 subd. 2 par. [a] (1967).

[14]*People v. Patterson*, 383 N.Y.S. 2d 573 (1976).

[15]Mich. Comp. Laws §768.39; Mich. Stat. Ann §28.1059 (1985).

It should be emphasized that a verdict of guilty but mentally ill (unlike not guilty by reason of insanity, diminished capacity, and extreme disturbance defense) neither mitigates criminal responsibility nor reduces the punishment. Defendants found guilty but mentally ill must serve their complete prison sentence as if they had simply been found guilty of the offense. The fact finder, however, knows by the verdict itself that the defendant will be hospitalized as long as the severity of his or her disorder warrants this intensity of treatment, at least within the parameters of the prison sentence.

3.5 Competency to be Sentenced

3.5.1 Competency to be Imprisoned in Germany

In Germany, competency to be sentenced is, more specifically, competency to be imprisoned. This is "the capacity of an accused or convicted to be able to live in a facility within the prison system, to endure loss of freedom without special or serious danger to health or life, and to recognize the meaning and purpose of serving a prison sentence." Determining whether someone is competent to be imprisoned is a legal decision that is only partially based on medical findings.

A distinction should be made between illnesses that appear *before* pretrial confinement and those illnesses that begin during imprisonment. The issue is regulated by the German Code of Criminal Procedure (Sec. 455 StPO):

1. The execution of a prison sentence is to be removed if the convicted person becomes mentally ill.
2. The same is true for other illnesses if danger to the convicted person's life is expected to be caused by imposition of the punishment.

In these cases, lifting the order for detainment is therefore urgent. If the mental illness appears during confinement, then an optional regulation applies according to the German Code of Criminal Procedure (Sec. 455(4) StPO): The prison authorities *can interrupt* the execution of a prison sentence if (1) the sentence causes decompensation of the mental illness, (2) because of an illness, execution of the sentence would endanger the life of the convicted person, or (3) the convicted person becomes seriously ill and the illness cannot be recognized or treated in a prison unit or prison hospital, and the illness will likely continue to exist for a substantial period of time. If public safety is a compelling reason for continued confinement, then interruption of the prison sentence is not permitted (authors' italics). If public safety concerns contraindicate release from prison, treatment of the mentally ill prisoner must be carried out in a prison hospital.

The above-mentioned regulations are valid for convicted criminals. Psychiatrically determined incompetency to be imprisoned during detainment for pretrial investigation is to be substantiated with an illness which necessitates acute treatment in a hospital, such as florid psychosis or serious psychotic depression. If there is also substantial danger to the defendant, then admission to the psychiatric clinic of a prison or to a special security hospital should be considered in accordance with the German Code of Criminal Procedure (Sec. 126a StPO). The same applies for acute, potentially life-threatening alcohol withdrawal syndrome (delirium tremens) and conditions of intoxication requiring intensive observation that cannot be provided in a prison infirmary or prison hospital. For such serious conditions, treatment is more sensibly accomplished in a hospital setting. In contrast, conditions such as personality disorders, substance abuse, reactive depression, suicidality, and claustrophobia do not constitute incompetency to be incarcerated.

3.5.2 Competency to be Sentenced in the United States

Incompetency to be imprisoned is recognized in the United States but rarely applied in practice. In principle, if continued confinement would cause the individual's condition to deteriorate, he or she could be found incompetent to be imprisoned (Sadoff 1988). If a prison detainee requires psychiatric hospitalization prior to trial, this is more typically handled by one of the following actions: dismissal of criminal charges followed by civil commitment, adjudication of incompetency to stand trial followed by hospitalization to restore fitness, or transfer to a hospital under police supervision. In none of these examples is the defendant found by a court to be incompetent to be imprisoned. Prison systems typically have prison hospitals for mentally ill offenders. A court hearing is required before a prisoner can be transferred to a psychiatric hospital,[16] not to determine competency to be sentenced, but rather to establish legal criteria for hospitalization which correspond to criteria for involuntary civil commitment of nonprisoners. If a prisoner or a jail inmate requires hospitalization for surgical or medical treatment, this is arranged without a hearing to address competency to be sentenced. Occasionally, prisoners can be released prior to having completed their sentence if they suffer from a chronic, debilitating illness where near death is foreseeable, but this is handled more as a formalized act of mercy.

Thus, while the German model emphasizing competency to be imprisoned serves to protect the

[16] *Vitek v. Jones*, 445 U.S. 480 (1980).

mentally ill prisoner's right to necessary hospital treatment, the U.S. model protects prisoners from treatment that may not always be desired by the prisoners themselves.

3.5.3 Competency to be Executed in the United States

The United States Supreme Court held in *Ford v. Wainwright*[17] that the execution of an incompetent person violates the Eighth Amendment of the United States Constitution, which prohibits cruel and unusual punishment. The Council on Ethical and Judicial Affairs of the American Medical Association expanded the AMA ethical code's prohibition against participation in legal executions to include treating a prisoner for restoration of competency to be executed. An exception to this prohibition is if treatment is necessary to relieve the prisoner from "extreme suffering" caused by his or her illness. Although testimony regarding competency is not unethical, "a physician cannot be compelled to provide medical testimony as it relates to legal competency for execution if it is contrary to the physician's personal beliefs" (Council on Ethical and Judicial Affairs 1996/1997, p. 12).

4 References

American Psychiatric Association (1994) Diagnostic and statistical manual of mental disorders, 4th edn. American Psychiatric Association, Washington, DC

Aschaffenburg G (1906). Das Verbrechen und seine Bekämpfung, 2nd edn. Winter, Heidelberg

Böker W, Häfner H (1973) Gewalttaten Geistesgestörter. Springer, Berlin Heidelberg New York

Brakel SJ, Parry J, Weiner BA (eds) (1985) The mentally disabled and the law. American Bar Foundation, Chicago

Briscoe I, Carson D, d'Orbân P, Grubin D, Gunn J, Mullen P, Stanley S, Taylor PJ (1993) The law, adult mental disorder, and the psychiatrist in England and Wales. In: Gunn J, Taylor PJ (eds) Forensic psychiatry: clinical, legal, and ethical issues. Butterworth-Heineman, Oxford, pp 21–117, 109–110

Ciccone JR (1994) Testamentary capacity and guardianship. In: Rosner R (ed) Principles and practice of forensic psychiatry. Chapman and Hall, New York, pp 252–257

Council on Ethical and Judicial Affairs (1996/1997) Code of medical ethics: current opinions with annotations. 10.06: Capital punishment. American Medical Association, pp 11–12

Davis S (1993) Changes in the Criminal Code provisions for mentally disordered offenders and their implications for Canadian psychiatry. Can J Psychiatry 18: 122–126

Diederichsen U (1994) Zivilrecht-Juristische Voraussetzungen. In: Venzlaff V, Foerster K (eds) Psychiatrische Begutachtung, 2nd edn. Fischer, Stuttgart, pp 485–600

Elliott C (1996) The rules of insanity: moral responsibility and the mentally ill offender. State University of New York Press, Albany

Felthous AR (1979) A competency to waive counsel: a step beyond competency to stand trial. J Psychiatry Law 7: 471–477

Felthous AR (1994) The right to represent oneself incompetently: competency to waive counsel and conduct one's own defense before and after Godinez. Ment Phys Disability Law Reporter 18: 105–112

Foerster K (1994) Psychiatrische Begutachtung im Zivilrecht. In: Venzlaff V, Foerster K (eds) Psychiatrische Begutachtung, 2nd edn. Fischer, Stuttgart, pp 602–620

Goldstein RL (1989) New York's "extreme emotional disturbance" defense: a hybrid creature of the law at the psycholegal interface. In: Rosner R, Harmon RB (eds) Criminal court consultation. Plenum, New York, pp 119–133

Grisso T, Appelbaum PS (1995) Comparison of standards for assessing patients' capacities to make treatment decisions. Am J Psychiatr 152: 1033–1037

Grubin DH (1991) Regaining fitness: patients found unfit to plead who return for trial. J Forens Psychiatry 2: 139–184

Gunn J, Taylor PJ (eds) (1993) Forensic psychiatry: clinical, legal, and ethical issues. Butterworth-Heineman, Oxford

Janzarik W (1995) Grundlagen der Schuldfähigkeitsprüfung. Enke, Stuttgart

Koch HG, Reiter-Theil S, Helmchen H (eds) (1996) Informed consent in psychiatry. Nomos-Verlagsgesellschaft, Baden-Baden

Kröber HL (1995) Konzepte zur Beurteilung der "schweren anderen seelischen Abartigkeit". Nervenarzt 66: 532–541

Kröber HL (1996) Kriterien verminderter Schuldfähigkeit nach Alkoholkonsum. NStZ 16: 569–596

Miller RA (1994) Criminal responsibility. In: Rosner R (ed) Principles and practice of forensic psychiatry. Chapman and Hall, New York, pp 198–215

Nedopil N (1996) Forensische Psychiatrie. Thieme and Beck, Munich

Perlin ML (1989) Mental disability law: civil and criminal. Michie, Charlottesville

Perlin M (1996) "Dignity was the first to leave": Godinez V. Moran, Colin Ferguson, and the trial of mentally disabled criminal defendants. Behav Sci Law 14: 61–81

Rasch W, Bayert R (1985) Der Mythos vom luziden Intervall – Zur Begutachtung der Testierfähigkeit. Lebensversicherungsmedizin 37: 2–8

Rose HK (1986) Psychiatrische Begutachtung im Zivilrecht. In: Venzlaff V (ed) Psychiatrische Begutachtung. Fischer, Stuttgart, pp 509–534

Rosner R (ed) (1994) Principles and practice of forensic psychiatry. Chapman and Hall, New York

Sadoff RL (1988) Forensic psychiatry: a practical guide for lawyers and psychiatrists. Thomas, Springfield

Saß H (1985) Ein psychopathologisches Referenzsystem zur Beurteilung der Schuldfähigkeit. Forensia 6: 33–43

Saß H (1987) Psychopathie – Soziopathie – Dissozialität. Zur Differentialtypologie der Persönlichkeitsstörungen. Springer, Berlin Heidelberg New York

Saß H (1991) Forensische Erheblichkeit seelischer Störungen im psychopathologischen Referenzsystem. In: Schütz H, Kaatsch H, Thomsen H (eds) Festschrift Schewe. Springer, Berlin Heidelberg New York, pp 266–281

Saß H (ed) (1993) Affektdelikte. Springer, Berlin Heidelberg New York

[17] 477 U.S. 399, 106 S.Ct. 2595 (1986).

Schneider K (1948) Die Beurteilung der Zurechnungsfähigkeit. Thieme, Stuttgart

Slovenko R (1973) Psychiatry and law. Little and Brown, Boston

Slovenko R (1995) Psychiatry and criminal culpability. Wiley, New York

Steadman HJ, McGreevy MA, Morrissey JP, Callahan LA, Robbins PC, Cirincione C (1993) Before and after Hinckley: evaluating insanity defense reform. Guilford, New York

Venzlaff V, Foerster K (1994) Psychiatrische Begutachtung, 2nd edn. Fischer, Stuttgart

Weiner BA, Wettstein RM (1993) Legal issues in mental health care. Plenum, New York

World Health Organization (1992) International statistical classification of diseases and related health problems, tenth revision, vol I. World Health Organization, Geneva

CHAPTER

17

J. Arboleda-Flórez

Treatment and Care of the Mentally Abnormal Offender

1 Introduction

For centuries, there has been a relationship between medicine and law, specifically in relation to the treatment and care of the mentally ill offender. Ancient texts and codes such as the Baba Ramma, the Deuteronimus or the Justinian Code already make reference to individuals who, because of particular characteristics regarding their mental make up, had to be afforded special protection under the law (Arboleda-Flórez 1989). *Furiosus satis ipso furore punitor* ("the madman is sufficiently punished by his madness") is a dictum attributed to Marcus Aurelius. In the Nicomachean Ethics, Aristotle (1941) exhorted legislators to consider, when assigning punishment for an illegal action, whether the action was carried out by a madman, whom he called "ignorant". In English law, Bracton may have been the first legal scholar to attempt to identify the degree of legal impairment needed to exculpate an offender. In the first systematic treatise on English Law, written in the thirteenth century, he stated that "an insane person is one who does not know what he is doing and is lacking in mind and reason." Bracton's influence is still felt today, as his concept of "does not know" is the basic formula in the McNaghten rule. This rule, according to which an accused may be found not guilty because of a mental disorder, was developed in England in 1843 and has remained, with some variations, one of the most important medico-legal concepts in the majority of the commonwealth countries (Guttmacher 1968).

More recently, the growing number of legal cases involving individuals suffering from mental disorders or conditions that affect their emotional well-being has given rise to a body of jurisprudence that touches on practically every facet of the interface between psychiatry and the law. The study of this interface is the subject matter of forensic psychiatry.

Forensic psychiatry, as a subspecialty of psychiatry, provides methodologies for the assessment of individuals who are presumed to have some mental condition and who are caught in the midst of criminal or civil law proceedings. Among other topics, forensic psychiatry is concerned with the assessment of fitness to stand trial, criminal responsibility or dangerousness in criminal law and testamentary capacity, civil liabilities or the assessment of competency in civil law (Arboleda-Flórez and Copithorne 1996).

Past the legal entanglements, however, a number of mentally ill individuals end up in the correctional system, either detained while awaiting trial or after adjudication. The latter would include those found not criminally responsible because of mental disorder (NCRMD) or those who were simply found guilty and sentenced to a term of imprisonment, regardless of the existence of a mental problem. In addition, because prison environments tend to be pathogenic, mental illness might develop de novo while the person is incarcerated. In other cases, the natural history of mental disorders may interact with long sentences so that the probability of an inmate developing a mental illness at a particular age means that the illness will materialize while the person is in prison. Equally, a person with a history of mental illness who was well at the time of being sentenced to a prison term may experience a relapse while in prison. Thus prison environments contain individuals who were mentally ill at the time of arrival and individuals who were well on arrival, but subsequently relapsed from a previous mental condition or developed a mental condition while in prison.

Corrections, a branch of the criminal justice system, involves "all agencies of social control which attempt to rehabilitate and neutralize deviant behaviour of adult criminals and juvenile delinquents for the protection of society" (Kruzich 1982). The study of how the mentally abnormal offender is handled within correctional settings is the purview of correctional psychiatry, which in the strictest sense refers to psychiatric practice in the corrections system (Travin 1994). More specifically, correctional psychiatry could be defined as the branch of forensic psychiatry that studies the incidence, prevalence, determinants and management of mental disorders in prisons, the response of correctional systems to the mentally ill offender and the relationship between criminality and mental illness. The epidemiology of mental disorders in judicial/correctional systems is covered in the chapter by Konrad (Vol. 3, Chap. 20). The present chapter will deal more specifically with the management of the mentally ill within the system, including community corrections, and issues pertaining to the relationship between mental illness and criminality.

2 Relationship Between the Legal and the Mental Health Systems

The mentally ill and, by extension, mental health professionals interact with legal systems at practically every junction of the different components of the system, be it law enforcement, administration of justice or correctional systems (Fig. 1). The close relationship between the two systems is dictated by the intrinsic needs of the population served by both psychiatry and corrections. In some communities, the transfer of patients back and forth between the two systems

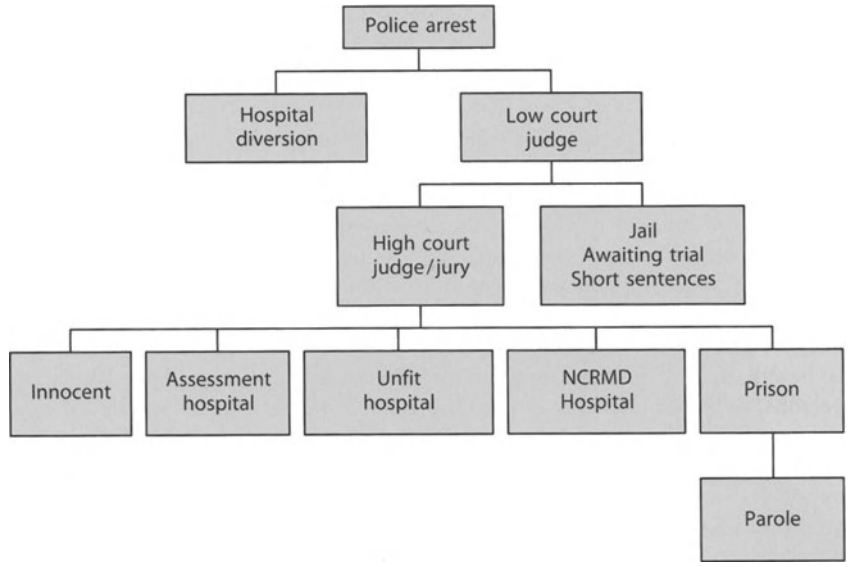


Fig. 1. Legal pathways of arrested individuals. *NCRMD*, not criminally responsible because of mental disorder

makes the correctional system, particularly jails, part and parcel of the mental health system.

2.1 Detention

In many jurisdictions, as a result of minor infractions or when public behaviour is obviously aberrant, police forces have the option to convey mentally abnormal individuals to a mental hospital or to a psychiatric emergency unit for an examination as an alternative to laying criminal charges. For example, in the Province of Alberta in Canada, the Mental Health Act (1992) gives specific powers to a police officer to bring the person for an examination to the emergency unit of a local hospital if the person was acting in a public place in such a manner that, in the opinion of the police officer, the person's behaviour was abnormal. Similar legislation exists in many other jurisdictions in Canada, in the United States (Godschalx 1984; Janus et al. 1980; Matthews 1970) and in other countries. In Jamaica and in the Commonwealth of Dominica, an island state in the Caribbean, the Pan American Health Organization has been involved in the training of policemen as mental health officers in order to extend the reach, the capabilities and the manpower of the mental health system. Policemen in these two countries are given ongoing training on basic mental health concepts, identification of mental problems among detainees, mental health resources and mental health legislation (Arboleda-Flórez and Crisanti 1996).

Powers vested on police officers under these circumstances allow them to circumvent the judicial system and to divert the person directly to the mental health system. Police officers become an extension of the mental health system. Studies have indicated that, given their familiarity with individuals who display abnormal behaviour, police officers can be relied upon to make this type of judgment (Holley and Arboleda-Flórez 1988; Bittner 1967; Teplin 1983).

The failure of de-institutionalization policies in the face of inadequate community resources makes it necessary for the mental health system to utilize other community agencies, such as the police force, to help the mentally ill person in the community. To allow police officers to be part of the mental health system helps prevent further criminalization of the mentally ill and is one of the ways of dealing with the mentally ill offender in the community. The role of the police officer, however, is limited to bringing the person to a hospital; it is the responsibility of clinicians at the hospital to make decisions on further management. The clinical team at the hospital, for example, could make arrangements for outpatient follow-up, the level of community supervision or admission to hospital, either voluntarily or by invoking the commitment powers vested on clinicians within the Mental Health Act or similar legislation.

2.2 Community Disposition

To prevent the revolving door phenomenon and, at the same time, maximize treatment opportunities and provide a modicum of protection to the community, several jurisdictions have implemented outpatient commitment legislation (Hiday and Scheid-Cook 1987, 1989; Scheid-Cook 1987; Miller 1992; Swanson

et al. 1997). Under this type of legislation, the court places an accused on probation on condition of treatment, or a physician discharges the patient to the community on a certificate that enjoins the patient to continue attending treatment and receiving medication for a specified period of time, usually 1–2 years. Case management, where a mental health specialist follows a few number of cases in the community and brokers their social and treatment needs, is usually a strong component of outpatient commitment. Patients are reported back to the court or brought back to the hospital if they fail to comply. Outpatient commitment is based on the premise that seriously mentally ill persons, because of delusions and other symptoms, do not have the ability to make decisions on their need for treatment and do not avail themselves of treatment opportunities. According to some authors, this failure to take advantage of treatment opportunities constitutes a denial of the patient's right to be treated and has negative impacts such as relapses or incarcerations (Bluglass 1993; Royal College of Psychiatrists 1993). Outpatient commitment is, therefore, promoted as an alternative to imprisonment and as a preventative step to avoid relapses or criminalization. This treatment alternative, however, has been criticized on the basis that it presents a threat to civil liberties (Fulop 1995).

2.3 Sentencing

Working arrangements between the police and local hospitals are difficult to implement as they require administrative mechanisms and previously organized payment systems (Borzecki and Wormith 1985; Arboleda-Flórez and Holley 1988). Thus, more frequently, and especially if the charge is of a serious nature, police follow the justice route and bring mental patients to local police lockups, jails or remands centres. Some of these patients may be found not fit to stand trial and sent for assessment and treatment to specialized institutions, from where they are supposed to return to court once their mental conditions stabilize. Later on, during their trial, some of these patients may be found to be not criminally responsible because of mental disorder and sent, according to local laws, for indeterminate periods of time to mental hospitals, special hospitals or special wings in a penitentiary (see Sect. 3). On the other hand, despite the presence of a mental condition, a mental patient may still be found guilty and sentenced to a prison term. These mentally ill offenders tend to suffer from chronic mental problems and are usually found in remand centres or local jails, either because they are detained while awaiting their legal disposition or because they are sentenced to short periods of incarceration. Medium- and maximum-security prisons usually house inmates who have been sentenced to long terms for offences somehow related to abnormal behaviour (antisocial, psychopathic), sexual deviations or addictions that would inevitably bring them into conflict with the law.

Research and the management and treatment of the mentally ill in prisons have a long tradition. For example, considerable research has been conducted in Germany since the 1850s on suicide, the prevalence of mental conditions among prisoners and the influence of the prison environment as a risk factor in psychiatric illness. Ganser syndrome (Ganser 1898) and the term "prison psychosis" originated from this research (Travin 1994).

3 International Perspective

3.1 Canada

In 1992, the Parliament of Canada enacted legislation to amend the dispositions contained in the Criminal Code regarding the management of the mentally abnormal offender. The new law, colloquially known as Bill C-30 (Government of Canada 1991), stipulates that accused individuals who appear to be mentally ill should undergo an assessment as ordered by the judge in no longer than 5 days to determine whether they are mentally ill and, if so, whether they are fit to stand trial. These assessments could be conducted on an in-patient or out-patient basis, depending on the severity of the suspected mental problem and/or the seriousness of the charge. At the request of the psychiatric facility, an assessment could be extended to 1 month and, in special cases, to a maximum of 2 months.

The courts cannot order treatment during the period of assessment, but if treatment is necessary, it is offered on the expectation that the person will consent. If the person refuses and is deemed to be representing a danger to self or others, the psychiatric facility can "commit" the person under the mental health act of the particular province, declare the person incompetent and proceed to forced treatment. The protection of rights is taken into account, and "committed" individuals can appeal their commitment order to an outside hospital tribunal. Following the assessment, the accused could be diverted to the mental health system entirely without any legal encumbrances or sentenced to a probation period on condition of treatment.

If the person is found not fit to stand trial, he or she is sent to a specialized mental hospital for treatment and

is expected to return to court to face the charges once improvement has been achieved. Those found to be not criminally responsible because of mental disorder are also sent to a specialized mental hospital for an indeterminate period of time, but their stay and progress are closely monitored by quasi-judicial tribunals established in each one of the provinces. These tribunals have the power to discharge individuals not criminally responsible because of mental disorder to the community under strict guidelines for treatment. Eventually, it is possible that an individual not criminally responsible because of mental disorder may be set free without any legal restrictions.

A novel disposition in this law is "capping", which means that individuals who are not criminally responsible because of a mental disorder cannot spend more time in a mental institution that the time they would have spent in a prison had they been found guilty of the charge. On the other hand, a controversial section of the law pertains to a finding of "mentally ill and dangerous". Prisoners who fulfill the criteria for this finding are committed to a mental hospital after expiration of their sentences. In Canada, the law does not make any reference to particular types of mental disorders such as psychopathy, alcoholism or drug addictions. Rather it has a general rule, a modification of the McNaghten rule (see Sect. 3.5), to determine the level of incompetence required before an offender can plead not criminally responsible.

Finally, the legislation allows for "hospital orders", but unfortunately, this disposition has not yet been enacted. The health care system and mental health professionals have expressed concerns about the cost and the risk of mixing these patients with "regular psychiatric patients".

3.2 Germany

In Germany, the law of 1975 stipulates that mentally disordered offenders convicted of an imprisonable offence be sentenced by means of a "hospital order" instead of incarceration. According to Article 63 of the Penal Code, the courts have the power to commit an offender to a psychiatric hospital, and under Article 64 to an institution for the treatment of addicts (Konrad 1993). The law does not prescribe the length of stay in the hospital, but the courts retain jurisdiction over these type of offenders and review their progress every year. Commitment of this nature is envisioned when the objective evidence satisfies the definition of the offence, when the condition is of long duration (chronic), when a link can be established between the particulars of the offence and the symptoms, when there is a probability of re-offending, especially if recidivism can be tied to the symptoms, and if the severity of future offences and hence the risk to society are assessed as considerable. These findings will also reduce the level of criminal responsibility ascribed to the accused.

Basically, German legislation requires that, in order to be found not criminally responsible, the offender has to possess a "severe mental abnormality". More specifically, however, criminal responsibility is assessed in two stages. Firstly, it has to be determined whether the accused was suffering from a "severe mental abnormality" at the time of the offence, and secondly, whether the accused had the capacity to judge the wrongfulness of the criminal act or had a capacity to act in accordance to his or her judgment (Rasch 1990). In fact, a third step seems required in that the expert is expected to determine the potential future risk the accused presents to the community. Treatment is provided from large specialized forensic hospitals or from small forensic wards located in general hospitals (Nedopil and Ottermann 1993).

As in Canada, the mental health system and forensic mental health professionals in Germany are concerned with the type and the large number of these patients in mental institutions. For example, it is considered that 27% of judicial committals are "unjustified" and dictated more by legal or social reasons than by clinical needs (Konrad 1993). The effect of hospital orders has been to treat "a heterogeneous population diagnosed with various psychiatric disorders and who have committed different types of criminal behaviour", including "personality-disordered and sexual deviants". This state of affairs has led forensic hospital directors to recommend that treatability and willingness to undertake treatment be made key factors to commit personality-disordered individuals to forensic facilities (Muller-Isberner 1996).

3.3 Italy

The new Italian Penal Code, which replaced the one in existence for 65 years, has kept the old provisions for total and partial insanity. Although it allows for a psychiatric perspective concerning the assessment of criminal responsibility (Ceretti and Merzagora 1994), it does not provide for procedures and factors to be considered in its determination. The Code establishes that a judge should decide whether to commit an offender for custody and treatment to a criminal psychiatric hospital. A very liberal section of the Code would allow for "house detention" if the sentence was less than 3 years of imprisonment and if the offender is suffering from "particularly serious health problems" (De Fazio 1996).

3.4 Japan

The Mental Health and Welfare Law of 1995 makes available to the mentally disordered those services previously available only to the physically disabled and the mentally retarded (Sakuta 1996a). Correctional facilities are staffed with physicians and other medical specialists, and some of these facilities are accredited as hospitals under the Medical Services Law (Sakuta 1996b). Japanese offenders may be found to be "not criminally responsible" and be fully acquitted, "partially criminally responsible" and receive a reduced sentence, or "fully criminally responsible" and punishable. Offenders found "not criminally responsible" on account of a mental condition are admitted to a psychiatric facility. Psychiatrists are usually charged with the responsibility of releasing these patients and setting up the discharge conditions. Unfortunately, these individuals are not required by law to report for treatment and tend to relapse and commit further offences (Sakuta 1996c).

3.5 United Kingdom

As a matter of policy, every effort is made in England and Wales to divert the mentally ill to the mental health system (Home Office 1990). However, most diversion is accomplished after conviction via hospital orders as an alternative to imprisonment. Baker (1996) indicates that The Insanity and Unfitness to Plead Act of 1991 establishes that those defendants who are found not criminally responsible should not be detained for an indeterminate period of time, as was the only possible disposition before, but rather an array of dispositions are now available and left to the decision of hospital authorities subject to consent by the Secretary of State.

The new dispositions still use the McNaghten rule, which states that "to establish a defense on the ground of insanity, it must be clearly proved that, at the time of the committing of the act, the party accused was labouring under such a defect of reason, from disease of the mind, as not to know the nature and quality of the act he was doing; or, if he did know it, that he did not know he was doing what was wrong" (*R v. McNaghten* 1843).[1] Under this rule, "disease of the mind" has been understood to have been caused by an "internal factor" (*R v. Quick* 1973),[2] and "wrong" means "legally wrong" (*R v. Windle* 1952).[3] The new law "equips the courts with the ability to tailor their response to the particular case before them in terms of treatment needs and the interests of public protection" (Baker 1996).

3.6 United States

The acquittal of John Hinckley by reason of insanity of the shooting of President Ronald Reagan caused a major shift in the population against the insanity defence. This has resulted in a variety of legal tests and sentencing dispositions that change from state to state. Thus, between 1982 and 1985, 34 states revised their insanity defence statutes (Callahan et al. 1987), and by 1995, Montana, Idaho, Utah and Nevada had abolished the special plea of insanity. Nevada, for example, created a "guilty but mentally ill" clause and provided involuntary treatment for these defendants, regardless of their competency, and in Colorado, persons found "not guilty by reason of impaired mental condition" are committed indefinitely to the state forensic hospital (Miller 1996b).

Following legal practice in the United States, management and treatment of the mentally disordered offender are decided not by national legislation, but by legal precedent. As such, legal bases exist that enjoin states to provide mental health services in jails and prisons and that stipulate the required components and standards for these services (O'Leary 1989). A convicted inmate's right to medical and psychiatric treatment in prison is guaranteed by the Eighth Amendment of the United States Constitution, which prohibits cruel and unusual punishment (Dvoskin 1994). In *Estelle v. Gamble* (1976),[4] this prohibition was constructed to mean an obligation to avoid "deliberate indifference" to serious medical needs of inmates. Based on the legal precedent established by *Ruiz v, Estelle* (1980),[5] Cohen listed six essential components for the planning of mental health services in prisons. These include the following: (1) systematic screening and evaluation, (2) treatment that is more than mere seclusion or close supervision, (3) participation of mental health professionals, (4) accurate, complete and confidential records, (5) safeguards and regulations for the use of psychotropic medications and (6) suicide prevention programmes (Cohen 1988). Several national organizations, notably the American Correctional Association, the American Public Health

[1] R v. McNaghten 1843, 10 Cl & Fin 200.
[2] R v. Quick 1973, QB 910.
[3] R v. Windle 1952, 3 All ER 1.
[4] Estelle v. Gamble 1976, 429 U.S. 97.
[5] Ruiz v. Estelle 1980, 53 F Suppl. 1265 (S.D. Texas).

Association and the Joint Commission on Accreditation of Health Care Organizations, have produced standards that, according to Anno (1994), are not perfect, but tend to complement each other.

4 Relationship Between Mental Illness and Criminality

Mental illness is expressed through behavioural manifestations. Mental conditions affect the cognitive, emotional and volitional aspects and functions of the personality. These are the very functions that the law considers essential to assess in order to adjudicate guilt, label the accused a criminal and proffer a sentence. When a mental condition is suspected in relation to a crime, the unstated assumption is that the condition preceded, and possibly, caused the crime. However, the clinician and the epidemiologist conducting prevalence studies and assessing the relationship between mental conditions and criminality have to keep in mind that mental illness may develop after a crime has been committed. The study of the relationship between mental illness and criminal offences (defined as unlawful acts, by commission or omission, leading to an arrest), therefore, is extremely difficult. This is more so if the researcher or clinician intend to establish a causality link between the two. Much work remains to be done in this area.

4.1 Relationship Between Behavioural Manifestations and Criminality

In general, mental conditions can be related to crime and this is a fact acknowledged by law. The problem, however, is not that the two elements converge, but what the degree of relatedness is between the two and whether a causal connection can be established. In regard to the degree of connectedness, for example, there are mental disorders whose very behavioural manifestations become, ipso facto, a criminal offense. In other words, the semiography of the disorder connotes a criminal act. Such is the case of most of the paraphilias (e.g. exhibitionism, voyeurism, paedophilia), pyromania, kleptomania and others. In these cases, it could be determined that the relationship between mental disorder and criminality is absolute.

Other disorders such as antisocial personality, borderline personality, pathological gambling and impulse control disorders connote a criminological element, but the degree of relatedness is not one to one in that their symptoms could be expressed without necessarily breaking the law. While not all alcoholics end up breaking the law, alcoholism carries a high risk of law-breaking in the form of victimization at the time of intoxication (Pihl and Peterson 1993). Other substance dependencies are known to lead to income-generating crimes in order to finance the habit. Results of the research by Hare and Hart (1993) suggest that psychopathy is strongly associated with a high risk for criminal and violent offences. On the other hand, many mentally ill individuals never commit a criminal offence and many types, perhaps the majority, of mental disorders do not necessarily lead to criminal offences, even despite the high prevalence of mental disorders in the general population.

4.2 Mental Illness and Violence

Lately, the literature seems to provide support for the proposition that there is more than a correlation between mental illness and criminality. For example, Hodgins (1993) followed up a 30-year birth cohort in Sweden and reported that a relationship existed between crime and mental disorder and between crime and intellectual deficiency, with men who had a mental disorder being 2.5 times more likely to have been registered for a criminal offence and four times more likely to have been registered for a violent offense than men not mentally ill or intellectually handicapped.

On the matter of mental illness and violence, Monahan (1992) has reasoned that factors such as socio-economic status and institutionalization should be considered integral components of mental disorders and should not be controlled when studying the association. Previously, these factors had been considered confounders and had been statistically controlled. This type of reasoning and research findings have led some authors to conclude that a strong, and potentially a causal, relationship exists between mental illness and violence (Link and Steuve 1996).

More specifically, on the issue of the degree of causality, studies that have linked mental illness to violence have been criticized on multiple methodological grounds, especially because of selection biases or lack of proper controls for confounding factors. Mental conditions co-morbid with alcohol and substance abuse, however, present a higher risk for violent offences. In a thorough review of the literature on mental illness and violence, Arboleda-Flórez et al. (1998) concluded that "as yet, there is no consistent evidence to support the hypothesis that mental illness, uncomplicated by substance abuse, is a significant risk

factor for violence or criminality, once past history of violence is controlled".

4.3 Potential Impact of a Finding of Causality

A conclusion that more than a statistical correlation exists between mental conditions and criminality, and especially between mental illness and violence, will undoubtedly increase the stigma against mental patients. Under the guise that they are dangerous to be released into the community, such a conclusion will lead to an increase in the number on mental patients sent to prison and in the length of time they remain incarcerated.

5 Systems Issues and Correctional Response

5.1 Burden of the Mentally Ill in Prisons

The proportion of individuals who are seriously mentally ill in prison, although small compared to the general number of prisoners, constitutes, nonetheless, a sizeable number at any given time (Arboleda-Flórez 1994). These individuals have to be housed in an institution that is not typically mandated, or funded, to provide clinical services. Resources for their care have to be placed in the balance with other health priorities and issues in the prison system, such as treatment for acquired immunodeficiency syndrome (AIDS), hepatitis, tuberculosis or sexually transmitted diseases.

Mental patients in prisons represent a burden to the system in that, apart from the need for regular clinical medical and nursing services, they require specialized correctional measures to protect them from suicidal behaviour, from being abused or assaulted by other inmates or from becoming assaultive towards others. Clearly, these patient/inmates require extra services and staff attention, which, given the type of institution, will never be of the quality to be expected from mental hospitals or psychiatric units in general hospitals.

5.2 Organizational Models

These difficulties and problems in the management of the mentally ill in prison have led to the development of several organizational models for the delivery of mental health services in the institutions (Table 1).

Table 1. Models for the delivery of mental health services in prisons

Ambulatory treatment	Mental patients remain with other inmates in the regular cells and tiers of the prison
Special wing within the prison	Mental patients are transferred to this wing, usually for the duration of their incarceration
Specialized security hospitals	Mental patients are transferred out to these hospitals, usually for the duration of their incarceration
Contractual arrangements with outside psychiatric facilities	Mental patients are transferred out to these hospitals or psychiatric units
Forensic community corrections	Every effort is made to prevent mental patients from entering the prison system or, if released from prison, to ensure that they do not go back

The first model opts to keep the patients with the rest of inmates. Ordinarily, this model is prompted by the lack of appropriate facilities in the outside community or the lack of appropriate resources within the system. Under this model, it would be expected that normal inmates will learn to live with the eccentricities and bizarre behaviour of the mentally ill and that the latter will learn to modify and control to some extent their behaviour given the response they receive from the other inmates. Needless to say, this model leads to abuse of the mentally ill and has a negative impact on the guards, who have to be constantly aware of the potential for disruption. In addition, the model presents a potential source of civil liability against the authorities in the prison.

The second model seeks to segregate the mentally ill in special units within the prison. These units are expensive because they have to be staffed, in accordance to their functions, with appropriate number of nurses and correctional officers. These units are often open to abuse of the rights of other inmates, because inmates who are not mentally ill but simply disruptive and unruly could be transferred to the "special unit" for "special handling", meaning being placed under chemical control with medication. These units, therefore, may become the repository of inmates who present disruptive disciplinary problems, whether or not caused by a mental condition. Thus these units can mix treatment needs of the mentally ill with punishment for unruly prisoners. As could be expected, the collusion between the aims of corrections (control and punishment) and the aims of psychiatry (treatment, care and rehabilitation of the mentally ill offender) has brought much criticism and disrepute to psychiatric interventions in prisons.

A third model, special security hospitals, is a variance of the specialized unit within the prison and is subject to the same pressures. These institutions, "maximum-security hospitals", however, offer the advantage that they could be used to house sexual offenders and highly violent individuals who are transferred there to receive specialized treatment or sometimes merely for security reasons. For example, sexual offenders are sent to specialized hospitals not only for treatment but to protect them from other inmates. Similarly, violent offenders are usually sent both for treatment and to keep them away from other inmates who cannot protect themselves from their assaultiveness. Many maximum-security hospitals are, in fact, built for this specific purpose.

Unfortunately, special hospitals tend to develop an affinity for the treatment of serious criminals, sexual offenders or violent individuals and tend to forget the chronic mental patient who also may become a target of abuse at the hands of those in the first group. The confluence between clinical treatment needs and treatment for criminality proper, and the usual lack of autonomy from the prison services, have made these institutions a matter of concern. In fact, several special security hospitals have been involved in controversy because of the use of unorthodox therapeutic modalities, especially behaviour modification techniques or psychosurgery.

Finally, a fourth model seeks to make contractual arrangements with local hospitals or mental hospitals to transfer the patients for the duration of their acute conditions. Arrangements are such that, after improvement has been obtained, the inmate will be returned to prison to conclude the sentence and will receive maintenance treatment for the remaining time. Usually, as they have recovered, these patient/inmates are housed with the rest of the inmates and attend weekly clinics, similar to ambulatory clinics, where they are followed up and provided with medication, usually depot neuroleptics. Apart from the obvious clinical reasons for the use of these medications, depot medications cannot be stolen from the patients and they cannot be used for trafficking by either the patients or the other inmates.

Every system has its flaws and virtues, but any one of them will eventually collapse when the number of patient/inmates outstrips the resources that have been made available within the prison system. In the end, the system is able to provide nothing else than a warehouse service.

This situation has converted prisons, especially the local jails, into modern-day asylums, with all the horrors of abuse and deprivations of the asylums of yesteryears (Torrey 1995). An alternative model would be to prevent imprisonment, especially detention in jails, of those mental patients who are usually charged with minor offences. Such a model is described below (see Sect. 7).

6 Treatment Modalities in the Institutions

In general, although psychological treatments are usually available, reconstructive psychotherapy is not often provided in correctional institutions. However, in specialized maximum-security hospitals a large array of psychotherapeutic and behavioural modification techniques and modalities are found. Some of these techniques have actually been pioneered in prisons. Biological approaches are favoured and include electroconvulsive therapy and even psychosurgery in highly specialized institutions for the criminally insane. Special hospitals provide specific medications for the treatment of sexual offenders, usually anti-androgens such as cyproterone or Depo-Provera (medroxyprogesterone acetate). Some countries allow for castration as a form of treatment for sexual offenders. Sexual offenders are usually also offered other treatment modalities in the form of cognitive therapies, re-socialization or behavioural modification techniques.

Treatment for mental conditions in prisons is largely based on medications. These, however, are usually restricted, especially when resources are limited, to the classic neuroleptics (chlorpromazine) or the classic anti-depressants (tricyclics). The treatment of violent offenders has progressed to include powerful medications including lithium, anti-convulsants such as carbamazepine, and the latest atypical neuroleptics such as risperidone and olanzepine. Because of the potential for trafficking and abuse of medications, as well as reported paradoxical reactions due to disinhibition of violent potential, many correctional facilities stay away from the use of anxiolytics or hypnotics.

It should be kept in mind, however, that treatment in prisons, and especially in maximum-security hospitals, is not only aimed at specific mental conditions, but often at criminality proper. This is a very important distinction in order not to confuse the reasons and aims of treatment and the reasons and aims of corrections. In prisons, it is not often clear what needs treatment or what is being treated, a mental condition or criminal behaviour. Worse, the aims of treatment may also confuse improvement from a mental problem with a push to prevent recidivism. Confusing the reasons for treatment and mixing up the aims of treatment has led to the introduction of ethically dubious "therapies" and abuse of prisoners. Finally, it should be kept in mind that overcrowding, lack of privacy, the overall need for security and the ever-present risk of

assaultiveness among inmates themselves can make the quality of treatment in prison substandard or below the standard expected in hospitals.

7
Forensic Community Corrections

Programmes have been established to divert the mentally abnormal offender from the justice system to the general mental health system. Steadman et al. (1995) reviewed a number of these programmes in the United States and concluded that there are two main types: pre-booking (police) and post-booking (court/jail). The former type is similar to that found in many other places and already described above, where the police have the option of bringing mentally disordered offenders to a psychiatric emergency unit if the offence is a minor one. The second type could be pre-arraignment, post-arraignment or mixed. These programmes usually have workable arrangements at the community level so that services are integrated between several agencies, including correctional institutions, mental health facilities, social services and the judiciary. Regular meetings with members of these different groups guarantee the involvement of each group and their interest in making the programmes successful. Most programmes are based on a "boundary spanner", or case manager, in charge of negotiating with institutions in different systems and of ensuring patient's compliance. The basic ingredient for a successful programme is the ability to integrate with the regular mental health and social systems in the area.

An example of this type of programme exists in Calgary, Canada, where psychiatrists from a major teaching hospital attend regular clinics in the prison system; once the person is released to the community, they provide follow-up from the hospital or by attending a half-way house for the mentally ill offender. This half-way house is operated by the Department of Justice. Community nurses and social workers act as case managers to make sure that patients do not lose their social support entitlements (e.g., welfare, apartments) while they are in prison. The programme also arranges to divert mental patients from the justice to the general mental health system in consultation with the courts.

In Canada, the federal government has just enacted Chapter 22, an act to amend the sentencing provisions of the Criminal Code. This act, already called the Community Corrections Act, stipulates, among other provisions, that the great number of offenders who present no danger to the public should be dealt with in the community rather than incarcerated. It is expected that this act will have a major impact on the management of the mentally ill offender who presents no danger to others (Government of Canada 1996).

8
Research on the Mentally Abnormal Offender

As research subjects, mental patients and prisoners are members of vulnerable populations. These populations are defined as those who, because of mental or physical incapacity, powerlessness induced by institutionalization or status (e.g., minors), need special protection if they are to be engaged as subjects of medical research.

Unfortunately, the record of medical experiments using prisoners is not pristine. Experiments such as those at the Willowbrook State School for the mentally defective, the Alabama Kilby and Draper Prisons or the Vacavillle Medical Facility testify to the blatant abuse of prisoners exposed to medical experimentation. In those instances, no ethical bounds were placed on the investigators and these, on their part, did not know or were oblivious to ethical constraints. For the reader on ethics, a sense of "not being fair", and of injustice, permeates such experiments (Arboleda-Flórez 1991). However, not all abuses of prisoners used as research subjects involve glaring trespasses of ethical boundaries; subtle abuses are also important. For example, confidentiality and informed consent issues, the balancing of risks and benefits and the lack of a clear delineation between therapy and experimental treatments are some of the issues that require much vigilance when treating or conducting research among prisoners.

These issues become magnified if the prisoner happens to also be a mentally ill person. Results from epidemiological studies purporting to substantiate peculiarities of mentally abnormal offenders, as a group, tend to further stigmatize them. Given that these studies may have methodological flaws, the use of their results to implement policies before they have been properly replicated damages the interest of these patients (Weisstub et al. 1995). Prisoners are often used in experiments of new psychoactive substances, clinical trials or experimental treatments. If some of these prisoners are also mentally ill, it would call into question the ethical tenets of such research, as these individuals often may not be completely competent, and hence will be unable to provide proper informed consent to participate.

Although a total ban of medical experiments in prisons is not advocated, given that competent prisoners should be able to exercise their will, Verdun-Jones et al. (1998) suggest mechanisms, such as properly constituted extramural ethical research boards, to make sure that both mentally ill offenders

and regular prisoners are properly protected from potential abuses during medical experiments.

9 Rights of the Mentally Abnormal Offender

Mentally abnormal offenders are in need of further legal protection over those accorded the average prisoner. The reasons why this should be an accepted expectation in any justice and prison system include the fact that mental conditions often lead to legal incompetence, that legal determinations regarding insanity acquittals impose special duties on the state and may represent further restrictions on the mentally ill offender and that findings of dangerousness due to mental illness may lead to indeterminate sentences. Laws stipulating such protection vary from country to country.

Internationally, the plight of prisoners, and in particular mentally ill ones, has been recognized by the United Nations, which has passed resolutions to the effect that prisoners have a right to health services similar to those enjoyed by individuals in the community at large and resolutions for the promotion of the rights of the mentally ill. Similarly, the World Health Organization (1996a) has produced a set of Guidelines for the Protection of Human Rights of Persons with Mental Disorders. A special feature of these guidelines is Principle 20 – Criminal Offenders – which pertains to proper safeguards and treatment needs of prisoners, especially those who are mentally ill. In addition, the World Health Organization (1996b) is preparing a set of Guidelines for Quality Assurance of Forensic Facilities, in which ethical issues are considered in relation to physical environment, security needs, administrative arrangements, staffing and the process of care in these facilities.

Fundamental to this protection are the principles that mentally ill prisoners should not be subjected to cruel or unusual punishment or torture, that proper nursing facilities be provided for treatment and that this be made available within the resources allocated for health needs in the correctional system of any country. Finally, it should be indicated that the Madrid Declaration of the World Psychiatric Association (1996) clearly prohibits as unethical the involvement of psychiatrists in the assessment of mental fitness for torture or death penalty.

10 Conclusions

Mentally ill individuals who suffer from serious and chronic mental conditions may act illegally and come into conflict with the law, which may lead to arrest and possible imprisonment. Other type of individuals may suffer from specific pathologies whose behavioural expression may lead to an automatic claim of criminality. When individuals in either of these groups end up in prisons, they pose a challenge with respect to their management and care, a challenge that prisons are often not able to meet. Several responses have been devised to deal with these inmates within correctional systems. Basically, these responses come down to four organizational models within the system and an extramural model of community corrections.

Mentally abnormal offenders, by virtue of their mental problem and of being in prison, have a dual vulnerability. Administrators, clinicians and researchers should all be vigilant and should err on the side of caution when called to manage these type of offenders.

11 References

Alberta Mental Health Act (1992) Queen's Printer, Calgary

Anno BJ (1994) Standards for the delivery of mental health services in a correctional setting. In: Rosner R (ed) Principles and practice of forensic psychiatry. Chapman and Hall, New York

Arboleda-Flórez J (1989) Problemas médicolegales de las terapéuticas psiquiatricas. In: Puppo Touriz H, Soiza Larrosa A, Puppo Bosch (eds) Medicina legal latino americana. Copygraf, Montevideo

Arboleda-Flórez J (1991) Ethical issues regarding research on prisoners. Int J Offend Ther Comp Criminol 35(1): 1–5

Arboleda-Flórez J (1994) An epidemiological study of mental illness in a remanded population and the relationship between mental illness and criminality. Doctoral dissertation, University of Calgary

Arboleda-Flórez J, Copithorne M (1996) Mental health law and practice. Carswell, Toronto

Arboleda-Flórez J, Crisanti A (1996) De policia a sanitarista mental: La experiencia en el Caribe. 10th World Congress of the World Psychiatric Association, Precongress Meeting in Segovia, Spain, August 1996 (book of abstracts)

Arboleda-Flórez J, Holley HL (1988) Criminalization of the mentally ill. II. Initital detention. Can J Psychiatry 33: 87–95

*Arboleda-Flórez J, Crisanti A, Holley H (1995) The effects of changes in the law concerning mentally disordered offenders: the Alberta experience with Bill C-30. Can J Psychiatry 40: 1–9

**Arboleda-Flórez J, Holley H, Crisanti A (1998) Understanding causal paths between mental illness and violence. Soc Psychiatry Psychiatr Epidemiol 33: S38–S46

Aristotle (1941) The Nichomachian ethics. In: McKeon R (ed) The basic works of Aristotle. Random House, New York

Baker E (1996) The law of insanity in England and Wales. Int Bull Law Ment Health 6: 19–22

Bittner E (1967) Police discretion in emergency apprehension of mentally ill persons. Soc Probl 14: 4279–4292

Bluglass R (1993) Maintaining the treatment of mentally ill people in the community. Br Med J 306: 159–160

Borzecki M, Wormith JS (1985) The criminalization of psychiatrically ill people; a review with a Canadian perspective. Psychiatr J Univ Ottawa 10(4): 241–247

Callahan LA, Mayer C, Steadman HJ (1987) Insanity defense reform in the United States – post Hickley. Ment Phys Disab Law Rep 11: 54–59

Ceretti A, Merzagora I (1994) Questioni sull'imputabilita. Cedam, Padova

Cohen F (1988) Legal issues and the mentally disordered prisoner. National Institute of Corrections, Washington

De Fazio L (1996) The Italian penal code reforms and the nature of criminal responsibility. Int Bull Law Ment Health 6: 24

Dvoskin JA (1994) The structure of correctional mental health services. In: Rosner R (ed) Principles and practice of forensic psychiatry. Chapman and Hall, New York

Fulop NJ (1995) Involuntary outpatient civil commitment: what can Britain learn from the U.S. experience? A civil liberties perspective. Int J Law Psychiatry 18(3): 291–303

Ganser S (1898) Über einen eigenartigen hysterischen Dämmerzustand. Arch Psychiatr 30: 633

Godschalx SM (1984) Effect of a mental health educational program upon police officers. Res Nurs Health 7: 111–117

Government of Canada (1991) An act to amend the Criminal Code (mental disorder) and to amend the National Defense Act and the Young Offender Act in Consequence thereof. Queen's Printer, Ottawa

Government of Canada (1996) An act to amend the Criminal Code (sentencing) and other acts in consequence thereof. Queen's Printer, Ottawa

Guttmacher MA (1968) The role of psychiatry in Law. Thomas, Springfield

Hare RD, Hart SD (1993) Psychopathy, mental disorder, and crime. In: Hodgins S (ed) Mental disorder and crime. Sage, Newbury Park/CA

Hiday VA, Scheid-Cook TL (1987) The North Carolina experience with outpatient commitment: a critical appraisal. Int J Law Psychiatry 10: 215–232

Hiday VA, Scheid-Cook TL (1989) A follow-up of chronic patients committed to outpatient treatment. Hosp Community Psychiatry 40(1): 52–59

*Hodgins S (1993) The criminality of mentally disordered persons. In: Hodgins S (ed) Mental disorder and crime. Sage, Newbury Park/CA

Holley HL, Arboleda-Flórez J (1988) Criminalization of the mentally ill. I. Police perceptions. Can J Psychiatry 33: 81–86

Home Office (1990) Provision for mentally disordered offenders, circular 66/90. Home Office, London

Janus SS, Bess BE, Cadden JJ, Greenwald H (1980) Training police officers to distinguish mental illness. Am J Psychiatry 137(2): 228–229

Konrad N (1993) The legal and psychological conditions in Germany required for commitment of convicted offenders by a criminal court to a psychiatric hospital, or to a special institution for treatment of addicts. Quad Psich Forense 2(1): 26–40

Kruzich JM (1982) Services for mentally ill offenders. In: Austin JM, Hershey WE (eds) Handbook on mental health administration. Jossey-Bass, San Francisco

**Link B, Steuve A (1996) Evidence bearing on mental illness as a possible cause of violent behaviour. Epidemiol Rev 17(1): 172–181

Matthews AR (1970) Observations on police policies and procedures for emergency detention of the mentally ill. J Criminal Law Criminol Police Sci 62(2): 283–295

Miller RD (1992) An update on involuntary civil commitment to outpatient treatment. Hosp Community Psychiatry 43(1): 79–81

Miller RD (1996a) Nevada abolishes the insanity defense, adopts GBMI. Newsl Am Acad Psychiatry Law 21: 50

Miller RD (1996b) Recent changes in the insanity defense in the United States. Int Bull Law Ment Health 6: 33–34

**Monahan J (1992) Mental disorder and violent behaviour. Am Psychol 47(4): 511–521

Muller-Isberner R (1996) Insane offender treatment in Germany: legislation, organization, and treatment programs. Int Bull Law Ment Health 6: 23

Nedopil N, Ottermann B (1993) Treatment of mentally ill offenders in Germany. Int J Law Psychiatry 16: 247–255

O'Leary WD (1989) Custodial suicide: evolving liability considerations. Psychiatr Q 60(1)

Pihl RO, Peterson JB (1993) Alcohol/drug use and aggressive behaviour. In: Hodgins S (ed) Mental disorder and crime. Sage, Newbury Park/CA

Rasch W (1990) Criminal responsibility in Europe. In: Bluglass R, Bowden P (eds) Principles and practice of forensic psychiatry. Churchill Livingstone, Edinburgh

Royal College of Psychiatrists (1993) Community supervision orders. RCP, London

Sakuta T (1996a) The new "Mental Health and Welfare Law" in Japan. Int Bull Law Ment Health 6: 28

Sakuta T (1996b) Administration and medical treatment in prison in Japan. Int Bull Law Ment Health 6: 28–29

Sakuta T (1996c) Mentally disordered offenders in Japan. Int Bull Law Ment Health 6: 29–31

Scheid-Cook TL (1987) Commitment of the mentally ill to outpatient treatment. Community Ment Health J 23(3): 173–182

Steadman HJ, Mcgreevy MA, Morissey JP, Callahan LA, Robbins PC, Cirincione C (1993) Before and after Hinckley: evaluating insanity defense reform. Guildford, New York

Steadman HJ, Morris SM, Dennis DL (1995) The diversion of mentally ill persons from jails to community-based services: a profile of programs. Am J Public Health 85(12): 1630–1635

Swanson JW, Swartz MS, George LK, Burns BJ, Hiday VA, Borum R, Wagner HR (1997) Interpreting the effectiveness of involuntary outpatient commitment: a conceptual model. J Am Acad Psychiatry Law 25(1): 5–16

*Teplin L (1983) The criminalization of the mentally ill: speculation in search of data. Psychol Bull 94(1): 54–67

Torrey EF (1995) Editorial: jails and prisons – America's new mental hospitals. Am J Publ Health 85(12): 1611–1613

Travin S (1994) History of correctional psychiatry. In: Rosner R (ed) Principles and practice of forensic psychiatry. Chapman and Hall, New York

Verdun-Jones SN, Weisstub DN, Arboleda-Flórez J (1998) Prisoners as subjects of biomedical experimentation: Examining the arguments for or against a total ban. In: Weisstub DN (ed) Research on human subjects, ethics, law and social policy. Pergamon, Oxford, Chap 26

WHO (1996a) Guidelines for the promotion of human rights of persons with mental disorders. Division of Mental Health and Prevention of Substance Abuse, Geneva

WHO (1996b) The forensic facility checklist. WHO, Geneva

World Psychiatric Association (1996) The Declaration of Madrid. (Approved by the General Assembly in Madrid on August 25, 1996). WPA Information Folder (1996–1999)

CHAPTER 18

H. Helmchen, J. Vollmann

Ethical Questions in Psychiatry

Translator: S. Henn

1 Introduction

Present awareness of the ethical implications of dramatic transformations in medicine and psychiatry is growing against the background of upheavals caused by profound changes in social structures resulting from new information technologies, global economic competition, limited social safeguards, the end of the cold war, and, especially in Germany, a renewed review of the past. Likewise, the reanimated discussion of euthanasia has to be seen in the context of the large increase in the number of individuals with severe residual problems following intensive medical treatment in life-threatening circumstances and of very old, dependent, and sick individuals due to longer life expectancies and in the context of patients' increasing desire for self-determination. Similarly, the issue of possible and appropriate care of the mentally ill in the face of current financial restrictions cannot be separated from the question of priorities in resource allocations in health care. In the light of rapid advances in decoding the human genome and in information technology, many are understandably calling for improvements to be made in data protection and for confidentiality to be strictly observed. However, the main question is how, in an increasingly regulated, formalized, and bureaucratic world, psychiatrists and other psychiatric caregivers can maintain and further develop their ability, not only to care adequately for the health and well-being of individual patients using their professional competence, but also to preserve each individual patient's dignity and right of self-determination by perceiving and respecting their wishes and interests.

Ethical questions in medicine and hence also psychiatry have become considerably more important in the last decade. This is shown by an increase in relevant publications (Table 1) and the adoption of a series of declarations by international committees such as the United Nations (UN) and the World Psychiatric Association (WPA) and regional committees such as the Council of Europe or of Latin America – Pan American Health Organization (PAHO)/World Health Organization (WHO) (Table 2); in addition, the law is becoming increasingly complex.

We will begin by presenting aspects of medical ethics that are important as a basis for ethical judgments of psychiatric problems, followed by the relevant declarations, guidelines, and legal norms that serve as a means of orientation and as a framework. These will then be illustrated by current and especially significant problems followed by frequent and typical ethical questions that arise in psychiatric practice.

Table 1. Periodicals

Name	Location	Comments
Bioethics	Clayton	Since 1987
Bulletin of Medical Ethics	London	Since 1985
Cambridge Quarterly of Healthcare Ethics	Cambridge	Since 1992
Christian Bioethics	Lisse (NL)	Since 1995
Ethik in der Medizin	Heidelberg	Since 1989
Hastings Center Report	Briarcliff Manor, NY	Since 1971
Informations- und Dokumentationsstelle Ethik in der Medizin (IDEM): Datenbank ETHMED	Göttingen	Database (1993)
Bioethicsline	Washington, DC	Since 1973
IRB: A Review of Human Subjects Research	New York	Since 1978
Journal international de Bioethique	Lyon	Since 1990
Journal of Clinical Ethics	Frederick	Since 1990
Journal of Medical Ethics	London	Since 1975
Journal of Medicine and Philosophy	Dordrecht	Since 1976
Kennedy Institute of Ethics Journal	Baltimore	Since 1991
Recht und Psychiatrie	Bonn	Since 1983
Zeitschrift für medizinische Ethik	Ostfildern	Founded in 1954 as Arzt und Christ
Zeitschrift für Medizinrecht	Heidelberg	Since 1985

Table 2. Codes and declarations in chronological order

Title	Year	Reference
The Nuremberg Code	1947	Sass (1989)
WMA Declaration of Helsinki (on Biomedical Research)	1964	Latest revision in Somerset West (1996)
UN Declaration on the Rights of Mentally Retarded Persons	1971	WHO (1996)
UN Declaration on the Rights of Disabled Persons	1975	WHO (1996)
WPA Declaration of Hawaii (on the Duties of Psychiatrists)	1977	Helmchen and Müller-Oerlinghausen (1978)
WPA Declaration of Hawaii II (on the Duties of Psychiatrists)	1983	WHO (1996)
WMA Declaration of Hong Kong on the Situation of the Elderly	1989	World Medical Association (1996)
WMA Declaration of Hong Kong on Persistent Vegetative States	1989	World Medical Association (1996)
PAHO/WHO Declaration of Caracas (on Restructuring Psychiatric Care in Latin America)	1990	WHO (1996)
UN Resolution 46/119 – The Protection of Persons with Mental Illness and the Improvement of Mental Health Care	1991	WHO (1996)
Council of Europe – Recommendation 1235 on Psychiatry and Human Rights	1994	WHO (1996)
WPA Declaration of Madrid (on the Duties of Psychiatrists)	1996	World Psychiatric Association (1998)
Council of Europe – Convention for the Protection of Human Rights and Dignity of the Human Being with Regard to the Application of Biology and Medicine: "Bioethics Convention"	1996	Council of Europe (1996)
WMA Statement on Ethical Issues Concerning Patients with Mental Illness	1995	World Medical Association (1996)
Grafenecker Erklärung	1996	Arbeitskreis zur Erforschung der "Euthanasie"-Geschichte (1996)
Nürnberger Erklärung	1996	International IPPNW Congress (1996)
Deutsche Gesellschaft für Psychiatrie, Psychotherapie und Nervenheilkunde (DGPPN) Stellungnahme zum Entwurf der Bioethikrahmenkonvention des Europarates vom 8.3.1996	1996	DGPPN (1996)
Stellungnahme zum Schutz nicht-einwilligungsfähiger Personen in der medizinischen Forschung	1997	Zentrale Ethikkommission bei der Bundesärztekammer (1997)
Entwurf der Richtlinie der Bundesärztekammer zur ärztlichen Sterbebegleitung und den Grenzen zumutbarer Behandlung	1997	Bundesärztekammer (1997)

2 Foundations of Medical Ethics

2.1 Definition of Terms

Whereas there is often no distinction made between the terms "ethics" and "morality" in common usage, a distinction is made in academic discussions. The terms "morality" (L *mores*) and "customs" are understood to refer to the area of moral phenomena, i.e. the essential, normative, and fundamental assumptions for the manner of existence of human beings (e.g. rules for behavior, standards of values, the meaning of ideas). Ethics (Gk *ethos*) as a scholarly discipline of philosophy analyzes, systematizes, and reflects upon these moral phenomena on a theoretical level. In contrast to metaethics, which makes no contents-related statements about the moral value of individual actions, their rules, or criteria for rules, but rather examines such statements in their linguistic form, the

philosophical investigation of practical or applied ethics concentrates on concrete moral problems (Höffe 1992; Patzig and Schöne-Seifert 1995). Depending upon the subject of investigation, we speak of economic ethics, media ethics, animal ethics, and medical ethics, for example, as subdisciplines of applied ethics. It should be emphasized here that we are not dealing with "special ethics" with specific values, rules, criteria, or reasoning strategies, but that generally valid ethical rules simply obtain in particular fields and problem applications (Ach and Gaidt 1993; Beckmann 1996). Whereas in German philosophical usage, the term *Medizinethik* ("medicine ethics") is used, in medicine the terms *medizinische Ethik* ("medical ethics") and *Ethik in der Medizin* ("ethics in medicine") are usual. In English-speaking countries, the term "bioethics" is used (see below). Branches of medical ethics are denoted by terms such as "doctors' ethics," "care ethics," "patients' ethics," etc. Since ethical conflicts in the medical field are increasingly less often able to be solved from the perspective of one professional group, the comprehensive concept of "health ethics" is widely used in the literature. In comparison, suggestions of terms such as "ethics of the healing professions" or "the theory of values" (Seidler 1986) have not gained acceptance.

2.2 Development and Current Discussion of Ethics in Medicine

Problems of medical ethics have long been handled in the medical profession in Germany largely as questions of professional ethics and medical law. Ethical actions were primarily communicated to prospective doctors as medical-collegial behavior on the basis of the Hippocratic tradition (Winau 1994; Wiesemann 1996), which cannot do justice to the many current ethical problems in medicine (Höffe 1987; Beauchamp and Childress 1994). Issues of medical ethics were only of marginal interest and were worked on by medical lawyers, medical historians, and psychiatrists, but above all by doctors who had a personal interest in them alongside or following retirement from their clinical duties. Here, practical experience gained in medical professional practice and moral intuition about the proper behavior of doctors played the decisive role, and traditional ethical principles based on the Hippocratic oath or Christian faith were accepted as a matter of course. However, current ethical dilemmas in modern multicultural societies require a rational, secular foundation for argumentation in medical ethics in order for assumptions to be able to be critically discussed. Here, ethical justification is not developed by a professional group alone or behind closed doors, but rather value decisions must be made transparent and be publicly justified (Steigleder and Mieth 1990; Tugendhat 1994; Schöne-Seifert 1996).

The development of a rational, secular, and liberal discussion of issues concerning medical ethics has become indispensable because of two developments. One is that medical advances have opened up possibilities for diagnosis and therapy for which the traditional principles of medical ethics have proven to be too general, making specific obligations necessary. Second, medical advances have been accompanied by an increasing plurality of value systems and the demand for more patient autonomy in Western societies. Because of varied individual attitudes toward values, problems of medical ethics, which frequently touch upon a person's private sphere, cannot be usefully settled by means of "professional ethics" or by a unified ethical theory.

Since the 1970s, in light of these new ethical questions in medicine, the new, academic discipline of bioethics has developed in the United States, a discipline in which different professional groups (e.g. doctors, philosophers, lawyers, theologians, social workers) work together as equal partners in an interdisciplinary manner (Reich 1994, 1995). In English usage, bioethics is understood as a subdiscipline of applied ethics and concerns itself with questions in medicine and biology (see above). Bioethics constitutes neither a "special ethics," in which only certain ethical rules or theories apply, nor is the term connected to a statement of content relating to concrete questions about the issues. In fact, the Anglo-American bioethics discussion is characterized by a large number of methods, theories, and opinions. While at the beginning of this development approaches of analytic philosophy, liberal contract theory, and value ethics were given priority, since the 1980s philosophical contributions from virtue ethics, feminist ethics, care ethics, communitarian philosophy, and casuistic and narrative models of ethics have gained in significance (Sass 1988). In the international debate, methodical approaches, theoretical models, and the problem of transferring theoretical medical ethics into practice are the subject of much controversy (Siegler et al. 1990; Birnbacher 1993; Bok 1996; Levi 1996), and there is neither a unified theory nor a homogeneous opinion.

Due to lack of knowledge of these developments, the term bioethics in Germany is used mainly as a negatively loaded slogan with which to discredit the opposition in ethical and political controversies. Instead of a discussion based on reasoned argument, it is insinuated that the so-called bioethicists allow themselves to be used to legitimate biomedical research and thereby threaten human rights and dignity. These scholars are accused of a dangerous faith in

progress, which would stand in fatal continuity to that biologistic-reductionistic thought tradition that led medicine to perpetrate the gravest crimes against human rights during the rule of National Socialism. The "anti-bioethics movement" has received wide public attention in Germany, especially with the controversy over the preference-utilitarian theses of the Australian philosopher Peter Singer concerning euthanasia (1984) and in the discussion of the so-called bioethics convention of the Council of Europe (Sass 1995; Schöne-Seifert et al. 1995; Reiter 1996; Vollmann 1996b). The term "bioethics" therefore has such strong negative connotations in Germany that, in contrast to the international academic discussion, it cannot currently be used as a neutral term.

2.3 Theories, Principles, and Rules in Ethical Actions

In considering the medical ethics of moral problems, four levels have to be differentiated: ethical theories, principles, rules, and individual cases.

On the primary level of ethical theories, an attempt is made to determine valid principles and rules which can systematize a great number of individual cases and ideally combine them in a consistent theoretical system. However, this has not yet succeeded, because all the theories proposed in medical ethics have their specific strengths and weaknesses and are invoked by various ethicists for different ways of looking at problems. Fundamental differences among ethical theories concern terminology, classification (which principles and rules?), establishment of methods (which rules in accordance with which principles?), weighting (which norm takes precedence in a case of conflict?), and the interpretation of rules in a concrete individual case. Aside from these differences, all theories that are candidates for a general medical ethics possess three common conditions, requiring (1) the possibility for the generalization of norms (the precept of universalizability), (2) the rejection of egoism as an ethical argument, and (3) the postulation that a rational analysis and foundation is of central significance in the solution of moral conflicts and that only these can be made the basis of generally binding norms (Höffe 1992; Patzig and Schöne-Seifert 1995).

Two main groups of modern ethical theories can be differentiated. Deontological theories (Gk *deon* = duty) judge an action to be moral if a recognized moral principle is followed, regardless of the effects this action has. These include Kantian ethics, which are also important in the debate on medical ethics, the central principle of universality of which states that if I cannot consistently wish that everyone in my situation acts in agreement with the norm by which I intend to be guided, then my way of acting is not morally correct. In practice, this "categorical imperative" can lead to situations that contradict our moral intuition. If, for example, it is generally recognized that we should tell the truth (as a moral duty), well-intended lies are not permissible, regardless of possible consequences in individual cases. Thus, according to the deontological understanding, a physician is obliged in every case to inform a patient completely, even of an unfavorable prognosis. On the other hand, teleological theories (Gk *telos* = end) judge the moral quality of actions on the basis of whether they are appropriate to promote a presupposed goal, such as happiness or well-being. Examples of this are utilitarian ethical theories (L *utilis* = useful), the majority of which are Anglo-Saxon in origin. In utilitarianism, the decisive moral criterion lies in whether the action promotes the welfare of all those affected in an optimal manner. Whereas classical utilitarianism defines hedonism as the maximization of well-being as pleasure and happiness or the absence of pain and unhappiness, in modern preference utilitarianism the degree to which the wishes, needs, and interests of all those affected by an action are taken into consideration is decisive. In practical medical ethics, however, considerable difficulties arise with the necessity to evaluate and weigh up the interests of the various individuals involved. Which "human beings" have relevant interests? How can personal interests be quantified and, if a conflict arises, how can they be weighed up against one another? These problems play an important role in current ethical discussion, e.g. in euthanasia, abortion, research on human embryos, etc.

Principles of medical ethics are derived from ethical theories and define the fundamental norms in the area of health. In spite of differences on the level of theories in the debate on medical ethics, there is much agreement on the so-called middle level of principles. For example, the right of a patient to preservation of life, self-determination, assistance, protection from bodily harm, respect for his or her person, truthfulness, and confidentiality is not disputed. In addition to these patients' rights, there are also certain duties of the physician, e.g. the precept of the well-being of the patient (*salus aegroti suprema lex*) or the medical principle to do no harm (*nil nocere*). Other duties include the duty to help, respect for the dignity and right of self-determination of the patient, the prohibition against killing, the duty to maintain confidentiality, and the principles of fairness, tolerance, and openness (Sass 1988). Various medical ethicists have proposed principles of medical ethics which are relevant to medical practice and which differ in variety, number, and the importance attached to them (Veatch 1981; Engelhardt 1986; Beauchamp and Childress 1994). The latter proposal has received the

most attention, which refers to respect for the patient's right of self-determination (autonomy), the principle of not causing injury ("nonmaleficence"), acting for the well-being of the patient ("beneficence"), and justice (fairness) as basic principles of medical ethics.

From these generally formulated principles, concrete rules are derived by which the physician should be guided.[1] The psychiatrist in clinical practice is generally confronted with ethical questions in individual cases (e.g. should I inform Mr. Smith about his diagnosis at the present time?). In order to act normatively correctly in individual cases, the physician is guided by rules (e.g. patients who are able to make their own decisions should be informed). These rules derive from a higher principle of medical ethics (e.g. everyone's right to self-determination should be respected), for which the ultimate basis is an ethical theory (e.g. Kantian philosophy).

2.4 Weighing Up Conflicting Values

Although there are still differences on the level of theory, the general acceptance of the principles of medical ethics cited here represents a step forward toward achieving a pragmatic consensus on essential norms in medical ethics. Unfortunately, the ethical problems in medicine cannot solved by this pragmatic agreement, however. The major ethical questions always appear when two or more ethical principles come into conflict with one another. In psychiatry, for example, the principle of respect for the patient's right of self-determination (autonomy) and the duty of the physician to act for the well-being of the patient ("beneficence") and to avert harm ("nonmaleficence") can come into conflict with one another if a patient needs to be restrained and treated against his or her expressed will. Other examples are suicide prevention, the benevolent lie, and the just allocation of resources in psychiatry and psychotherapy, which can bring the principle of justice into competition with the well-being of the individual patient.

Agreement in the various ethical theories on the middle level (principles) is therefore only superficially satisfying. In clinical practice, in addition to a knowledge of principles of medical ethics, a physician requires a method by which he or she can judge divergent principles in an individual case. There is no generally recognized approach stipulated by medical ethics for this; instead, each case has to be decided individually by evaluating competing ethical values. Criteria and methods for taking decisions about weighing up values against one another need to be developed, because ethical considerations are necessary not only to take a decision in the individual case but also for subsequent action (Schmidt 1989; Sass and Viefhues 1991; Gillon and Lloyd 1994). In this context, the excessive emphasis of the principle of autonomy coupled with the simultaneous undervaluation of the principles of beneficence and justice has repeatedly been criticized in American ethics. This one-sidedly and short-sightedly concentrates on self-determination, while neglecting the welfare of the patient and the complex consequences for other patients and individuals in a welfare state community (Holm 1994, 1995). Faced with the need to take concrete decisions and actions in an individual case, the physician is dependent upon clinical decision-making, in which not only principles of medical ethics, but also moral intuition and values also play a role. In this context, medical ethics has experienced a revival in the justification of medical actions by virtue ethics on behalf of the well-being of the patient (Pellegrino and Thomasma 1993), as has the casuistry of medical ethics for ethical assessment in individual cases (Jonsen and Toulmin 1988). In view of the fact that a generally binding definitive moral justification is unlikely, Rawls (1975) proposed an interplay between a rational basis and intuitive experience, a reflective balance, in which, in the last resort, a responsible decision needs to be taken in the individual case. Rationally based principles of medical ethics can make an important contribution in the identification, organization, and thoughtful debate of ethical problems. However, not even such a comprehensive ethical system can relieve the treating physician of responsibility for the moral decision in an individual case.

3 Framework: Declarations, Guidelines, Legal Norms

The first statement made by the WPA on ethical questions in the profession was the 1977 Declaration of Hawaii (on the Duties of Psychiatrists; Helmchen and Müller-Oerlinghausen 1978). It was intended to give psychiatrists moral support in conflicts of loyalty in contemporary society and help with decisions. The background, made clear in the first paragraph of the declaration, was the political abuse of psychiatric concepts, knowledge, and techniques that was recognized in the 1970s in countries such as the Soviet

[1] The area of rules concerning medical ethics can be further differentiated. The German Federal Board of Physicians (*Bundesärztekammer*) distinguishes four categories: (1) memorandum, (2) recommendation or opinion, (3) standards, and (4) guidelines (Bachmann and Heerklotz 1997).

Union, Romania, Chile, and South Africa (Helmchen 1986). The declaration specified the fundamental ethical principles of respect for the dignity of patients and their right to self-determination and of action solely in the best interest of the patient. The latter was also supported and promoted in the book *Principles of Biomedical Ethics* by Beauchamp and Childress (1994), a book which later became very influential. Here, they formulate the elements of informed consent, underline the responsibility to maintain confidentiality, comment in particular on forensic examination and involuntary measures, demand the possibility of independent review of such involuntary measures, and require psychiatrists not to abuse the opportunities their profession allows them, and, especially, to refrain from participation in involuntary measures in the absence of psychiatric illness. The emphasis of this last rule, in particular, makes clear the resistance to the political abuse of psychiatry that was necessary at the time. This focus also highlights the fact that other important ethical problems, such as the just allocation of resources in psychiatric care and research, were barely addressed. The Soviet psychiatrists avoided discussion of psychiatric abuse by withdrawing from the WPA before the 1983 World Congress in Vienna (Helmchen 1986). At the same congress, certain binding formulations and the particular emphasis on individual autonomy in the Hawaii Declaration were somewhat moderated by rewording (Hawaii/II; WHO 1996). A series of additional statements by the WPA, prompted not least by an antipsychiatric tendency in the so-called Daes Report (by a working group of the UN Commission on Human Rights), were accepted by the 1989 World Congress in Athens as a Charter for the Rights of the Mentally Ill. The efforts made by psychiatrists and health care politicians in many countries over the last three decades to bring about a fundamental improvement in the care of the mentally ill and a growing awareness of their fundamental rights led in 1991 to UN Resolution 46/119 on the Protection of Persons with Mental Illness and the Improvement of Mental Health Care (*Principles for Policy on Mental Health*; WHO 1996). This is the first UN document drafting human rights for the mentally ill and their right to treatment. They are to be procedurally guaranteed by means of what are in part very detailed regulations. Psychiatrists are to be able to refer to them in their dealings with their administrative bodies and even governments. The governments themselves are to give effect to the principles of this resolution through appropriate legislative, judicial, administrative, educational, and other measures. The first principle stated is the right of every individual to the best available mental health care. It is to be a component of the health and social welfare system. With this and especially with principles 8 and 9 of the formulated standards for care and treatment, which are oriented to the health needs of the patients and to their receiving equal treatment to those who are physically ill, the issues of quality control and the just allocation of resources are gaining more significance.

These central themes of the 1990s also are specifically expressed in the 1996 WPA Declaration of Madrid. At the very beginning, in its first point, it emphasizes that psychiatrists are to be engaged in the just allocation of health care resources if they wish to meet their commitment to the best possible care of their patients in accordance with scientifically proven knowledge and ethical principles. The second guideline explicitly states that psychiatrists must keep themselves informed about scientific developments and that those trained in research should push back the scientific frontiers of psychiatry. The ethical standards to be observed in this context are cited in detail in guideline 7. It is emphasized that the mentally ill are particularly vulnerable as research subjects and that particular care therefore needs to be exercised in order to protect their autonomy and their mental and physical integrity.

Three international declarations that formulate ethical standards for human research, and thus also apply to research with the mentally ill, are particularly significant for psychiatric research. These are the 1947 Nuremberg Code (Sass 1989), which was developed at the time of the trial of the National Socialist physicians and served as a basis for guiding the judgment passed by this international court, the 1964 Helsinki Declaration (on Biomedical Research; most recent amendment made in Somerset West, South Africa, in 1996), which the World Medical Association (WMA) put forward as medical guidelines for biomedical research that are still valid today, and the 1996 Bioethics Convention of the Council of Europe, which is important for international law. The special problem regarding the inclusion of research patients who are not able to give their consent is also addressed in this convention; since this is a very topical and fundamental issue, it will be treated separately in the next section.

The ethical standards formulated in these declarations are in accord with international agreement, but they do have different binding force depending on the national law and cultural context concerned, as a comparison concerning informed consent among various European countries shows (Koch et al. 1996). The laws relevant to psychiatry are treated in earlier chapters (see Chaps. 15, 16, this volume). In Germany, these are above all the *Betreuungsgesetz* (BtG, Law on Care), valid since 1992, and the *Gesetze für psychisch Kranke* (PsychKG, Laws on the Mentally Ill) in each individual *Land*.

4
Special Problems

Medical actions are shaped by medical socialization, based on experience, and stabilized by feelings. They lead to complex intuitions (biases) constituted by implicit pictures of human beings, physicians, and one's self, without which decisive actions in the reality of daily life are scarcely possible. Rational and instructive argumentation can lead to counterintuitive inconsistencies in this context. Nevertheless, the challenges of modern medicine force discussion about these contradictions, even if it is repeatedly endangered by emotional blocks and by intellectual temptations. It should serve to elucidate and justify the various positions and attempt to find permissible, adequate, and practical solutions.

4.1
Psychiatric Research

Psychiatric research is primarily research with human beings. The main reason for this lies in the fact that there are few adequate animal models for most mental disorders. However, in the case of mental disorders, the ability to give consent and the validity of such consent is much more questionable than with other illnesses. Before a patient is included in a research project, his or her ability to give consent must therefore be examined and, if this ability to consent is lacking, it must be considered whether research with this patient is ethically justifiable and legally permissible at all and, if so, under what conditions.

As early as 1900, the Prussian education authority issued a short, but clear instruction to the directors of clinics concerning the principles to be followed in patient research. The supreme court of the German Reich had already given its view on the voluntary nature of patient participation in research projects as early as 1894 and reinforced this view in 1906. In 1931, the German Reich Ministry of the Interior issued rules for clinical research (Vollmann and Winau 1996). In spite of these clear instructions regarding the voluntary nature of patient involvement and weighing up benefits and risks, German physicians carried out criminal experiments under National Socialism which were condemned in Nuremberg in 1947. As a result, explicit rules for the conduct of human experimentation were formulated, and – as the Nuremberg Code – these then became the basis for the adoption of the 1964 Helsinki Declaration by the WMA. Self-determined participation of patients or probands has since been regarded as the fundamental prerequisite for all human research.

The right to self-determination is universally guaranteed in Germany in the constitution (*Grundgesetz*, Basic Law, Art. 1). Only the law governing the manufacture and prescription of medicines deals with this and other conditions for conducting research, however. It is the only German law that is specifically relevant to research. However, it only regulates experiments with medication. Further decrees and guidelines have been issued by the Federal Minister of Health, the Federal Public Health Department and the Federal Institute of Medication and Medical Products (BfArM), and the *Bundesärztekammer* (German Federal Board of Physicians), and court decisions have been made, as a result of which clinical research has become increasingly legally regulated in recent years. Since the revision of the Helsinki Declaration in Tokyo in 1975, ethics commissions have been formed in Germany in nearly every medical faculty and in the boards of physicians of the individual *Länder*. According to German medical law, all physicians must consult such an ethics commission before they begin a research project with patients (Helmchen 1995b).

4.1.1 Informed Consent

The general validity of the principle of self-determination has more than one reason. One source is the European tradition, which is considerably influenced by Kant's philosophy, according to which a person shall never be viewed and treated merely as a means, but always also as an end. Another root of this principle stems from the ethical and political tradition, particularly in the United States, a tradition which is constituted by the self-determination of each citizen and emphasizes the right to freedom of the individual against all forms of authority (state, church, and other institutions). It is the right to decide on matters affecting one's own sickness or health that exemplifies the perception of individual responsibility and self-determination.

The legal theory of informed consent – as developed over the last three decades principally in the United States – can claim general validity as the means to legitimate all medical interventions. The relationship between patient and physician is traditionally understood from the medical point of view primarily as a relationship based on trust and often with only implicit or presumed consent. However, the explicit consent of the patient has acquired a central position in present legislation and court decisions, not least because of physicians' omissions (Vollmann, in press). For that reason, the legal rules concerning the representation of patients who are actually or legally unable to give consent themselves are oriented toward the precept of consent (presumed consent, consent through a care

giver or legal representative; see Chaps. 15, 16, this volume, Part 2).

This development has many and diverse roots. The main focus in modern medicine has increasingly expanded to cover not only acute treatments but also the long-term treatment of chronic illnesses or now even concentrates primarily on the latter. Successful long-term treatments with all their burdens – especially in ambulatory psychiatry – are not possible without the responsible participation of the patient just as in the case of clinical research. In addition, human rights, or rather basic civil rights such as self-determination and patient dignity, are now even more strongly asserted as the patient becomes more dependent upon treatments which cannot be self-determined, something which can happen in modern medicine and its often standardized courses of treatment.

Especially the procedures required by scientific methodologies can limit the individuality of research patients. The nature of gains in scientific knowledge, especially in the testing of hypotheses, goes beyond the individual (and, as a result, at least also uses the individual for the sake of another). Knowledge is gained from more than one patient and exceeds the experience of each single psychiatrist. In order to ensure the scientific comparability of results, observations as well as diagnostic and therapeutic methods must be standardized. In accordance with convictions prevalent today, this is most likely in clinical trials under controlled conditions. Control means the reduction and standardization of context variables, the systematic variation of intervening variables, and repetition. Important techniques for the control of objective influences include randomization. Blind trials, including the use of placebos, control for subjective influence.

Such objectification is compatible with the dignity of the patient only if the patient accepts it, i.e. if the individual patient consents freely and personally following appropriate explanation of the methodologically conditioned dependence. The patient should also be informed about such methodological dependencies precisely because it establishes trust, without which there can be no accepted dependence. The temporary external control must therefore be legitimated by valid consent. Thus the greater the extent to which the individual patient is objectified, i.e. to which he or she is treated as an object, the more certain the validity of the consent must be.

In contrast to the clear validity of ethical and legal standards for consent as a prerequisite to every medical intervention, clear, generally accepted, and practical criteria and methods for determining the ability to give consent are currently still not well known. Grounds for this may lie in the generally accepted subordination of the ability to consent, because the physician as a rule does not question the acceptance or nonrefusal of a proposed intervention that is in the best interest of the patient. Moreover, a mentally ill patient who does not have clear cognitive or behavioral disorders might experience doubt expressed in his or her ability to consent as discrimination and an impairment of the relationship of trust to the physician. Generally, the ability to consent is only called into question in practice if a patient refuses an urgent or life-saving intervention without a recognizable or understandable reasons. In research, however, this question is more important for the very reason that the physician needs to be certain that consent is really valid.

Assessment of the ability to give consent is not easy, requires experience, and is burdened with uncertainty. The key to this assessment is informing the patient, because the ability to give consent is not a universal characteristic, but a relational one, i.e. it can only be assessed in relation to concrete facts in the here and now. Thus the patient must receive sufficient, specific information in a way that is appropriate to his or her ability to understand. "Not everyone is equally capable of understanding the same explanation of a treatment plan. A person is more likely to give valid consent if the explanation is appropriate to the level of his assessed ability" (UK Mental Health Code of Practice, revised in 1993). Analysis of the processing of this information forms the basis of judgment. Criteria for this judgment of the ability to give consent were developed above all in the United States by Appelbaum and Grisso (1995) and in Germany by a working group on "Research Needs and the Problems of Consent in the Mentally Ill" (Helmchen and Lauter 1995).

Two levels of assessment were proposed using the latter (Amelung 1995). First, the psychiatrist must decide whether the affected patient has any psychiatric disorder at all. If this is the case, four functions should then be examined (Helmchen 1995a):

1. The ability to understand a particular fact ("what?"); this is assumed if the patient can repeat in his or her own words what it is about, e.g. trial versus standard treatment, randomization, placebo control, voluntary agreement to participate.
2. The ability to process particular information appropriately ("why?"); this appears given if the patient can substantiate the decision, e.g. with hope for greater chances for healing in spite of certain risks.
3. The ability to assess the information personally without illness-dependent distortion ("why me?"); this should be the case if this assessment is associated with understandable value convictions or concepts of illness that the patient has.

4. The ability of the patient to determine his or her own will on the basis of understanding, processing, and assessment ("if and how?"), i.e. if the patient can clearly express his or her decision.

4.1.2 Informed Consent and the Classification of Research Models

The special importance of consent following information about the research is based on the methodologically required "objectification" of the patient in a research plan that does not exclusively serve the patient, in use for the sake of others, and, finally, in contrast to pure care of the sick, the greater uncertainty, opportunities, and risks of a research project. In such a case, therefore, it is not a matter of an obvious, established standard treatment, but of an innovative diagnostic or therapeutic attempt or even a nontherapeutic experiment. Accordingly, the law raises the standard for informing the patient as the basis for valid consent and limits the possibilities of surrogate consent by a legal representative. Similarly, the threshold of grounds for investigating the ability to give consent should be lowered and the assessment standards raised. The threshold values for these extensive actions depend not only on the type of research intervention and its significance to the well-being of the individual patient, but also on the type and intensity of the potential risks.

If the patient will also be treated by the research intervention, then the project is classified as a therapeutic trial. The individual patient has the opportunity of receiving effective therapy (from which it is expected that it will be more effective and/or safer than the current therapy), and the researcher has the opportunity to improve or increase knowledge. However, if the researcher is pursuing research with a patient solely to win knowledge, then this patient has no personal benefit from it. This is then classified as human experimentation without potential individual benefit or, in the Helsinki Declaration, as nonclinical research. This difference determines the type and extent of consent after the patient has been informed and corresponds to international guidelines for ethical implementation of clinical research.

The therapeutic trials mentioned thus include the "last" trial and innovative "pilot treatment trials" in particular cases, as well as experimentally implemented research treatment with patient groups according to a research plan, i.e. a controlled clinical trial in the strict sense of the term. Consequently, the boundaries between pure therapy and pure experiment can become blurred (Fig. 1). It is therefore the responsibility of the research psychiatrist to recognize these boundaries, because the assignment of a research project to one of various models has different consequences with respect to the information required and the benefit–risk ratio.

Benefits and risks can be quite different in type and intensity. A taxonomy of benefits and risks might therefore be helpful (Table 3).

4.1.3 Research with Patients who Lack Competence

Ethics commissions have increasingly concerned themselves with research projects in which patients are included who are not able to give valid consent after receiving appropriate information. This applies not only to research projects in pediatrics and psychiatry, but also to projects in neurology and neurosurgery, and above all to those in anesthesiology and intensive care medicine. Ethics commissions sometimes appear uncertain if they seek legal advice or if their judgments contradict those of other ethics commissions involved in the same projects, e.g. in multicenter projects. This applies particularly to research projects which do not deal with therapeutic research in the narrow sense. Clearly, therefore, there is a need for clear and binding rules of assessment. In Germany, the working group of the medical ethics commissions has been working intensively for a long time on this theme, and the *Zentrale Kommission zur Wahrung ethischer Grundsätze in der Medizin und ihren Grenzgebieten* ("Central Commission for the Safeguarding of Ethical Principles in Medicine and Adjacent Fields") at the German Federal Board of Physicians, Central Ethics Commission for short, has just published a position on this matter (Zentrale Ethikkommission bei der Bundesärztekammer 1997). The subject has been in the public eye since 1994 in particular as a result of the discussion of the Council of Europe's "Convention for the Protection of Human Rights and Human Dignity in View of Biological and Medical Applications": Bioethics Convention. Above all Art. 17.2 of the Convention of the European Assembly, adopted on September 26, 1996, involving a common legal framework for human biomedical research, was controversial and debated right up to the end. In accordance with a series of preconditions, which are formulated in the preceding Arts. 17.1, 16, and 5, this article attempts to set up criteria for authorized exceptions to research which "does not have the potential to produce results of direct benefit to the health of the person concerned." This position was supported primarily by the Dutch jurist Roscam Abbing (1994) with the following arguments:

- Without nontherapeutic research, the very conditions that make the patient unable to give consent could scarcely be investigated and treated.

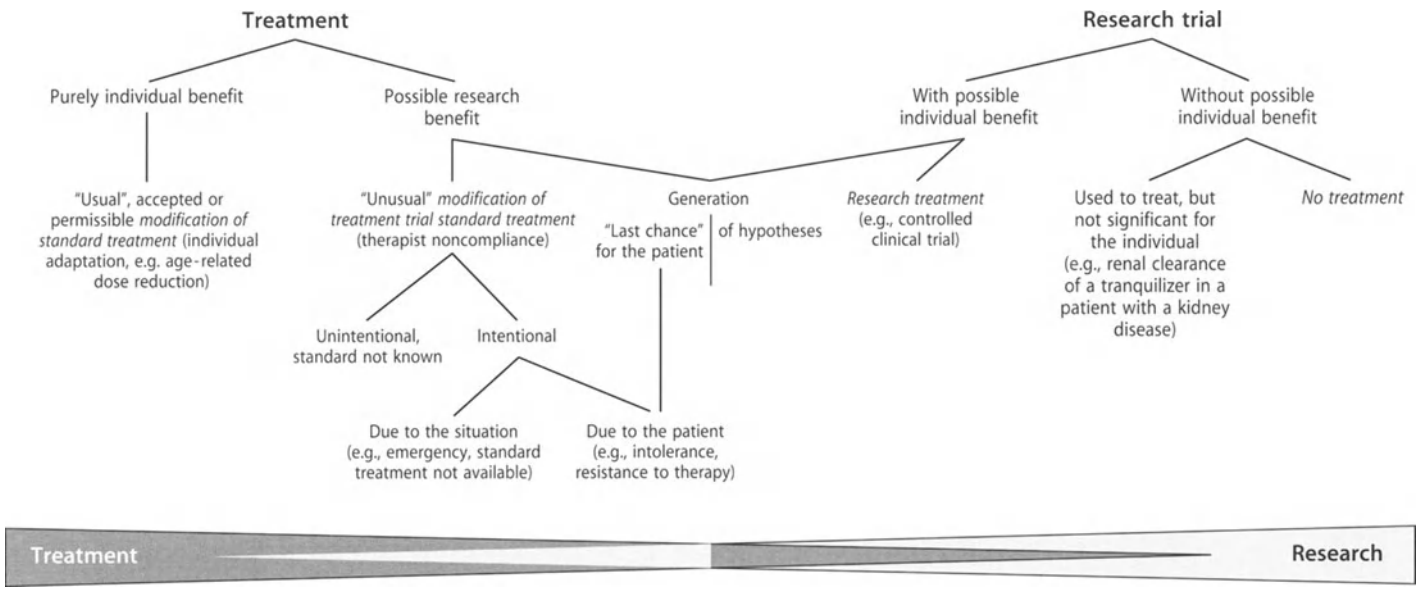


Fig. 1. Types of clinical research. The term "treatment" includes diagnostic, preventive, and therapeutic measurements. According to the *Arzneimittelgesetz* (German Medical Preparations Law, AMG Sect. 40.4), not only therapeutic drug trials, but also diagnostic and preventive ones are permissible, at least with children. Accordingly, these three different research categories can be viewed as being of potential value to the individual. (From Helmchen 1994)

Table 3. Classification of research risks and benefits

Research risks	Research benefits
No risk	No benefit
Minimal risk	Society benefits exclusively (trial *without* possible benefit to individual): - By increasing and validating present knowledge - By discovering qualitatively new knowledge
Slight increase of the minimal risk Clearly greater than minimal risk	The individual and society benefit (trial *with* possible benefit to individual): - By measurably improving existing standards - By discovering new therapeutic (and diagnostic) measures

- The research contributes to the health of other patients groups.
- Society has potentially great interest in the research in view of burdens placed upon the nearest relatives and health care costs.
- Solidarity can be expected from every citizen.

While medical research without direct potential benefit to the person who is incapable of giving consent is viewed in other European countries as permissible according to the exceptions formulated in Art. 17.2, (e.g. the 1988 French research law, Fagot-Largeaut 1996; the proposal by the British Law Commission 1993/1995), under German law (Taupitz et al. 1997) it is questionable or not permissible. It is also criticized or strongly disapproved of by the German public, as became clear in a joint hearing of the Federal Lower House of Parliament's Committees on Law, Health, and Education, Science, and Research on May 17, 1995 and at the Congress of Physicians on "Medicine and Conscience – 50 Years after the Nuremberg Trial Against Doctors," October 25–27, 1996 in Nuremberg (Stein 1996). The main arguments of critics of research without direct benefit to patients incapable of consent are the following:

- There is no real need for this research.
- This research contradicts the constitutionally guaranteed respect for human dignity.
- This kind of research – due at least in part to the criminal human experimentation carried out by Germany under the National Socialists and especially in view of crimes committed against the mentally ill (Klee 1983; Mitscherlich and Mielke 1960) – cannot be controlled and could be the signal of a "dam bursting."

As a consequence of this public debate, Germany was one of three European nations who abstained in the vote over the aforementioned 1996 Convention in the Council of Europe (de Wachter 1997).

The Central Ethics Commission took the following position (Zentrale Ethikkommission bei der Bundesärztekammer 1997) on the problem:

A particular ethical dilemma occurs in research in which the affected persons themselves are not expected to benefit, but other persons in the same age group or affected by the same illness or disorder can at least gain from the discoveries. Here, on the one hand, we have the ban on subjecting a person without his or her consent to measures which benefit another and do not also serve his or her own interests (the ban on using a person as a means to an end). On the other hand, we have the ethical conviction that slight risks may be required of a person if others can thereby be greatly helped.

It is clear that no one – competent or not – can be obliged to help a group of future patients by participation in a scientific investigation, even if the benefit for these patients is great and the risks to the patient minimal. However, inclusion of individuals who are unable to give consent in such an investigation appears to be defensible if, in addition to the observance of further protective criteria, the legal representatives, being familiar with the represented individuals (especially with their earlier lifestyle and views or explicit earlier statements), have sufficient grounds to be able to assume their willingness to take part in the investigation and no contradictory expressions of intent of the affected individuals exist.

A central point of this argument is that this research must be connected to the illness which led to the patient's inability to give consent. This criterion is very important for adults in order to exclude patients who are unable to give consent from ethically unjustifiable research for the exclusive benefit of others, as non-clinical biomedical research is defined in Sect. III of the Helsinski Declaration. Accordingly, individuals with dementia, for example, are excluded from research projects which are not concerned with dementia.

This criterion does not, however, apply to minors, who – independent of a particular illness – are still not able to consent because of their age. The French version of this criterion stipulates that the research project must be valuable to individuals of the same age

or to those suffering from the same illness or similar disabilities (Fagot-Largeaut 1996). An Irish version contains an obviously broader definition, whereby benefit must at least be expected for other patients (Casey 1996).

The Central Ethics Commission has formulated the following protective criteria:

1. "The research project cannot be carried out on a person able to give consent instead." This means that there is no alternative research strategy to answer the research question (Fagot-Largeaut 1996). If, however, the research question can be answered using patients able to give consent, then this research is not permissible with patients unable to give consent. Psychiatric examples of such research questions include the following:
   a) Projects with patients with dementia who are unable to give consent which cannot be conducted with patients with dementia who are still able to give consent as a result of their only slight development of dementia, because the exact diagnosis during the early course of the illness cannot be determined with the necessary certainty or because factors determining the progression of late stages of the illness are different from those in early stages.
   b) Can the assumption of differences in care needs for patients in late stages of dementia between those still living at home on the one hand and those who are institutionalized on the other be verified through epidemiological observation and questioning, which, if confirmed, would have consequences for the specification and improvement of care?
   c) An answer to the question of whether the pattern of psychopathological disorders and the speed of progression of later stages of dementia differs between the "young old" (i.e. up to 85) and "old old" patients (i.e. over 85) would be important in solving the problem of whether, in older patients, dementia is only a matter of accentuated aging or an expression of brain disease (Reischies and Schaub 1997; see also Vol. 2, Part 2, Chap. 10).
   d) In late stages of dementia, does the brain also have a regenerative potential, which could then serve as the basis of a specific therapy, a question which could possibly be answered by a defined pattern of particular neurotrophins, e.g. nerve growth factors, or by metabolites of particular cerebral proteins seen in venous blood or with magnetic resonance imaging.
   e) Particular knowledge of special brain functions, e.g. the basis of verbal skills, which can only be gained by the neuropsychological investigation of aphasic patients, i.e. from patients with speech disorders following brain injuries or strokes. These disorders can impair language understanding and, consequently, the conditions for giving valid consent after being informed.
   f) An investigation into the validation of assumed ability to give consent can only be undertaken if not all patients are actually able to give consent. Thus results with patients in a memory clinic who agree to more precise psychological tests of their ability to consent would show that a proportion of the patients are not in fact able to give consent.
   g) In the report by the Meijers Commission appointed by the Dutch government, a series of additional examples of the necessity for nontherapeutic research with incompetent patients are cited (Meijers et al. 1995).
2. "Fundamental information about the recognition, explanation, prevention, or treatment of an illness can be expected from the research project." Purely replicative ("me too") research or research designed exclusively to devise hypotheses is therefore not defensible. In their assessment, the ethics commission should also take into account the need for the research. For example, the need is urgent in the case of dementia, because dementia which progresses over many years entails continuous suffering for the patient and closest relatives. In addition, the causes of dementia cannot yet be treated and, in the long run, the disease presents an increasing burden on public health services because of its steadily increasing frequency, something highlighted by the continuing discussion about the law governing nursing care insurance and its practical implementation.
3. "Defensible risks in relation to expected benefits can be expected in the research project." This criterion should be used only for research projects which benefit the patients involved themselves, if not at the time of the research, then at least in the later course of their illness or in a later recurrence of it. With research projects in which benefits cannot be expected by the affected patient, but only by those in the same age group or with the same illness, this criterion must be complied with more strongly, "so that in all cases minimal risks

and burdens are expected from the research project."[2]

4. "The legal representative has given effective consent provided that he or she is familiar enough with the person he or she is representing in order to have sufficient grounds to assume that this person is willing to participate in the investigation." A helpful guide to the patient's wishes could therefore be a so-called patient directive, prepared at a time when the patient was able to give consent, or a power of attorney for health affairs. The caregiver should not simply disregard thoughts of the patient in solidarity with other patients simply because the ability to give consent has been lost. Another possibility mentioned in the Dutch legal draft in this context is that the participant included in a research project should be informed according to his or her present level of understanding (Berghmans 1995).
5. "The person concerned does not show rejection in his or her behavior." This means that the investigation must always be discontinued if the patient, particularly a patient unable to give consent, indicates that the investigation is so unpleasant that he or she wishes it to be discontinued. This criterion is specified in the Dutch proposals such that the research investigation shall not be continued if the patient unable to give consent protests and this protest can be judged as different from the usual behavior observed in the affected group of individuals unable to give consent (Berghmans 1995).
6. "The relevant ethics commission has judged that the research project is appropriate." In controversial cases, at least, the ethics commission should substantiate their assenting vote in light of specified criteria and should recommend prompting a judicial decision by their vote, either directly or for the (urgent) appointment of a caregiver.

[2]The Central Ethics Commission explains that the concept of "minimal risk" is difficult to define, "but can be specified by distinguishing degrees of risk and giving a list of examples. Medical specialty associations and ethics commissions can contribute to this. In each case, the objective risk and subjective burden or complaints are distinguished (e.g. magnetic resonance imaging holds no objective risks, but can quite possibly be a subjective burden, which leads to the investigation being interrupted). In particular, there is a great deal of individual variation with respect to subjective suffering and large differences among groups. In the view of the commission, an example of a "minimal" risk would be if small amounts of body fluids or tissue can be obtained during operations that are necessary anyway and thus entail no additional risk for the patient. Certain physical tests (e.g. sonograms, transcutaneous tissue measurements) and particular psychological tests (e.g. questionnaires, tests, behavioral observations) also fall into this group" (Zentrale Ethikkommission bei der Bundesärztekammer 1997).

In summary, extrapolating the corresponding regulations of the *Arzneimittelgesetz* (German Medical Preparations Law) to general medical research, nontherapeutic research with adult patients unable to give consent is not currently regarded as legally permissible in Germany. Nontherapeutic research with potentially immediate benefits, such as diagnostic research in the narrow sense for patients included in such a research project, is seen as ethically justifiable and should also be legally permissible for adult patients unable to give consent, as it is for minors. Nontherapeutic research with only indirect benefits for the patient (group 2 in the statement issued by the Central Ethics Commission) or benefiting only the group of patients with that illness (group 3 in the statement issued by the Central Ethics Commission) appear to be legally defined exceptions under the ethically justifiable prerequisite that strictly defined protective criteria are fulfilled and the patient is exposed to no risk. Finally, research of use to others in the sense of Sect. III of the Helsinki Declaration is not ethically justifiable with patients unable to give consent (Table 4).

Medical research expected to have only questionable or no individual benefit with patients unable to give consent (often incorrectly equated with nontherapeutic research) is and remains a difficult and controversial problem. This is reflected in the fact that such research is legally permissible under certain conditions in some countries such as France and probably also in England, but not in other countries such as Germany. From a medical point of view, such research can only be ethically justified if its necessity is determined according to defined criteria, patient needs for protection are fulfilled, so that the research has no more than minimal risks and negligible unpleasantness for the person participating, and refusal by the patient is accepted. There is a large international consensus on these criteria (Dresser 1996; Helmchen 1998a,b,d; Keyserlingk et al. 1995; Law Commission 1995). In addition, concepts and definitions of benefit and – especially minimal – risks and criteria, rules, and procedures for weighing benefits and risks against one another and, in addition, for the evaluation of individual as opposed to societal benefits need to be developed. Moreover, the question of who should make such evaluations needs a satisfactory answer.

The discussion about so-called nontherapeutic research with individuals unable to give consent strengthens fears of a medical science with inhuman characteristics. Because this research, more than other medical interventions, carries the risk of using a person as an instrument for others, it is highly relevant to the fundamental human right to personal dignity. Public discussion is therefore necessary, not only because respect for human dignity is associated with openness and efforts to understand others, but also

Table 4. Differentiation between case groups according to the position statement of the Zentrale Ethikkommission (Central Ethics Commission)

Case group	Definition	Examples
1	Treatment with present potential benefit to individual	Use of a new antidementia drug which is expected to be more effective and/or have fewer side effects than currently used antidementia medication
2	Intervention with future potential benefit to individual	Diagnostic measures relating to pathogenesis as foundation for development of therapeutic measures which can benefit the patient with a long or relapsing course of illness
3	Research without direct benefit to individual patient, but with potential benefit for the group of patients with the same illness or of the same age	Is the course of late-stage dementia determined by completely different factors than in the early stages? A positive answer would increase the perspective from which to search for treatment possibilities relating to causes in late-stage dementia
4	Nonclinical biomedical research (corresponding to Sect. III of the Helsinki Declaration)	Pharmacokinetic research with a medication which is irrelevant to patients with dementia

because in an open society an understanding of the opinions present in the society is a prerequisite for socially appointed decision-makers to define the basic conditions for such research. In this discussion, for example, philosophers should clarify the relationship between utilitarian and deontological ethics and its practical significance, while medical researchers have the task of representing the individual and societal need for such research, providing specific examples of its consequences and the variety of real benefits, risks, and burdens together with the ethical justification of the research.

To conclude from conceptual errors and the criminal abuse of psychiatry, e.g. "solving the social question" (Dörner 1988) by the "extermination of lives of no human value" (Binding and Hoche 1920), that any deviation from the Nuremberg Code is an inevitable step into the abyss – as, for example, the 1996 Grafenecker Statement (Grafenecker Erklärung 1996; Arbeitskreis zur Erforschung der "Euthanasie"-Geschichte 1996) suggests – does not do justice to the following problems (Helmchen 1998a,b):

1. The medical motivation for this research comes from the immediate experience of the patients' suffering, e.g. from the suffering that dementia causes in patients, their relatives, and caregivers. Physicians wish to treat such conditions more effectively than is presently possible and, therefore, in certain cases also include patients unable to give consent in research investigations, since there is no other way to reach this goal. Thus it is correct that patients unable to give consent are indeed especially weak and vulnerable, because their illness (and not the physician!) robs them of the possibility of being aware of their rights. It is for this very reason that physicians strive to achieve improvements, to heal the illness or at least relieve it to such an extent that the patient recovers the ability to exercise his or her rights. Physicians, however, do not wish – as is implied by some critics – to make these patients "available" for research because they are weak and defenseless, but rather to include them in research under defined protective measures because their illness is so severe and difficult to treat that doctors are called upon to do something about it. This argumentation is also supported by the medical professional directive according to which the physician also has to serve the population (Ärztekammer Berlin 1990) and by the constitutional law guaranteeing freedom of research in Art. 5. Nevertheless, it must be stressed that these duties and rights are secondary to the primary duty of the physician "to serve the health of the individual" (Ärztekammer Berlin 1990).
2. These individual and societal desires for progress in the treatment of serious illnesses have found their expression in the fact that, according to the leading ethical assessment of medical research generally recognized today (the Helsinki Declaration and its revisions), the inclusion of patients unable to give consent in research is viewed as ethically defensible. This appears to be a contradiction to the Nuremberg Code, but can be understood in the historical context in that the Nuremberg Code of 1947 was aimed at condemning "research" with individuals who were not informed, were not voluntary, and were deceived in Germany under the National Socialists and its authors therefore did not mention research with patients unable to give consent

because they did not have them in mind. Furthermore, the necessity for therapeutic research with such patients could not be foreseen from the state of scientific development at that time (see also Meijers et al. 1995).

3. To place a taboo on research with patients unable to give consent hinders the development of criteria and methods which can recognize the special vulnerability of these patients and guarantee their protection and the observance of ethical standards. However, this is exactly what appears to be required. On the one hand, the need for research on serious illnesses is growing, and with it those circumstances which can lead to the inability to give consent (e.g. brain trauma, stroke, dementia, intoxication) or which appear in children who have a developmentally dependent inability to give consent. The example of the broad public conviction of benefit from azidothymidine (AZT) trials in the treatment of acquired immunodeficiency syndrome (AIDS) in the 1980s in the United States, which drove patients and physicians to forge inclusion criteria, makes it clear that precautions to protect vulnerable individuals are evaded if they are not adequate to the problem (Levine 1996). Furthermore, without criteria for the ability to give consent, such consent will be assessed very liberally on the basis of the clinical impression and accepted as given or the threshold determined as an indication for the explicit investigation of the ability to give consent will be placed so high that its absence cannot even be determined (see, e.g. example *f*, p. 327). On the other hand, the globalization of research and the worldwide use of its results carry the danger that a regionally absolute exclusion of all use of a person as an instrument could lead to the continuing use of individuals in this way elsewhere (Rössler 1996). It would then certainly be ethically questionable, but unavoidable due to open borders, to use such results of foreign research in Germany.
4. History has shown that the proclamation of ethical standards alone is not sufficient. In spite of the previously mentioned instructions issued in 1900 and 1931, the crimes committed under National Socialism occurred, and even after the Nuremberg Code of 1947, seriously unethical research projects were carried out and published (e.g. Beecher 1966; Faden 1996). Continuing discussion of the increasing multiplicity of serious ethical problems in both medical research and practice is therefore imperative. Steps in this direction include the establishment of ethics commissions and public discussion. Although there seems to be no alternative for the time being to the institution of expert committees (Rössler 1996), it is still an important step that these experts contribute their arguments in an understandable form to the public discussion and concentrate on significant questions (see statements issued by the Scientific Advisory Board of the German Federal Board of Physicians, the Central Commission for the Safeguarding of Ethical Principles in Medicine and Adjacent Fields, and specialist associations).

4.2 Suicide and Euthanasia

Psychiatrists are confronted with the issue of suicide and euthanasia in a number of ways: (a) in the overwhelming majority of cases, suicide occurs against a background of a mental disorder (see Vol. 6, Chap. 9); (b) a mental illness may also be important in a person's request for help with suicide or in dying; and (c) psychiatrists are increasingly consulted in the assessment of the ability to give consent when a patient asks to die by a life-sustaining treatment being discontinued. The central ethical problem here is that the duty of the physician, the guarantor of life, to act in the best interest of patients (in this case possibly a mentally ill patient unable to give consent) conflicts with respect for the right to self-determination of the patient.

However, as a profession, psychiatrists have had particular experience with the thousandfold abuse of euthanasia in the National Socialist past and therefore have a special obligation to concern themselves with this subject now (Lauter and Meyer 1992), after mentally ill patients have also been affected by physician-assisted suicide (see the comments on the Chabot case below).

Recent developments in the international debate over euthanasia show an increasing acceptance toward forms of directly assisted dying. In the Netherlands, there has been a legal ruling since 1993 (and de facto since 1992) according to which, while retaining the criminal ban on killing on demand and assisted suicide, physicians are assured of exemption from punishment if they follow a formally designated process and actively end the life of a patient whose unbearable suffering has led to their expressing a wish to die (van der Wal et al. 1996). In the state of Oregon (USA), a law was passed by public referendum in 1994 permitting physician-assisted suicide for seriously ill patients who had no chance of recovering. Similar legal reforms were only narrowly defeated in the states of California and Washington. These legal bans on physician-assisted suicide in individual states were upheld as compatible with the Constitution by the U.S. Supreme Court in

June, 1997.[3] In 1996, in the Australian Northern Territory, active euthanasia by a physician was permitted under particular preconditions. However, this ruling was later repealed on constitutional grounds by the Australian House of Representatives.

4.2.1 Definitions

In the controversial medical ethics debate about euthanasia and assisted dying, central terms are used differently. Terms will therefore be defined before their content is discussed. The expression "euthanasia" comes from classical Greek antiquity (Gk *eu* = good, *thanatos* = death), whereby a good, easy, gentle, and honorable death was meant. Euthanasia was a term in philosophy for a long time, and only in the 18th century did it find its way into medicine as related to medical help with dying and to providing support for the dying. Euthanasia was understood positively as a medical task to make the death of the dying person as easy as possible. Euthanasia in the medical literature of the 18th and 19th centuries meant help with dying, providing support for the dying, and relieving and assuaging death. At the same time, it was constantly emphasized that medical measures should never lead to a shortening of life. Such a wish on the part of the dying person was refused by the doctor. Only with the development of Darwinism at the end of the 19th century was euthanasia also discussed as the disposability of human life from the point of view of medical advances, "genuine humanity," and social aspects (Winau 1984; Vollmann and Dörries 1996). In Germany, the term euthanasia was abused by the National Socialists as a code name for the systematic and state-run murder of disabled and mentally ill children and adults (so-called "lives without value"; Winau 1984). Because of this abuse, the term "euthanasia" in Germany has such negative connotations that, in contrast to international usage, it should not be used (von Lutterotti 1992; Winau 1993). Other authors argue against this, saying that the National Socialism abuse of the term cannot be grounds to renounce it (Wassermann 1993), especially since it is common in international discussion.

The international use of the term euthanasia corresponds essentially to that of the German term *Sterbehilfe* ("help in dying"). Especially in the medical literature on *Sterbehilfe* (euthanasia), a distinction is made between an "active" and a "passive" form. According to this distinction, passive euthanasia is understood as the physician reacting passively toward the biological dying process or the fundamental illness and allowing the illness to run its natural course. This includes the omission, discontinuation, and removal of medically possible treatment measures. Whether artificial nutrition is a part of these medical treatment measures or represents an essential component of the care of a dying patient is a matter of controversial discussion.[4] Active euthanasia, on the other hand, is defined as active and intentional medical intervention in order to accelerate the occurrence of death. This includes, for example, deadly injections and administration of medications which intentionally and immediately kill the patient. However, this common distinction is disputed. Since so-called passive euthanasia includes not only a (passive) omission, a nontreatment, but also the (active) removal of therapies, a legal differentiation of indirect and direct euthanasia has been proposed (Wassermann 1993).

Obviously, neither the kind of medical treatment (active/passive) nor its intention and consequences (direct/indirect) are sufficient for a sound, normatively relevant differentiation of euthanasia. In addition to an extensive legal attempt at classification according to type of treatment and success, motivation, condition of the patient, insight and ability to give consent, purposefulness, and potential group of people involved in the deed (Eser 1976), the following classification was suggested from the medical perspective. Euthanasia is divided into two main groups: (1) help in the process of dying and (2) help in order to die. Help in the process of dying means the help and relief which is given to the dying person, i.e. to the patient who already is in the process of dying. This can be subdivided into euthanasia without the shortening of life, by allowing a person to die, by shortening life as a side effect, and by the intentional shortening of life. In comparison, help in order to die means that support is given to a person who is not in the dying process, but who for other reasons no longer wishes to live. Here, a distinction is made between ending a life subjectively felt to be without value through assisted suicide or death on demand, and ending a life which, from the view of a third party, has become without value (Winau 1993). In addition, it must be noted that, according to the Christian religion, there is no objectively worthless life ("the sanctity of God-given life"), and the murder of mental patients under

[3] Vacco, Attorney General of New York, et al. vs. Quill et al., decision of the Supreme Court of the United States, June 26, 1997; Washington et al. vs. Glucksberg et al., decision of the Supreme Court of the United States, June 26, 1997.

[4] This distinction in medical ethics has practical significance in the treatment and care of patients in a persistent vegetative state. In Germany, under the jurisdiction of the Federal Supreme Court (BGH 1 StR 357/94), the room for discretion in decision-making has increased for relatives and physicians of incurable patients lying in comas, and the significance of the presumed will of the patient is emphasized.

National Socialism – referred to as "mercy killing" – was prepared for using this concept, which is why such "killing out of pity" is currently unthinkable in Germany from the physician's point of view (Lauter and Meyer 1992; Winau 1993).

On the other hand, in the Anglo-Saxon discussion of medical ethics, euthanasia is often differentiated not according to the treatment or intention of the physician, but on the basis of the patient's wishes. Voluntary euthanasia occurs through a third party following the express wish of a patient who is capable of self-determination. It can happen either through assisted suicide or through killing at the express request of the affected person, e.g. in cases in which the suffering person is no longer physically able to commit suicide (Singer 1984). Whether the actual act of killing is active or passive, direct or indirect, is less significant in this differentiation. The ethical legitimation emphasizes much more the right to self-determination of the afflicted person, whereas the fundamental distinction from the medical and legal viewpoint between treatment (active) and omission (passive) is seen as ethically insignificant and is therefore the subject of controversy.[5]

Nonvoluntary euthanasia refers to the killing of a person who is not able to understand the decision between life and death, e.g. patients in a persistent vegetative state or severely deformed infants. Euthanasia is always nonvoluntary if the affected person either never had the ability to choose between life and death or, as a self-determining individual, never expressed a wish regarding such a situation. Under certain circumstances, nonvoluntary euthanasia is ethically justified from a preference utilitarian perspective (Singer 1984).[6]

An interdisciplinary, international discussion about euthanasia from the viewpoint of medical ethics therefore cannot draw on a unified terminology. More frequently, the different definitions of terms are used in parallel, and the particular conceptual criteria for definitions must be defined. In the clinical practice of the various specialities of somatic medicine, issues concerning the medical ethics of refraining from or withdrawing treatment at the end of life play the dominant role. Since the majority of mental illnesses, with the exception of organic psychoses, do not lead to a natural death, fewer questions of so-called passive euthanasia arise in psychiatric practice. Psychiatrists are much more increasingly confronted with questions concerning the active ending of a life. According to the new euthanasia laws in the U.S. state of Oregon and in the Australian Northern Territory, a psychiatrist has to be brought in for consultation by a physician before the active ending of a life either generally or in any case of doubt. In the Netherlands, the Supreme Court requires the consultation of at least two psychiatrists before physician-assisted suicide in the case of psychiatric patients. Hence, in the following sections, the problem areas of death on demand and medical assistance with suicide ("physician-assisted suicide") will be addressed in more detail.

4.2.2 Death on Demand

In Germany, any active killing of a seriously ill person with the goal of eliminating unbearable pain is a criminal offense, even if it is done at the express, earnest wish of an autonomous person capable of giving consent (although death on demand fulfills the requirements for mitigation of a sentence, Sect. 216 *Strafgesetzbuch*, StGB[German Penal Code]). Only in extreme exceptions can an emergency be assumed, which may partly justify the act and partly excuse it or may lead the person concerned being exempted from punishment (Eser 1992; Herzberg 1996). In practice, however, this provision of criminal law is extremely rarely used in Germany, and a considerable number of cases presumably go unrecorded (Schreiber 1995).

The overwhelmingly majority of lawyers agree that human life is accorded exceptional legal significance. Demands for legal exemptions from punishment for killing another person, e.g. to make possible a death with human dignity, will come up against the argument of the danger of the "dam bursting" in the protection of life. According to this argument, the protection of life by the state would be overcome and human life would be at the disposal of others. No one, it claims, has both a right to life and a simultaneous claim to end this life as he or she pleases, and if exceptions to the ban on killing were permitted, the protection of the ill, the severely impaired, and those whose lives were coming to an end would be endangered. Elderly and suffering individuals might then be pressurized, directly or indirectly, to request their own death, thus actually serving the interests of a third party (e.g. relatives). Indirectly, the self-determination of the patient would in this way be limited due to the

[5]On the controversy about the moral equivalence of treatment and omission, see Beauchamp (1989), Pellegrino (1989), Rachels (1989), Birnbacher (1990a), Thomas (1993), Bartlett (1995), Cartwright (1996), and Fuchs and Lauter (1997b).

[6]In contrast, involuntary euthanasia is understood as the killing of a person who is able to agree to his or her own death but does not do this because he or she wishes to live or was not asked before the killing. In both cases, involuntary euthanasia is unanimously rejected (Singer 1984). Involuntary euthanasia is a contradiction in terms (*contradictio in adjecto*) and is actually murder. The protest against Singer is probably substantiated even in his own terms, since in Germany this is understood as exactly what the National Socialists did.

unrestricted granting of self-determination for the end of life by another. In addition, advocates of this view point out the practical problems relating to evidence that would arise if death on demand were to be liberalized (Schreiber 1995). In 1986, an alternative outline for a euthanasia law (Baumann et al. 1986) was published, which was worked on by leading lawyers in collaboration with physicians, but it did not lead to any changes in the law. In the draft, the option of not serving the sentence associated with killing on demand in situations of extreme suffering was proposed, in addition to a legal clarification of the mutually agreed discontinuation of treatment and the measures used to lessen suffering with their risk of shortening life (indirect euthanasia). A change in the criminal law can scarcely be expected in the future either, because the abuse of euthanasia in National Socialist Germany has made public discussion of euthanasia in Germany more difficult (Wassermann 1993; Schreiber 1995). In addition, the highest jurisdiction in the Federal Republic (the Federal Supreme Court) has so far been very restrained in extending the discretion that can be exercised by physicians regarding euthanasia.

The German Federal Board of Physicians first published their *Guidelines on Euthanasia* in 1979, and these were revised in 1993. The guidelines refer exclusively to "individuals who are dying," for whom "the irreversible illness or the traumatic injury is taking an unfavorable course and death will occur in a short time." Only in such cases is the physician permitted to refrain from taking further treatment measures. This refers solely to the omission or discontinuation of therapeutic measures, whereas an "intentional shortening of life by artificial interventions in order to accelerate the occurrence of death," even at the request of the patient, is currently rejected (Bundesärztekammer 1979). German medical conferences have repeatedly and decisively rejected the various public demands to legalize so-called mercy killing, most recently in 1996 in the face of developments in Australia and in the Netherlands. In the opinion of the Deutscher Ärztetag (German Medical Society), a physician cannot be obliged to fulfill a person's wish for "mercy killing." Such a demand would destroy the relationship of trust between physician and patient (Bundesärztekammer 1988, 1993). The recently revised guidelines introduced by the Federal Board of Physicians in 1998 for medical support for the dying also changes nothing in the present unambiguous rejection of active euthanasia by the German medical community.[7] In the traditional way in which the medical profession sees itself, any active killing of a patient, even at his or her request, is rejected, since it stands in direct opposition to the medical precept to help and do no harm. Medical participation in active killing would seriously change the inner stance and professional identity of the physician and his or her public image.[8] Death on demand by means of a lethal injection destroys the relationship of trust between physician and patient and is not compatible with the physician's respect for human dignity. Death demanded by a patient should much more be understood as a call for help in coping with death (Lasch 1985; see also Matouschek 1989; Dichgans 1992; von Lutterotti 1993). Petitions and public protests through physician initiatives make clear the overwhelming rejection of active euthanasia in Germany.[9] In view of the decriminalization of death on demand in the Netherlands, the dangers of abuse connected with this are being vigorously pointed out by psychiatrists. Death on demand can be influenced by strong external factors, e.g. interests of other people, or by mental illnesses, which can influence the self-determination of the affected person. In the first case, societal opinions can be cited, as they find

[7]In the new formulation of the euthanasia guidelines, there is a greater consideration of patient directives and the value placed upon the individual, self-determining patients' wishes with respect to dying, and the question of indirect euthanasia, e.g. the cessation of artificial nutrition for patients in a persistent vegetative state (Bundesärztekammer 1998).

[8]The widespread, deeply rooted image which people have of physicians can be illustrated by several quotations. The German psychiatrists Fuchs and Lauter: "The price for the certainty of a quick death is the imposition made on another to carry out the killing. This price shall be paid exclusively by a profession, whose duty it is particularly to maintain life, heal illnesses, and lessen suffering" (Fuchs and Lauter 1997a). The WPA in its Madrid Declaration: "The first and noblest duty of the physician is to promote health, to lessen suffering, and to protect life ... The psychiatrist should realize that the opinions of a patient in a mental illness such as depression could be distorted. In such situations, it is the task of the physician to treat the illness and not to contribute to the death of the patient" (World Psychiatric Association 1998). The American anthropologist Margaret Mead: "The job of the physician, to save life, is of priceless value to humanity and must be protected against the constant attempts to harness the physician into activities of death. The public must be careful that this responsibility of the physician for life is kept. Measures and techniques, such as get discussed under the heading of euthanasia, must remain under the initiative and control of the laity – the medical profession should not be compromised by participation in them" (Mead 1963).

[9]See, e.g. the advertisement in *Die Zeit* on May 5, 1995: "Physicians' Initiative Against Active Euthanasia," the joint action "Europe Against Euthanasia" of the German Hartmann Association and the Dutch Medical Society (NAV) in 1996, or the actions of the German Society for Social Psychiatry and the "in part very emotionally led discussion of different population groups" (Bundesärztekammer 1997). On the Singer debate, see also Vollmann (1989), Hegselmann and Merkel (1991), Schöne-Seifert and Rippe (1991), Singer (1990), Singer and Kuhse (1994), and Vollmann (1996b).

expression in the present euthanasia discussion. Direct or indirect moral pressure can be engendered in expectations made on the affected person, who in light of present economic bottlenecks, e.g. in the medical treatment and care of older people, in effect place the interests of a third party over what are only presumed to be individual wishes. Second, illness-determined influences, particularly mental disorders, e.g. depression, can change the way a patient expresses his or her will, so that there is no real self-determined will of the affected person (Barocka 1992; Helmchen 1992; Lauter and Meyer 1992). With these arguments, as well as reference to the euthanasia debate in the Weimar Republic, in which influential psychiatrists (e.g. Hoche) participated, and haunted by the catastrophic consequences for the mentally ill during National Socialism, psychiatrists in Germany are arguing against any step toward a liberalization of active euthanasia. The demand for euthanasia in the 20th century may spring from the liberal demand for more self-determination, emancipation, freedom, and progress, mixed with an almost religious belief in science and the possibility of abolishing human suffering, and from the fact that death is a taboo. The approach which derives from that – quantifying human suffering and regarding a life as no longer having human dignity or being worth living above a certain level – is a dangerous continuation of earlier demands for euthanasia, which inevitably led down a slippery slope away from the protection of human life and toward a form of medicine which despised and destroyed individuals. It is on the very basis of the murder of mental patients with the cooperation of psychiatrists under the National Socialists that the new euthanasia movement in Germany needs to be opposed firmly and unambiguously, especially by physicians (Dörner 1988; Bastian 1990; Lauter and Meyer 1992; Daub and Wunder 1994). In contrast, Anglo-Saxon psychiatrists are increasingly placing the patient's right to self-determination at the end of his or her life in the foreground and are calling for the legal regulation of active euthanasia. Psychiatrists should participate actively and play an important role in the development of the controls required here (Helme 1993).

4.2.3 Physician-Assisted Suicide

Suicide is legally understood as a conscious and deliberately chosen killing of oneself by means of a definite and purposeful action. Accordingly, a patient who refuses what is probably a life-saving treatment and hence gives free rein to the deadly course of the illness is not considered to be attempting or committing suicide (Kaiser 1992). In German, the expression *Selbstmord* ("self-murder") is frequently used in addition to the word "suicide." However, on the basis of the legal situation in murder, the term is incorrect and normatively prejudicial. Therefore, instead of "self-murder" the term *Selbsttötung* ("self-killing") is more correct (Wassermann 1993). In contrast to other countries,[10] in Germany, neither suicide nor the incitement and assistance to perform suicide are criminal offenses (Wassermann 1993, Schreiber 1995). In medical practice, however, problems frequently arise in defining and drawing distinctions. On the one hand, the seriousness of the death wish needs to be considered. Has the person concerned really openly and autonomously wished for death, or is it much more a case of an attempted suicide, intended as a cry for help? This is frequently a difficult differentiation to make and brings with it the question of the ability for a self-determined will, which can be restricted to a great extent by influences from outside or by mental illness. On the other hand, is the person contemplating suicide faced with a society (or medical profession) that overwhelmingly rejects his or her behavior and views it as socially deviant? Legal questions result for clinical practice as to the measures that are allowed or even medically required for the prevention of suicide, if necessary even against the expressed will of the person. Physicians can thus be caught in an ethical and legal dilemma between euthanasia desired by a terminally ill patient, assistance with suicide desired by someone who is seriously suffering (respect for patient self-determination), and the failure to give assistance (professional duty of the physician). In clinical practice, legal problems frequently exist relating to the distinction between assisted suicide, which is permitted, and the criminal offense of killing on demand (Sect. 216 StGB). According to the highest judicial interpretation of the Federal Supreme Court, the question of who actually controlled the act leading to death is considered to be the crucial factor.[11] In spectacular cases of active

[10] Under Austrian criminal law, anyone involved in suicide, and under Swiss law anyone involved out of selfish motives can be sentenced. In England and Wales, suicide was a criminal offense until 1961 (Kaiser 1992). Since the Suicide Act of 1961, suicide is no longer a criminal offense, although a third party who takes an active part in a suicide is liable for prosecution. In contrast, in Scotland participation in suicide is not a criminal offense (Neeleman 1996; Eser and Koch 1991). According to the Dutch penal code, not only death on demand, but also assisted suicide is a criminal offense (Arts. 293, 294 nlStGB). The wording of the law, valid since 1886, speaks in favor of an unrestricted ban on killing, whereby little room remains for euthanasia. The present liberal practice of euthanasia is made possible by a broad interpretation of general reasons for exemption from punishment and by the decision by criminal prosecution authorities not to prosecute in all cases (Scholten 1991).

[11] See BGHSt 19, 135.

euthanasia on behalf of severely physically disabled patients wishing to die, the public has been very much made aware of how difficult it may be to differentiate between assisted suicide, which is not a criminal offense, and killing on demand, which is (Kaiser 1992).[12]

The criminal assessment of noninterference in a suicide is even more complex. If a person committing suicide becomes unconscious and therefore loses control over the course of events he or she has set in motion, the attending physician, who has a particular obligation toward this person, can be charged with the criminal offense of murder by omission if he or she does not provide the help that is necessary, possible, and reasonable at this stage at the latest (Kaiser 1992). Court decisions generally regard suicide as an accident that needs to be averted, even if the person concerned has acted on his or her own free responsibility. As a result of these decisions passed by the Federal Supreme Court, three constellations arise in the physician's clinical practice:

1. If a patient is under medical or even psychiatric treatment for suicidal tendencies and a suicide attempt is the result of the patient's lack of clear responsibility, the physician has a duty to prevent the suicide. The same applies to minors or patients with an impaired ability to determine their own will.
2. If the physician encounters an emergency situation in which an unconscious patient unknown to him or her has attempted suicide and there is no additional relevant information, the physician must assume that the suicide is "pathological" and take appropriate life-saving medical measures.
3. If, however, an apparently clear desire for suicide is known to the physician, e.g. as a consequence of previous information from the patient or through a patient's directive, according to the decisions taken by the Federal Supreme Court, the physician may not merely comply with the patient's wishes alone. Instead, he or she has "to make a responsible decision on whether or not to intervene, even if such an intervention will not necessarily be successful."[13]

This legal decision is controversial, however, among German criminal lawyers and has come in for sharp criticism in some cases. In differentiating between assisted suicide (not a criminal act) and killing on demand (a criminal act), according to the Federal Supreme Court what matters is that the person attempting suicide does not go to a doctor to tolerate death by the doctor's hand, but that this individual remains in charge of events until the very end. The prevailing legal opinion regards it as decisive whether the event ultimately proves to be an action arranged by the person concerned on his or her responsibility or one arranged by another person (Kaiser 1992; Wassermann 1993). A further point open to criticism in the decision passed by the Federal Supreme Court is the classification of all suicide attempts as accidents, even if the suicide wish was reached autonomously and on the patient's responsibility and the physician has given the patient assistance exempt from punishment beforehand (Schreiber 1995). This decision declares the will of the patient to be irrelevant. However, the decision concedes to the physician an autonomous discretion to appraise the situation, in which the patient's right to self-determination represents only one of several factors (Eser 1985; Gropp 1985). This is even more astonishing, as the Federal Supreme Court rulings in the case of so-called normal patients usually rejects the idea that the physician has a professional duty to prolong life following the onset of unconsciousness if permission to abandon treatment has been granted (Schreiber 1995). This inequality in the way that patient autonomy is treated by the highest court for a self-determining suicidal patient and a so-called normal patient does not appear justified, because the basic law of an ideologically neutral country does not derive from prior assumptions, however well-founded, but from the right to self-determination of the individual citizen (Wassermann 1984, 1993), while in the Federal Supreme Court's decision "the right to self-determination of the person wishing to die ... falls by the wayside" (Gropp 1985).

The Federal Board of Physicians goes beyond these criminal legal requirements by condemning any involvement by a physician in suicide as "nonmedical" (Bundesärztekammer 1993). German medical conferences have repeatedly described interventions to end a life and assisted suicide at the wish of a patient as professionally unethical, without acknowledging the ethical differences between the two acts (Bundesärztekammer 1988). The rejection of any form of

[12] A good example is that of the surgeon Dr. Hackethal, who caused a sensation in Germany the mid-1980s by helping seriously ill patients commit suicide by giving them access to poison. Hackethal was acquitted by the Higher Regional Court of Munich in 1987, because his actions did not exceed assisted suicide and thus did not constitute a criminal offense (OLG Munich NJW 1987, 2940). This decision acquires fundamental significance since, in contrast to the decision by the Federal Supreme Court, it gives priority to the patient's right to self-determination, even for suicide (Schreiber 1995).

[13] BSHSt 32, 367, 387; decision of the Federal Supreme Court of July 4, 1984 in the case of Dr. Wittig (Az. BGH 3 StR 96/84). The physician did not initiate any life-saving measures for an unconscious patient following ingestion of medication in order to commit suicide, since he was aware from conversations and a letter from the patient of her wish to die, respected this wish, and considered medical recovery very unlikely on the basis of the severe intoxication.

physician-assisted suicide is in accordance with the revised Swiss guidelines on euthanasia and with the stance of the WMA.

This rejection is even more understandable in psychiatry, because suicidal potential and mental disorder frequently occur concurrently in practice (see Vol. 3, Part 2, Chaps. 9, 10). Moreover, in Germany, psychiatry is burdened with the experience of the involvement of psychiatrists in the registration, systematic sterilization, and murder of mental patients under National Socialism. On the basis of this historical experience, it is understandable that the overwhelming majority of German psychiatrists reject any form of active euthanasia, including physician-assisted suicide (Helmchen 1986; Dörner 1988). Medical ethos obliges all psychiatrists to attempt to prevent the patient's suicide in the doctor–patient relationship. The one-sided overvaluation of the autonomous freedom of decision could all too easily lead to indifference on the part of the physician. Instead, suicide prevention presents a humane duty incumbent on medicine, especially psychiatry, that should not be questioned by the propagation of the right to commit suicide (Bron 1986). In difficult economic times, in particular, what is presumed to be self-determination on the part of the individual can become the suicide of the sick, disabled, and elderly under the influence of social and economic factors and societal pressure (Heinrich 1992; Lauter and Meyer 1992). With such purposefully weighed and progress-oriented thinking, the danger exists of inhumane elimination of so-called useless and unproductive individuals, as in National Socialist Germany, "the final solution of the social question." Today the danger exists again in Germany that mentally ill and weak individuals will be isolated by the decline of moral and material solidarity. Thus many regard the present euthanasia debate as taking a dangerous, misguided course in social, ethical, and political terms (Dörner 1988).

The mere possibility of abuse is an insufficient argument to counter this rejection of physician-assisted suicide, which stems from the historical experience of abuse. The argument frequently cited in this connection, the "slippery slope,"[14] (Fuchs and Lauter 1997b) claims that making physician-assisted suicide permissible would mean that the active killing of the mentally ill could no longer be controlled at all. If the previous medical taboo against killing falls away, all mentally ill patients are at risk of being killed by their treating psychiatrists. The Dutch development is cited as evidence, according to which the number of deaths on demand in recent years has increased considerably and the concept of "unbearable suffering" has been extended, so that within 1 year 0.8% of all deaths were attributed to nonvoluntary euthanasia. Philosophical authors, in particular, but also physicians in the Netherlands and the United States argue that the danger of the "slippery slope" does not hold on the following grounds: through prevailing legal circumstances, which permit a physician-assisted suicide in a democratic society, it is possible to publicly control the current practice of physician-assisted suicide of severely ill patients. Different social groups, including the medical profession, are in a position to draw up exact guidelines with which to prevent an abuse more effectively than could be done previously (high number of unrecorded cases). Accordingly, American physicians have proposed concrete criteria and regulations for physician-assisted suicide (Quill et al. 1992; Miller et al. 1994). Especially in comparison to active euthanasia, physician-assisted suicide would be less in danger of abuse and could, as the Dutch example shows, be accepted by a large majority of people. Due to the fact that it is "only" a physician-assisted suicide, the patient retains decisive control over the event, which is simply competently supported by the physician. Hence physician-assisted suicide represents a form of voluntary euthanasia (see above) in which the dignity and autonomy of the person is also respected in the last part of life (Birnbacher 1990b; Ach and Gaidt 1994; review in Schöne-Seifert 1996, pp. 604–613). The frequently expressed argument that physician-assisted suicide would damage the integrity of medicine can also be countered by the argument that this form of euthanasia for severely suffering, dying patients represents medical help shaped by a respect for humanity (Momeyer 1995; van der Maas et al. 1996; van der Wal et al. 1996; for reviews, see Fins and Bacchetta 1994, Groenewoud et al. 1997). Of course, it must be considered that, in the special case of psychiatry, physician-assisted suicide by a psychiatrist would fundamentally change the present image of the profession and the way it views itself. Based on their clinical experience as psychiatrists, the authors doubt that medical responsibility to prevent suicide in mentally ill patients is compatible in practice with an active support of patient suicide.

In this context, the case of psychiatrist Dr. Chabot, who helped a depressed patient in the Netherlands to commit physician-assisted suicide, has been critically discussed in psychiatry (Fuchs and Lauter 1997b).

In the autumn of 1991, the psychiatrist Boudewijn E. Chabot helped his 50-year-old patient Netty Boomsma to commit suicide by providing her with a deadly dose of medication, which Mrs. Boomsma took in his presence. The patient, a social worker, no longer

[14]The terms "bursting dam", "stray path," and "inclined plane argument" are used synonymously in German for the English expression "slippery slope argument" (see Guckes 1997).

wanted to live following her divorce and the death of both her sons (her older son committed suicide in 1986 aged 20, and her younger son died of cancer in 1991). Since both her children had been the focus of her life, she saw no hope and meaning for the future whatsoever, viewed her life as hopeless and without any prospects, and repeatedly expressed to her psychiatrist the wish to die. Dr. Chabot had a number of psychiatric conversations with Mrs. Boomsma over a period of about 5 weeks, totaling about 24 h. He diagnosed a pathological grief reaction (adjustment disorder DSM-III-R), which had existed for 5 years, with predominantly depressed mood, but without the presence of a psychiatric illness in the narrow sense, e.g. a severe depression or personality disorder. The mood of the patient was overwhelmingly depressed, although it was capable of change, so that the patient could exhibit self-perspective and humor. Mrs. Boomsma's ability for self-determination was not impaired by her mental state in Dr. Chabot's judgment. She refused both psychopharmacological and psychotherapeutic treatment aimed at making her able to grieve, because she knew from the death of her first son what this entailed and was not prepared to go through it again. Although they had no personal contact with the patient, the four colleagues consulted by Dr. Chabot agreed with him on the basis of the patient's records and his report that no actual psychiatric illness was present, but that the patient suffered persistent and unbearable mental torment. The patient did not suffer from a "hopeless suffering" in the sense of a severe, physical illness leading to death, as euthanasia practice in the Netherlands had required up to that point, nor was she in a physically life-threatening situation.

In June 1994, the Supreme Court of the Netherlands found Dr. Chabot guilty, but refrained from prosecution. In the reasons given for its decision, the court stated that the psychiatrist was not sentenced because of his assistance in his capacity as a physician in the suicide of a mentally ill woman, but because he did not observe the operative rule of a personal consultation by a professional colleague. In this context, the Dutch Supreme Court stated that a physician could also cite an emergency in circumstances in which a patient was not physically suffering and thus not dying. This would justify physician-assisted suicide, since here the extent and prognosis of the suffering and not its cause (physical/mental) were decisive. In the present case, the court recognized for the first time the hopelessness of the situation of a mentally ill patient and explicitly approved physician-assisted suicide. In 1995, a disciplinary court for physicians in Amsterdam reprimanded Dr. Chabot because he had not attempted sufficient psychiatric treatment of the patient. The psychiatrist was permitted to continue to practice. This decision of the disciplinary court, composed of a judge and four physicians, should cause him to "reflect."[15]

Case-related psychiatric comment is difficult, because the facts are known exclusively from Dr. Chabot's point of view and the published court judgment. In spite of the restricted basis for judgment, from a psychiatric viewpoint, both the uncertainty with respect to the assessment of the grief reaction as having the status of an illness (not a real mental illness, but requiring treatment) and the short observation period of only 5 weeks in a depressed patient and the short period of 4–5 months after the death of the second son for the loss determining the abnormal grief reaction are striking. As a rule, this short period is not sufficient for an adequately reliable psychiatric prognosis of outcome. Furthermore, it must be asked whether a (renewed) psychotherapeutic and psychopharmacological treatment trial should not have taken place immediately before the physician-assisted suicide. This requirement seems all the more worthy of consideration since a pathological grief reaction with depressive mood can be successfully psychiatrically treated (Fuchs and Lauter 1997b). On the other hand, it does not appear appropriate in view of the respect demanded for the right to self-determination, since the patient had clearly rejected all forms of treatment. However, the decisive issue of the ability to give consent remains open and is only emphasized by the uncertainty with respect to the evaluation of illness and need for treatment of (as well as the ability to treat) the abnormal condition.

In addition, the case raises fundamental questions about the doctor–patient relationship in mental disorders and in relation to the way in which the psychiatric profession is understood. In clinical practice, a suicidal tendency usually presents as a temporary symptom of a mental illness. With improvement in the state of mental health, the suicidal tendency disappears in the majority of cases. From this psychiatry derives a general treatment and protective role toward each patient wishing to commit suicide. According to Fuchs and Lauter, the psychiatrist must take a clear position against suicide and for the life of the patient in each case. Otherwise, a therapeutic relationship is not possible because communication with the suicidal patient is established from the outset on a contradiction. The psychiatrist, as "the final representative of life and the community" and "the affirmation of life," therefore has to carry out the task "as a person who makes the patient able to experience his or her own

[15]On the judgment in the Dr. Chabot case, see Oglive and Potts (1994), Griffiths (1995), Klotzko (1995), and Fuchs and Lauter (1997b).

value again" and to "find a way back into the community with the patient" (Fuchs and Lauter 1997b). However, it remains to be examined to what extent a teleological mind-set lies at the basis of this assessment of suicide as an illness-dependent action directed against one's own worth, nature, and the community. There are certainly justifiable reasons for this, but in a secular and pluralistic discussion on the medical ethics of values, no universal validity can be claimed. In addition, the value of the patient and of his or her place in society is postulated without addressing the decisive argument to justify suicide, the individual's right to self-determination, in a differentiated way. The fundamental problem of the medical ethics of physician-assisted suicide lies between the patient's right of self-determination (principle of autonomy) on the one hand and the duty of the psychiatrist to foster the well-being of the patient (principle of beneficience) and not to cause harm (principle of nonmaleficence) on the other. Both arguments therefore have to be thoroughly analyzed and carefully considered, which is not possible if suicide is devalued a priori (Momeyer 1995; Diekstra 1996).

From the psychiatric point of view, a real practical problem in the assessment of the medical ethics of physician-assisted suicide lies in establishing as securely and objectively as possible the autonomous will of the patient. The problem for the psychiatrist in terms of medical ethics and the law is less the psychiatric diagnosis (nosological assignment), and more the judgment of the patient's ability for self-determination. The Supreme Court of the Netherlands established in the Chabot case that the ability for self-determination can be destroyed or impaired by both mental and physical disorders, and the differentiation between physical or nonphysical disorders is thus not relevant in assessing whether assisted suicide is permissible (Griffiths 1995). Above all, psychiatric experience and knowledge have shown that the ability to give consent are not necessarily impaired even in patients with mental illnesses. This is confirmed by recent empirical results of psychiatric research on informed consent (for a review, see Vollmann and Helmchen 1997). Consequently, the presence of a mental disorder does not automatically rule out a self-determined and voluntary wish to die.

The evaluation of informed consent in assisted suicide is extraordinarily difficult in depressed patients, because the unambiguous establishment of what the patient "really" wants is, in practice, often hardly possible. Criteria for establishing the presence of self-determination that have a clearly defined content, are differentiated, and are generally recognized are lacking. In addition to the multitude and poor definition of basic criteria (criterion variance), there is also the problem of the determination of these criteria in a particular patient in clinical practice, as different psychiatrists can arrive at different assessments (observation variance; Helmchen 1992). In a mental disorder, the self-determined request for assisted suicide can be impaired by external and internal factors, such that an autonomous decision by the patient is not possible. While human decisions can in general be influenced by a multitude of external factors (e.g. opinions and values of relatives, treating physicians and caregivers, socioeconomic factors, media, the "spirit of the times"), in depressed patients the existential dependence upon the environment can reach such an extent that a request determined by another person can be masked by an expressed suicide wish. (The same naturally applies to the reverse case of a suicide wish that is made taboo by relatives or psychiatrists.) In this context, the ethically paradoxical and dangerous situation arises that, by calling for permissibility, availability, and feasibility of suicide through individual self-determination, the fundamental inviolability of human life would be abandoned and human life would thus become vulnerable not only to one's own, but also to other interests (Helmchen 1992). As an example of the particular impairment of self-determination by external factors, the so-called suicide epidemics can be mentioned, e.g. in religious sects or in the wake of publications such as Goethe's *Die Leiden des jungen Werther* or presentations of suicide in television programs (Schmidke and Häfner 1986; Simkin et al. 1995).

In addition to these external factors, internal ones, such as mental disorders, can impair the ability for self-determination. In a large empirical study, 23.9% of depressed patients were not able to give their consent in therapeutic decisions (Grisso and Appelbaum 1995). Since the threshold for the ability for self-determination in an assisted suicide must be placed higher than with therapeutic measures, it can be assumed that the number of depressed patients unable to give consent is even higher. In addition, a connection between depressive symptoms and the wish for physician-assisted suicide in patients in the final stages of physical illnesses has been proven in empirical studies. In these patient groups, 58.8% of the patients with a desire to die suffered from depression, in contrast to only 7.7% of those not wishing to die, whereby the desire to die fluctuated strongly over time (Chochinov et al. 1995). The inner impairment of a self-determined request to commit suicide correlates with the severity of the depression and is largely independent of its medical genesis (e.g. reaction to serious physical illness or in the framework of a grief reaction, as in the Chabot case, versus "endogenous" or organic brain causes). It is, however, unclear how the degree of depressive psychosis can be established independently of symptoms of physical illness and suicidal tendency,

and it has not been established at what stage of the development of depression the self-determined desire for suicide becomes questionable (Brown et al. 1986; Cassel and Meier 1990; Conwell and Caine 1991; Ganzini et al. 1993; Pohlmeier 1995). These difficult, complex questions require comprehensive, interdisciplinary, psychiatric, ethical, and legal study. Until the unsolved questions of the ability for self-determination of patients with psychiatric disorders desiring suicide have been explored more precisely and empirically, assisted suicide is not ethically defensible from the psychiatric point of view.

5
Common Ethical Problems in Psychiatric Practice

The ethical principles presented at the beginning are essential not only for the assessment of the particularly important problems with which the public is mainly preoccupied, for which two examples were discussed above. Consideration of these principles is also fundamentally significant for medical activity in daily practice. Although the principles are probably usually considered, this is done so more implicitly than consciously. Indeed, this might not always suffice to protect the physician against the pressures of the immediate situation, against personal motivation that is not considered, and above all against dangerous requests resulting from the spirit of the times, ideologically dominated on the one hand and pluralistically noncommittal on the other. In such a context, ethical principles can become unclear or lose their binding character, especially when the physician has to make decisions in a necessarily broad gray zone of individual discretion between the norms of what is unambiguously correct and the what is certainly wrong. The traumatic insights gained in analyses of the practice of the National Socialist euthanasia program, which showed a wide spectrum ranging from active resistance to such murder of the mentally ill to its support, may provide a warning in this context against too much self-assuredness (Schmuhl 1987; Kersting 1996; Helmchen 1998b).

The principle of autonomy as a general human right cannot be accepted without question in a medical context, since it can clearly come into conflict with the original principle in medical ethics of well-being (*salus aegroti*). This significance is not readily apparent in the now diverse practice of the courts either, which imposes an obligation on the physician to inform his or her patient of the nature of any planned treatment, since only an adequately informed patient can make a legally valid decision to consent to a medical procedure. Instead, it only becomes clear if this duty to inform becomes a critical component of medical treatment and facilitates the patient's ability to make an autonomous decision, thereby investing the doctor–patient relationship with trust and pointing out to the patient his or her own individual responsibility.

While the danger arising from laying the sole responsibility ultimately on the patient is now seen more often, paternalism was above all a risk of the previous psychiatric practice. This constitutes the most important problem specific to psychiatry that adherence to the principle of self-determination entails. It is obvious that mental illnesses – more than physical illnesses and even more than the emotional distress caused by the vicissitudes of life – can impair or even destroy a person's ability for self-determination, his or her ability to understand and want something. Since its beginnings, psychiatry has had to devote itself to this illness-related limitation, the loss of self-determination, with its recognition and even more so with its consequences, such as danger to self or others, compulsory admission and treatment, capacity to make a will, legal competence, recognition of guilt, and ability to provide consent (see Chaps. 15–17, this volume). In addition to the relevant legal questions, the physician is always called upon in a particular case to recognize correctly the boundary between the ability for self-determination which is already limited and that which is still retained in order to avoid the harmful consequences of a false assessment. This applies to an even greater extent in the case of patients who are clearly psychiatrically ill, who in many respects can be assumed to have retained the competence to make decisions and who, incidentally, should not be patronized and discriminated against.

5.1
Necessary Treatment Without Patient Consent

If a psychiatric illness leads to loss of the ability to make decisions, i.e. a loss of inner freedom with the consequence of presenting an immediate danger to self or others, then the patient must be prevented from presenting such a danger by withdrawing his or her external freedom, unless less radical measures are able to remove the danger. Such confinement against the will of the patient is legal; from the medical point of view, in addition to averting danger by physically preventing it, it aims primarily at eliminating the causes of the danger by treating the underlying psychiatric illness. Problems usually arise if the danger entailed by the illness is not so clear that the use of force appears to be necessary if the patient rejects the indicated treatment (Helle 1993). The physician is then confronted with the ethical question of how best to undertake the obligation to do everything for the well-

being of the patient without disregarding his or her autonomy, even if this autonomy is reduced on account of the illness. Many psychiatrists now patiently try to convince the patient of the necessity for therapy. However, discharging such a patient without treatment (although certainly not without the offer of resuming treatment at the wish of the patient), a procedure that is now often followed, makes many psychiatrists doubt whether they have correctly assessed the risk of extending the patient's suffering and the danger related to the illness against the risk of destroying by force any opportunity to develop trust on which to base voluntary treatment. Thus the question as to the influence that society-based or other concrete situation-specific conditions connected with the physician's discretionary powers have on his or her appraisal of risks and benefits, e.g. if the patient's motivation for therapy can be described as paternalistic manipulation or if the explanation of alternatives, such as nonvoluntary committal, are experienced by the patient as threats. The same is true if, in so-called psychiatric testaments, the psychiatrist is threatened with punishment, e.g. a lawsuit based on breaking confidentiality, if he or she wants to establish a clear legal basis for the necessary treatment, by submitting an application to a court to issue a mandate for treatment. This also applies to situations in which, in light of the tighter health care budget and the demands for a more just distribution of resources, the length of hospital stays are strictly limited and hospital stays that are not therapeutically useful are not paid for. In each case, the psychiatrist should be clear about the arguments determining the ethical considerations in order to win the trust of the patient through competence and responsibility or to justify the existing trust. Incidentally, the psychiatrist is also well advised to document the arguments comprehensively, since ethical decisions can be queried after the fact or even in court.

5.2 Informing Patients About Tardive Dyskinesias

Tardive dyskinesias are motor disorders that first appear late in the course of long-term treatment with neuroleptic drugs. They are its most serious risk because they are relatively common, only partially reversible, and so far difficult to treat. Reliable predictors of the risk are not known. However, the effectiveness of the long-term use of neuroleptic medication as a prophylaxis against relapse and in symptom suppression is clear. Useful individual response predictors are, however, not known either, with the result that the therapy response to long-term neuroleptic therapy in each patient can fluctuate between complete effectiveness without undesirable side effects and inadequate effectiveness with tardive dyskinesias. In view of the uncertainty of predictions of individual risk in an established therapy whose effectiveness makes it the standard one, a particularly careful appraisal of benefits and risks and informed consent are necessary at the beginning of long-term medication. It is a question of quality assurance not to let an acute neuroleptic drug treatment turn into a long-term one without comment, but rather to establish the indication independently and to explain this new treatment to the patient, including the different goals and risks; in addition, the course of treatment needs to be followed closely, examining therapeutic effectiveness, looking out for undesirable side effects of the medication, especially tardive dyskinesia, and checking that dosage is appropriate. Should tardive dyskinesia then appear, the benefit–risk appraisal needs to be repeated in the continuing course of treatment and alternative therapies considered anew, even if all psychiatrists probably know patients in whom neuroleptics cannot be discontinued despite the appearance of tardive dyskinesia, because the suffering experienced by these patients if their psychosis were to remain untreated (or if atypical neuroleptics were administered that were unable to treat the psychosis adequately) would be incomparably greater than that caused by the tardive dyskinesia (Helmchen 1991).

Fundamental criteria indicating long-term medication with neuroleptics are the persistence of symptoms or a considerable risk of relapse. Understandably, in the former case, the subjective pressure of suffering often motivates the patient to turn to symptom suppression using long-term medication; moreover, relief is usually experienced immediately. However, a patient will usually only decide in favor of long-term medication on the basis of its ability to prevent relapse if he or she has already experienced a traumatizing relapse. As a last resort, if frequent relapses owing to noncompliance put the patient at risk of long-term hospitalization, a neuroleptic depot treatment is indicated in order to compensate somewhat for the fluctuating insight on the part of the patient that the treatment is necessary. Particularly with such patients, as a result of their variable limited ability to give consent, information and self-determination sometimes have limits. Even if the majority of acute psychotic patients who have been treated against their will later consider this treatment to be correct after their symptoms have subsided (Schwartz et al. 1988), the physician will still need to examine closely in each case the particular conditions of an individual patient for indications for a neuroleptic depot treatment based on noncompliance. For example, if noncompliance is caused by a rejection of treatment because the patient is afraid of developing tardive dyskinesia, the physician will be more likely to have to accept this than if

the noncompliance is a result of a delusion of poisoning. In practice, the psychiatrist is often faced with the following dilemma: on the one hand, after carefully assessing the risk of tardive dyskinesia, the motives for depot medication as the best alternative for the patient must be examined without crossing the boundary of paternalistic care for the patient; on the other hand, the psychiatrist may have to initiate a treatment program ordered by a court if a patient's ability to give consent is found to be limited, even if the effectiveness of such treatment is doubtful on an outpatient basis. It is the experience gained with such patients that substantiates the fear of many psychiatrists that too far-reaching an advocacy of the patient's right of self-determination or of a legally indisputable substitute for the decision-making powers which are impaired or destroyed by the illness is actually not in the interest of the patient (Gutheil et al. 1980).

5.3 Information and Justice in the Case of Suboptimal Therapy

The development of so-called atypical neuroleptics that do not carry the risk of tardive dyskinesia entails a further ethical dilemma for the physician: widespread prescription of these neuroleptics for long-term treatment runs counter to the rules imposed on physicians by insurance companies regarding the economic prescription of medication (with consequent threats of recourse), because these drugs are considerably more expensive than standard neuroleptics and therefore place an additional burden on the recently introduced budgeting program, i.e. they limit the possibility of providing other patients with optimal treatment.

Generally formulated utilitarian ethics are gaining in importance in comparison to egalitarian ethics. Compared with the egalitarian precept that all patients should be treated equally, guided exclusively by the needs in an individual case, a precept that has determined medical actions thus far, utilitarian ethics mean, for example, that the available budget is to be used in the best interests of all patients (for whom this budget is valid), even if that means that an individual patient does not receive optimal treatment (Smith and Morrissy 1994). At the same time, it cannot be ignored that, in view of the great need and failing resources in many parts of the world, such utilitarian considerations are inevitable and even triage decisions cannot be avoided. Thus the insufficient aid is only sent to a group that is selected on the basis of certain criteria, such as the best prognosis, as in the allocation of organs in transplantation medicine, for example.

At the same time, physicians face the question of whether to explain to individual patients the suboptimal treatment, allowing the patients themselves to attempt to obtain alternatives, or not to explain it because these alternatives are regarded as not being able to be realized for a particular patient and could therefore merely further weaken the motivation for therapy, which is frequently already limited in psychiatrically ill patients. However, in the assessment of alternatives for the patient, physicians must be conscious of the increasing danger that they may no longer be solely representing the interests of an individual patient, but also other interests (i.e. their own) (Smith and Morissy 1994), e.g. if physicians are given (even partial) control over the use of resources and hence over rationing through the cost controls of budgeting or "managed care," which are now being implemented everywhere (Schlesinger 1995). This can be seen from the example of the introduction of budgeting in 1991 for part of the medical general practice (FHP) in the National Health System (NHS) in England (Smith and Morissy 1994). A further question remains open in this context, i.e. whether the allocation of resources is more just if it is carried out by the physician, who is more familiar with the problems of the individual patient, by the politician determining the regulations governing this allocation, or by insurance company employees, who allocate resources according to the regulations in a more formal manner (Vollmann and Dörries 1996).

These questions have only been addressed in part here and are becoming increasingly important in view of the current reform of our health services into health economies (Tischler and Astrachan 1996). A comprehensive analysis of the resulting influences on ethical principles and foundations of medical care remains to be carried out.

5.4 Medications with Addictive Potential

Methadon has been used for a long time in Germany and is increasingly administered as a heroin substitute treatment for heroin addiction. Recently, heroin itself has also been proposed for these indications (its use as an analgesic has long been permitted in England; Strang and Gossop 1996). Although this has the advantage of lessening the individual's suffering, there is the disadvantage of continued dependency, even if this dependency is medically controlled. The necessary benefit–risk assessment corresponds in principle to that facing the physician for many other treatments (Strang and Gossop 1996). However, this assessment is complicated by two problems: doubt about the ability of the dependent person to give consent and the context of society. The physician cannot simply comply with the patient's desire to suffer less by relieving withdrawal symptoms, since it is not certain

that the patient can meet the requirements for valid consent by correctly assessing benefits and risks. The craving for the addictive substance, which constitutes the dependency, probably limits if not destroys the ability of the dependent person to make decisions. Theoretically, the necessary initiation of compulsory treatment in such a patient might run into considerable practical problems, in particular the threshold of effort required by the dependent person for such a treatment might become considerably higher. As for the second problem, the physician has to verify the degree to which the social objectives associated with substitute treatment come into conflict with the individual well-being of the patient. The welfare of the patient and his or her freedom from suffering is understood much more in the sense of the social goals of decriminalization and resocialization and defined in terms of health policy as serving the common good, so that the unlimited implementation of a substitute treatment, while maintaining public order, cannot aspire to liberate the patient from dependency by achieving abstinence.

However, the social milieu can also tempt the dependent person to abuse substitution therapy, in particular leading him or her to misunderstand total substitution with methadon or even heroin as a "hedonistic signal." The nontherapeutic use of psychotropic substances is certainly widespread in our society, with substances ranging from tobacco, alcohol, hashish, and amphetamines to LSD, heroin, cocaine, and designer drugs such as ecstasy (Gouzoulis-Mayfrank et al. 1996), substances which can be used for relaxation, for pleasure, and for the intensification of mental experiences and achievements, in short to enhance the quality of life. Such psychotropic hedonism as an expression of personal freedom is the opposite pole from a pharmacological calvinism (Klerman 1972), which argues against apparent solutions to complex personal and social problems using psychopharmacological products and urges social control of these substances because of their prevalent and long-term negative consequences, including dependency. The physician therefore has to deal with the question of how to handle people's requests for drug prescriptions, which they seek to combat tiredness, listlessness, bad mood, anxiety, poor ability to concentrate, sleeplessness, a physically poor condition, etc., problems, however, which do not correspond to a diagnosable illness, but are probably connected with a perceptible life stress. Naturally, the question can be raised whether the physician is responsible for eliminating subjective complaints that have only a tenable link with illness just because people seek a doctor's help (Häfner 1997). The situation has become unclear for the physician since 1947, when the WHO proclaimed subjective well-being as a health goal. This effectively pushes everyday distress into the domain of medical diagnoses and has fostered the use of psychiatric treatment in problems of daily living. For this reason, a "social profile" of a drug next to its "biological profile" can be anticipated (Winckelmann 1989).

If a physician affirms his or her responsibility for such treatment, the question regarding appropriate measures must be answered. Naturally, the physician cannot be forced to prescribe a medication against his or her better judgment, no more than a patient can be forced to take it. However, to what extent should physicians be prepared to make compromises on what they regard as the best mode of treatment on the basis of their knowledge and experience, e.g. consultation or psychotherapy, in order to comply with the wishes of the patient? Which criteria should be used to make such decisions? Merely because the patient who is seeking treatment and perhaps also needs help would otherwise receive no treatment? Or because the physician might lose the patient? The duty to inform the patient can be particularly helpful for the physician in this context, as this duty does not merely cover any possible undesirable side effects (including abuse and dependency), but also treatment alternatives, such as psychotherapy or measures involving social therapy. A conversation to inform the patient will then contain elements of consultation and can itself bring significant relief.

5.5 Psychotherapy

Psychotherapy illustrates more subtle restrictions to the self-determination of the patient, but ones which might have greater consequences. Psychiatric therapies in general, especially if they are effective, can have a considerable impact on the patient's life. That is also true of certain forms of pharmacotherapy, as can be shown, for example, by changes in family relationships in the course of successful lithium prophylaxis in preventing relapse in patients with affective psychoses (Rüger 1977). However, this is more likely with psychotherapies, such as dynamically oriented psychotherapies, which not only aim at eliminating symptoms, but also examine the personality structures that produce these symptoms. For a long time, the ethical implications of these concepts seemed to play no role in psychotherapy, and the so-called humanistic therapies, in particular, were implicitly seen as ethically flawless. Only recently were these implications critically discussed (Bloch 1996).

A fundamental question is how much the therapy goals that are defined – not only in psychodynamic therapies, but also in other psychotherapies such as the deconditioning (and sometimes manipulative) techniques of behavior therapy – are influenced by the

therapist's image of the person? "Does it involve conformity, in order to adapt optimally to the social milieu, as if the meaning of human life consists in adaptation and conformity to others? Or is the goal the maximal development of the potential of the patient, so that the criteria for a healthy existence lie only in the individual person?" (Ritschl 1989). At any rate, therapists should continually review their "ethical counter-transference" with respect to their own values (Holmes 1996). This holds especially for psychoanalytic psychotherapies in which dependency on the part of the patient is necessary in the context of his or her "counter-transference neurosis."

In addition to the issue of goal setting, this highlights the asymmetric power relationship between the psychotherapist and the patient in therapy and is cause to reflect on the ethical use of this power, e.g. the balancing act between the desire to give the patient hope and the risk of being thought to be omnipotent (Karasu 1991; Helmchen 1998c). "Yes, the real trouble with psychotherapy lies in its non-explicit nature" (Goldberg 1977, cited in Karasu 1991).

Great demands are made upon the integrity of the therapist, not only because of "the proximity of this field of the healing arts to the sanction-free sphere" (Ritschl 1989), but above all to protect the therapist from abusing the dependency of the patient. This does not primarily mean sexual abuse, which according to psychotherapists themselves is committed by 5%–10% of therapists (Gartrell et al. 1986; Moggie et al. 1992) and is therefore nonetheless alarming, even if they themselves assess it differently. It refers instead to the psychological processes by which the therapist steps onto to this slippery slope (Gabbard 1996). However, the main focus are the many other forms of "emotional exploitation" (Birnbacher and Kottje-Birnbacher 1996), especially narcissistic abuse, under which all those interactions and constellations of relationships between therapist and patient are understood "which primarily serve the narcissistic gratification of the therapist and which hinder or at least make more difficult the development of the 'true self' of the patient" (Reimer 1991). Similar problems were revealed with so-called "recovered memory therapy," i.e. the search for specific suppressed memories (Merskey 1996).

This therapy, in particular, illustrates the need to explain to the patient before the beginning of psychotherapy the variety of possible undesirable effects and negative consequences, such as a psychopathological deterioration, the destruction of familial relationships, financial burdens, or the stress of legal disputes (Merskey 1996). In general, providing the patient with appropriate information before the beginning of a psychotherapy is fraught with difficulties, because it might already be part of the psychotherapeutic process itself, i.e. informing the patient that only a continuous change of personality can lead to an elimination of symptoms and, moreover, that the risk of failure might be associated with long-lasting negative effects as a consequence of unresolved transference or unresolved resistance; the patient's condition might even become worse than at the beginning of therapy. This risk seems especially high in patients with ambivalence, which lies at the core of their illness and might also be connected with a limited ability to give consent.

Some further considerations shall be briefly outlined. The observance of the duty to maintain confidentiality, which is absolutely vital in psychotherapy, can be jeopardized by psychotherapeutic consultative activity, in group therapy, during supervision of a psychotherapeutic training, or by publication of case histories (Vollmann 1996a). The ethical implications of breaking confidentiality if a (suicidal) patient presents a danger to him- or herself or to others were illustrated particularly clearly by the Tarasoff decision (Gurevitz 1977; Roth and Meisel 1977). A new ethical problem stems from the changes in our medical system, which make maintenance of confidentiality more difficult, since direct payment for psychotherapy is generally replaced by third-party (insurance) payment.

In addition to an unethical financial exploitation of patients, a further dilemma of increasing ethical relevance is currently developing as a result of "managed care" structures, since the psychotherapist is caught between the interests of the patient and those of the organization by which he or she is paid. Psychotherapists also increasingly have to answer questions concerning "the implementation of a professionally qualified and competent psychotherapy, the documentation of its effectiveness, the exact indication and scope, in which these conditions fall within the sphere of the medical model and which thereby fulfill the criterion of medical necessity" (Chodoff 1996). Prompted by the questionable effects of the recovered memory movement, the state of New Hampshire (USA) introduced the draft of a Truth and Responsibility in Mental Health Practice Act in 1995, according to which the psychotherapist has to ask for consent following explanation about the prospects of the psychotherapy which is offered. Otherwise, he or she cannot receive reimbursement from a third-party payer. The explanation preceding the consent has to include the following: the condition to be treated, the expected benefits, side effects, and risks of the treatment offered for this condition, the citing of at least two research publications showing adequate safety and effectiveness of the proposed psychotherapy, and accepted alternatives to the therapy. The beginnings of such a quality control in psychotherapy have developed in Germany by means of investigating the indications using an expert review process. The American Psychological Society recognizes in this

sense cognitive and behavioral therapies, interpersonal therapy, psychodynamic short-term therapies, and psychoeducative methods, but not, however, supportive psychotherapy, psychoanalytic long-term therapy, and group therapy (Mersky 1996). In order to avoid extreme regulations such as the legal draft mentioned, psychiatrists and psychotherapists must themselves develop, explicate, and conscientiously maintain high ethical standards. Such standards could be formulated as follows:

In view of the special (personal, nonrational) quality of the therapeutic relationship between the psychotherapist and patient, in consideration of the special difficulties in explaining the therapy appropriately to the patient before it begins in order to obtain the patient's consent, and in addition maintaining confidentiality under all circumstances and, not least, financing the psychotherapy in a changing health care system, the psychotherapist is called upon to develop sensitivity and to preserve it in face of the ethical implications of the view of human beings on which his or her therapy is based and against the use of power in the therapeutic relationship with its dependency of the patient upon the physician, especially with respect to the risk of emotional or narcissistic (and financial) exploitation of the patient.

Further examples could be cited, e.g. the ethical problems of genetic consultation, which is rapidly gaining in importance, e.g. in patients with Huntington's disease and their possibly presymptomatic relatives, in prenatal diagnosis, in genetic research into complex genetic psychiatric diseases such as schizophrenia and depression (see Chaps. 3, 4, this volume, Part 2; Chaps. 3, 18, this volume, Part 1), and the multiplicity of ethical questions which arise in dealing with those with age-related dementia (Berghmans 1997). It may be hoped, however, that the few examples already discussed in detail here have made clear the importance of ethical reflection in daily medical practice and will invite further questions, as the ones already discussed can be helpful.

Problems in medical decision-making have thus been addressed not only by making dogmatic statements – derived from principles of medical ethics – but also by using questions to examine them, in order to give the psychiatrist the ability to find the way which is the most likely to be comply with the principles and rules of medical ethics for each individual patient.

6 References

Ach JS, Gaidt A (eds) (1993) Herausforderung der Bioethik. Frommann-Holyboog, Stuttgart

Ach JS, Gaidt A (1994) Am Rande des Abgrunds? Anmerkungen zu einem Argument gegen die moderne Euthanasie-Debatte. Ethik Med 6: 172–188

*Amelung K (1995) Probleme der Einwilligungsfähigkeit. Recht Psychiatr 13: 20–28

*Appelbaum PS, T Grisso (1995) The MacArthur treatment competence study. I. Mental illness and competence to consent to treatment. Law Human Behav 19: 105–126

Arbeitskreis zur Erforschung der "Euthanasie"-Geschichte (1996) Grafenecker Erklärung 1996. Dr. med. Mabuse 21/102: 40–43

Ärztekammer Berlin (1990) Berufsordnung. Amtsblatt für Berlin. 14. September 1990, p 1884

Bachmann KD, Heerklotz B (1997) Der Wissenschaftliche Beirat der Bundesärztekammer. Dtsch Ärztebl 94: A582–588

Barocka A (1992) Psychiatrie vom Zeitgeist bedrängt? Altes und Neues zur Euthanasiefrage. Forum Interdisz Forsch 10: 73–83

Bartlett ET (1995) Differences between death and dying. J Med Ethics 21: 270–276

Bastian T (ed) (1990) Denken – Schreiben – Töten. Zur neuen "Euthanasie"-Diskussion. Hirzel, Stuttgart

Baumann J, Bochnik HJ, Branneck AE et al (1986) Alternativentwurf eines Gesetzes über Sterbehilfe (AE-Sterbehilfe). Entwurf eines Arbeitskreises von Professoren des Strafrechts und der Medizin sowie ihrer Mitarbeiter. Thieme, Stuttgart

**Beauchamp TL (1989) Antwort auf Rachels zum Thema Euthanasie In: Sass HM (ed) Medizin und Ethik. Reclam, Stuttgart, pp 265–286

Beauchamp TL, Childress JF (1994) Principles of biomedical ethics, 4th edn. Oxford University Press, New York

Beckmann J (ed) (1996) Fragen und Probleme einer medizinischen Ethik. de Gruyter, Berlin

*Beecher KK (1966) Ethics and clinical research. N Engl J Med 274: 1354–1360

Berghmans RLP (1995) Research with decisionally incapacitated subjects: the status quo and debate in the Netherlands. Institute for Bioethics, Maastricht

Berghmans RLP (1997) Dementia, care and ethics. A brochure for informal carers of dementing persons. European Alzheimer Clearing House, Brussels

Bernal y Del Rio V (1980) Psychiatric ethics. In: Kaplan HI, Freedman AM, Sadock BJ (eds) Comprehensive textbook of psychiatry, 3rd edn. Williams and Wilkins, Baltimore, pp 3216–3231

Binding K, Hoche A (1920) Die Freigabe der Vernichtung lebensunwerten Lebens. Ihr Maß und ihre Form. Meiner, Leipzig

Birnbacher D (1990a) Ist die Unterscheidung zwischen aktiver und passiver Sterbehilfe ethisch bedeutsam? In: Atrott HH, Pohlmeier H (eds) Sterbehilfe in der Gegenwart. Roderer, Regensburg, pp 25–40

Birnbacher D (1990b) Selbstmord und Selbstmordverhütung aus ethischer Sicht. In: Leist A (ed) Um Leben und Tod. Suhrkamp, Frankfurt am Main, pp 395–422

Birnbacher D (1993) Welche Ethik ist als Bioethik tauglich? In: Ach JS, Gaidt A (eds) Herausforderung der Bioethik. Frommann-Holzboog, Stuttgart, pp 45–67

Birnbacher D, Kottje-Birnbacher L (1996) Ethik in der Psychotherapie und der Psychotherapieausbildung. In: Senf W, Broda M (eds) Praxis der Psychotherapie. Ein integratives Lehrbuch für Psychoanalyse und Verhaltenstherapie. Thieme, Stuttgart, pp 499–506

Bloch S (1996) Ethics and psychotherapy. Am J Psychother 50: 257–258

*Bloch S, Chodoff P (eds) (1991) Psychiatric ethics, 2nd edn. Oxford University Press, New York

Bok S (1996) At the juncture of theory and practice. Remarks on receiving the Henry Knowles Beecher Award. Hastings Cent Rep 26/3: 5–8

Bron B (1986) Ethische und juristische Probleme des Suizidproblems. Fortschr Neurol Psychiatr 54: 232–239

Brown JH, Henteleff P, Barakat S, Rowe CJ (1986) Is it normal for terminally ill patients to desire death? Am J Psychiatry 143: 208–211

Bundesärztekammer (1979) Richtlinien für die Sterbehilfe. Dtsch Ärztebl 76: 957–960

Bundesärztekammer (1988) Weissbuch. Anfang und Ende menschlichen Lebens. Medizinischer Fortschritt und ärztliche Ethik. Deutscher Ärzte-Verlag, Köln

Bundesärztekammer (1993) Richtlinien der Bundesärztekammer für die ärztliche Sterbehilfe. Dtsch Ärztebl 90: 1791–1792

Bundesärztekammer (1997) Entwurf der Richtlinie der Bundesärztekammer zur ärztlichen Sterbebegleitung und den Grenzen zumutbarer Behandlung. Dtsch Arztebl 94: A1342–A1344

Bundesärztekammer (1998) Grundsätze der Bundesärztekammer zur ärztlichen Sterbebegleitung. Dtsch Ärztebl 95: 1690–1691

Cartwright W (1996) Killing and letting die: a defensible distinction. In: Dunstan GR, Lachman PJ (eds) Euthanasia: death, dying and the medical duty. Br Med Bull 2: 354–361

Casey PR (1996) National report Ireland. In: Koch HG, Reiter-Theil S, Helmchen H (eds) Informed consent in psychiatry. Nomos, Baden-Baden, pp 151–170

Cassel CK, Meier DE (1990) Morals and moralism in the debate over euthanasia and assisted suicide. N Engl J Med 323: 750–752

Chochinov HM, Wilson KG, Enns M, Mowchum N, Lander S, Levitt M, Clinch JJ (1995) Desire for death in the terminally ill. Am J Psychiatry 152: 1185–1191

Chodoff P (1996) Ethical dimensions of psychotherapy: a personal perspective. Am J Psychother 50: 298–310

Conwell Y, Caine ED (1991) Rational suicide and the right to die. N Engl J Med 325: 1100–1102

Council of Europe – Directorate of Legal Affairs (1996) Convention for the protection of human rights and dignity of the human being with regard to the application of biology and medicine: convention on human rights and biomedicine. Council of Europe, Strasbourg

Daub U, Wunder M (1994) Des Lebens Wert. Zur Diskussion über Euthansie und Menschenwürde. Lambertus, Freiburg

DGPPN (Deutsche Gesellschaft für Psychiatrie, Psychotherapie und Nervenheilkunde) (1996) Stellungnahme zum Entwurf der Bioethikrahmenkonvention des Europarates vom 8.3.1996. Nervenarzt 67: 888–889

Dichgans J (1992) Der Arzt und die Wahrheit am Krankenbett. Arzt Christ 38: 13–23

Diekstra RFW (1996) Sterben in Würde: über das Für und Wider der Beihilfe zum Suizid. In: Anschütz F, Wedler HL (eds) Suizidprävention und Sterbehilfe. Ullstein-Mosby, Berlin, pp 179–203

Dörner K (1988) Tödliches Mitleid. Zur Frage der Unträglichkeit des Lebens oder: die Soziale Frage: Entstehung, Medizinisierung, NS-Lösung, heute, morgen. Van Hoddis, Gütersloh

Dresser R (1996) Mentally disabled research subjects. The enduring policy issues. JAMA 276: 67–72

Engelhardt HT Jr (1986) The foundations of bioethics. Oxford University Press, New York

Eser A (1976) Lebenserhaltungspflicht und Behandlungsabbruch aus rechtlicher Sicht. In: Auer A, Menzel H, Eser A (eds) Zwischen Heilauftrag und Sterbehilfe. Zum Behandlungsabbruch aus ethischer, medizinischer und rechtlicher Sicht. Heymanns, Cologne, pp 75–147

Eser A (1985) Sterbewille und ärztliche Verantwortung – zugleich Stellungnahme zum Urteil des BGH im Fall Dr. Wittig. MedR 3: 6–17

Eser A (1992) Sterbehilfe – Recht. In: Eser A, von Lutterotti M, Sporken P (eds) Lexikon Medizin, Ethik, Recht. Herder, Freiburg, pp 1095–1101

*Eser A, Koch HG (eds) (1991) Materialien zur Sterbehilfe. Eine internationale Dokumentation. Max-Planck-Institut für ausländisches und internationales Privatrecht, Freiburg

Faden R (ed) (1996) The human radiation experiments. Oxford University Press, New York

Fagot-Largeaut A (1996) National report France. In: Koch HG, Reiter-Theil S, Helmchen H (eds) Informed consent in psychiatry. Nomos, Baden-Baden, pp 67–96

Fins JJ, Bacchetta MD (1994) The physiscian-assisted suicide and euthanasia debate: an annotated bibliography of representative articles. J Clin Ethics 5: 329–340

Fuchs T, Lauter H (1997a) Euthanasie: Kein Recht auf Tötung. Dtsch Ärztebl 94: 180–182

Fuchs T, Lauter H (1997b) Der Fall Chabot: Assistierter Suizid aus psychiatrischer Sicht. Nervenarzt 68: 878–883

Gabbard GO (1996) Lessons to be learned from the study of sexual boundary violations. Am J Psychother 50: 311–322

Ganzini L, Lee MA, Heintz RT, Bloom JD (1993) Depression, suicide, and the right to refuse life-sustaining treatment. J Clin Ethics 4: 337–340

Gartrell N, Herman J, Olarte S, Feldstein M, Localio R (1986) Psychiatrist–patient sexual contact: results of a national survey. I. Prevalence. Am J Psychiatry 143: 1126–1131

Gillon R, Lloyd A (1994) Principles of health care ethics. Wiley, New York

Goldberg C (1977) Therapeutic partnership: ethical concerns in psychotherapy. Springer, Berlin Heidelberg New York

Gouzoulis-Mayfrank E, Hermle L, Kovar KA, Saß H (1996) Die Entaktogene 'Ecstasy' (MDMA), 'Eve' (MDE) und andere ringsubstituierte Methamphetaminderivate. Eine neue Stoffklasse unter den illegalen Designerdrogen? Nervenarzt 67: 369–380

Griffiths J (1995) Assisted suicide in the Netherlands: the Chabot Case. Mod Law Rev 58/2: 232–248

*Grisso T, Appelbaum PS (1995) Comparison of standards for assessing patients' capacities to make treatment decisions. Am J Psychiatry 152: 1033–1037

Groenewoud JH, PJ van der Maas, G van der Wal, MW Hengeveld, AJ Tholen, WJ Schudel, A van der Heide (1997) Physician-assisted death in psychiatric practice in the Netherlands. N Engl J Med 336: 1795–801

Gropp W (1985) Suizidbeteiligung und Sterbehilfe in der Rechtsprechung. Zugleich eine Besprechung des BGH-Urteils im Fall Wittig vom 4.7.1994. NStZ 5: 97–103

Guckes B (1997) Das Argument von der schiefen Ebene. Schwangerschaftsabbruch, die Tötung Neugeborener und Sterbehilfe in der medizinischen Diskussion. Fischer, Stuttgart

Gurevitz H (1977) Tarasoff: protective privilege versus public peril. Am J Psychiatry 134: 289–92

Gutheil T, Shapiro R, Clair LST (1980) Legal guardianship in drug refusal: an illusionary solution. Am J Psychiatry 137: 347–351

Häfner H (1997) Was tun mit Krankheiten, die keine sind? Subdiagnostische Störungen und unversorgte Morbidität. Münch Med Wochenschr 139: 158–160

Hegselmann R, Merkel R (eds) (1991) Zur Debatte über Euthanasie. Beiträge und Stellungnahmen. Suhrkamp, Frankfurt am Main

Heinrich K (1992) Gefährdung und Gefährlichkeit der Psychiatrie. Fortschr Neurol Psychiatr 60: 349–355

Helle J (1993) Patienteneinwilligung und Zwang bei der Heilbehandlung untergebrachter psychisch Kranker. Medizinrecht 11/4: 134–139

Helmchen H (1986) Ethische Fragen in der Psychiatrie. In: Kisker KP, Lauter H, Meyer JE, Müller C, Strömgren E (eds) Psychiatrie der Gegenwart, vol 2, 3rd edn. Springer, Berlin Heidelberg New York, pp 310–368

Helmchen H (1991) Aufklärung über Späthyperkinesen. Nervenarzt 62: 265–268

Helmchen H (1992) Tötung auf Verlangen aus psychiatrischer Sicht. Fundam Psychiatr 6: 58–62

Helmchen H (1994) Ethics in psychiatric research. Arch Psychiatr Diagn Clin Eval (Jap) 5: 391–402

Helmchen H (1995a) Kriterien und Konsequenzen von Einwilligungsunfähigkeit. In: Toellner R, Wiesing U (eds) Wissen – Handeln – Ethik. Strukturen ärztlichen Handelns und ihre ethische Relevanz. Fischer, Stuttgart, pp 117–127

Helmchen H (1995b) Ziele, Beratungsgegenstände und Verfahrensweisen medizinischer Ethikkommissionen. Ethik Med 7: 58–70

Helmchen H (1998) The mutual patient-psychiatrist communication and the therapeutic contract. Compr Psychiatry 39: 5–10

Helmchen H (1998a) Forschung mit nicht-einwilligungsfähigen Kranken. Bericht und Verhandlungen der Berliner-Brandenburgischen Akademie der Wissenschaften 5: 9–30

Helmchen H (1998b) Research with competent patients. A current problem in light of German history. Eur Psychiatry 13[Suppl 3]: 93–100

Helmchen H (1998c) The mutual patient-psychiatrist communication and the therapeutic contract. Compr Psychiatry 39: 5–10

Helmchen H (1998d) Research with patients wincompetent to give informed consent. Curr Opin Psychiatry 11: 295–297

*Helmchen H, Lauter H (eds) (1995) Dürfen Ärzte mit Demenzkranken forschen? Thieme, Stuttgart

Helmchen H, Müller-Oerlinghausen B (eds) (1978) Psychiatrische Therapieforschung. Springer, Berlin Heidelberg New York

Helme T (1993) "A special defence": a psychiatric approach to formalising euthanasia. Br J Psychiatry 163: 456–466

Herzberg RD (1996) Sterbehilfe als gerechtfertigte Tötung im Notstand? NJW 46: 3043–3049

Höffe O (1987) Medizinische Ethik. In: Görres-Gesellschaft (ed) Staatslexikon. Herder, Freiburg, pp 1070–1074

Höffe O (ed) (1992) Lexikon der Ethik, 4th edn. Beck, Munich

Holm S (1994) American bioethics at the crossroads. A critical appraisal. Eur Phil Med Health Care 2/2: 6–23

Holm S (1995) Not just autonomy – the principles of American biomedical ethics. J Med Ethics 21: 332–338

*Holmes J (1996) Values in psychotherapy. Am J Psychother 50: 259–273

International IPPNW Congress (1996) Nürnberger Erklärung. Berl Ärzte 2/97: 24

Jonsen A, Toulmin S (1988) The abuse of casuistry. Oxford University Press, New York

Kaiser G (1992) Suizid – Recht. In: Eser A, von Lutterotti M, Sporken M (eds) Lexikon Medizin, Ethik, Recht. Herder, Freiburg, pp 1139–1148

Karasu T (1991) Ethical aspects of psychotherapy. In: Bloch S, Chodoff P (eds) Psychiatric ethics. Oxford University Press, Oxford, pp 135–165

Kersting FW (1996) Anstaltsärzte zwischen Kaiserreich und Bundesrepublik. Das Beispiel Westfalen. Schöningh, Paderborn

*Keyserlingk EW, K Glass, S Kogan, S Gauthier (1995) Proposed guidelines for the participation of persons with dementia as research subjects. Perspect Biol Med 38/2: 319–361

Klee E (1983) 'Euthanasie' im NS-Staat: Die 'Vernichtung lebensunwerten Lebens'. Fischer, Frankfurt am Main

Klerman GL (1972) Psychotropic hedonism versus pharmacological calvinism. Hastings Cent Rep 2: 1–3

Klotzko AJ (1995) CQ Interview: Arlene Judith Klotzko and Dr. Boudewijn Chabot discuss assisted suicide in the absence of somatic illness. Camb Q Health Ethics 4: 239–249

**Koch HJ, S Reiter-Theil, H Helmchen (eds) (1996) Informed consent in psychiatry. Nomos, Baden-Baden

Lasch HG (1985) Der Arzt und das Sterben. Gießener Universitätsbl 1/85: 5–16

Lauter H, Meyer JE (1992) Die neue Euthanasie-Diskussion aus psychiatrischer Sicht. Fortschr Neurol Psychiatr 60: 441–448

Law Commission (1993) Mentally incapacitated adults and decision-making: medical treatment and research. Consultation paper no 129. HMSO, London

*Law Commission (1995) Mental incapacity: HC paper no 189. HSMO, London. Bull Med Ethics, March: 13–18

Leist A (1996) Das Dilemma der aktiven Euthanasie. Gefahren und Ambivalenzen des Versuchs, aus Töten eine soziale Praxis zu machen. Humanitas, Dortmund

Leist A (ed) (1990) Um Leben und Tod. Moralische Probleme bei Abtreibung, künstlicher Befruchtung, Euthanasie und Selbstmord. Suhrkamp, Frankfurt am Main

Levi BH (1996) Four approaches to doing ethics. J Med Philos 21: 7–39

*Levine RJ (1996) Proposed regulation for research involving those institutionalized as mentally infirm: a consideration of their relevance in 1996. IRB 18/5: 1–5

Lutterotti M von (1992) Sterbehilfe – Medizin. In: Eser A, von Lutterotti M, Sporken P (eds) Lexikon Medizin, Ethik, Recht. Herder, Freiburg, pp 1086–1095

Lutterotti M von (1993) Grenzen ärztlicher Behandlungspflicht und passive Sterbehilfe. Z Med Ethik 39: 3–14

Matouschek E (ed) (1989) Arzt und Tod. Verantwortung, Freiheiten und Zwänge. Schattauer, Stuttgart

Mead M (1963) From black and white magic to modern medicine. Proc Rudolf Virchow Med Soc (New York) 22: 130–131

Meijers LCM, de Boer J, Glaudemans-van Gelderen I, Haveman MJ, van der Maas PJ, Schroten E, Visser HKA, Jansen I, Wortmann SFM, Berlihn J (1995) Committee "Medical Experiments With Incapacitated Persons" to the Ministry for Health, Welfare, and Sport and the Ministry of Justice, The Hague

Merskey H (1996) Ethical issues in the search for repressed memories. Am J Psychother 50: 323–335

Miller FG, Quill TE, Brody H, Fletcher JC, Gostin LO, Meier DE (1994) Regulating physician-assisted death. N Engl J Med 331: 119–123

*Mitscherlich A, Mielke F (1960) Medizin ohne Menschlichkeit: Dokumente des Nürnberger Ärzteprozesses. Fischer, Frankfurt am Main

Moggi F, Bossi J, Bachmann KM (1992) Sexueller Mißbrauch in therapeutischen Beziehungen. Nervenarzt 63: 705–709

Momeyer R (1995) Does physician assisted suicide violate the integrity of medicine? J Med Philos 20: 13–24

Neeleman J (1996) Suicide as a crime in the UK: legal history, international comparisons and present implications. Acta Psychiatr Scand 94: 252–257

Oglive AD, Potts SG (1994) Assisted suicide for depression: the slippery slope in action? Br Med J 309: 492–493

*Patzig G, Schöne-Seifert B (1995) Theoretische Grundlagen und Systematik der Ethik in der Medizin. In: Kahlke W, Reiter-Theil S (eds) Ethik in der Medizin. Enke, Stuttgart, pp 1–9

Payk TR (ed) (1996) Perspektiven psychiatrischer Ethik. Thieme, Stuttgart

Pellegrino ED (1989) Withholding and withdrawing treatments: ethics at the bedside. Clin Neurosurg 35: 164–184

Pellegrino ED, Thomasma DC (1993) The virtues in medical practice. Oxford University Press, New York

Pohlmeier H (1995) Depression und Selbstmord. Parerga, Düsseldorf

Pöldinger W, Wagner W (eds) (1991) Ethik in der Psychiatrie. Wertbegründung – Wertdurchsetzung. Springer, Berlin Heidelberg New York

Quill TE, Cassel CK, Meier DE (1992) Care of the hopelessly ill. Proposed clinical criteria for physician-assisted suicide. N Engl J Med 327: 1380–1384

Rachels J (1989) Aktive und passive Sterbehilfe. In: Sass HM (ed) Medizin und Ethik. Reclam, Stuttgart, pp 234–264

Rawls J (1975) Eine Theorie der Gerechtigkeit. Suhrkamp, Frankfurt am Main

Reich WT (1994) The word "bioethics": its birth and the legacies of those who shaped it. Kennedy Inst Ethics J 4/4: 319–335

Reich WT (ed) (1995) Encyclopedia of bioethics, 2nd edn. Simon and Schuster Macmillan, New York

Reimer C (1991). Ethik der Psychotherapie. In: Pöldinger W, Wagner W (eds) Ethik in der Psychiatrie – Wertebegründung, Wertedurchsetzung. Springer, Berlin Heidelberg New York, pp 127–147

Reischies FM, Schaub RT (1997) Epidemiologische Verlaufsuntersuchungen der Demenz. In: Rösler M, Retz W, Thome J (eds) Alzheimer-Krankheit. Deutscher Studienverlag, Weinheim, pp 58–66

Reiter J (1996) Bioethik und Menschenwürde. Ethische Aspekte der Bioethikkonvention des Europarats. Stimmen der Zeit 214/9: 579–589

Ritschl D (1989) Psychiatrie. In: Eser A, von Lutterotti M, Sporken P (eds) Lexikon Medizin, Ethik, Recht. Herder, Freiburg, pp 842–846

Roscam Abbing HDC (1994) Medical research involving incapacitated persons; what is legally permissible? Institute for Private Law, Department of Health Law, Faculty of Law, University of Utrecht

Rössler D (1996) Zur Diskussion über die Bioethik-Konvention. Ethik Med 8: 167–172

Roth LH, Meisel A (1977) Dangerousness, confidentiality, and the duty to warn. Am J Psychiatry 134: 508–11

Rüger U (1977) Intrapsychische und familiendynamische Prozesse vor der manifesten Erkrankung und während der Lithium-Therapie einer endogenen Depression. Z Psychosomat Med Psychoanal 23: 329–353

Sass HM (1988) Bioethik in den USA. Methoden, Themen, Positionen. Springer, Berlin Heidelberg New York

Sass HM (ed) (1989) Medizin und Ethik. Reclam, Stuttgart

Sass HM (1995) Bioethics in German-speaking western European countries (Austria, Germany, and Switzerland) 1991–1993. In: Lustig BA (ed) Bioethics yearbook, vol 4. Kluwer, Dordrecht, Netherlands, pp. 247–268

Sass HM, Viefhues H (1991) Güterabwägung in der Medizin. Ethische und ärztliche Probleme. Springer, Berlin Heidelberg New York

Schlesinger M (1995) Perspectives: ethical issues in policy advocacy. Health Aff (Millwood) 14: 23–29

Schmidt H (1989) Entscheidungsfindung, ärztliche – Ethik. In: Eser A, von Lutterotti M, Sporken P (eds) Lexikon Medizin, Ethik, Recht. Herder, Freiburg, pp 303–314

Schmidtke A, Häfner H (1986) Die Vermittlung von Selbstmordmotivation und Selbstmordhandlung durch fiktive Modelle. Die Folgen der Fernsehserie "Tod eines Schülers". Nervenarzt 57: 502–510

Schmuhl HW (1987) Rassenhygiene, Nationalsozialismus, Euthanasie. Von der Verhütung zur Vernichtung "lebensunwerten Lebens" 1890–1945. Vandenhoeck and Ruprecht, Göttingen

Scholten HJ (1991) Niederlande. In: Eser A, Koch HG (eds) Materialien zur Sterbehilfe. Eine internationale Dokumentation. Max-Planck-Institut für ausländisches und internationales Privatrecht, Freiburg, pp 451–500

Schöne-Seifert B (1996) Medizinethik. In: Nida-Rümelin J (ed) Angewandte Ethik. Die Bereichsethiken und ihre theoretische Fundierung. Ein Handbuch. Kröner, Stuttgart, pp 553–648

Schöne-Seifert B, Rippe KP (1991) Silencing the singer. Antibioethics in Germany. Hastings Cent Rep 21/6: 20–27

Schöne-Seifert B, Sass HM, Bishop LJ, Bondolfi A (1995) History of medical ethics: German-speaking countries and Switzerland. In: Reich WT (ed) Encyclopedia of bioethics, 2nd edn. Simon and Schuster Macmillan, New York, pp 1579–1589

Schreiber HL (1995) Behandlungsabbruch und Sterbehilfe. In: Deutsche Sektion der Internationalen Juristen-Kommission (ed) Lebensverlängerung aus medizinischer, ethischer und rechtlicher Sicht. Müller, Heidelberg, pp 129–145

Schwartz HI, Vigniano W, Perez CB (1988) Autonomy and the right to refuse treatment – patients' attitudes after involuntary medication. Hosp Community Psychiatry 39: 1049–1054

Seidler E (1986) Bioethik oder Ethik der Heilberufe? MMG 11: 258–263

Siegler M, Pellegrino ED, Singer PA (1990) Clinical medical ethics. J Clin Ethics 1/1: 5–9

Simkin S, Hawton K, Whitehead L, Fagg J, Eagle M (1995) Media influence on parasuicide. A study of the effects of a television drama portrayal of Paracetamol self-poisoning. Br J Psychiatry 167: 754–759

Singer P (1984) Praktische Ethik. Reclam, Stuttgart

Singer P (1990) Bioethics and academic freedom. Bioethics 4: 33–44

Singer P, Kuhse H (1994) Bioethics and the limits of tolerance. J Med Phil 19: 129–145

Smith LFP, Morissy JR (1994) Ethical dilemmas for general practitioners under the UK new contract. J Med Ethics 20: 175–180

Steigleder K, Mieth D (eds) (1990) Ethik in den Wissenschaften. Attempo, Tübingen

Stein R (1996) Mißbrauch der Medizin. Bericht über den Kongress "Medizin und Gewissen" in Nürnberg vom 25.–27.10.1996. Berliner Ärzte 12/96: 29–31

Strang J, Gossop M (1996) Heroin prescribing in the British system: historical review. Eur Addict Res 2: 185–193

*Taupitz J, Fröhlich U (1997) Medizinische Forschung mit nicht-einwilligungsfähigen Personen. VersR 22: 91–918

Thomas H (1993) Sind Handeln und Unterlassen unterschiedlich legitimiert? Ethik Med 5: 70–82

Tischler GL, Astrachan BM (1996) A funny thing happened on the way to reform. Arch Gen Psychiatry 53: 595–963

Tugendhat E (1994) Vorlesungen über Ethik. Suhrkamp, Frankfurt am Main

Van der Maas PJ, van der Wal G, Haverkate J et al (1996) Euthanasia, physician assisted suicide and other medical practices involving the end of life in the Netherlands 1990–1995. N Engl J Med 335: 1669–1705

Van der Wal G, PJ van der Maas, JM Bosma, BD Onwuteaka-Philipsen, DL Willems, I Haverkate, PJ Kostense (1996) Evaluation of the notification procedere for physician-assisted death in the netherlands. N Engl J Med 335: 1706–1711

Veatch RM (1981) A theory of medical ethics. Basic, New York

Vollmann J (1989) Ärztliche und moralische Pobleme der Sterbehilfe. Überlegungen zu Peter Singers Praktischer Ethik. Fundam Psychiatr 3: 203–209

Vollmann J (1996a) Informed Consent des Patienten zur Publikation von Kasuistiken. Neue Richtlinien des "International Committee of Medical Journal Editors" (Vancouver Group). Nervenarzt 67: 122–126

Vollmann J (1996b) Why does bioethics develop differently in Germany? Analysis and commentary. Eur Phil Med Health Care 4/1: 13–20

Vollmann J (in press) Das Informed Consent-Konzept als Politikum in der Medizin. Patientenaufklärung und Einwilligung aus historischer und medizinethischer Perspektive. In: Kettner M (ed) Angewandte Ethik als Politikum. Suhrkamp, Frankfurt am Main

Vollmann J, Dörries A (1996) Dem Einzelnen oder dem Ganzen verpflichtet? Ethische Überlegungen zur ärztlichen Verantwortung. Z Ärztl Fortbild 90: 527–532

Vollmann J, Helmchen H (1997) Aufklärung und Einwilligung (Informed Consent) in der klinischen Praxis. Dtsch Med Wochenschr 122: 870–873

Vollmann J, Winau R (1996) Informed consent in human experimentation before the Nuremberg code. Br Med J 313: 1445–1447

Wachter AM de (1997) The European Convention on Bioethics. Hastings Cent Rep 27/1: 13–23

Washington et al Vs. Glucksberg et al (1997) Decision of the Supreme Court of the United States. June 26, 1997

Wassermann R (1984) Das Recht auf den eigenen Tod. In: Winau R, Rosemeier HP (eds) Tod und Sterben. de Gruyter, Berlin, pp 381–412

Wassermann R (1993) Begriffsbestimmung: Sterbehilfe, Sterbebeihilfe, Euthanasie, unterlassene Hilfeleistung, fahrlässige Tötung aus juristischer Sicht. Z Ärztl Fortbild 87: 13–18

WHO (ed) (1996) Guidelines for the promotion of human rights of persons with mental disorders. WHO, Geneva

Wiesemann C (1996) Ist der Hippokratische Eid noch zeitgemäß? Arzt, Patient und Gesellschaft in der Medizin der Neuzeit. In: Frewer A (ed) Zur ethischen Kultur in der Humanmedizin. Palm and Enke, Erlangen, pp 13–24

*Winau R (1984) Die Freigabe der Vernichtung lebensunwerten Lebens. In: Winau R, Rosemeier HP (ed) Tod und Sterben. de Gruyter, Berlin, pp 25–50

Winau R (1993) Begriffsbestimmung: Sterbehilfe, Sterbebeihilfe, Euthanasie, unterlassene Hilfeleistung, fahrlässige Tötung aus ärztlicher Sicht. Z Ärztl Fortbild 87: 19–21

Winau R (1994) The Hippocratic oath and ethics in medicine. Forensic Sci Int 69: 285–289

Winckelmann HJ (1989) Selbstmedikation und der therapiebestimmende Patient. In: Heinrich K, Linden M, Müller-Oerlinghausen B (eds) Werden zu viele Psychopharmaka verbraucht? Thieme, Stuttgart, pp 141–152

World Medical Association (ed) (1996) Handbook of declarations. World Medical Association, Ferney-Voltaire

World Psychatric Association (1998) Declaration of Madrid, 1996. Nervenarzt 69: 454–455

Winslade WJ (1989) Ethics in Psychiatry. In: Kaplan HI, Sadock BJ (eds) Comprehensive Textbook of Psychiatry, 5th edn. Williams & Wilkins, Baltimore Hongkong London Sydney, pp 2124–2131

Zentrale Ethikkommission bei der Bundesärztekammer (1997) Stellungnahme "Zum Schutz nicht-einwilligungsfähiger Personen in der medizinischen Forschung". Dtsch Arztebl 94: B811–B812

CHAPTER

19

H.J. Walton

Psychiatric Education and Training

1 Introduction

This chapter deals with the training of a psychiatrist from entry to medical school through readiness to become a consultant, and finally to lifelong learning throughout professional life. It presents a paradigm of current good practice pertaining to the continuum of specialist training in psychiatry. The educational programme described is of necessity schematic and general, because of the great variation to be expected among varying countries and different cultures. However, the basic principles set out will apply in any setting.

Psychiatry is a speciality in the field of medicine which is often seen as the archetypal profession. Education in psychiatry is therefore a component of the professionalisation process affecting every doctor. Relevant to it are all the canonical aspects of medical education: access (application to medical school, selection procedures, dropout and wastage), curriculum, teaching methods, educational objectives,assessment (and examinations), evaluation and accreditation. Then follows, in turn, the newly graduated generic doctor, specialist training as a psychiatrist and finally continuing education to maintain professional competence throughout life.

The sequence stated above replaces the earlier medical school paradigm, which was that of vocational training. Instead, a university education in medicine results in a generic doctor, who can then enter postgraduate training appropriate to a particular specialty; followed finally by ongoing continuing medical education (CME). Doctors now trained in developed countries no longer have to rely on their medical school education as the basis for their professional competence.

In developing countries, the paradigm has to be different. The graduate from medical school is required immediately on graduation to start medical practice, under more or less supervision, and – if the numbers of doctors are limited – has to do his or her work through colleagues such as nurses or paramedicals and also traditional healers, whom the doctor instructs and monitors. Medical students need to receive additional training in psychiatry during the medical school period, since psychiatric disorders form such a large part of the work of primary care doctors. This additional psychiatric instruction in developing countries should extend across the curriculum as a whole. It should also continue, after graduation, as part of in-service training and CME (World Psychiatric Association/World Federation for Medical Education 1998).

Medicine, surgery and psychiatry, Reil said in the last century, constitute the three main branches of medicine. Psychiatrists treat only a small portion of any nation's psychiatric morbidity. A great bulk of psychiatric disorder is managed, adequately or otherwise, by doctors whose psychiatric training often is confined to their exposure to psychiatry as medical students.

2 Undergraduate Education: The Medical School Curriculum

Psychiatric teaching is called for throughout the undergraduate curriculum for all medical students, regardless of the branch of medicine they go on to enter (Walton 1986). The broad goals in the teaching of psychiatry to undergraduate medical students, relevant to all future doctors, are the following in order of priority:

1. Communication skills and empathy help medical students become more aware of patients' emotional responses and help students to create the interpersonal relationship between clinician and patient necessary in the medical interview. Both the patient's and the student's subjective responses are involved.
2. The second most important aspect of teaching sets out to make students aware of scientific knowledge about behaviour, so that they can distance themselves appropriately from the personal problems of patients (showing "detached concern"), augmenting their common-sense understanding with empirical knowledge. Behavioural science teaching (notably psychology and sociology) is intended to promote such an objective approach to the patient on the part of the clinician.
3. Medical students frequently experience the same anxieties about mental patients that are customary in their society. Teaching of psychiatry should enable them to overcome prejudices and preconceptions and relate clinically to a wide range of patients with psychiatric illness and disorder. To do so, students need skills in history-taking, in examining the mental state and in psychiatric interviewing of a patient repeatedly over a period of time.
4. Descriptive psychiatry, dealt with in the psychiatric textbooks, cannot be satisfactorily learned unless medical students have proper access to patients *at the same time* as the theoretical teaching is given. This clinical training, for learning clinical skills as well as knowledge, can be extended by videotaped presentation of psychiatric patients, ward rounds,

case conferences, simulated patients and the many other clinical teaching procedures which are available.

5. A great proportion of ill health is psychiatric. The psychiatric instruction given in medical schools has to provide introductory knowledge about the main psychiatric treatment methods and skills, because much the greater part of the mental health needs of the community is met not by psychiatrists, but by general practitioners and specialists in other branches of medicine.

2.1 Psychiatrists as Medical School Teachers

Psychiatrists must of course concern themselves with the total curriculum of the medical school: they must serve on the educational committees planning the curriculum and implementing it, and in their departmental teaching they must naturally promote the institutional objectives of the medical school.

In carrying out these institutional objectives, and in gaining recognition for the teaching of psychiatry itself, psychiatrists have traditional disadvantages. Medical school teachers in general tend to view psychiatry as a backward branch of medicine. They are further dismayed by the ideological differences among psychiatrists (one a biologically oriented doctor favouring the medical model, the other a humanist concerned with psychodynamic processes and practising some or other form of psychotherapy). Medical school teachers are the more disconcerted when these contrasting ideologies tend to be applied to the patients they refer to psychiatrists; it can seem relatively arbitrary how any patient will be diagnosed and treated, appearing to depend too much on the outlook of psychiatrist to whom he or she is referred. Colleagues of the psychiatrist in other branches of medicine find the doctrinal differences among psychiatrists confusing.

The many investigations by sociologists of medical schools, classically those of Merton et al. (1957) at Cornell and Chuval (1980) in Israel, have helped to demonstrate that the status of psychiatry, of teachers of psychiatry and of psychiatric patients in medical schools is not at the level of, say, medicine or surgery. This has a seriously adverse effect on the teaching of psychiatry.

Medical students acquire their professional values and outlook, in large measure, by modelling themselves on leading and respected teachers in the medical school. Students react to teachers in three different ways, by active identification, active rejection or inactive orientation. The third of these responses, bland disregard, is now greatly less discernible in attitudes of contemporary medical students towards psychiatry than it was a generation earlier, when psychiatric instruction was less and poorer.

2.2 Student Attitudes to Psychiatry

It has been found that "psychologically minded" medical students show a greater interest in and a more positive attitude towards psychiatry than organically orientated students. Medical students' preference for psychiatry as a future career is little affected by psychiatric instruction during their training. It has long been known that certain learning experiences in psychiatry, particularly the clinical clerkship when in some types of unit (Eagel et al. 1979), can change attitudes towards psychiatry in a positive way. Despite such favourable responses, very few medical students commit themselves to a future career in psychiatry.

2.3 Psychiatry in the Curriculum

That psychiatry should occupy a major part in the basic medical curriculum is now generally agreed, for three reasons. Firstly, the general approach in contemporary medicine stressing the unity of body and mind is important in the whole of medical practice. Secondly, skills that are learned in psychiatry are important for all doctors, e.g. the ability to form a good relationship with a patient, to assess the mental state and to impart distressing information. Thirdly, psychiatric illness is common among patients seen by specialists in other branches of medicine; for example, among medical outpatients, about 15% of those given a medical diagnosis have an associated psychiatric disorder, and up to 40% of those given no specific diagnosis in fact have a psychiatric disorder (Clare and Lader 1982). Psychiatric disorders are even more frequent among patients seen in general practice. That all future doctors must know about psychiatric disorder is not only because they are common; in addition, their management involves much medical time and resources and gives rise to many serious incidents.

The World Psychiatric Association and the World Federation for Medical Education (1998) have compiled a *Core Curriculum in Psychiatry for Medical Students*. It sets out the minimum psychiatric instruction that will be required by medical students who, after qualification, will enter further training as specialists or general practitioners. Incidentally, in many countries, general practice is now also designated as a specialty, for which a further period of training

is required after graduation and in which future general practitioners extend their psychiatric skills. In countries which have no formal training for general practitioners, the teaching in psychiatry at medical school needs to be supplemented by a module containing the additional material that is essential for management of the psychiatric morbidity encountered in the community.

The psychiatry programme in the basic medical curriculum should be extended through all the years of the undergraduate teaching programme rather than concentrated in isolated separate courses. The sciences related to psychiatry of course include the biological sciences integral to the medical curriculum. That said, science teaching in medicine has been seriously criticised for being concentrated in a so-called preclinical phase instead of being extended throughout the curriculum; in addition, departmental interests have led to the biological sciences being fragmented, at a time when – in place of division according to disciplines – integration of knowledge is the hallmark of advances in the sciences relevant to medicine. *New biology* is an integrated view of whole structures, whether at molecular or macroscopic level. Medical schools have not responded to the paradigm shift in science. Medical schools have imperatively to examine their entire educational programme in order to integrate science teaching properly with clinical instruction. Traditional basic science teaching in the medical schools stands charged with being outmoded, restricted, simplistic and responsible for much of the information overload which mars medical curricula (Marston and Jones 1992).

Reform in this regard is a matter for the medical school as a whole. With respect to the teaching of psychiatry, two main components are to be specified, behavioural science and psychiatry itself.

2.4 Behavioural Science Teaching

Since the 1960s, behavioural sciences (also referred to as psychological medicine, psychology and sociology in relation to medicine) have been included in the undergraduate curriculum, often directed by the department of psychiatry. However, the behavioural sciences remain a problematic area. There has been widespread failure of the teaching of these subjects with considerable criticism from students and from other preclinical and clinical teachers. The following general points are important:

1. Behavioural science teaching should begin early in the curriculum, preferably in the first or second year.
2. Clinically experienced medical (psychiatrists) and non-medical (psychologists, sociologists) teachers should be actively involved in the planning of curricula and teaching.
3. Such teaching should be strictly relevant to medical practice and under no circumstances consist of a condensed general psychology and sociology course.
4. Behavioural science should be a compulsory and not an optional subject.
5. It should be examined along with other preclinical subjects and count towards professional examinations.
6. Separate behavioural science departments are not favoured, and teaching should be inter-departmental with a university department of psychiatry taking an active part in the organising and coordinating of the course.
7. Teachers should be encouraged to introduce clinical cases into their teaching from the start.
8. Teaching should be centred around clearly stated educational objectives.

Behavioural science teaching has often not succeeded in that students sometimes regard it as somewhat irrelevant or discordant with their main concerns as future doctors. The courses have failed at times to be perceived as convincing or a substantial addition to the curriculum.

2.5 Objectives of Behavioural Science Teaching

In each subdiscipline of behavioural science, specific requirements can be formulated.

2.5.1 Psychology

By the end of the course, each student should:

1. Have acquired a basic overall knowledge of the findings, methodologies and theories of psychology which are relevant to the practice of medicine
2. Be aware of how a patient's emotions, attitudes, values and experiences influence his or her response to illness and its treatment
3. Have a knowledge of learning processes and their relevance to medicine
4. Possess normative data of the main aspects of psychological development of humans from birth to old age
5. Possess skills relevant to effective doctor–patient communication and particularly to interviewing

6. Have knowledge of techniques of assessment used to assess the reliability and validity of investigation procedures and therapeutic trials
7. Have attitudes to development that will enable him or her as a doctor to see each patient as a complete human being living in his or her own social environment

2.5.2 Sociology

By the end of the course, each student should be able to:

1. Outline the various attempts to define and measure health and discuss their merits and limitations
2. Describe and discuss the importance of social institutions such as the family, the community, the economy and the law on health and medical practice
3. Describe and discuss the problems of equity and inequality in the provision and utilisation of health services, particularly when classified by age, gender, social class and region
4. Describe some of the more important changes in society and in the practice of medicine which have affected health and disease and the development of social policy
5. Describe the main reasons for the development of the welfare provisions, discuss the advantages and disadvantages of a country's health service and appreciate the problems of planning for change within a health service
6. Describe the social (and sociological) factors that influence the process of becoming ill and the doctor–patient relationship and the effect that ill health and hospitalisation have on the lives of patients and their families
7. Discuss critically the role of preventive medicine and health education and the role of self-help groups
8. Discuss critically the process of medical professionalisation
9. Describe some of the research methods used to evaluate health and medical practice

2.5.3 Biostatistical Sciences

The aims of the course in the probabilistic and information sciences are:

1. To explore ways of making valid deductions from medical data
2. To familiarise students with the basic statistical terminology as found in the medical literature, e.g. *The Lancet* and the *British Medical Journal*
3. To introduce the concepts necessary both for designing and for analysing comparison and experimental studies in medicine
4. To teach methods of accessing and organising medical knowledge and computer literacy

2.6 Clinical Psychiatry

Instruction in clinical psychiatry occurs with the other main clinical subjects. The components are clinical attachment(s), lectures and small-group tutorials.

Lectures are an effective way of promoting factual knowledge if well presented, and an efficient way of teaching large groups of students, but they should not constitute the main form of psychiatric teaching. Lewin (1948) established how experiences created in groups could stimulate learning and modify attitudes. Small-group teaching is expensive and time-consuming (more teachers are of course required than is the case in lectures to large classes), but must continue to be given a high priority by teachers in order to enable students to develop the professional attitudes relevant to psychiatry (Walton 1968). A combination of lectures and small-group teaching is desirable and is most effective.

Small-group tutorials, important in all medical teaching and learning, are especially important whenever *attitudes* are in question, e.g., respect for psychiatric patients or willingness to view mental illness as a valid clinical concern. Teachers who are uninformed about small-group teaching and group processes are of course unqualified, but that has not precluded them from being assigned to provide tutorial teaching, often lamentably. Small-group tutorials are a fundamental component of problem-based learning, a major innovation in contemporary medical education (Tosteson 1994) which medical schools worldwide are eager to introduce.

2.7 Clinical Clerkship

There is evidence that the type of the unit in which medical students do their clerkship in psychiatry influences their attitude to mentally ill patients. The clinical attachment may include experience in general practice, general hospitals and in the community, and at least a proportion of the clinical attachment should be in a general hospital psychiatric unit or in a psychiatric hospital.

The following recommendations may be made:

1. A full-time clerkship of 8–10 weeks' duration should be undertaken.
2. Wherever possible, the clerkship should be on a single general psychiatry unit, and students should not spend their whole clerkship entirely on a highly specialised unit.
3. Students should have both in-patient and out-patient contact.
4. They should have the opportunity to regularly see psychiatrists at work interviewing and treating patients.
5. Students should have the opportunity to interview patients themselves and be involved in the decision-making about clinical management.
6. Students should be encouraged to see the patient in the context of his or her family setting and background. They should have the opportunity to visit a patient's home and assess such a setting.

A prominent aim of psychiatric teaching should be to find instructional approaches fostering positive attitudes to all patients, including psychiatric patients. The consensus is that such attitudes are provided through a clerkship. The indications are that it needs to be 8 weeks at least in duration (World Psychiatric Association/World Federation for Medical Education 1998). Empirical studies demonstrate that attitudinal changes can be achieved and, moreover, are sustained.

2.8 Objectives

In the various areas of clinical psychiatry, different objectives can be formulated. In terms of general objectives, the student should be able to:

1. Conduct a diagnostic interview, including the taking of a psychiatric history and carrying out a mental state examination
2. Relate a patient's symptoms to his or her past experiences, personality and social circumstances
3. Give an account of his or her emotional responses to patients of different kinds and the way in which these can influence his or her judgment and hence the patient's management
4. Give an account of patients' emotional responses to doctors and the way these can influence the presentation of illness
5. Outline the main principles of, and indications for, counselling and psychotherapeutic intervention

In the field of organic psychiatry, the student should be able to:

1. Distinguish between organic and non-organic (functional) psychiatric disturbance
2. Describe and recognise common organic psychoses (acute and chronic), e.g. delirium tremens, senile dementia
3. List the common causes of confusional states and dementia in different age-groups
4. Describe the management of acute confusional states
5. Describe the management of dementia in and out of hospital

With regard to functional syndromes, the student should be able to:

1. Distinguish between a depressed mood and depressive illness and describe in detail the management of the latter
2. Diagnose schizophrenic and related psychoses, outline the management of acute attacks and describe the management of the chronic illnesses in the community
3. Describe and recognise common symptoms of neuroses
4. Define and recognise the signs and symptoms of psychiatric illness important in differentiating the major syndromes, e.g. flight of ideas, passivity experiences
5. Describe and recognise the features of normal and abnormal grief and outline their management
6. Recognise the diverse clinical presentation of alcohol dependence and describe the syndrome
7. Describe the management of alcohol and drug dependence
8. Describe and recognise the common forms of psycho-sexual disorder and outline the principles of their management
9. Discuss the common causes of acute emotional disturbance in different age-ranges and social groups and outline the principles of crisis management, especially in relation to parasuicide
10. Assess the risk of suicide in depressed patients
11. Describe and recognise the common psychological problems of childhood and adolescence
12. Describe and recognise the common psychological problems of old age and outline the social services available for their management

As far as treatment methods and agencies are concerned, the student should be able to:

1. Outline the psychiatric effects of drugs commonly used in medical practice, including corticosteroids, anti-hypertensive agents, opiates, oral contraceptives, barbiturates, sulphonamides and anti-convulsants
2. List the main indications, contraindications and unwanted effects of neuroleptics, anti-depres-

sants, etc. in common use, e.g. phenothiazines, tricyclic anti-depressants, monoamine oxidase inhibitors (MAOIs), benzodiazipines, lithium salts
3. Outline the principles of behaviour modifications and their main clinical applications
4. Discuss some clinically important concepts in psychodynamic approaches, e.g. unconscious conflict and defence mechanisms such as projection and denial
5. List and describe the main agencies for the care and rehabilitation of the psychiatrically ill and mentally handicapped in the community

Further objectives include the student being able to:

1. Describe the main psychiatric disorders found in children and the methods for investigating and treating these conditions
2. Describe the social and psychological problems of the mentally handicapped
3. Describe and recognise the common psychological reactions to physical illness
4. Outline the psychological mechanisms which can produce somatic symptoms and influence the course of physical illness
5. Describe the common associations between crime and mental illness
6. Outline the conditions under which it is legitimate to detain patients in hospital and treat them against their wishes

2.9 Communication Skills

During psychiatry teaching, and ideally also earlier in their preclinical career, students can benefit from the teaching of interview techniques. The effectiveness of such teaching has been amply demonstrated (Sanson-Fisher and Maguire 1980).

The value of allowing the student to watch or listen subsequently to his own interviews has been emphasised. Videotape provides an ideal medium, but listening to simple audio recordings also inculcates skills in perceiving and paying appropriate attention to a multitude of cues which are evident in even a brief interview.

Feedback from a tutor can help the student to understand the interpersonal process; it is as effective when given in a group as when given individually and is much more effective than when students are working on their own. Training in eliciting the history from a patient increases students' ability in essential skills, such as the elementary steps of introducing themselves to the patient and in orienting the patient to the task in hand.

Medical students (as well as postgraduates) can gain from training in interview skills through the use of video recordings, not only of history-taking but also extended to mental state examinations. Attention can be focused on the technique of eliciting the phenomena as well as on analysis of the significance of the signs and symptoms thus detected.

Video recordings of any interview occurring during routine clinical work can form a good basis for teaching (Westberg and Jason 1994). Everyday concerns of psychiatric management, such as the prescription of electroconvulsive therapy (ECT), the decision to permit weekend leave or the review of a weekend at home, all prove a valuable focus for learning and discussion.

Family therapy and individual psychotherapy training can also be augmented greatly by the use of videotape. An experienced therapist can benefit considerably from a regular review of a fragment of one of his or her interviews, and the benefit of such instruction to medical students is correspondingly large. Skills in observing and responding to verbal and non-verbal cues can be sharpened through systematic review of videotapes, particularly of the students' own interactions with patients assigned to them.

2.10 Examinations

The assessment methods used should always be congruent with the objectives of the course. Only too often the methods of assessment are discordant, e.g. multiple choice examinations are given predominant emphasis in some countries, although in theory teachers claim they are *not* concerned with factual knowledge alone. To assess clinical skills and attitudes, the professional examination needs to be composite, e.g. essay questions, short answer questions, an oral examination, continuous assessment during the clinical attachment. In addition, the objective structured clinical examination (OSCE), patient management problems (PMP), simulated patients and role-playing (e.g. the examiner taking the role of a parent enquiring about prognosis in schizophrenia) are all increasingly used.

Psychiatry forms part of the professional examination in many medical schools; it is sometimes part of a composite final examination, when only some students are given a psychiatry clinical examination.

Continuous assessment during the clinical attachments is an important means of evaluation. Colleagues in district hospitals are asked to play a large role in psychiatry teaching and they should of course also be involved in the assessment procedures. There may be

merit in supplementing the clinical examination with the writing up of case histories. Most medical schools now make extensive use of multiple choice questionnaire (MCQ) examinations; a joint MCQ in use among a number of universities in partnership provides resources for more sophisticated construction and makes it possible to draw interesting and informative comparisons between differing patterns of psychiatric teaching.

It is necessary to stress the dominating impact of examinations, because examinations drive the curriculum. Students learn to pass examinations, whatever contrary emphases teachers may seek to convey when specifying learning objectives.

2.11
Conclusions on Basic Education in Psychiatry

The objectives of the psychiatry department must of course be congruent with the overall objectives specified by the faculty of medicine for its medical curriculum. Efforts to reform curricula commonly fail because of failure to change the committee structure of the medical school in keeping with new educational goals. The curriculum cannot be left in the hands of departmental heads. Appropriate governance is essential, namely a curriculum committee answerable to the dean and not to departments, and given its own resources such as budgets and autonomy (Bloom 1988). Departments are then free to pursue their specific non-educational goals, such as research and clinical work, to which in any case they give greater priority (Abrahamson 1996).

Each university hospital must have a psychiatry unit, called upon by the curriculum committee to participate in the educational programme and advise (and provide) its psychiatric component.

Psychiatry, like all branches of medicine, is urgently confronted by the necessity to re-orient the curriculum, so that not only curative medicine is taught, but also prevention of illness and promotion of health. The obstacles to curricular change in medical schools are notorious. Prominent among them is the apathy – as far as education is concerned – of the majority of medical teachers: it seems almost impossible to get a critical mass of teaching staff to become educationally informed, interested and involved in the education of medical students. Medical teachers commonly think they are equipped to teach simply by virtue of their specialist expertise in their research or clinical discipline; they do not respect or know the findings of medical education research and are ignorant of the medical education literature (Miller 1980). They have very often not had personal instruction about how to teach.

When students themselves are studied, they repeatedly convey that they learn best when active student participation is encouraged, when the emphasis is on applied problem-solving (rather than learning of factual materials), when instruction is centred on the students and when teachers are humanistically oriented.

Teaching and learning methods have been revolutionised by electronic information technologies, problem-based learning and simulated (standardised) patients. Curricular time has been freed for special-interest options of students. Above all, the medical school phase in the continuum of medical education has to be disencumbered of information overload and passivity-inducing requirements for the memorisation of facts. Scope must be given for critical thinking, and the development of clinical competence made a priority. Educational leadership is essential, certainly not provided by deans at all commonly, and students have to be valued partners at every level, recognised as able to assist actively in monitoring the quality of their education.

3
The Generic Graduate

Young doctors are not yet equipped for independent practice on graduation. The goal of medical school is to produce a pluripotential doctor, who must then have postgraduate training to acquire the abilities needed for specialist (or generalist) practice. Prior to beginning such postgraduate training, the graduate is in the phase of transition from medical student to doctor.

In Britain, for example, doctors do not become registered practitioners for 1 year after leaving medical school. They are granted provisional registration by the General Medical Council and are required to work as pre-registration house officers receiving "general experience" in general hospitals, under the supervision of the medical school from which they graduated. This mandatory period of graduate clinical training, when the graduate becomes a clinician, is generally acknowledged as the most problematic of all (Richards 1992).

It is variously labelled in different countries: the pre-registration year, senior house officer training, the junior hospital doctor phase, etc. The junior doctor becomes a member of a consultant team in hospital. The notorious defects have been insufficient clinical supervision, insufficient feedback from consultants, excessive hours and poor training standards with inadequate educational provision. In many countries, legal steps have been taken to reduce working hours of junior doctors. By general consent, overwork and low

pay, with junior doctors misused as pairs of hands rather than doctors in pre-specialist training, is no longer acceptable.

Junior hospital doctors, when surveyed, report that the style of teaching on ward rounds and outpatient clinics markedly influence their ability to learn, their self-confidence and their acquisition of clinical competence. They are characteristically sympathetic to patients having psychiatric disorders complicating the organic disease for which they are hospitalised. However, junior doctors do not get into psychiatric hospitals; unless the general hospitals in which they work have psychiatry liaison services or psychiatric units, they receive no psychiatric instruction and no encouragement to consider training as a psychiatrist.

Experience in psychiatry during the pre-registration year would broaden the doctor's education. It could act to counterbalance influences which prejudice recruitment to psychiatry and are hostile to patients with psychiatric disorders. This could be best achieved by an improved liaison psychiatric service in general hospitals.

4
Postgraduate Training

Postgraduate training is in transition at present, not only nationally but worldwide. Indeed, the whole of medical education globally is undergoing massive change (World Federation for Medical Education 1994). In part, this process of re-evaluation and reform results from the altering roles, responsibilities and relationships of doctors: an upheaval so extensive and multifarious that the World Summit on Medical Education at Edinburgh in 1993 was designated: "The Changing Medical Profession".

The Recommendations (World Federation for Medical Education 1994) of the summit gave particular emphasis to the necessity of constant awareness that medical education is now viewed as a continuum. Basic medical education certainly is preparation for postgraduate training; both phases, in turn, are preparation for CME throughout each doctor's entire subsequent medical career.

Emphasis on the continuity of medical education is explicit in Recommendation 15 of the Summit, which calls for a holistic view of postgraduate medical education: "Mission statements of medical schools should specify the types of graduates to be produced, so that the competences assigned to postgraduate education and specialist training programmes are explicit."

The Recommendation continues: "There is need for a holistic view in planning for the broad fields of postgraduate education, with policy-making mechanisms that can support production of balanced numbers of generalists and specialists. The postgraduate training programmes need to be carefully related to the local context in which they will be practised, and linked with undergraduate and continuing educational programmes."

The reference to the local context is crucial. It recognises the importance of culture and history in determining patterns of health care. Appropriately, therefore, each country in Europe is separately engaged in a process of re-orienting postgraduate medical education. Psychiatry is one of the specialties undergoing profound change (Caldicott 1996). Moreover, there are major differences between countries in their organisational frameworks for funding, delivering and assessing postgraduate medical education. This being the phase of specialist training or at least its initiation, postgraduate training is the sector of medical education of particular significance to the specialty organisation in each country, keeping very much in mind that general practice now also has the status of a specialty. The psychiatry component of general practice is of the utmost importance, because of the epidemiological fact that the majority of psychiatric patients are seen not by specialist psychiatrists but by generalists.

4.1
Structure and Implementation of Training

In almost all countries, the Ministry of Health is concerned with postgraduate training, which takes place mainly in the hospitals of the country's health service. Universities are associated with health services in the postgraduate training programmes. For example, 20 psychiatric hospitals in the Netherlands have psychiatric training approval, eight of which are university hospitals.

Speciality organisations customarily set standards and hold examinations in postgraduate education. Their particular role differs considerably between countries. In the United Kingdom, for example, the Royal Colleges and their Higher Training Committees have major responsibilities. Each College, including the Royal College of Psychiatrists, through its Higher Training Committee produces its own assessment standards. The Colleges set formal competitive examinations to regulate entry into higher specialist training. During specialist training, the Higher Training Committee vets the quality of the training posts and the educational progress of each trainee doctor. This is achieved through inspection visits and regular formative assessment of the trainee.

Postgraduate training in psychiatry involves a sequence of posts, providing a wide range of clinical experience and making up a rotational training scheme, usually in one geographical area and often in one approved hospital complex. The rotational training scheme includes a university department or hospital and allows experience in the various branches of psychiatry, including (in addition to adult general psychiatry) child and adolescent psychiatry, mental handicap, geriatric psychiatry, psychotherapy, general hospital (liaison) psychiatry and other special fields.

The academic content should include an introduction to general psychopathology, pharmacology, genetics, interviewing, diagnosis and classification, organic disorders, drug and alcohol dependence, schizophrenia, mood disorders, neurotic disorders, personality disorders, learning disability, psychiatric disorders of childhood and adolescence, old age psychiatry, forensic psychiatry, drug treatments, counselling and interpretative psychotherapies, behavioural and cognitive therapies, rehabilitation and community care.

The specialist examination is taken at differing stages of training in different countries and is also phased differently in the various specialties in each country. In order to sit the U.K. Royal College of Psychiatrists examination, a candidate must have completed 3 years of psychiatric training and must be in a training post. After such general training, trainees may choose to enter additional training for a psychiatric subspecialty; in the Netherlands, for example, an additional 2 years in child psychiatry is required to obtain the Certificate of Completion of Specialist Training (Centraal College 1994).

In addition to the specialist examination, which, as indicated above, may be during the course of training, a number of colleges in the United Kingdom have introduced formal summative assessment at the end of specialist training. Continuous assessment, either as an examination procedure or feedback to trainees about the standard of their clinical performance, or both, is of vital importance. Such formative assessment of each doctor in training includes positive feedback about performance, specific help about any weakness and planning of future training. Summative appraisal, where knowledge, skills and attitudes are evaluated by college-appointed external assessors, will ensure that college-set standards are being met, so that the award of specialist grading may be made. Unlike other specialties in the United Kingdom, the Higher Training Committee for General Practice has statutory authority independent of the Royal College of General Practitioners. This is evidence of more direct involvement of the government in determining the duration and context of training for general practice than is the case for any other specialty.

Educational objectives are as necessary in postgraduate psychiatry are they are in basic medical education and have to be stated in the broad categories of knowledge, clinical skills and attitudes. Customarily, educational programmes are stated in little more than course descriptions. Much more specific educational objectives have to be specified, setting out the knowledge, skills and attitudes expected of trainees in all of the academic courses and aspects of clinical psychiatry included in the psychiatry postgraduate programme (Walton 1986, pp. 77–85).

4.2 Medical Specialities

Countries differ in the designation of separate specialities. Psychiatry is variably recognised; even in the European Union, where medical education and practice is regulated by binding legislation in the form of directives adopted in 1975, the specialty is not uniformly defined in all member states. (The specialties recognised in Europe are set out on pp. 11 and 12 of the European Specialist Medical Qualifications Order 1995). When psychiatry is one the recognised specialties in a country, there are differences in training requirements, including duration. For example, it has been said that psychiatric training in the United Kingdom takes 6–7 years (Caldicott 1996), but only 4 years in Portugal, Greece or Belgium (with no formal examination in Belgium).

4.3 National Differences in Standard Setting

In broad terms, five different patterns of responsibility for standard setting can be identified within Europe:

1. National medical associations, as in Norway and Portugal, take the leading responsibility. In other words, the medical profession itself provides and dispenses the postgraduate training in the country. This is also the case in Germany and in certain Central European countries.
2. Designated professional bodies, distinct from the national medical association, may have major responsibilities, as in the United Kingdom, where the Royal Colleges approve training posts and assess the competence of doctors in training.
3. The government (Ministry of Health), together with the regional medical administration, may take the main role. Sweden is an example of this arrangement.
4. Universities may be the chief responsible bodies, as in Finland and Southern Europe.

5. Postgraduate institutes in some countries are, or were, the chief postgraduate authority, as in Central and Eastern Europe.

In Europe, within each country, there has been steady evolution of postgraduate training over the past 20 years, with formal definition, recognition and support given to the educational programmes (Karle et al. 1993). Postgraduate medical training has secured the confidence of the public on grounds of the medical care provided by accredited specialists.

In the United Kingdom, the statutory basis for the medical education infrastructure has been regarded as a major strength of the national system. The General Medical Council, the body responsible for registering qualified medical practitioners, extended its remit to specify doctors regarded by the General Medical Council as specialists. However, the overriding influence of international law in European Law resulted in a challenge by the European Commission to the system of specialist accreditation in the United Kingdom, as it appeared to discriminate against the recognition of specialist training in other states of the European Economic Area (EEA) and thereby restrict the free movement of doctors. In responding to this challenge, the U.K. government set up the Working Group, whose work culminated in the Calman Report (Department of Health 1993). In meeting the legal requirements, the Working Group availed itself of the opportunity to review not only statutory provisions for specialist registration, but also the entire postgraduate training system in the United Kingdom.

The duration of postgraduate training has been shortened. Planned and structured rotational training programmes have been created to give each trainee an adequate breadth of experience. Doctors will work as specialists in the National Health Service (NHS) sooner than they do at present. The United Kingdom has approximately 30% fewer doctors per head of population compared with other EEA states. For the new arrangements for training to work, many more specialist posts are required. These are drastic changes indeed in the country's medical education system and thus in the structure of its medical services – and all as a result of the European Union's calling into question the country's specialist certification system and thus its medical manpower structure.

The United Kingdom, therefore, is one European country providing an illustration of the rapid, major re-orientation which its postgraduate training system is undergoing. The country has been powerfully influenced by external forces, in this instance the international legislation to which national medical education is subject in the United Kingdom as a member state of the European Union.

4.4 International Context

The European Union is the one region in the world where medical education is controlled by international legislation. The sectoral directives of the European Commission, such as those of 1975 for medical training, promised to be a catalyst for professional improvement and a guarantee of quality. The directives require provision for the free inter-country movement of doctors. The Advisory Committee on Medical Training (ACMT) is the statutory body in Europe for development of common standards and ongoing review of medical education, which in 1978 and 1982 defined and widely publicised criteria for coordinated training to occur in remunerated posts in both university centres and general teaching hospitals; other criteria include a predominance of training over service provision, the existence of a safe system of supervision and instruction in the prescribed curriculum by a range of training techniques. Directive 93/16/EC defines the minimum duration of recognised training demanded in each specialty to achieve specialist status.

So that women, in particular, are not penalised, training arrangements for doctors with domestic or other special responsibilities have to be specified; otherwise, parents who have interrupted their training for family reasons would be precluded from re-entering medical practice. European Union directives on postgraduate medical education now also specify the minimum number of hours of work in part-time training (for general practice, at least 60% of the hours worked by those employed full-time). Similarly, doctors training part-time in hospital specialties will be required to extend their already lengthy training on a pro-rata basis in order to qualify for specialist registration.

Reliance on the duration of educational experience as a guarantee of competence is of course highly questionable. In addition, in every country, professional competence must also be directly assessed by methods of performance appraisal and by formal examinations.

Doctors in postgraduate training have other professional needs which training programmes must cater for. They require, for example, to learn the limits of their own competence, to identify and remedy deficits in their knowledge and skills, to learn when and how to seek the assistance of others and to develop effective working relationships with professional colleagues and others. Structured training programmes are therefore obligatory, incorporating modern educational practice (Borman and O'Grady 1997).

5
Continuing Medical Education

Maintenance of competence throughout professional life is now obligatory and accepted as such by employing authorities, doctors themselves and the general public, who require quality medical care and accountability by doctors for the clinical service they provide. CME consists of those educational activities undertaken by practising doctors to maintain and update their clinical competence.

It is the personal responsibility of every doctor to remain competent. This ultimate obligation of individual doctors calls for proper means to assess constantly the quality of their work and effectiveness of their professional services, such as a system of performance review. Agreed standards are necessary, and computerised medical records greatly enhance audit, the process of examining one's clinical performance in relation to peers.

Being related to performance, CME goes beyond information and knowledge and focuses on the carrying out of actual clinical responsibility and services. The material included in CME, over and above lectures, case presentations, conferences, scientific meetings and conferring among clinical colleagues and their hospital firms or clinical teams, must be based on understanding how adults learn: they are oriented to tasks they undertake themselves and to the personal interests they pursue. Material must be relevant to the practitioner's own daily responsibilities. Old-style courses, in which doctors were brought together to sit through series of lectures haphazardly given by outside experts, are not congruent. Mere attendance at CME courses for which credit is given cannot be assumed to improve professional competence.

There is widespread agreement that the CME system in a country must be closely integrated with the health care system and related to the health needs of the population. Some countries are already imposing sanctions on doctors who do not engage in approved CME activities. Self-evidently, differences in the CME system between countries are to be expected, and indeed greatly varying patterns are compounded by national realities in the health care services, the varying professional bodies with authority for medical education and training, and the forms of statutory regulation of the medical profession.

Medical journals and other medical scientific publications always were and continue to be paramount in disseminating new knowledge and promoting best clinical practice (Vysohlid and Walton 1990). Special continuing education courses were first introduced, in some Central European countries, for doctors practising in state health services in 1878. As increasing numbers of countries introduced social insurance, furthermore, continuing improvement of their employees' professional competence also came to be expected. This direct intervention and involvement of the state, to a lesser or greater extent, in promoting new knowledge and experience for doctors in the public health service is the original model of CME.

Before the Second World War, opportunity was provided for CME in practically all European countries, with great differences in the providers, the organisation and the methods of activities, which – it goes without saying – were not nationally monitored or evaluated. Coordination and cooperation among undergraduate, postgraduate (specialty studies) and CME was non-existent.

5.1
International Cooperation

The World Health Organization (WHO) was founded as a health care agency of the United Nations system in 1948. The regionalised structure of WHO made it possible to deal with comprehensive health problems in terms of specific regional needs, conditions and circumstances.

In Europe, with its many and varied systems of health care, its very different needs, conditions and opportunities, the intentions for unified actions were difficult, if not impossible, to realise. Although some minor problems were recognised, partly solved and innovative approaches initiated, by the end of this period it became recognised that health personnel had not been trained to perform tasks which were essential for providing necessary services and care to the population as a whole. In the absence of long-term planning for training of health care personnel, and of doctors in particular, comprehensive planning, education and management of health personnel remained only partially achieved.

A new impetus for the attainment of greatly improved health for all people by the year 2000 "permitting them to lead a socially and economically productive life" was accepted in 1977 by all member states of the World Health Assembly, the health parliament of the world. In 1978, the Alma Ata Conference on Primary Health Care specified a concept of "primary health care" as the key to achievement of the health goals (World Health Organization 1978).

The European Regional Committee of WHO, within this framework, adopted for the first time a broad health policy for the European region in 1980. In 1984, the same body took the great step of deciding that this regional policy should be strengthened by adopting 38

European regional targets to be attained for resolving identified defects.

5.2 Educational Reform

A decisive initiative was taken by the World Federation for Medical Education, the international body recognised as representative of all stages of medical education, in 1984. The World Federation for Medical Education planned and conducted, for the first time, a world-wide assessment of all stages in medical education and training,. The assessment was based on agreed six main themes (World Federation for Medical Education 1986), which were first discussed at the country level of respective national associations for medical education; the outcomes were then reported to the six Regional Conferences, organised by the World Federation for Medical Education and the Regional Associations. The outcomes of these worldwide actions were used as the basis for the World Conference on Medical Education held in Edinburgh in 1988, with the close cooperation and support of WHO, the United Nations International Children's Emergency Fund (UNICEF), the United Nations Educational, Scientific and Cultural Organization (UNESCO) and the United Nations Development Programme, the U.K. government participating with the World Federation of Medical Education as hosts. The recommendations of the World Conference are summarised in the Edinburgh Declaration (World Federation of Medical Education 1988), translated into all main languages and now widely known and adopted (World Health Assembly 1989). The detailed recommendations of the World Conference are specified in the International Collaborative Programme for Reorientation of Medical Education, now being implemented at global, regional, national and institutional levels.

The reorientation called for by the International Collaborative Programme has many statutory and political implications requiring the active understanding and involvement of the governmental authorities of all countries at the highest level. Principle 8 of the Edinburgh Declaration called for ministerial consultations involving both Ministers of Health and Ministers of Education to be held in all six regions. The European ministerial consultation was held in 1988; the Ministers of Health and of Education of Portugal were the hosts, together with the European Office of WHO and the World Federation for Medical Education; UNESCO was actively associated. This European ministerial consultation, again for the first time, brought together the Ministers of Health and of Education in order to decide how best to implement in close cooperation the recommendations of the World Conference on Medical Education, their report designated by them as the Lisbon Initiative (Government of Portugal, European Office of the World Health Organization and the World Federation for Medical Education 1989).

In 1993, the World Federation for Medical Education called the World Summit on Medical Education, again in Edinburgh. The principles of the Edinburgh Declaration were re-affirmed, and the summit recommendations (World Federation for Medical Education 1994) specifically required CME systems in all countries to ensure professional competence of all medical doctors was updated throughout professional life. Particularly mandatory was the requirement to view the three phases of medical education as a continuum. Moreover, the undergraduate, postgraduate and CME phases had to be interrelated with regard to objectives, content, methods and assessment.

Of greatest moment have been the directives of the European Union which promote the free movement of doctors among the member countries in Western Europe. At the same time, massive, unanticipated and extraordinarily swift political changes have swept across the countries of Eastern Europe, immeasurably complicated by warfare and even genocide of unprecedented ferocity.

5.3 Range of Systems in Europe

The political transformation of the Central and Eastern European countries since the demise of communism has found expression in urgent approaches to other countries for advice and assistance. As these countries change to a market economy, their pressing need is for education and training support, in addition to advice about insurance systems or other methods for financing clinical services, on which training depends.

In all countries of Europe, some systems for CME exist, with the exception of very small countries such as Monaco, which send participants for CME to neighbouring countries. The many national variations are strikingly different in approach, and some of the priority problems can be cited:

1. There is often a lack of any close, working cooperation between the medical education system and the health sector, with neither related to the CME sector, especially with regard to health care needs.
2. A multiplicity of CME providers is the rule, e.g., the Health Service, the State (central, regional and district levels), professional associations, trade unions and private providers, without there being informed cooperation in planning, imple-

menting and evaluating CME programmes and activities.
3. Resources are lacking – both financial allocation and material provision – and planning is fragmented (as follows from the points above).
4. Much discussion over a long period, which is still unresolved, has centred on the conflict between voluntary versus compulsory CME.
5. CME is most commonly considered as merely a means to transmit information, and not recognised as an indispensable agent for change, as the WHO targets and the Edinburgh Declaration insists.
6. CME is now always adapted to identified and actual needs of the health problems of the population and to the health services, particularly in relation to primary health care and ambulatory settings. There has been very slow progress in active, distance and multi-professional learning and teaching activities; very little active involvement occurs by participants in the planning of CME programmes and in the related teaching activities.
7. The assessment of CME which is undertaken is concerned mostly with the teaching and learning process, and not with the outcome in improving the health care provided.

In the following, the variety of CME systems in the health service organisations existing in Europe by the end of the 1980s will be outlined in brief. The World Federation for Medical Education, together with the European Regional Office (EURO) of WHO in Copenhagen and UNESCO, conducted a project exploring and promoting the status in Europe of CME, with the collaboration of the Association for Medical Education in Europe (Walton 1993).

A survey was conducted by obtaining information from each National Association for Medical Education in the European countries. The national profiles have been reported in detail (Walton 1994) and provide a preliminary guide to knowledge about current arrangements in Europe.

There is nearly unanimous agreement among medical educators, practitioners and administrators everywhere that CME is essential in the interests of delivering good medical care. Nevertheless, there is equally clear diversity throughout Europe regarding the current needs, provision, aspirations, planning and legislation of CME.

There are grey areas, in the first place, in the definition of the boundary between CME and postgraduate training. This is most clearly seen in the case of general practice, where the requirement, in many countries, for formal postgraduate training in primary health care has resulted in the better provision of CME. On the other hand, the provision of postgraduate certification for specialists has sometimes led to a view of their being seen to carry an implication of qualification for a lifetime career, with consequent neglect of provision of formal CME. This view is not common, however, and numerous specialties in many countries have given particular emphasis to CME programmes.

Most national governments and the regional government (the European Union), with the support of medical associations and professional societies, are involved in sponsoring and financing CME, setting up mechanisms and providing resources. In about half the countries in Europe, legislation has been passed regarding it. Several national respondents referred to the controversy over the nature and purpose of such legislation. On the one hand, it can be *permissive*, giving every doctor the right and means to participate in CME, but without compulsion. On the other hand, CME can be made *mandatory*. In either option, permissive or mandatory, an optimistic view can be taken that everyone who attends CME activities will learn and practise what they have learned (implied when credits are conferred on the basis of attendance). The alternative licencing approach is to insist that all doctors participating in CME must demonstrate, in addition (usually every 5 or 6 years), that their knowledge and skills have kept pace with advances in modern medical care; otherwise they lose registration.

Naturally, such a re-licensure requirement raises anxieties, and therefore opposition, even among many competent practitioners who fear that they could lose their livelihood through a system that has not been adequately validated. This is probably why some informants considered a valuable role could be taken by the World Federation for Medical Education in encouraging research to evaluate CME programmes before those same programmes become used to evaluate and pass judgement on doctors. It is noteworthy that in countries where there is legislation regarding re-certification for certain groups, there were no worries expressed over such dangers.

At the practical level, the main barriers to CME identified in most countries are lack of finance, lack of other resources and limitation of time.

The provision of clinical information made available by modern information technology is the basis now needed for CME and is already in use in several countries. Many informants pointed out the relative unavailability of CME for doctors working in rural areas and distant from the main teaching areas. There is urgent need to extend awareness about effective techniques of CME and to make generally known why some of the CME practices currently in use are not valid and have to be discarded.

All doctors now have to accept personal responsibility for maintaining professional competence

throughout their working lives through CME. This third, and longest, phase of medical education is the most important, and both basic medical education and postgraduate training must now be planned and implemented as preparatory for lifelong CME.

6 References

Abrahamson S (1996) Essays on medical education. University Press of America, Lanham/MD

Bloom SW (1988) Structure and ideology in medical education: an analysis of resistance to change. J Health Soc Behav 29: 294–306

Borman E, O'Grady P (1997) Policy paper on postgraduate training: Permanent Working Group of European Junior Hospital Doctors. Med Educ 31: 3–8

Caldicott F (1996) Training in psychiatry in Europe. Adv Psychiatr Treat 2: 141–142

Centraal College (1994) Opleidingseisen psychiatrie. Med Contact 8: 269–279

Chuval J (1980) Entering medicine: the dynamics of transition. Pergamon, Oxford

Clare AW, Lader M (eds) (1982) Psychiatry in general practice. Academic, London

Department of Health (1993) Hospital doctors: report on the Working Group on Specialist Medical Training. Her Majesty's Stationery Office, London

Eagel PF, Marcos JR, Cancro R (1979) Medical students' attitudinal changes associated with the psychiatric clerkship. J Psychiatr Educ 3: 180

El Gaili DE, Hamad TA, Magzoub MEMA (1996) The teaching of mental health in a community-based medical school. Educ Health 9: 353–358

Government of Portugal, European Office of the World Health Organization and the World Federation for Medical Education (1989) The Lisbon initiative. Med Educ 23: 206–208

Her Majesty's Government (1995) European Specialist Medical Qualifications Order. Her Majesty's Government, London

*Karle H, Nystrup J, Walton HJ (1993) Medical specialisation in Europe: the way forward. Med Educ 27: 299–303

Lewin K (1948) Resolving social conflict. Harper, New York

Marston RQ, Jones RM (eds) (1992) Medical education in transition. Robert Wood Foundation, Princeton (Commission on Medical Education: The sciences of medical practice)

Merton RK, Reader G, Kendall P (1957) The student physician. Columbia University Press, New York

*Miller G (1980) Educating medical teachers. Harvard University Press, Cambridge/MA

Richards P (on behalf of the Council of Deans of the United Kingdom Medical Schools and Faculties) (1992) Educational improvement of the preregistration period of general clinical training. Br Med J 304: 625–627

Sanson-Fisher R, Maguire P (1980) Should skills in communication with patients be taught in medical schools? Lancet 8193: 523–526

Tosteson D (1994) Problem-based learning. Med Educ 28[Suppl 1]: 108–111

Vysohlid J, Walton HJ (1990) Development of continuing medical education in Europe: a review. Med Educ 24: 406–412

Walton HJ (1968) Different methods for teaching medical students. Proc R Soc Med 61: 109–115

**Walton HJ (ed) (1986) Education and training in psychiatry: a case study in the continuity of medical education. Oxford University Press, London

Walton HJ (1993) Project on continuing medical education in Europe. Postgrad Med J 69[Suppl 2]: 68–69

*Walton HJ (1994) Continuing medical education in Europe: a survey. Med Educ 28: 333–342

Walton HJ (1997) Small group methods in medical teaching. Med Educ 31: 457–464

Westberg J, Jason H (1994) Teaching creatively with video. Springer, Berlin Heidelberg New York

World Federation for Medical Education (1986) Six major themes. Med Educ 20: 378–389

*World Federation for Medical Education (1988) The Edinburgh Declaration. Lancet 8608: 464

*World Federation For Medical Education (1994) Proceedings of the World Summit on Medical Education. Med Educ 28[Suppl 1]: 1–171

World Health Assembly (1989) World Health Assembly Resolution WHA 42.38, 19 May 1989

World Health Organization (1978) Primary health care, Alma Ata. Report of the International Conference. World Health Organization, Geneva

World Health Organization (1985) Targets for Health for All. European Health for All Series, no. 2. World Health Organization, European Office, Copenhagen

**World Psychiatric Association/World Federation for Medical Education (1998) Core curriculum in psychiatry for medical students. International Center for Mental Health, Mount Sinai School of Medicine, New York

Part 1
Psychiatry in Specific Situations and Periods of Life

CHAPTER

1

U.M. Staudinger, P.B. Baltes

Lifespan Developmental Psychology

Translator: S. Goss

We would like to acknowledge the many valuable discussions with our colleagues from the Max Planck Institute for Human Development and the Research Network on Successful Midlife Development of the MacArthur Foundation (chair: O. Brim). We also wish to thank Hans Lauter and Hanfried Helmchen for their helpful comments on an earlier version of this contribution.

1 Introduction

Does human development extend across the entire life course or is it completed at early adulthood? Is development completely genetically "preprogrammed"? Is there great variability in the developmental trajectories of different individuals or in different domains of functioning? Can development be both latent and manifest? Is it possible to improve developmental trajectories? Questions such as these are addressed within the field of lifespan developmental psychology. In the following sections, we will try to provide answers to some of these questions.

In order to gain an accurate impression of human development, it is essential to approach it from a systemic perspective. It is not enough, for example, for psychologists to study only psychological development, for physicians and biologists to consider the development of physical functioning alone, or for sociologists and historians to analyze only the social embeddedness of development and how this evolves over time. Physical, psychological, and sociohistorical factors interact and condition each other in the process that we observe as human development.

2 Three Basic Assumptions About the Effects of Biological and Cultural Systems of Influence on Human Development

In lifespan developmental psychology, certain assumptions are made about the effects of biological and sociohistorical or cultural systems of influence on human development (P.B. Baltes 1997; P.B. Baltes et al. 1998). These assumptions are illustrated in Fig. 1 and will be elaborated on briefly below. It should be noted that the specific form of the three functions depicted is not decisive. What is critical is the *overall direction* and the *reciprocal relationship* between the three functions.

2.1 Evolutionary Selection Benefits Decrease Across the Lifespan

The first assumption about the effects of biological and cultural systems of influence on human development is that the selection benefits resulting from biological evolution display a negative age correlation (e.g. Finch 1996). As a consequence, with increasing age the human genome contains more deleterious genes and thus more dysfunctional genetic expressions than in younger years.

The main reason for the "evolutionary neglect of old age" lies in the midlife location of biological selection. The reproductive system, a central component of natural selection, uses fertility and parenthood – events which typically occur in early adulthood – to ensure the selective transmission of the gene. Consequently, over evolutionary history, genetic selection operated *primarily* on the first half of life. Moreover, due to the shorter life expectancy in early human evolution, selection processes could not operate as frequently to begin with when it came to the second half of life. Most people died before possible negative genetic attributes became manifest.

One concrete illustration of the age-related biological weakening of the human organism is the existence of late-life illnesses such as Alzheimer's disease (for further examples, see Martin et al. 1996). This disorder does not typically become manifest before the age of 70. After this age, however, the number of new diagnoses increases markedly (e.g. Helmchen et al. 1999). Following the representation shown in Fig. 1a, Alzheimer's disease is at least in part a late-life disease because reproductive fitness-based selection processes had only a very slim chance, if any, to eliminate it.

Biological aging is, of course, influenced by other factors which individually or conjointly further reduce the biological functioning and plasticity of the organism across the lifespan (e.g. Finch 1996; Martin et al. 1996). Many of these age-related biological losses are associated with the mechanisms of ontogenesis itself, e.g. natural wear and tear, information losses within systems (entropy costs), and the cumulative increase in genetic mutations.

These various considerations about the role of genetic and biological factors in lifespan development result in a converging conclusion which is generally undisputed. Evolutionary selection and the ontogenetic biology of the aging process mean that, with time, the human body becomes increasingly susceptible to dysfunction and marked by symptoms of decline.

2.2 Need for Culture Increases with Age

The second assumption is that cultural development has to be able to compensate for biological processes of decline in order for adaptive development to extend on into advanced old age. With increasing age, the need for supportive cultural resources to help maintain previous levels of functioning becomes more and more pronounced (see Fig. 1b). This view of "culture as compensation" is a main tenet in many evolutionary theories in cultural anthropology (e.g. Elwert 1994).

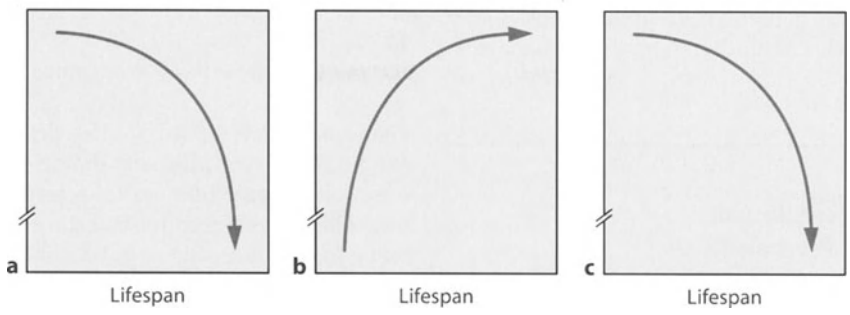


Fig. 1a–c. Three basic assumptions which play a decisive role in regulating the nature of ontogenic development. a Evolutionary selection benefits: decrease with age. b. Need for culture: increases with age. c Efficiency of culture: decreases with age. (After Baltes 1997)

In this context, culture includes all of the psychological, socially interactive, material, technological, institutional, and symbolic (knowledge-based) resources which humans have developed over the millennia. These resources, passed on from generation to generation, have made human development and the life course possible as we know them today (Durham 1991).

Human ontogenesis was therefore only able to achieve increasingly higher levels of functioning (e.g. a longer lifespan and the ability to read and write) because of the concurrent development and dissemination of culture and the associated structures of societal opportunity. If human development is to extend further and further into old age, it will be necessary for additional sociocultural forces and resources to emerge.

Culture and ontogenesis are inextricably intertwined, and the increase in average life expectancy during the twentieth century illustrates the extent of their mutual dependence. Life expectancy rose from an average of approximately 45 years in 1900 to about 75 years in 1995. This marked increase was not caused by changes in the genetic makeup of *Homo sapiens* that occurred over this period of time. On the contrary, it was primarily societal, cultural, and technological progress which contributed to the significant increase in life expectancy.

2.3 Efficiency of Culture Decreases Across the Lifespan

Figure 1c illustrates the third assumption, according to which there is an age-related decrease in the effectiveness or *efficiency* of cultural factors and resources. The human organism is defective from birth onward and thus in need of cultural support in order to survive. However, with age (the age of onset is unclear), the relative efficiency of psychological, social, material, and technical interventions wanes. This is conditioned primarily, but not exclusively, by biological processes of decline.

Cognitive learning potential provides us with an illustrative example of the reduced efficiency of cultural factors over the life course. The older adults are, the more time, practice, and cognitive support it takes them to attain the same level of learning success (e.g. Kliegl and Baltes 1991). Moreover, when it comes to peak levels of performance, older adults are unable to attain the same results as younger adults, even after extensive training. The same can be said for plasticity on the neurobiological level (Magnusson 1996). Neuronal plasticity continues to exist across the entire lifespan, but it becomes impaired with age. The age-related reduction in pharmacologically relevant adaptivity, as shown in pharmacokinetics and pharmacodynamics, is another good example (see Coper and Schulze 1994).

This third assumption might be called into question, or its force diminished, by the following two objections:

1. The notion that culture- and knowledge-based symbolic systems are inherently different from, or even more effective, than biological and physical systems. For instance, fewer lifespan developmental losses in information content (entropy costs) are observed for symbolic systems than for biological processes, which implies that symbolic systems should be able to function longer and more effectively during ontogenesis (P.B. Baltes et al. 1998). However, we should not overlook the fact that the operation of symbolic systems is inalienably tied to a minimum degree of biological functioning (in order to impart knowledge, for example, an individual must first be able to retrieve and formulate that knowledge).
2. The term "efficiency" refers to an evaluation of performance which cannot be applied to certain facets of development, such as the meaning of life and other personal meaning systems (e.g. Dittmann-Kohli 1995; Filipp and Aymanns 1996). This is indeed a serious line of argument. However, it is unlikely that it will alter the ontogenetic *direction* of the efficiency function outlined,

although it will perhaps have implications for the extent to which this third assumption holds.

3 Consequences for Lifespan Developmental Psychology

These three assumptions form a theoretical framework which has certain consequences for the further development of a psychological theory of human ontogeny across the lifespan.

3.1 Three Categories Organize the Two Systems of Lifespan Developmental Influences

The first consequence of this theoretical framework is that development occurs under both facilitating and limiting conditions. According to P.B. Baltes (1987), these systems of opportunities and constraints can be categorized into the following three groups:

1. Age-graded
2. History-graded
3. Non-normative

Some biological and environmental influences have a strong relationship to chronological age and are therefore fairly predictable in their temporal course. For example, the age-related clustering and sequential ordering of biological and cultural challenges has been summarized in the concept of developmental tasks (e.g. Havighurst 1972). The theory of developmental tasks shows that at particular stages of life, we are faced not only with particular biological and cultural opportunities, but also with challenges and constraints (e.g. at the age of 65, symptoms of physical decline, retirement, loss of social contacts, more freedom in planning leisure).

Biological and environmental influences on development may also be history-graded, meaning that they depend on historical time. The influence of pharmacology and nutrition on health and of the education system on cognitive development are examples of such determinants.

Finally, however, there are also non-normative biological and environmental systems of developmental influences. The term "non-normative" is used here in the statistical sense, i.e. these influences do not follow a predictable course, but instead differ from individual to individual. They include idiosyncratic life experiences, such as being involved in an accident or winning a lottery.

3.2 Lifespan Development as a Multifaceted Dynamic System

Figure 1 also allows us to infer that attainments and functions that are primarily biologically based should evince a different lifespan trajectory than those functions that mainly feed on cultural resources. Biology-based functions should reveal decline from early adulthood onward, while culture-based functions should show stability – and under favorable conditions lasting growth – well on into late adulthood. Any theory of ontogenetic development which postulates that developments in later adulthood are "generally positive" or "generally negative" across all domains of functioning must be regarded as false. It would be just as erroneous to assume that child development only consists of gains (see P.B. Baltes 1987). On the contrary, lifespan developmental psychology works on the assumption that development is multidimensional, multidirectional, multicausal, and multifunctional throughout life (M.M. Baltes and Carstensen 1996; P.B. Baltes et al. 1998).

By *multidimensional* we mean that development does not consist of a unitary process, but that it occurs simultaneously in different domains of functioning. For example, an individual's intelligence does not develop independently of his or her personality and social relations. And as we will see below, subfacets of broader domains of functioning, such as intelligence (e.g. reaction speed and knowledge), can develop differently, depending on whether they are primarily based on biology or culture.

Development is *multidirectional*, i.e. not only is variability of development observed in different domains of functioning on the intraindividual level, but there are considerable interindividual differences in developmental trajectories. Development does not consist exclusively of growth or loss on either the intra- or the interindividual level, but involves both components at all points of ontogenesis. Intellectual performance, for example, tends to show losses with time, whereas features of personality and social behavior tend to remain stable. To put it simply, development can also be described as a dynamic balance between gains and losses. However, the balance of developmental gains to losses does become less positive across the lifespan. Thus it is important to differentiate between developmental trajectories specific to particular domains of functioning and the overall balance of developmental gains and losses.

What constitutes a gain in ontogenetic change and what represents a loss is a topic of both theoretical and empirical inquiry (e.g. M.M. Baltes and Carstensen 1996; P.B. Baltes and Baltes 1990). For example, great differences can be observed depending on whether an

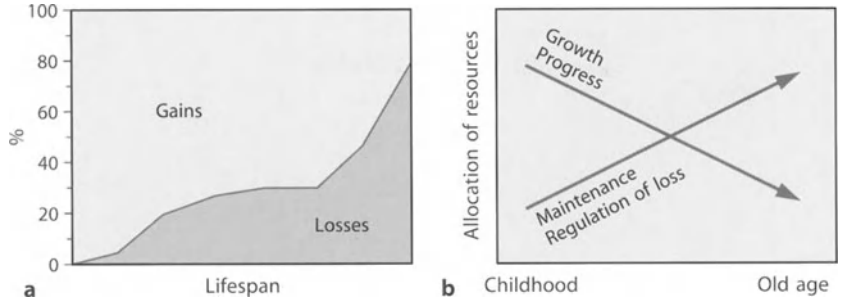


Fig. 2. a With increasing age, losses come to predominate. (After Heckhausen et al. 1989). **b** The relative allocation of resources into three adaptive developmental goals: lifespan changes in the investment of resources into growth, maintenance, and the regulation of losses. (After Staudinger et al. 1995)

external view of the developmental outcome is taken, using objective criteria, or whether the internal view of the developing person is taken, using subjective criteria. Cultural context and age also play a significant role.

The intrinsically relative nature of what is classed as a gain or a loss is encapsulated in the concept of the *multifunctionality* of development. The notion of multifunctionality denotes that there is no single criterion for ontogenetic adaptivity or for what is to be considered a gain or a loss. Investigations on the topic of autonomy versus dependency provide a telling example of such multifunctionality (see M.M. Baltes 1996). It has been shown that the advent of dependent behavior in old age should not always be seen as a loss. Within the everyday ecology of older adults, dependency can also imply gains such as increases in social contact and close relationships to caregivers.

3.3 Developmental Reserves and Their Allocation Across the Lifespan

The third consequence of the assumptions illustrated in Fig. 1 pertains to the latent potential of development. According to the third assumption presented above, the efficiency of culture decreases with age. The aging organism thus displays a reduced plasticity. In other words, there is not only an age-related decrease in biologically based functioning and the level of functioning, but in the potential for adaptivity and repair.

It is important to distinguish between the development of the functional status and the development of adaptivity reserves in lifespan developmental psychology. Only when both of these levels of development are taken into consideration can a comprehensive conception and understanding of lifespan development be arrived at. The performance status attained by an individual at a given point in time cannot be equated with his or her maximum performance potential. Taking developmental reserves into account allows us to consider possibilities for intervention and improvement, i.e. the ways of shaping development.

In lifespan developmental psychology, a distinction is made between three major functions of developmental potential and reserves: growth, maintenance/recovery, and regulation of nonreversible losses (see also Staudinger et al. 1995). In this context, growth means all behaviors involved in reaching higher levels of functioning or adaptive capacity. Maintenance and recovery (resilience) refer to the adaptive goal of maintaining levels of functioning in the face of new challenges or losses. The regulation of loss means adaptive behavior aimed at securing and coming to terms with lower levels of functioning when maintenance of former levels is no longer possible.

Losses in functioning and the conjoint reduction in adaptive capacity, or reserves, suggest that there is a systematic pattern in the *relative* allocation of reserves to these three adaptive functions across the lifespan (see Fig. 2b; P.B. Baltes et al. 1998; Staudinger et al. 1995). In childhood, the major share of resources is invested in growth and the quest for higher levels of functioning; during adulthood, the predominant allocation is toward maintenance and recovery (resilience). In old age, an ever-increasing proportion of the available resources has to be directed toward the regulation of irreversible losses.

3.4 Illustrating Lifespan Theory: Empirical Evidence on Intellectual Development

In the following, intellectual functioning is used as an empirical example to illustrate the tenets of lifespan psychology presented above (see P.B. Baltes et al. 1998; Lindenberger and Baltes 1994). Ever since the early work of Hebb and Cattell, a distinction has been made between two main components of intelligence (compare multidimensionality of development): biology-regulated fluid intelligence and experience-based crystallized intelligence.

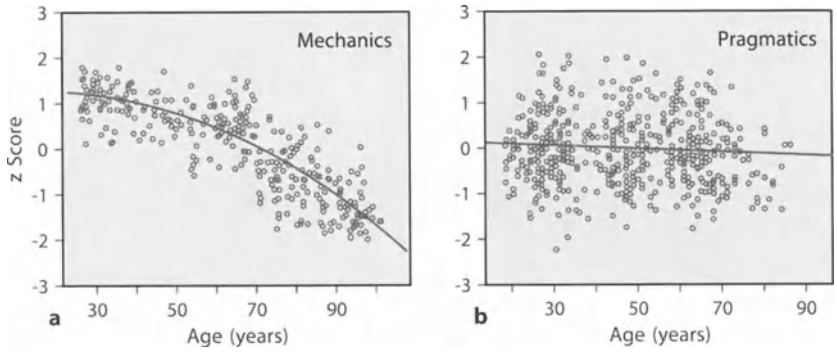


Fig. 3a,b. Pragmatics and mechanics of intelligence: adult age gradients for two exemplary kinds of task. The figure displays empirical data; each *dot* represents the performance of one study participant. (After Lindenberger and Baltes 1995)

In lifespan developmental psychology, this psychometric theory of intelligence has been combined with perspectives from process-oriented cognitive psychology, and a distinction has been made between the fluid *mechanics* and the crystallized *pragmatics* of intelligence (e.g. P.B. Baltes 1987).

By analogy to computer language, the mechanics of intelligence correspond to the "hardware" of the mind, i.e. the brain's neurophysiological information-processing basic operating system (Barkow et al. 1992). The mechanics of intelligence include elementary processes of information processing such as information input, visual and motor memory, and basic perceptual-cognitive processes such as discrimination, comparison, and categorization, as well as the application of these processes in working memory. Because of the close connection between the mechanics of intelligence and neurobiological development, lifespan theory postulates that the life-course trajectory of the mechanics takes the form of an inverted U, with the decline beginning in young adulthood, if not earlier.

Staying with computer language, the pragmatics of intelligence can be understood as the culture- and knowledge-based "software" of the mind. They reflect culturally acquired bodies of declarative and procedural knowledge. Typical examples of the pragmatics of intelligence include reading and writing skills, language, and professional expertise, but also the knowledge about the self and the world which is required in the planning, conducting, and interpretation of life.

Figure 1 suggests that the ontogenetic trajectory of pragmatics can show stability and even growth for longer than that of mechanics, pragmatics being primarily determined by sociocultural factors (compare multidirectionality of development). How far the trajectory can be extended and refined is, in essence, dependent on two conditions. The first is the availability of cultural resources and their translation into programs of cognitive enhancement. The second is the mechanics of intelligence and their development across the lifespan. The mechanics are important because they provide the basic operations on which pragmatic performance is based. For example, if an individual is no longer able to retrieve the relevant knowledge from his or her memory in response to a question, this knowledge is of little further use to him or her.

Most empirical studies of intelligence come to conclusions which are consistent with both this dual-component model and the three basic assumptions of lifespan development. Marked differences are observed in the lifespan trajectories of different aspects of cognitive functioning, depending on whether the task involved is based on biology or culture (Fig. 3).

Thus abilities that involve the mechanics of intelligence, such as working memory, speed of information processing, and attention span, typically show negative age gradients even in early adulthood (Lindenberger and Baltes 1994). However, trajectories for pragmatic components of intelligence, including vocabulary tasks, dimensions of social intelligence, and expert knowledge about life and the world such as wisdom, often remain stable into later stages of life. The lifespan trajectory of pragmatics, which we have studied with reference to the prototypical concept of wisdom (see Fig. 3), profits from certain personality characteristics and experiences, the processing of these, and exchange with other people.

In our wisdom studies, for example, older participants performed as well as younger adults up to at least the age of 75. After this age, however, performance levels seemed to decline in wisdom tasks as well (Staudinger and Baltes 1994). In advanced old age, the biologically determined limitations of the mechanics of intelligence impose increasing constraints on what culture and individuals can accomplish.

Findings on the interdependence between the two components are of particular importance. There is ample evidence that the mechanics and pragmatics of

intelligence interact. For instance, the pragmatics of cognition can be used to offset losses in the mechanics. It has been shown that older office workers who continue to be excellent typists compensate for their reduced reaction time (mechanics: decrease in information-processing speed) by reading further ahead in the text to be typed (pragmatics: experience) than their younger colleagues who have a faster reaction time (Salthouse 1991).

Investigations carried out within the framework of testing-the-limits research show that intellectual developmental reserves also decrease with age (Kliegl and Baltes 1991). The testing-the-limits method is an experimental approach which, not unlike the use of stress tests in medicine, examines the individual at his or her upper limits of functioning, where differences in performance are more pronounced. While older individuals do improve their performance when they undergo training, they are not able to equal the peak performances of younger persons, even after long-term training programs. It should be noted that the age gradient for the mechanics of intelligence remains identical when members of different social classes are compared. Only the level of functioning at the start of the process of decline distinguishes between people from various backgrounds (Lindenberger and Baltes 1995). In other words, culture is not able to guard against cognitive decline.

Further support for the assumption that the mechanics of intelligence are mainly biologically determined is provided by the finding that there is an extremely strong association between sensorimotor functioning and intelligence. All negative age differences in general intellectual functioning disappear when interindividual differences in sensorimotor functioning are controlled for with simple measures of visual and auditory acuity as well as balance. We believe that this powerful connection between sensorimotor functioning and intelligence in old age has a common cause (a so-called third-variable hypothesis), i.e. the age-related changes in neurophysiological brain functioning which concurrently impair the sensory, sensorimotor, and intellectual abilities (P.B. Baltes 1997).

4
A General Model of Development: Selective Optimization with Compensation

Human development across the lifespan is thus characterized by a reduced level of reserves on the one hand and by a combination of growth, stability, and losses in various domains of functioning on the other, with stability and losses coming to predominate with age. In recent years, lifespan developmental psychologists have proposed models describing general developmental processes that make it possible to deal with such conditions in an adaptive way (see, e.g. Brandtstädter and Greve 1994). One of these models is that of selective optimization with compensation (SOC) developed by P.B. Baltes and M.M. Baltes (1990; see also Heckhausen and Schulz 1995).

The precise definition of the terms selection, optimization, and compensation is governed by the theoretical framework and specific domain of functioning in which they are to be applied at any one time. The following general descriptions of these processes hold when an action-theoretical approach is taken:

- Selection refers to the *direction*, the goal, or the outcome of development (e.g. an individual may define himself in the following five phrases: good story-teller, good father, athlete with plenty of stamina, loyal friend, and faithful husband. Should serious physical impairment occur in old age, he can exclude the domain "athlete with plenty of stamina" and concentrate on the remaining domains instead).
- Optimization refers to the *resources* or means which allow developmental goals or developmental outcomes to be realized (e.g. individuals may set themselves the goal of getting a promotion at work and focus their effort on achieving this).
- Compensation involves an adaptive *response to the loss of means* (resources), with the aim of maintaining the functional status (e.g. people may write notes when they realize that their memory is deteriorating).

SOC theory proceeds from the assumption that any process of development involves a unique combination of selection, optimization, and compensation. We further assume that, on account of the biological and cultural factors which condition development, there is a lifespan script for these combinations of selection, optimization, and compensation. For example, as is to be expected on the basis of the assumptions described above, selection and compensation retain more significance than optimization with increasing age (Freund and Baltes 1998).

The use of action-theoretical perspectives in this taxonomy of the SOC model might suggest that SOC is always a process with intention and rationality, but this is not so. Each of the three components can be active or passive, internal or external, conscious or unconscious. Moreover, during ontogenesis, the focus of the components may shift. For instance, forms of behavior which initially emerged as compensatory processes may later be used as strategies of optimization (P.B. Baltes 1997).

5 Adaptive-Productive Functions of Self and Personality

The various empirically identified adaptive-productive functions of self and personality (summarized briefly below) are easily integrated into this general model of development (see Staudinger et al. 1995). Measures of functioning from the domain of self and personality (e.g. subjective well-being, personality characteristics, control beliefs, coping strategies) display a lifespan trajectory that is more akin to the pragmatics than the mechanics of intellectual functioning, i.e. they are primarily characterized by stability. Furthermore, and perhaps contrary to expectations, there is no increase in depressivity and anxiety in old age (e.g. Helmchen et al. 1999). We should not forget that considerable reserves exist in the domain of pragmatics, even in advanced old age (Staudinger et al. 1995). This is illustrated by the fact that older people are able to profit from therapy (e.g. Radebold 1989).

This lifespan trajectory of self and personality functioning lends support to the notion of an "orchestrating" function of self and personality where the activation and management of internal and external developmental resources is concerned. By orchestrating function we mean that certain processes and structures related to self and personality are able to offset losses experienced by the aging person in many domains of functioning, balance out the losses through selection and compensation, and optimize the situation once more. In this context, we speak of the psychological resilience of the self (e.g. Staudinger et al. 1995). Such processes and structures include goal selection, compensatory coping styles, a rich and differentiated self-image, self-efficacy beliefs, conscientiousness, and openness to new experiences (see Staudinger et al. 1995).

6 The Fourth Age: A New Challenge for Science and Society

With increasing age, however, even the processes of selection, optimization, and compensation can no longer completely offset losses. Impairments become increasingly extensive and irreparable. This has recently lead to this stage of life being described as "the fourth age" (P.B. Baltes 1997). The losses and limitations experienced by the very old – those aged over 80 – are far more extensive than those observed in young old age or "the third age" (see P.B. Baltes and Mayer 1999).

In the years and decades to come, the very old will present a formidable challenge for society. Take the prevalence rates of almost 60% for Alzheimer's disease in the over-90 age-group, for example. In a large, representative study carried out in America, Crimmins et al. (1996) studied the percentage of sickly and dysfunctional years of life remaining in old age. They found that for 70-year-olds, 20% of the remaining lifetime is marked by illness and inactivity. For 90-year-olds, this rose to 60%. Hence, the oldest of the old do not survive simply because their functional status is superior to that of their deceased counterparts. The physical functioning of the survivors is also seriously impaired, but not yet to the extent that death ensues.

Moreover, the increase in risk for the fourth age is not observable only where physical variables are concerned, but also when the focus is on psychological variables such as intelligence, the self, personality, and social behavior (Smith and Baltes 1999; Staudinger et al. 1999). With the exception of intelligence, the negative age effects of each individual psychological variable were, in general, relatively small. However, the psychological age gradients for the 70- to 100-year age-group in the Berlin Aging Study all sloped in the same direction, namely downward, in the sense of increasing losses.

When these rather weak age effects for individual variables are aggregated into multivariate profiles, a succinct and rather negative picture of psychological development in old age develops. From the psychological perspective, advanced old age appears to be a period of life which is increasingly characterized by stress and challenge. It appears that advanced old age, the fourth age, represents a kind of boundary situation in which the limits of psychological resilience are frequently overstepped.

7 Concluding Remarks

Although the optimization of human development becomes increasingly difficult in advanced old age, we do not wish to leave the impression that achieving a positive balance of gains and losses in all ages of life is not possible in principle. Neither the biological framework nor the current state of cultural development ought to be viewed as a fixed representation of "the" nature of human aging. States of deficit and limitations, such as those characteristic of the fourth age of life, are powerful catalysts for scientific, cultural, and psychological innovation. Furthermore, despite age-related losses in developmental reserves, there is latent potential for the enhancement

of functioning in advanced old age as well (e.g. training gains, receptiveness to therapy). In this spirit, we need to keep in mind that the future of (our) aging is not something we simply enter, but something that we help create.

8 References

Baltes MM (1996) The many faces of dependency in old age. Cambridge University Press, New York

*Baltes MM, Carstensen LL (1996) The process of successful ageing. Ageing Soc 16: 397–422

Baltes PB (1987) Theoretical propositions of life-span developmental psychology: on the dynamics between growth and decline. Dev Psychol 23: 611–626

Baltes PB (1997) On the incomplete architecture of human ontogeny: selection, optimization, and compensation as foundation of developmental theory. Am Psychol 52: 366–380

Baltes PB, Baltes MM (1990) Psychological perspectives on successful aging: the model of selective optimization with compensation. In: Baltes PB, Baltes MM (eds) Successful aging: perspectives from the behavioral sciences. Cambridge University Press, New York pp 1–34

Baltes PB, Mayer KU (eds) (1999) The Berlin Aging Study: aging from 70 to 100. Cambridge University Press, Cambridge

*Baltes PB, Lindenberger U, Staudinger UM (1998) Life-span theory in developmental psychology. In: Lerner RM (ed) Handbook of child psychology, vol 1. Theoretical models of human development. Wiley, New York, pp 1029–1143

Barkow JH, Cosmides L, Tooby J (eds) (1992) The adapted mind: evolutionary psychology and the generation of culture. Oxford University Press, New York

Brandtstädter J, Greve W (1994) The aging self: stabilizing and protective processes. Dev Rev 14: 52–80

Coper H, Schulze G (1994) Arzneimittelwirkungen im Alter: Bedingungen – Besonderheiten – Folgerungen. In: Baltes PB, Mittelstraß J, Staudinger UM (eds) Alter und Altern: Ein interdisziplinärer Studientext zur Gerontologie. de Gruyter, Berlin, pp 204–230

Crimmins EM, Hayward MD, Saito Y (1996) Differentials in active life expectancy in the older population of the United States. J Gerontol Soc Sci 51B: 111–120

Dittmann-Kohli F (1995) Das persönliche Sinnsystem: Ein Vergleich zwischen frühem und spätem Erwachsenenalter. Hogrefe, Göttingen

Durham WH (1991) Coevolution: genes, culture and human diversity. Stanford University Press, Stanford/CA

Elwert G (1994) Alter im interkulturellen Vergleich. In: Baltes PB, Mittelstraß J, Staudinger UM (eds) Alter und Altern: Ein interdisziplinärer Studientext zur Gerontologie. de Gruyter, Berlin, pp 260–282

Filipp SH, Aymanns P (1996) Bewältigungsstrategien (Coping). In: Adler RM, Herrmann JM, Köhle K, Schonecke OW, von Uexküll T, Wesiack W (eds) Psychosomatische Medizin. Urban and Schwarzenberg, Munich, pp 277–289

Finch CE (1996) Biological bases for plasticity during aging of individual life histories. In: Magnusson D (ed) The life-span development of individuals: behavioral, neurobiological and psychosocial perspective. Cambridge University Press, Cambridge, pp 488–511

Freund AM, Baltes PB (1998) Selection, optimization, and compensation as strategies of life-management: correlations with subjective indicators of successful aging. Psychol Aging 13: 531–543

Havighurst RJ (1972) Developmental tasks and education. McKay, New York

Heckhausen J, Schulz R (1995) A life-span theory of control. Psychol Rev 102: 284–304

Heckhausen J, Dixon RA, Baltes PB (1989) Gains and losses in development throughout adulthood as perceived by different adult age groups. Dev Psychol 25: 109–121

Helmchen H, Baltes MM, Geiselmann B et al (1999) Psychiatric illnesses in old age. In: Baltes PB, Mayer KU (eds) The Berlin Aging Study: aging from 70 to 100. Cambridge University Press, Cambridge, pp 167–196

Kliegl R, Baltes PB (1991) Testing the limits: Kognitive Entwicklungskapazität in einer Gedächtnisleistung. Z Psychol [Suppl 11]: 84–92

Lindenberger U, Baltes PB (1995) Kognitive Leistungsfähigkeit im Alter: Erste Ergebnisse aus der Berliner Altersstudie. Z Psychol 203: 283–317

Lindenberger U, Baltes PB (1994) Aging and intelligence. In: Sternberg RJ (ed) Encyclopedia of human intelligence, vol 1. Macmillan, New York, pp 52–66

*Magnusson DL (ed) (1996) The life-span development of individuals: behavioural, neurobiological and psychosocial perspectives. Cambridge University Press, Cambridge

Martin GM, Austad SN, Johnson TE (1996) Genetic analysis of ageing: role of oxidative damage and environmental stresses. Nat Genet 13: 25–34

Radebold H (1989) Psychotherapie bei älteren Menschen. In: Kisker KP, Lauter H, Meyer JE (eds) Psychiatrie der Gegenwart, vol 2. Springer, Berlin Heidelberg New York, pp 234–248

Salthouse TA (1991) Theoretical perspectives on cognitive aging. Erlbaum, Hillsdale/NJ

Smith J, Baltes PB (1999) Trends and profiles of psychologcial functioning in very old age. In: Baltes PB, Mayer KU (eds) The Berlin Aging Study: aging from 70 to 100. Cambridge University Press, Cambridge, pp 197–226

Staudinger UM, Baltes PB (1994) Psychology of wisdom. In: Sternberg RJ (ed) Encyclopedia of human intelligence, vol 2. Macmillan, New York, pp 1143–1152

Staudinger UM, Marsiske M, Baltes PB (1995) Resilience and reserve capacity in later adulthood: potentials and limits of development across the life span. In: Cicchetti D, Cohen D (eds) Developmental psychopathology, vol 2. Risk, disorder, and adaptation. Wiley, New York, pp 801–847

Staudinger UM, Freund A, Linden M, Maas I (1999) Self, personality, and life regulation: facets of psychological resilience in old age. In: Baltes PB, Mayer KU (eds) The Berlin Aging Study: aging from 70 to 100. Cambridge University Press, Cambridge, pp 302–328

CHAPTER

2

H. Remschmidt and E. Fombonne

Developmental Psychopathology

1 Developmental Psychopathology as an Integrative Discipline

There can be no doubt that the developmental perspective is of great importance for the understanding of psychiatric disturbances in children and adolescents. Developmental physiology, developmental neurology, and developmental psychology are basic sciences of child psychiatry. The developmental perspective can be looked upon as a kind of bridge between the different disciplines or as a unifying concept (Eisenberg 1977), integrating different scientific and practical approaches to normality and psychopathology, not only for children, but also for adults. Though there is general agreement about this view, we are far away from a comprehensive and substantial theory of development that would be able to integrate earlier and recent knowledge – and at the same time be open to new hypotheses and results.

Yet the developmental perspective is not a new one. Looking through the history of child psychiatry and psychopathology, we found one of the first hints of the developmental aspect in the German textbook by Emminghaus (1887) entitled *Psychic Disturbances of Childhood*. After complaining that there is no systematic and general symptomatology of childhood psychoses, Emminghaus writes that it is the task of psychopathology to study the anomalies of the mind through all developmental stages and to differentiate normal from pathological psychic processes (Emminghaus 1887, p. 4).

Other textbooks written by Homburger (1926), Kanner (1935), and Tramer (1941) include the developmental perspective, followed by books by Achenbach (1974, 1982), Rutter (1980), Remschmidt and Schmidt (1992), and the *Handbook of Developmental Psychopathology* edited by Cicchetti and Cohen (1995).

Nevertheless, the emergence of developmental psychopathology was characterized by "adultomorphism," i.e. a tendency to understand psychiatric disorders of childhood as predecessors of analogously named disturbances in adults.

According to Sroufe and Rutter (1984, p. 18), the basic perspectives of developmental psychopathology are as follows:

1. The discipline "is concerned with development and is therefore closely wedded to the whole of developmental psychology." Logically, developmental psychopathology thus has to use the methods, theories, and perspectives of developmental psychology.
2. The second topic is the focus on psychopathology, which means looking primarily at developmental deviations.
3. The third element is the integrative perspective, including and combining biological, psychological, and psychosocial approaches with respect to all structures and functions in the growing child.

Taking these elements into consideration, we may define developmental psychopathology as

> ...the study of the origins and course of individual patterns of behavioral maladaptation whatever the age of onset, whatever the causes, whatever the transformations in behavioral manifestation, and however complex the course of the developmental pattern may be (Sroufe and Rutter 1984, p. 18).

Another definition, given by Rolf and Read (1984, p. 9) runs as follows:

> The term 'developmental psychopathology' can be defined as the study of abnormal behavior within a context of measuring the effects of genetic, ontogenetic, biochemical, cognitive, affective, social or any other ongoing developmental influences on behavior.

Both definitions stress the integrative aspect, and from this point of view, developmental psychopathology "is the product of an integration of various disciplines the efforts of which had previously been separated and distinct" (Cicchetti 1984, p. 1).

Bearing these definitions in mind, we can say that the developmental perspective includes processes such as growth, maturation, and learning as well as the interactions between and among these influences. However, one has also to realize that not all influences on children have to do with development; accidental influences also exist, i.e. impairments and unfavorable conditions that are not developmental factors as such, although they do influence development in a profound way. The changes in development are age related and consist of transformations in the structure and the function of an organism, including quantitative and qualitative changes, the latter emerging with the formation of new structures and functions.

The view of development would not be complete without including the role and nature of experience. According to Gottlieb (1976), one may distinguish three distinct roles of experience in the development of behavior:

1. Experience can *maintain* or *preserve* behavioral states.
2. Experience can *facilitate* development by accelerating its rate without changing its course or by increasing the terminal level of proficiency achieved.
3. Experience can *induce* new forms of behavior which will directly reflect the configuration of the stimulus event.

Gottlieb concludes from a review of the literature that there is evidence for the roles of maintaining and facilitating experience, but that at the moment there is no evidence for the role of inducing experience. This issue might have to do with sensitive or critical periods in development, known from research with animals, where new kinds of behavior are induced. The only behavior in children that could be related to this issue is the development of social bonds and attachment in infants.

2 Levels for the Study of Behavior

Behavior can be studied on several levels that represent normal and abnormal behavior of different complexity. In Table 1, eight levels are presented, beginning with very simple ones and ending with very complex ones that take into consideration interactions of individuals, family structures, and also structures that might be influential over several generations.

With respect to a developmental perspective, normal and abnormal behavior on the different levels can show patterns of continuity and discontinuity. Thus it has been shown in several studies that dissocial behavior and conduct disorders that begin early during individual development remain stable over wide age spans and therefore follow a pattern of continuity. In contrast, internalizing disorders such as anxiety states and neurotic disorders show a high tendency of spontaneous remission and therefore represent a pattern of discontinuity.

Table 1. Levels of behavioral organization (from Remschmidt 1992)

Level	Normal behavior	Abnormal behavior
Molecular	Behavior traits	Symptoms
Molar	Patterns of behavior traits	Syndromes
Functions	Levels of functioning	Levels of dysfunction
Adaptation processes	Patterns of adaptation	Patterns of maladaptation
Individual	Personality	Disturbance of personality
Interactional	Personality in context	Personality disturbances in context
Family	Normal family function and structure	Family dysfunction and deviant family structures
Generations	Normal structure of generations	Abnormal structure of generations

From the viewpoint of developmental psychopathology, future research should pay more attention to more complex functions than to symptoms and syndromes. Patterns of adaptation and maladaptation in different age-groups and developmental stages and behavioral reorganization seem to be very important issues for future studies. Behavioral reorganization at different developmental levels, for instance, is a key issue without which certain new abilities or sometimes also a disability cannot be fully understood. Kagan (1981, 1984) has suggested that

> ...many instances of developmental change can be characterized by the replacement of an old structure or process by a new one, with little or no connection between the two hypothetical structures. The suggestion implies that some structures or processes vanish (Kagan 1981, p. 68).

Examples of such reorganization are, according to Piaget (1950, 1983), reaching the level of abstract thinking or the hormonal changes at puberty and their subsequent effects on body and psychological functions. By reaching these levels of functioning, new abilities are acquired that cannot be understood in terms of previous abilities and that represent discontinuity. In the same context, so-called turning points in development may be discussed (Rutter 1996). These are often dependent on individual experiences that are able to restructure past experiences and positively influence the self-concept of children. From these examples, it should be clear that the interaction of biological and psychological influences has to be studied longitudinally, not on the simple levels of behavior traits, symptoms, and syndromes, but rather on higher levels of functioning, including the processes of restructuring and reorganization at different age levels.

3 Milestones of Normal Development: Implications

Several authors have defined the stages of development in relation to different ages. Well-known systems of this kind are the developmental model of Sigmund Freud (1948) and Jean Piaget (1983).

Table 2 describes four developmental models, those of Freud and Piaget and those described by Erikson (1980) and Jersild and Holmes (1935). These models are based upon different perspectives. While the Freudian system is based on the psychosexual stages of development, Erikson concentrates his system on psychosocial stages, Piaget on cognitive ones, and Jersild and Holmes on affective ones. The last column of the table describes different psychopathological

Table 2. Anthony's proposed diagnostic classification for developmental psychopathology (Anthony 1970)

Age (years)	Psychosexual stages (Freud 1948)	Psychosocial stages (Erikson 1980)	Cognitive stages (Piaget 1983)	Affective stages (Jersild and Holmes 1935)	Psychopathology
0–1 1/2	Oral	Basic trust vs. mistrust	Sensorimotor	Fears of: dark, strangers, being alone, sudden noise, loss of support	Autism, anaclitic depression, feeding and sleeping problems
1 1/2–3	Anal	Autonomy vs. doubt, shame	Symbolic	Separation, desertion, sudden movements	Symbiosis, negativism, constipation, shyness and withdrawal, night terrors
3–5	Genital, oedipal	Initiative vs. guilt	Intuition, representational	Animals, imaginary creatures, injury	Phobias, nightmares, speech problems, enuresis, encopresis, anxiety states
6–11	Latency	Industry vs. inferiority	Concrete, operational	School failure, ridicule, loss of possessions, disfigurement, disease, death	School problems, school phobias, obsessions, conversion symptoms, tics
12–17	Adolescent; recapitulation of earlier conflicts	Identity vs. role confusion	Formal, operational	Being different physically, socially, intellectually; sexual fears; loss of face	Identity diffusion, anorexia nervosa, delinquency, schizophrenia

disorders that can be related to the relevant age-groups or developmental stages.

The problem of such a classification is that not all of these stages can be substantiated by empirical studies, e.g. the so-called latency stage in the Freudian system has no empirical evidence. Beyond that, there exists a high variability with regard to age, and finally, the different theories of development do not also correlate with regard to age. For instance, a child could have reached the symbolic stage according to the cognitive theory of Piaget and still be at the oral stage according to the Freudian system. Therefore, modern developmental theories are mainly based on functions (e.g. physical growth, cognitive development, moral development, identity, self-concept) in relation to developmental tasks. A developmental task can be defined according to Havighurst (1972, p. 2) as one

> . . .which arises at or about a certain point of life of the individual, successful achievement of which leads to his happiness and success at later tasks while failure leads to unhappiness in the individual, disapproval by society, and difficulty with later tasks.

Theories of development "attempt to describe, predict, and explain specific changes in behavior over some part of the life-span. To do so, they must generate testable predictions, i.e. ones that can be confirmed or disconfirmed" (Jaffe 1998, p. 47). However, not all theoretical assumptions are and have been subject to empirical testing. Nowadays, there is general agreement that developmental changes depend upon the interaction between hereditary factors and environmental influences, taking into account the fact that the individual is not a passive object of the influences but also shapes and creates his or her own environment.

Current developmental psychopathology has now shifted away from the stage models toward the process models of development that emphasize these interactions between individuals and the large variety of environmental influences and social contexts, at the same time paying attention to the individual as an active participant in his or her own development (Lerner 1996; Wachs 1996; Zahn-Wexler 1996).

These modern trends of developmental sciences also have implications for developmental psychopathology, which still relates psychopathological conditions to the classical stage models instead of the process models that are based on a continuous development of the different functions, paying special attention to the process of interaction.

4 Research Fields and Strategies

In the following section, we would like to refer to some research fields and strategies in developmental psychopathology that seem important to us and promising for the near future.

4.1 Sex Differences and Individual Differences

There is a great amount of knowledge about sex differences in relation to psychopathological disorders

and also of individual differences of general development (for reviews, see Steinhausen 1992; Costello and Angold 1995). The sex differences concern physical state, growth, maturation, development of differential abilities such as speech and language functions, spatial abilities, and most psychopathological disorders (e.g. hyperkinetic syndrome, developmental delays, autism, aggressive and dissocial behavior).

Until puberty, most psychopathological conditions are more frequent in boys than in girls, although the boy–girl relation is very different for the various syndromes. For instance, the hyperkinetic syndrome is up to nine times more frequent in boys than in girls (Cantwell 1977), whereas in autism the relation is 3.7:1 across 20 studies (Fombonne 1998). Tourette's syndrome is more frequent in boys than in girls (3:1; Shapiro et al. 1978). Obsessive–compulsive disorder is also more frequent in boys than in girls (Rapaport 1986).

Although these differences are very clear and well established, we do not really know their cause. This question is therefore open for research, and several hypotheses have been offered:

- Some differences might be connected with a different maturation of the hemispheres in boys and girls. Histological studies of infant brains by Conel (1939–1959) gave evidence for the assumption that selected regions of the cortex are more mature in girls than in boys during the time from birth to the second year of life. A faster maturation may also apply to the left hemisphere and the language functions, which are more advanced in girls than in boys until puberty (Kelly 1985).
- Psychological and psychosocial factors could also be important for some of the above-mentioned sex differences. It is a well-established fact that socialization and education are different in boys and girls. This may explain the proneness of girls toward depression during adolescence in the light of differential socialization patterns (Sroufe and Rutter 1984).
- There could also be an interchange between biological factors such as maturation of the hemispheres and educational and psychosocial influences.
- It might also be possible that some differences between boys and girls are not in fact real differences, but rather caused by a different time pattern of maturation (Kelly 1985). In line with this hypothesis are the results published by Waber (1976, 1977) which demonstrate differences in psychological profiles between early and later maturers. Following this line, it could very well be that early acquired achievement structures could enhance future achievements, finally causing a new or adifferent quality of achievement structure. This hypothesis is connected with another open question, namely, the progressive development from quantitative to qualitative changes during development. Finally, one should mention that the androgenes play an important role in influencing maturation with long-lasting differences in neuropsychological functioning between early maturers and late maturers (Jones 1965; Katchadourian 1977; Rubinow and Schmidt 1996).

A very important issue in both individual and sex differences is aggression. As to childhood aggression, biological factors are as important as psychological ones. Among the biological factors, the relationship between testosterone levels and aggressive behavior in adolescent boys has been studied by Olweus et al. (1988). They found that the circulating levels of testosterone in the blood had a direct influence on provoked aggressive behavior, measured by self-reports. High levels of testosterone intensified impatience and irritability in boys, thus increasing their proneness to aggressive–destructive behavior. On the other hand, psychological factors are also of great importance, e.g. a "deviation from usual attributional processes." According to Bobbitt and Keating (cited in Sroufe and Rutter 1984), "the attributional error is a potentially dysfunctional social cognitive skill that mediates aggressive activity in these boys."

The understanding of sex differences with respect to aggression in boys and girls would also be the key to our understanding of delinquency. This issue also leaves us with the open question of why there is so much continuity in aggression and dissocial behavior in boys.

4.2 Continuity and Change of Behavior

The second research field is that of continuity and discontinuity of behavior in children with respect to the developmental perspective (Rutter 1984, 1995). Although we know the different types of disorders, e.g. a continuous type remaining stable in childhood and diminishing toward adulthood and a type of newly manifested disorder during adolescence, the cause for these different types remains unknown. Further on, the continuity of psychopathological entities from childhood to adulthood depends also (a) on the research strategy (e.g. follow-up, follow-back study), (b) the sources of information (personal investigation, interview, records, questionnaires), (c) the type of sample (clinical populations, general populations), (d) the type of disorder (neurosis, conduct disorder), (e) the outcome variables, and (f) the diagnostic criteria (Garber 1984).

Irrespective of these methodological considerations, the following continuities have been found in several studies: There is a high continuity in pervasive

developmental disorders such as early infantile autism, Asperger's syndrome, and Rett's syndrome (Van Acker 1997; Gillberg and Steffenberg 1987; Gillberg 1991; Klin and Volkmar 1997). There is also a high continuity concerning externalizing behavior patterns (Kazdin 1986; Taylor 1989; Verhulst 1992) and aggressive behavior in boys (Olweus 1979). In addition, a substantial proportion of boys characterized by hyperactivity–impulsivity and attention deficit problems tend to have continuing behavior problems such as antisocial behavior or personality and substance abuse (Gittelman et al. 1985) or aggressive behavior and delinquency (Farrington et al. 1990; McGee et al. 1984a,b).

In a delinquency study, Remschmidt et al. (1984) were able to demonstrate that positive life events (successful professional career, friends, trust in an adult person) were associated with stopping a delinquent career. In relation to this issue, the concept of "turning points during the individual life course" (Rutter 1996) could be useful. Nevertheless, here too it has yet to be demonstrated *why* there is continuity in some children and not in others. With respect to the continuity issue, those children belonging to a certain diagnostic category of the continuity type (e.g. conduct disorders, hyperkinetic syndrome) are especially interesting if they are able to stop the behavior in question on the way to adulthood.

A somewhat neglected issue is that of discontinuity. With the exception of certain developmental disorders that are connected with maturation processes, such as enuresis, encopresis, and some transient speech disturbances, not much is known about discontinuity. It is, for example, unclear why a certain proportion of children with depression do not develop a depressive disorder of the adult type. The same applies to a small proportion of children and adolescents with conduct disorders or substance abuse. It may very well be the case that major life changes induce positive turning points in favor of an undisturbed development of adolescents (Rutter 1995).

4.3 Research on Risk Factors and Protective Factors

Research on risk factors has been one of the most important issues in developmental psychopathology during the last two decades. Using the prospective longitudinal approach, different developmental courses of children at risk were studied and compared with matched control groups. The most interesting question in this context is the development of those subjects at risk who do and do not develop the disorder in question (Garmezy 1974; Garmezy et al. 1984; John et al. 1982; Robins 1978; Robins and Rutter 1990).

This approach gives us not only a better understanding of the effect of risk factors, but also new insights into the process of development as such. Finally, from this study we can draw conclusions for primary prevention, which is an almost completely unsolved problem at the moment. In this context, the following open research questions emerge:

1. What causes are at work in facilitating the manifestation of disorders in children at risk?
2. Which risk factors are most important and how do they interact with each other and with different other influences during the course of development?
3. Which influences stop the continuation of a disorder or a behavior pattern that has been in action for a certain period of time? Examples are dissocial behavior, delinquency, different anxiety states, social phobia, and depression.

The last question can be solved only by longitudinal prospective research in groups with a different set of risk factors and compared with carefully matched control groups. One problem in this context is the natural course of several disorders, which is presently unknown.

Risk factors and protective factors interact with each other during the course of individual development. At the moment, we have by far more knowledge from empirical research on risk factors than on protective factors. The type and nature of protective factors may be very different. We can distinguish protective factors in the child itself (e.g. high self-esteem, favorable temperamental features), within the family (e.g. high family cohesion, good relationship between the parents), and others that have to do with the presence of favorable influences in the external system. As we all know, some protective factors exist and have been studied over the past years, e.g. favorable temperament, good marital relationship of the parents, high self-esteem of the child, and female sex, which is a protective factor against psychiatric disorders, at least until puberty. Nevertheless, many questions remain open for research.

4.3.1 General Mechanisms in the Context of Which Protective Factors Are at Work

There are, of course, hypotheses characterized by the issues of learning, maturation, identification, self-concept, interaction, and attribution. For instance, there is evidence that early secure relationships are favorable influences for the development of high self-esteem. The same applies to successful experiences concerning a high achievement level at early age. High self-esteem and a positive self-concept are in this sense

modified according to the experiences of a child (Rutter 1985; Masten and Garmezy 1985).

As far as the mechanism of action is concerned, two possibilities are likely (Rutter 1995):

1. Children shape and select their environment to some extent, and the capability to do that seems to be connected with certain protective factors.
2. Children also vary in their mode and intensity of response to risk environments, which also seems to be modulated by protective factors such as temperament or favorable experiences in early life.

4.3.2 Coping Mechanisms at the Different Ages and Developmental Stages in Relation to the Developmental Tasks

The issue of coping mechanisms is one of the most promising in all developmental psychopathology. Table 3 gives an overview of the potential relationship between psychopathology and developmental tasks.

The study of coping mechanisms in relation to developmental tasks may lead us to a better understanding of both normal developmental processes and pathological ones. In the clinical field, we can learn a great deal from our patients by asking them questions and observing them while they carry out developmental tasks. Once again, it seems most interesting whether, at certain stages of development, key turning points are at work in changing the course of development and making a child more resistant in the face of adverse factors. This issue, put forward by Rutter, has not yet been investigated, but might be a fruitful idea for our understanding of unexpected developmental changes.

Table 3. Potential relationship between psychopathology and developmental tasks (Garber 1984)

Psychopathological disorder	Developmental task
Separation anxiety	Object permanence, attachment and dependency
Depression	Differentiation of self, self-esteem, social comparison
Suicide	Concept of death, time perspective (future)
Conduct disorder Undersocialized	Moral development
Aggressive	Perspective-taking, empathy
Impulsivity	Delay of gratification
Oppositional disorder	Autonomy, individuation
Schizoid disorder	Peer relations, friendship patterns

4.3.3 Interaction of Risk Factors and Protective Factors

It is too simple an idea that a complementary relationship exists between risk factors and protective factors. Both kinds of influences are not independent of children's personality, their experience, their surroundings, and the development of their self-concept. Thus the study of the interaction of protective and risk factors has to control for many other influences. It may also be dependent on the intensity of risk factors whether protective factors are facilitated or not.

4.4 Research on Prediction

Another important task of developmental psychopathology is to study the predictive power of specific patterns of behavioral or emotional organization within the context of general development. If we look at the large number of influences, issues, and problems present during the course of individual development, we might arrive at the very pessimistic conclusion that prediction will not be possible at all. Several studies (Robins 1978; Bohman 1978; Kohlberg et al. 1972) show that this is not the case.

Why is this so? We think there is a hierarchy of behavioral patterns with a very different predictive power. Successful prediction of later behavior from earlier patterns depends very much on the behavior pattern that was chosen as the predictor.

The following general results can be put forward (Kohlberg et al. 1972; Sroufe and Rutter 1984):

- Early maladaptation predicts later maladaptation (e.g. school failures, poor peer relationships, pronounced antisocial behavior, overdependency in preschool age).
- Early competence and maturity predict later competence.
- Absence of problems and symptoms does *not* predict later competence or maladaptation.
- The "strongest predictors likely will be adaptational failures defined in age-appropriate terms" (Sroufe and Rutter 1984, p. 24).
- There is also evidence that prediction is better when using longitudinal parameters rather than cross-sectional ones.

Important questions for research in prediction may be the following ones:

- Detecting behavior patterns that have a high predictive value in different age and developmental

stages concerning the natural course of behavior and the course of behavior disorders
- Studies on the prediction of therapy success derived from initial variables at the beginning of therapy or from variables characteristic of the individual, the individual's family, and the disorder
- The different predictive value of behavior patterns at different ages and developmental stages.

All of us know that language acquisition and intelligence is of high predictive value in infantile autism, but what are the optimal predictors in these children after they have reached the age of 6? Then other predictors are important, although they are not as efficient as the two mentioned concerning the very early stages of development (for a review, see Gillberg 1991). In the case of psychiatric disorders, other predictors than during normal development are important. Nevertheless, there is a relationship between developmental tasks (as steps in normal development) and symptomatology.

In addition to these considerations and examples, several other predictors of later psychopathology are known:

- Conduct disorders in childhood carry a high risk for depressive symptoms in adult life (Harrington et al. 1990, 1991; Rutter 1991). Further on, adverse parenting in childhood is also associated with depressive disorders in adult life (Rutter 1995).
- Obsessive–compulsive disorders in childhood and adolescence show a high persistence in adulthood. A worse outcome is associated with the presence of tics, parental mental disorders, and an unsuccessful treatment with clomipramine (Leonard et al. 1993).
- Premorbid nonanorexic eating disorders in childhood are associated with a poor prognosis in anorexia nervosa (Remschmidt et al. 1990). Anorexia nervosa in childhood and early adolescence predisposes to nonspecific eating disorders and anxiety states in later life (Herpertz-Dahlmann et al. 1996).
- Hyperkinetic disorders in childhood are associated with an increased rate of antisocial personality disorders, an increased rate of substance abuse, and criminality in adult life (Klein and Mannuzza 1991; Mannuzza et al. 1993).
- Dyslexia in childhood shows a high rate of persistence in adolescence and adulthood and is highly associated with conduct disorders in adolescence with a diminishing tendency in adulthood (Maughan et al. 1994; Rutter 1995).
- Early-onset schizophrenia has a poor prognosis and reveals a high persistence in adulthood (Remschmidt et al. 1994).

4.5 Classification and Categorization of Disorders and Behavior Patterns from the Developmental Perspective

Bearing the above-mentioned research fields and issues in mind, it becomes very clear that a developmental framework for classification is essential (Graham and Skuse 1992; Garber 1984; Sroufe and Rutter 1984). Progress was made when a developmental axis was included in the ICD-9 (World Health Organization 1975) as well as in the ICD-10 multiaxial classification system (World Health Organization 1992) and in DSM-IV (American Psychiatric Association 1994). However, we think we need to go further and include the following research problems (Remschmidt 1989; Graham and Skuse 1992):

1. To further investigate the validity of criteria for use with adults and to develop special criteria that are specific for childhood disorders. It is quite clear that several criteria from adult psychiatry cannot be applied to psychopathological conditions in childhood. For instance, the concept and criteria of personality disorders, of depression, and of anxiety disorders are quite different from what is observed in childhood: age-specific aspects, phase of life orientation, longitudinal course, and interactional aspects are left out (Graham and Skuse 1992). On the other hand, specific criteria for childhood such as hyperkinetic disorders, developmental disabilities, autism, and pervasive developmental disorders that do not have a counterpart in adult psychiatry are still ill-defined and need a further subdivision according to the results of recent investigations and ongoing research (Rutter 1995; Remschmidt et al. 1998; Rispens et al. 1998).
2. To include the issue of continuity and discontinuity of childhood disorders in the classification systems. This issue, discussed earlier in this chapter, raises an important problem with regard to psychiatric nosology. Recent investigations have demonstrated that several psychiatric disorders in childhood and adolescence reveal a high level of comorbidity. Thus the question arises of whether the comorbid disorder is really a separate psychopathological condition or whether it belongs to the core syndrome. A good example is childhood depression, which shows a high comorbidity with conduct disorder and anxiety states (Harrington 1993) and also much continuity as far as depressive symptoms are concerned. In other disorders, e.g. Gilles de la Tourette's syndrome (GTS), the symptomatology changes in relation to age. A quite substantial subgroup shows the following sequence of symptoms from childhood to adulthood: hyperactive

behavior, motor tics, vocal tics, obsessive-compulsive symptoms, and/or obsessive-compulsive disorder (Leckman and Cohen 1988). From this example, it needs to be considered whether GTS is a nosological entity over the entire life span or whether there is a transition into another syndrome in a certain phase of life. This issue has yet to be solved.

3. To change the focus of classification from more limited and isolated behaviors toward patterns of adaptation. These are, of course, different in different developmental stages. A good example (and probably the only one) in this field is the classification presented by Ainsworth concerning the scheme of infant-caregiver attachment during early infancy (Ainsworth et al. 1978): secure attachment, anxious-avoidant attachment, and anxious-resistant attachment. It has been found that these "patterns" are stable for a long time and have predictive value for later emotional behavior. Efforts are now being made to include such patterns of adaptation into classification systems concerning behavior in preschool age (Greenspan and Lourie 1981; Sroufe 1983).
4. Using the idea that not only symptoms or syndromes are important for classification, but also patterns of adaptation, the idea of including the kind and pattern of relationships in a classification system also emerged. This is a very typical and important issue in developmental psychopathology, because relationships are clearly dependent on age and developmental stage. We are still at the beginning of this issue, and from the perspective of the classification of relationships, old problems of psychopathology such as continuity-discontinuity and normality-abnormality have to be seen in a new light.
5. When taking up these new issues, it is also important to further develop empirically derived classification systems and taxonomies. We would like to mention in this context the proposal made by Achenbach (1982, 1995), which uses not only the general categories of "internalizing and externalizing behavior," but also other dimensions and in particular different sources of information (child, parents, and teachers). Thus, finally, in the classification of behavior patterns, the context in which they are observed is very important (e.g. at home, at school, or with peers). A very good example of these different perspectives is provided by the hyperkinetic syndrome. A hyperkinetic child may not be hyperactive in a dyadic communication and situation, but very hyperactive in the classroom together with other children. Thus the issue of "situational context" seems to be very important. Only a few studies exist that focus on these contextual approaches in classification.

4.6 Research on the Nature of the Developmental Process Itself

A comprehensive discussion of research on the nature of the developmental process itself is beyond the scope of this chapter. Nevertheless, we would like to say a few words on this fundamental issue in developmental psychopathology. First, it has to be clear which influences are specifically connected with development and which are not. Second, we have to define very carefully the salient issues of developmental processes in different age-groups. One example is the approach of developmental tasks, but there might be others as well. Third, it is important to differentiate between quantitative and qualitative changes during development and to focus on the relationship between, or the transition from, quantitative to qualitative dimensions. Piaget (1950, 1983) proposed the process of equilibration for the transition from quantitative to qualitative changes. Fourth, we should study the antecedents of psychopathology in close relation to age-defined adaptation processes. This kind of view is not static but dynamic and includes everything discussed before. In doing this, we must be aware of the fact that development is not entirely a progressive phenomenon, but also includes regressive patterns. By this, we do not mean the psychoanalytic concept of regression, but the fact that, at a certain moment in development, both new and progressive tendencies and earlier patterns of adaptation and maladaptation can be observed simultaneously without the necessity of falling back on an earlier developmental stage, as proposed by psychoanalysts.

Finally, the concept of turning points seems to be important and worthy of future study for both normal and abnormal development. In summarizing research on the nature of the developmental process itself, it is necessary to have a relatively universal model or theory of development that is able to include the issues already mentioned and is suitable for theoretically guided longitudinal research.

5 Implications of Developmental Psychopathology for Diagnostics

The influence of the developmental perspective on childhood psychopathology has increased in the successive revisions of the DSM and ICD classifications. This is most obvious for disorders in which age of onset, developmental level, and outcome in adult life are key features of the definition. As an example of the

former, the ICD-10 separation anxiety disorder is defined as "an abnormal continuation of developmentally appropriate separation anxiety" with an onset before the age of 6, which postulates a close link between normal and pathological development. The need to take developmental level into account is exemplified in the diagnosis of autism, where the emphasis is placed on qualitative abnormalities in the definition of communication, social interaction, and play/interests abnormalities and where clinical descriptors for each area of dysfunction have been chosen to illustrate clinical presentations across the life span. Reliance on longitudinal research is shown in the use of diagnostic categories such as depressive conduct disorder (ICD-10) to account for the different prognostic implications for adult adjustment of childhood depression, whether or not it occurs in conjunction with conduct disturbances. However, current nosologies are far from being developmentally sensitive for most disorders. The diagnosis of depression, for instance, is made using adult criteria with little acknowledgment of phenomenological differences in younger, less cognitively mature individuals. Attempts to devise sets of diagnostic criteria specific to childhood depression have so far failed to demonstrate a consistent superiority (Garber 1984). Similarly, the validity of several diagnostic categories used in child psychiatry has not been adequately tested due to the relative scarcity of outcome studies and data on response to treatment.

The assessment of psychopathology needs to be based on a sound knowledge of what constitutes normal development. Thus an important contribution made by child psychiatric epidemiology over the last 30 years was to demonstrate that many symptoms such as fears, phobias, and nervous habits are reported in large numbers of nonreferred children (Rutter et al. 1970; Achenbach and Edelbrock 1981) and cannot be used alone to differentiate between normal from pathological development. In addition, the assessment of psychopathology in children requires complex information-gathering and -processing strategies. In contrast to adult psychiatry, child interviews do not always provide the main source of data needed to assess psychopathology. Rather, children's problems are generally assessed using a variety of procedures and of informants. This reliance on multiple sources of data reflects the situational variations of children's behavior and the cognitive immaturity of young subjects who cannot report on their symptoms accurately. The agreement between informants is, however, far from perfect. In their meta-analysis, Achenbach et al. (1987) showed that the mean correlation was 0.60 between two parents or two teachers, but only 0.28 between two different types of informants (i.e. parent and teacher) and 0.22 between children's self-reports and other informant ratings. These findings clearly illustrate the necessity to rely on multiple sources of data to assess children's behavior and emotions, as no single informant is likely to provide a comprehensive picture of them. Taking this issue one step further, Achenbach (1993) has recently devised an assessment paradigm which is based on cross-informant syndrome constructs, i.e. clusters of symptoms which occurred in the majority of several age and sex groups according to several informants. The eight syndromes thus identified provide an empirical and quantitative system particularly useful to assess current psychopathology and to measure its changes over time.

More generally, it has been recognized that both dimensional and categorical approaches to the measurement of child psychopathology are needed to avoid scientific investigation being stopped prematurely in this field. Even with disorders which represent extreme departures from normal development, the question arises as to where to draw the boundary between indicators of susceptibility to the disorder and expressions of normal development. Thus, to take the example of autism, it is now recognized that genetic susceptibility is indexed in relatives of autistic probands by a broader phenotype, the features of which overlap with expressions of normal variations in development; the measurement of such susceptibility, crucial for etiologic and genetic research, is indeed more amenable to a dimensional model than to a categorical one (Piven et al. 1997; Smalley 1991). Several other areas of psychopathology also illustrate the need to be attentive to subthreshold expressions of psychological dysfunction and of maladaptive behaviors which are not captured by classical, dichotomous, diagnostic approaches and yet have an important psychopathological significance.

For example, with regard to eating disorders, there is a group of adolescents (boys and girls) who meet the weight criteria of anorexia nervosa (BMI below 17.5 according to ICD-10), but are not anorexic in terms of the other criteria. This observation raises the question of whether these individuals are subthreshold anorexics. The same applies to restraint eaters who reveal low blood leptin levels like anorexics, but do not suffer from this disorder according to the ICD-10 or DSM-IV criteria (von Prittwitz et al. 1997).

When assessing psychiatric disorders in children and adolescents, it is now recognized that, in addition to a specific pattern of symptoms, there should be evidence of impairment of functioning (Bird et al. 1990; Weissman et al. 1990). This is consistent with a dynamic view of psychopathology which emphasizes the need to assess several areas of functioning at the same time and to gauge the interference of symptoms with the capacity to perform developmentally appropriate social and cognitive tasks for an individual. Of course, the current measures of impairment are still

too crude, and an important task on the research agenda is to devise subtler, age-referenced, and culturally sensitive measures of impairment.

Current nosologies still rely, however, on a medical and adult-oriented model of "psychiatric disorder" which assumes that dysfunction is entirely defined by its symptom constituents and is based within the individual. To some extent, the emphasis placed on reliability of psychiatric diagnosis, as exemplified by 20 years of use of symptom-orientated interviews and research diagnostic criteria, has contributed to this tendency. While there is no doubt that this emphasis on replicable and rigorous measurement approaches was and will remain necessary, especially when compared to previous theory-driven and unfalsifiable assessment views (A. Freud 1965), it could be argued that there is a danger of reifying psychopathology in uncritically adopting a "disorder" formulation of child psychopathology.

Thus the validity of most of these disorders remains to be firmly established, particularly regarding the identification of distinct etiologic processes at their origin (Jensen and Hoagwood 1997).

In addition, current classifications of child psychiatric disorders do not adequately reflect the fact that disorders, however they are defined, represent complex transactions between an individual and his or her environment and that the historical, cultural, and familial background of the subject is relevant both to the onset, maintenance, and cessation of maladaptive behaviors. Most experienced clinicians weigh the symptom pattern against some form of appreciation of the meaning and background of the individual child; however, these clinical procedures are highly idiosyncratic and remain poorly operationally defined and standardized. Timid attempts to reflect the contextual and transactional nature of child psychopathology can, however, be observed in the multiaxial systems currently in use in which psychosocial situations, life events, and family factors are described, together with a measure of overall impairment. Nevertheless, these contextual variables are mostly treated as accompanying features which need to be covaried out rather than as playing a full role in the etiologic process and current display of dysfunctional behaviors and leading to appropriate intervention. Developmental psychopathology requires models of a higher order of complexity than those currently available in an individual-based approach to disorders. These models will be developed as more data become available from longitudinal studies integrating different domains and contexts of measurement. Perhaps the most important need at this stage is to maintain a critical stance toward our available assessment and diagnostic approaches of psychopathology, in order to avoid their premature closure and to allow for further theory development.

6 Implications of Developmental Psychopathology for Therapy

The consideration that complex transactions exist between different areas of functioning and various social contexts and across developmental stages should ideally lead to fine-tuned multimodal interventions. Indeed, child psychiatry practice has long been characterized by such multifaceted approaches combining community, school-based, family, and individual measures. The particular use of these ingredients certainly reflects the developmental orientation of child psychiatry treatments, as seen, e.g. in the fact that most interventions in preschool children would consist largely in family- or school-based interventions, whereas individual therapies used in isolation are mostly seen among mature, psychologically open-minded adolescents. Similarly, cognitive–behavioral therapy is used with various age-groups, with behavioral techniques used predominantly among the youngest groups and cognitive techniques among the more mature subjects. The fortune of interventions based on family therapy principles or parent–child interactions also reflect the transactional and interactive nature of models used to treat child psychopathology.

Much less is known, however, about the optimal sequences, combinations, and particular timing of these interventions for any given problem. Thus little research has been conducted so far on the relative merits of combinations of interventions in child samples, such as the combination of psychotherapy and drug treatment. In addition, and for obvious reasons, research into the efficacy of intervention has generally followed a paradigm where one variable is manipulated (i.e. the drug or the number of therapy sessions), while other variables are hopefully maintained constant or equal across treatment modalities. These useful randomized interventions allow for a test of the absolute efficacy of one treatment component, but they do little to assess the role of the context in the processes of change. Furthermore, the assessment of intervention has tended so far to focus on symptom reduction as a single measure of outcome, whereas symptom levels as such might be less important to overall outcome than changes in the pattern or direction of symptoms or than parallel increases in social skills, self-esteem, or academic achievement. For example, the Perry preschool project, conducted as part of the assessment of the head-start program, offers a good illustration of the need to adopt an approach to outcome evaluation which includes both a long-term perspective and a wide-ranging strategy to assess change. Thus poor children randomly allocated to an

active preschool program showed improved educational achievement and employment opportunities in their late teens, as well as lower rates of crime, delinquency, and, for females, of teenage pregnancies than controls (Berrueta-Clement et al. 1984). Furthermore, most evaluative research has been centered on the individual, with few studies incorporating measures of change in the broader social context of the intervention, i.e. the family or the school. Recognizing these limitations, Hoagwood et al. (1996) have recently developed a model of outcomes for mental health care in children which, in addition to symptoms and diagnoses, incorporates adaptive functioning, quality of life of the child and his or her family, environmental settings, and broader system functions as elements in outcome measurement. While the empirical value of this model remains to be established, its heuristic value is certain and would appeal to most clinicians and researchers in the field.

7 Conclusions

Though the developmental perspective is not new, developmental psychopathology has only recently become an interdisciplinary and integrative field of research for the understanding of normal and pathological development as well. Progress has been made during the past two decades in several respects: regarding individual and sex differences of certain disorders, continuity and change of behavior, risk factors and protective factors, and turning points in development. Moreover, the current classification systems include the developmental perspective on a special axis devoted to developmental disorders. Nevertheless, there are still huge gaps in our knowledge in nearly all relevant fields that need be filled in the future. On the other hand, the available knowledge has not yet been applied to diagnostic procedures, therapeutic interventions, or preventive measures.

Drawing analogies to adults is still a prevailing trait in many fields of clinical and developmental psychopathology and child and adolescent psychiatry. Developmental psychopathology has to have a broad scope, including biological, psychological, and psychosocial dimensions of development in an integrative view. This view should lead us in the future to the identification of broader patterns of functioning in different environmental settings and to the conceptualization of therapeutic strategies that take into account the multiple influences on children and their active role during normal development and developmental crises.

Finally, more sophisticated methods of evaluation have to be developed that take into account all these influences far beyond the narrow symptom-oriented approach.

8 References

Achenbach TM (1974) Developmental psychopathology. Ronald, New York

Achenbach TM (1982) Developmental psychopathology, 2nd edn. Ronald, New York

Achenbach TM (1993) Empirically based taxonomy: how to use syndromes and profile types derived from the CBCL/4–18, TRF, and YSR. Department of Psychiatry, University of Vermont, Burlington

Achenbach TM (1995) Developmental issues in assessment, taxonomy and diagnosis of child and adolescent psychopathology. In: Cicchetti D, Cohen DJ (eds) Developmental psychopathology, vol 1. Wiley, New York, pp 57–80

Achenbach TM, Edelbrock C (1981) Behavioral problems and competencies reported by parents of normal and disturbed children aged four to sixteen. Monogr Soc Res Child Dev 46: 1 (serial no 188)

Achenbach TM, McConaughy SH, Howell CT (1987) Child/adolescent behavioral and emotional problems: implications of cross-informant correlations for situational specificity. Psychol Bull 101: 213–232

Ainsworth M, Blehar M, Waters E, Wall S (1978) Patterns of attachment. Erlbaum, Hillsdale

American Psychiatric Association (1994) Diagnostic and statistical manual of mental disorders, 4th edn. (DSM-IV). American Psychiatric Association, Washington, DC

Anthony EJ (1970) The behavior disorders of childhood. In: Mussen PH (ed) Carmichael's manual of child psychology, vol 1. Wiley, New York

Berrueta-Clement JR, Schweinhart LJ, Barnett WS, Epstein AS, Weikart DP (1984) Changed lives: the effects of the Perry Preschool Program on youths through age 19. High/Scope, Ypsilanti (Monographs of the High/Scope Educational Research Foundation, no 8)

Bird HR, Yager TJ, Staghezza B, Gould MS, Canino G, Rubio-Stipec M (1990) Impairment in the epidemiological measurement of childhood psychopathology in the community. J Am Acad Child Adol Psychiat 29: 796–803

Bohman M (1978) Some genetic aspects of alcoholism and criminality. A population study of adoptees. Arch Gen Psychiat 35: 269–276

Cantwell D (1977) The hyperkinetic syndrome. In: Rutter M, Hersov L (eds) Child psychiatry: modern approaches. Blackwell, Oxford

Cicchetti D (1984) The emergence of developmental psychopathology. Child Development 55: 1–7

Cicchetti D, Cohen DJ (eds) (1995) Developmental psychopathology. Wiley, New York

Conel J (1939) The postnatal development of the human cerebral cortex, vol 1. Harvard University Press, Cambridge, MA

Conel J (1941) The postnatal development of the human cerebral cortex, vol 2. Harvard University Press, Cambridge, MA

Conel J (1947) The postnatal development of the human cerebral cortex, vol 3. Harvard University Press, Cambridge, MA

Conel J (1951) The postnatal development of the human cerebral cortex, vol 4. Harvard University Press, Cambridge, MA
Conel J (1955) The postnatal development of the human cerebral cortex, vol 5. Harvard University Press, Cambridge, MA
Conel J (1959) The postnatal development of the human cerebral cortex, vol 6. Harvard University Press, Cambridge, MA
Costello EJ, Angold A (1995) Developmental epidemiology. In: Cicchetti D, Cohen DJ (eds) Developmental psychopathology, vol 1. Wiley, New York, pp 23–56
Eisenberg L (1977) Development as a unifying concept in psychiatry. Br J Psychiatry 131: 225–237
Emminghaus H (1887) Die psychischen Störungen des Kindesalters. Laupp, Tübingen
Erikson EH (1980) Elements of a psychoanalytic theory of psychosocial development. In: Greenspan SI, Pollock GH (eds) The course of life: psychoanalytic contributions toward understanding personality development. I. Infancy and early childhood. NIMH Mental Health Study Center, Adelphi, pp 11–61
Farrington DP, Loeber R, Van Kammen WB (1990) Long-term criminal outcomes of hyperactivity-impulsivity, attention deficit, and conduct problems in childhood. In: Robbins L, Rutter M (eds) Straight and devious pathways from childhood to adulthood. Cambridge University Press, Cambridge, pp 62–81
Fombonne E (1998) Epidemiological surveys of autism. In: Volkmar F (ed) Autism and pervasive developmental disorders. Cambridge University Press, Cambridge
Freud A (1965) Normality and pathology in childhood. International Universities Press, New York
Freud S (1948) Vorlesungen zur Einführung in die Psychoanalyse (1916/17). Gesammelte Werke, vol 11. Imago, London
Garber J (1984) Classification of childhood psychopathology: a developmental perspective. Child Dev 55: 30–48
Garmezy N (1974) Children at risk: the search for the antecedents of schizophrenia. I. Conceptual models and research methods. Schizophr Bull 8: 14–90
Garmezy N, Masten AS, Tellegen A (1984) The study of stress and competence in children: a building block for developmental psychopathology. Child Dev 55: 97–111
Gillberg C (1991) Outcome in autism and autistic-like conditions. J Am Acad Child Adol Psychiatry 30: 375–382
Gillberg C, Steffenberg S (1987) Outcome and prognostic factors in infantile autism and similar conditions: a population-based study in 46 cases followed through puberty. J Aut Dev Dis 17: 273–287
Gittelman R, Mannuzza S, Shenker R, Bonagura N (1985) Hyperactive boys almost grown up. I. Psychiatric status. Arch Gen Psychiatry 42: 937–947
Gottlieb G (1976) The roles of experience in the development of behavior and the nervous system. In: Gottlieb G (ed) Studies in the development of behavior and the nervous system, vol 3. Academic, New York
Graham P, Skuse D (1992) The developmental perspective in classification. In: Schmidt MH, Remschmidt H (eds) Developmental psychopathology. Hogrefe and Huber, Lewiston
Greenspan S, Lourie RS (1981) Developmental structuralist approach to the classification of adaptive and pathologic personality organizations: infancy and early childhood. Am J Psychiatry 138: 725–735
Harrington R (1993) Depressive disorder in childhood and adolescence. Wiley, Chichester
Harrington R, Fudge H, Rutter M, Pickles A, Hill J (1990) Adult outcome of childhood and adolescent depression. I. Psychiatric status. Arch Gen Psychiatry 47: 465–473
Harrington R, Fudge H, Rutter M, Pickles A, Hill J (1991) Adult outcome of childhood and adolescent depression. II. Links with antisocial disorder. J Am Acad Child Adol Psychiatry 30: 434–439
Havighurst RJ (1972) Developmental tasks and education, 3rd edn. McKay, New York
Herpertz-Dahlmann B, Wewetzer C, Schulz E, Remschmidt H (1996) Course and outcome in adolescent anorexia nervosa. Int J Eating Dis 19: 335–345
Hoagwood K, Jensen PS, Petti T, Burns BJ (1996) Outcomes of mental health care for children and adolescents. I. A comprehensive conceptual model. J Am Acad Child Adol Psychiatry 35: 1055–1063
Homburger A (1926) Vorlesungen über die Psychopathologie des Kindesalters. Springer, Berlin Heidelberg New York
Jaffe ML (1998) Adolescence. Wiley, New York
Jensen PS, Hoagwood K (1997) The book of names: DSM-IV in context. Dev Psychopathol 9: 231–249
Jersild AT, Holmes FB (1935) Children's fears. Teacher's College, Columbia University (Child Development Monograph no 20)
John R, Mednick S, Schulsinger F (1982) Teacher reports as a predictor of schizophrenia and borderline schizophrenia: a bayesian decision analysis. J Abnormal Psychol 91: 399–413
Jones MC (1965) Psychological correlates of somatic development. Child Dev 36: 899–911
Kagan J (1981) The second year: the emergence of self-awareness. Harvard University Press, Cambridge, MA
Kagan J (1984) The nature of the child. Basic, New York
Kanner L (1935) Child psychiatry. Blackwell, Oxford
Katchadourian H (1977) The biology of adolescence. Freeman, San Francisco
Kazdin AE (1986) Conduct disorders in childhood and adolescence. Developmental clinical psychology and psychiatry. Sage, Beverly Hills
Kelly DD (1985) Sexual differentiation of the brain. In: Kandel ER, Schwartz JH (eds) Principles of neural science, 2nd edn. Elsevier, New York, pp 781–782
Klein RG, Mannuzza S (1991) Long-term outcome of hyperactive children: a review. J Am Acad Child Adol Psychiatry 30: 383–387
Klin A, Volkmar FR (1997) Asperger's syndrome. In: Cohen DJ, Volkmar FR (eds) Handbook of autism and pervasive developmental disorders, 2nd edn. Wiley, New York, pp 94–122
Kohlberg L, Lacrosse J, Ricks D (1972) The predictability of adult mental health from childhood behavior. In: Wolman B (ed) Manual of child psychopathology. McGraw Hill, New York
Leckman JF, Cohen DJ (1988) Descriptive and diagnostic classification of tic disorders. In: Cohen DJ, Bruun RD, Leckman JF (eds) Tourette's syndrome and tic disorders: clinical understanding and treatment. Wiley, New York, pp 4–19
Leonard HL, Swedo SE, Lenane LC et al (1993) A two to seven-year follow-up study of 54 obsessive-compulsive children and adolescents. Arch Gen Psychiatry 50: 429–439
Lerner RM (1996) Relative plasticity, integration, temporality, and diversity in human development: a developmental contextual perspective about theory, process and method. Dev Psychol 32: 781–786
Mannuzza S, Klein RG, Bessler A, Malloy P, Lapadula M (1993) Adult outcome of hyperactive boys. Arch Gen Psychiatry 50: 565–576

Masten AS, Garmezy N (1985) Risk, vulnerability and protective factors in developmental psychopathology. In: Lahey BB, Kazdin AE (eds) Advances in clinical child psychology, vol 8. Plenum, New York

Maughan B, Hagell A, Rutter M, Yule W (1994) Poor readers in secondary schools. Reading Writing Interdiscipl J 6: 125–150

McGee R, Williams S, Silva PA (1984a) Background characteristics of aggressive, hyperactive, and aggressive-hyperactive boys. J Am Acad Child Psychiatry 23: 280–284

McGee R, Williams S, Silva PA (1984b) Behavioral and developmental characteristics of aggressive, hyperactive and aggressive-hyperactive boys. J Am Acad Child Psychiatry 23: 270–279

Olweus D (1979) Stability of aggressive reaction patterns in males: a review. Psychol Bull 86: 852–875

Olweus D, Mattsson A, Schalling D, Löw H (1988) Circulating testosterone levels and aggression in adolescent males: a causal analysis. Psychosom Med 50: 261–272

Piaget J (1950) The psychology of intelligence. International Universities, New York

Piaget J (1983) Piaget's theory. In: Mussen PH (ed) Handbook of child psychology. I. History, theory, and methods, 4th edn. Wiley, New York, pp 1–25

Piven J, Palmer P, Jacobi D, Childress D, Arndt S (1997) Broader autism phenotype: evidence from a family history study of multiple-incidence autism families. Am J Psychiatry 154: 185–190

Rapoport JL (1986) Childhood obsessive-compulsive disorder. J Child Psychol Psychiatry 27: 289–295

Remschmidt H (1989) Developmental psychopathology as a theoretical framework for child and adolescent psychiatry. In: Schmidt MH, Remschmidt H (eds) Needs and prospects of child and adolescent psychiatry. Hogrefe and Huber, Toronto, pp 3–24

Remschmidt H (1992) The interaction of biological and psychosocial influences in developmental psychopathology. In: Remschmidt H, Schmidt MH (eds) Developmental psychopathology. Hogrefe and Huber, Lewiston, pp 17–25

Remschmidt H, Schmidt MH (eds) (1992) Developmental psychopathology. Hogrefe and Huber, Lewiston

Remschmidt H, Höhner G, Walter R (1984) Kinderdelinquenz und Frühkriminalität. In: Göppinger H, Vossen R (eds) Humangenetik und Kriminologie. Kinderdelinquenz und Frühkriminalität. Enke, Stuttgart

Remschmidt H, Wienand F, Wewetzer C (1990) The long-term course of anorexia nervosa. In: Remschmidt H, Schmidt MH (eds) Anorexia nervosa. Hogrefe and Huber, Toronto, pp 127–136

Remschmidt H, Schulz E, Martin M, Warnke A, Trott GE (1994) Childhood onset schizophrenia: history of the concept and recent studies. Schizophr Bull 20: 727–745

Remschmidt H, Schulte-Körne G, Hennighausen K (1998) What is specific about specific reading disorder? In: Rispens J, Van Yperen TA, Yule W (eds) Perspectives on the classification of specific developmental disorders. Kluwer, Dordrecht, pp 83–104

Rispens J, Van Yperen TA, Yule W (eds) (1998) Perspectives on the classification of specific developmental disorders. Kluwer, Dordrecht

Robins L (1978) Sturdy childhood predictors of adult antisocial behavior: replications from longitudinal studies. Psychol Med 8: 611–622

Robins L, Rutter M (eds) (1990) Straight and devious pathways from childhood to adulthood. Cambridge University Press, Cambridge

Rolf J, Read PB (1984) Programs advancing developmental psychopathology. Child Dev 55: 8–16

Rubinow DR, Schmidt PJ (1996) Androgenes, brain, and behavior. Am J Psychiatry 153: 974–984

Rutter M (1980) Developmental psychiatry. Heinemann, London

Rutter M (1984) Psychopathology and development. I. Childhood antecedents of adult psychiatric disorders. Aust NZ J Psychiatry 18: 225–234

Rutter M (1985) Resilience in the face of adversity. Protective factors and resistance to psychiatric disorder. Br J Psychiatry 147: 598–611

Rutter M (1991) Childhood experiences and adult psychosocial functioning. In: Bock GR, Whelan J (eds) The childhood environment and adult disease. Ciba Found Symp 156: 189–208

Rutter M (1995) Relationships between mental disorders in childhood and adulthood. Acta Psychiatr Scand 91: 73–85

Rutter M (1996) Transitions and turning points in developmental psychopathology: as applied to the age span between childhood and mid-adulthood. Int J Behav Dev 19: 603–626

Rutter M, Garmezy N (1983) Developmental psychopathology. In: Heatherington EM (ed) Socialization, personality, and social development, vol 4, 4th edn. Wiley, New York, pp 775–911

Rutter M, Tizard J, Whitmore K (1970) Education, health and behaviour. Longmans, London

Shapiro AK, Shapiro ES, Bruun RD, Sweet TRD (1978) Gilles de la Tourette's Syndrome. Raven, New York

Smalley SL (1991) Genetic influences in autism. Psychiatr Clin North Am 14: 125–139

Sroufe LA (1983) Infant-caregiver attachment and patterns of adaptation in pre-school: the roots of maladaptation and competence. In: Perlmutter M (ed) Minnesota symposia in child psychology, vol 16. Erlbaum, Hillsdale

Sroufe LA, Rutter M (1984) The domain of developmental psychopathology. Child Dev 55: 17–29

Steinhausen HCH (1992) Sex differences in developmental psychopathology. In: Remschmidt H, Schmidt MH (eds) Developmental psychopathology. Hogrefe and Huber, Lewiston

Taylor E (1989) Externalizing disorders: priorities for future research. In: Schmidt MH, Remschmidt H (eds) Needs and prospects of child and adolescent psychiatry. Hogrefe and Huber, Toronto

Tramer M (1941) Lehrbuch der allgemeinen Kinderpsychiatrie. Schwabe, Basel

van Acker R (1997) Rett's syndrome: a pervasive developmental disorder. In: Cohen DJ, Volkmar FR (eds) Handbook of autism and pervasive developmental disorders, 2nd edn. Wiley, New York, pp 60–93

Verhulst FC, Koot HM (1992) The stability of externalizing behavior in an epidemiological sample. In: Schmidt MH, Remschmidt H (eds) Developmental psychopathology. Hogrefe and Huber, Lewiston

von Prittwitz S, Blum WF, Ziegler A, Scharmann S, Remschmidt H, Hebebrand J (1997) Restrained eating is associated with low leptin levels in underweight females. Mol Psychiatry 2: 420–422

Waber DP (1976) Sex differences in cognition: a function of maturational rate? Science 192: 572–574

Waber DP (1977) Sex differences in mental abilities, hemispheric lateralization and in the rate of physical growth at adolescence. Dev Psychol 13: 29–38

Wachs TD (1996) Known and potential processes underlying developmental trajectories in childhood and adolescence. Dev Psychopathol 32: 796–801

Weissman MM, Warner V, Fendrich M (1990) Applying impairment criteria to children's psychiatric diagnosis. J Am Acad Child Adol Psychiatry 29: 789–795

World Health Organization (1975) International classification of diseases, 9th revision. World Health Organization, Geneva

World Health Organization (1992) The ICD-10 classification of mental and behavioural disorders, clinical descriptions and diagnostic guidelines. World Health Organization, Geneva

Zahn-Wexler C (1996) Environment, biology, and culture. Implications for adolescent development. Dev Psychopathol 32: 571–573

CHAPTER

3

H. Remschmidt, M.H. Schmidt

Child and Adolescent Psychiatry and Psychotherapy as a Clinical and Scientific Discipline: An Introduction

Translator: S. Henn

1 Development of the Field as an Independent Medical Discipline

According to the currently valid guidelines for specialty training issued in 1984 by the German Federal Chamber of Physicians (Bundesärztekammer), the specialty of child and adolescent psychiatry is defined as follows:

"Child and adolescent psychiatry and psychotherapy comprises the diagnosis, non-surgical treatment, prevention, and rehabilitation of mental, psychosomatic, developmental, and neurological illnesses and disorders, as well as abnormalities of behavior and conduct in children and adolescents."

This definition is generally valid in most European and other countries, although there are variations specific to the different countries.

Table 1 provides a survey of the current status of child and adolescent psychiatry in 31 European countries. The table shows, among other things, that the greatest concentration of specialists (i.e. the largest number of physicians practicing child and adolescent psychiatry per population under 20 years of age) is in Switzerland, followed by Finland, France, Sweden, Iceland, Norway, and Italy. At the same time, the table also reveals variations in population structure (variable percentage of the population under 20 years of age) and gives the year of founding of the specialty society in each country, as well as the number of university and other child and adolescent psychiatric clinics and departments. Further details of the current status of child and adolescent psychiatry in Europe can be found in Remschmidt and van Engeland (1999).

In historical terms, child and adolescent psychiatry has two roots: adult psychiatry and neurology on the one hand, and pediatrics on the other. Important contributions also come from psychology, from various branches of the social sciences and jurisprudence, and from the practical experience of youth and social welfare services.

Child and adolescent psychiatry and psychotherapy have derived important contributions from neurology with respect to the development of the central nervous system, the concept of vulnerability, and considerations of differential diagnosis. "Vulnerability" refers to the heightened susceptibility of children and adolescents with pre-existing brain damage to harmful environmental influences. Special considerations of differential diagnosis arise because many psychogenic illnesses present with somatic manifestations.

Child and adolescent psychiatry, like adult psychiatry, shares objectives and points of contact with psychopathology, which has acquired great importance today in the form of developmental psychopathology (Remschmidt 1989; Remschmidt and Fombonne 1999; see also Chap. 2, Part 1 of this volume). Child and adolescent psychiatry also converges with adult psychiatry in family psychiatry (i.e. the care of mental illness both in children and in adults in a family context), and in the care of adolescent patients, many of whom go on to be treated by adult psychiatrists when they reach adulthood (see Remschmidt and Mattejat 1994). Of special importance for child and adolescent psychiatry are the children of mentally ill parents, especially psychotic parents.

Finally, child and adolescent psychiatry converges with pediatrics in the field of infant psychiatry, which has become very important in recent years; in the care of chronically physically ill and handicapped children; and in the field of psychosomatic disease. In recent years, the collaboration of pediatricians with child and adolescent psychiatrists has led to considerable progress in all three of these areas, with respect to the scientific understanding and treatment of various illnesses.

Also of great significance for child and adolescent psychiatry and psychotherapy is clinical psychology, especially neuropsychology, which may be described as an extension of neurology with psychological techniques. Recently, it has been shown that many disorders can be better understood with the aid of neuropsychological examination techniques and better treated with neuropsychologically based therapies.

The specialty of child and adolescent psychiatry has emerged from the confluence of all of these specialties and has integrated and structured the knowledge gained from each to meet the specific needs of mentally ill children and adolescents and their families. A separate medical specialty has thus arisen, and the American child psychiatrist Leo Kanner, who was the first to describe early childhood autism, has traced its development up to the present as follows:

The first decade of our discipline was characterized by "thinking about children", the second by "doing things to children", the third by "doing things for children" and the fourth decade by "working with children" (Kanner 1957, p. 15).

This historical overview reveals both a gradual process of emancipation not only for the child, but also for the specialty as a whole. Mentally ill and handicapped children now have a right to specialty treatment.

2 Clinical Aspects and Treatment

The majority of studies agree that the incidence of mental disorders and illnesses in children and

Table 1. Current status of child and adolescent psychiatry in 31 European countries. (According to Remschmidt and Van Engeland 1999)

Country	Type of society	Year established	University departments (*n*)	Other child and adolescent psychiatric clinics (*n*)	Child and adolescent psychiatrists (*n*)	Population (in 1000)[c]	Population under 20 years of age (in 1000) (*n*)	(%)	Population under 20 years of age per child and adolescent psychiatrist
Austria	CANP	1974	1	8	65	8134	1859	23	28,600
Belgium	CAP	1961/1976	4	19	300	10,175	2385	23	8000
Bulgaria	CAP	1993	3 (0 chairs)	9	46	8240	1939	23.5	42,000
Croatia	CAP	1990	3	3	35	4672	1125	24	32,000
Czech Republic	CAP[a]	1960	4 (0 chairs)	13	116	10,286	2524	24.5	21,700
Denmark	CAP[a]	1953	4 (1 chair)	12	141	5334	1263	23.6	9000
Estonia	CAP[a]	1973	1	2	20	1421	0.372	26	18,600
Finland	CP	1956	5	19 child, 15 adolescent	196	5149	1286	25	6600
France	CP	1937	33	120	2000	58,805	15,010	25.5	7500
Germany	CAPP	1940	26	145	781	82,079	17,323	21	22,000
Great Britain	CAP	1971	16 (18 chairs)	60	547	58,970	15,036	25.5	27,500
Greece	CP	1983	3	22[d]	160	10,662	2477	23	15,500
Hungary	CAP	1990	1	7	55	10,208	2499	24.5	45,000
Iceland	CP	1980	1 (0 chairs)	0	10	0.271	0.085	31	8500
Ireland	CAP	1983	2	12	36	3619	1135	31	31,500
Italy	CNP	1959	24	15	1200	56,783	11,297	20	9400
Latvia	CAP[a]	1950	1	3	26	2385	614	25.7	23,600

Table 1 *(Continued)*

Country	Type of society	Year established	University departments (*n*)	Other child and adolescent psychiatric clinics (*n*)	Child and adolescent psychiatrists (*n*)	Population (in 1000)[c]	Population under 20 years of age (in 1000) (*n*)	(%)	Population under 20 years of age per child and adolescent psychiatrist
Lithuania	CAP	1996	2	3	60	3600	0.998	27.7	16,600
Luxembourg	–	–	0	1	4	0.425	0.101	23.7	25,000
Netherlands	CAP	1948	7	19	257	15,731	3800	24	14,800
Norway	CAP	1957	4	60	130	4420	1130	25.5	8700
Portugal	CAP	1989	0	3	99	9928	2412	24	24,300
Rumania	CANP	1992	3	10	200	22,396	6023	26.9	30,000
Russia	CAPP	1992	0	99	1300	146,861	40,326	27.5	31,000
Serbia	CAP, DANP	1979	4	14 outpatient	57	10,526	2957	28	51,800
Slovakia	CAP	1971	3 (0 chairs)	5	113	5393	1575	29	14,000
Slovenia	CP	1979	2	12 outpatient	24	1972	0.475	24	19,800
Spain	CAP	1978	1	17	200	39,134	8739	22	43,000
Sweden	CAP	1956	6 (4 chairs)	24	282	8887	2166	24	7700
Switzerland	CAPP	1957	5	11	315	7260	1662	23	5300
Ukraine	CAP[a]	1995	1 (2 chairs)[b]	40	438	50,125	13,153	26	3000

CANP Child and Adolescent Neuropsychiatry; *CAP* Child and Adolescent Psychiatry; *CAPP* Child and Adolescent Psychiatry and Psychotherapy; *CNP* Child Neuropsychiatry; *CP* Child Psychiatry; *DANP* Developmental Age Neurology and Psychiatry

[a]Section.

[b]Research institutes outside of universities.

[c]Source: International Programs Center (IPC), U.S. Census Bureau, estimates for 1998.

[d]Including 15 child guidance positions.

adolescents taken together lies between 8% and 15%, and at least 5% of children and adolescents up to age 18 have been reported to require treatment. None of the methodologically sophisticated studies available to date have revealed a lower rate (Remschmidt and Walter 1989, 1990). Various studies also agree, however, that only a small percentage of the children and adolescents in need of treatment are actually treated (Remschmidt and Walter 1990).

With this problem in mind, the German Federal Government's Expert Commission for the Model Program in Psychiatry (Expertenkommission der Bundesregierung 1988) has directed that the following guiding principles for care should be put into practice:

1. Equal footing with other patient groups
2. Integration into medicine
3. Community approach
4. Appropriateness of care

The placement of mentally ill children and adolescents on an equal legal footing with those who are physically ill requires the integration of child and adolescent psychiatry into medicine as both a a clinical and a scientific discipline.

The institutions responsible for the care of mentally disturbed and ill children and adolescents are described in Chap. 6 (Part 1 of this volume).

3 Research and Scientific Societies

Research in child and adolescent psychiatry makes use of the methods of adult psychiatry, neurology, and other disciplines and has also recently developed a certain degree of methodological independence. This is especially true of methods used to assess the dependence of disorders on development, the role of the family, and special features of the forms of therapy that are practiced.

In research, there is common ground with the entire field of developmental medicine and psychology, with various disciplines in the neurosciences, and to a degree with pediatrics, education science, special education, and criminology.

Worldwide trends in research in child and adolescent psychiatry are characterized by three developmental tendencies:

Increasing emphasis on the biological basis of mental illness at all ages (the key words here are molecular biology, genetics, and imaging techniques)

Greater attention to all aspects of development (key words: developmental psychology, developmental psychopathology, developmental neurology, developmental biochemistry, and developmental pharmacology)

The establishment of rationally based, effective, and efficient methods of diagnosis and treatment for use in clinics and practices (key words: standardized diagnosis, development of guidelines, quality assurance, evaluation, and consideration of quality of life)

With regard to therapy, it is internationally accepted that there is no single theory or treatment approach that can be used to characterize or to treat all, or even most, mental and behavioural disorders; rather, disorder-specific approaches to treatment (treatment programs) must be developed which contain multiple components and are specific to the disorder at hand while still being adaptable to the individual features of each case.

We have made the following proposals concerning the further development of research in child and adolescent psychiatry (Schmidt and Remschmidt 1989; Remschmidt 1998):

1. Basic research on the etiology and pathogenesis of mental disorders in children and adolescents: This area includes studies using genetic and molecular biological methods as well as the new imaging techniques. Studies of this type have been performed to date in a number of countries for early childhood autism, Gilles de la Tourette syndrome, eating disorders, dyslexia, enuresis, and a number of other illnesses.
2. Multilevel research on the stability and change of psychiatric illness in children and adolescents: Because the stability and change of disease can only be understood in a network of individual, biological, and social pathogenetic factors and processes and their boundary conditions, they must necessarily be studied with a multilevel approach.
3. Research on externalized disorders (with expansive, aggressive, and disruptive manifestations): This question is particularly important, not least because of the high stability of these disorders and the enormous problems they create for the family, in school, and in society at large.
4. Research on developmental disorders: These are not only frequent, they are also a major risk factor for the development of secondary psychiatric disorders. In the current classification systems (ICD-10, DSM-IV), they are grouped into two categories:
   - Pervasive developmental disorders, which also include early childhood autism, atypical autism, Rett syndrome, and Asperger syndrome, and
   - Specific developmental disorders, which include the disturbed development of speech and language, scholastic skills, motor functions, and combinations of the above.

Important knowledge has been acquired in recent years for both groups of disorders.

5. Research on interventions and care: It is very important for the methods used to treat children, adolescents, and their families to be tested for effectiveness, efficiency, and possible side effects. In this context, effectiveness is understood as the ability to induce remission, while efficiency means a favorable cost–benefit ratio.

Research and development also partly depend on organizational alliances and must always be viewed from an international perspective. The professional societies play an important role here. The three international professional societies for child and adolescent psychiatry and the section of child and adolescent psychiatry of the World Psychiatric Association (WPA) have different emphases: the International Association for Child and Adolescent Psychiatry and Allied Professions (IACAPAP) and the child and adolescent section of the WPA address the whole spectrum of child and adolescent psychiatry, while the International Society for Adolescent Psychiatry (ISAP) focuses on adolescence and the World Association for Infant Mental Health (WAIMH) on the psychiatry of infants.

In addition to these international professional societies, there are other societies covering large regions of the world, such as the European Society for Child and Adolescent Psychiatry (ESCAP), the Asian Society for Child and Adolescent Psychiatry and Allied Disciplines (ASCAPAP), the Federation of Latin American Societies for Child and Adolescent Psychiatry (FLAPIA), the American Academy of Child and Adolescent Psychiatry, which cooperates closely with the Canadian society, and the Australian and New Zealand College of Psychiatrists, which has a section of child and adolescent psychiatry. Many of these societies are umbrella organizations for the national professional societies (e.g. ESCAP, IACAPAP, and ASCAPAP).

4 Undergraduate Medical Education, Specialty Training, and Continuing Medical Education

Child and adolescent psychiatry in Germany is incorporated into undergraduate medical education either in the form of lectures and courses specifically devoted to the subject or as a part of pediatrics and psychiatry. This is not yet the case in all other countries in and outside Europe.

Specialty training takes 5 years in Germany, of which 3 must be in child and adolescent psychiatry and 1 in either psychiatry or pediatrics. There are considerable differences between the European countries and the rest of the world in this regard. In 1993, a child and adolescent psychiatry and psychotherapy section was set up within the European specialty association, the Union Européenne des Médecins Spécialistes (UEMS). This section is dedicated, among other things, to standardizing specialty training in child and adolescent psychiatry and psychotherapy in Europe and has already developed a European specialty curriculum.

Continuing medical education (CME), which has been introduced as a requirement in a number of European countries, is also very important.

5 Ethical Problems

Besides general ethical questions that concern all clinical disciplines, there are a few problems that are quite specific to child and adolescent psychiatry and psychotherapy. One fundamental problem is that all important decisions affecting minor children are made by their parents. Although one can but agree with this arrangement in principle, there will always be cases in which parents do not act in the best interest of their children, especially when questions of custody, of physical, mental, or sexual abuse, or of consent to medical treatment are at issue. Child and adolescent psychiatrists are usually consulted in these cases and asked to apply their expertise toward the determination of a legal decision.

There is no universal rule about whether children and adolescents should be consulted about decisions concerning them; on a case-by-case basis, the child's wishes should be taken into account depending on his or her age. In this connection, the vague legal concept of the child's well-being is taken as a starting point, although it, too, has no universally valid definition, and must be specified individually (Remschmidt and Mattejat 1996).

Important ethical questions in everyday clinical practice relate to the treatment of children and adolescents, to their wishes and those of their parents, to commitment to closed wards in appropriate institutions, and to the use of biological and psychotherapeutic treatment methods. In addition, the use of ineffective or uneconomical treatment methods must be considered not only with regard to quality assurance, but also from an ethical viewpoint.

In the scientific area, controlled clinical studies in children and adolescents present a special problem, both legally and ethically. According to Helmchen and Müller-Oerlinghausen (1978), there are three kinds of clinical studies: (1) therapeutic experiments, (2) therapeutic trials, and (3) studies of established therapies.

The last category causes no problems in child and adolescent psychiatry, but therapeutic experiments, which are designed to increase knowledge about therapeutic interventions but not to help individual patients, create a genuine ethical dilemma, namely that of having to include children who are unable to consent to research that might not benefit them. Doing so may be necessary if we are to avoid the ethically questionable use of therapies that have been tested only in adults and therefore may be ineffective, or even harmful, in children (see also Chap. 18, Vol. 1, Part 2).

6 Outlook

The future of child and adolescent psychiatry cannot be predicted independently of social developments, and prognostication is therefore difficult. Regardless of social trends, however, the following three factors will be very important to the future development of child and adolescent psychiatry (Remschmidt 1988):

1. Lasting support for research: Only through research and development will it be possible to assess patients appropriately and treat them effectively. To achieve these ends, efforts are needed in several areas, ranging from basic and advanced training in methodology for young scientists, to interdisciplinary projects and grant programs, to regional and national research associations. It should be borne in mind that research into mental disorders in children and adolescents is not only a humanitarian task, but also has great economic importance. The development of preventive measures and the establishment of etiological treatment methods would certainly contribute to a major reduction of costs in the health care sector.
2. Improvement of undergraduate medical education, specialty training, and continuing medical education in view of the latest scientific findings: Efforts in this direction begin in the medical schools and must then be integrated into specialty training and then, finally, flow into CME.
3. Further development of interdisciplinary systems of care and establishment of cooperative links to the allied disciplines (pediatrics, psychiatry, education) and to youth services: Of particular importance with respect to cooperation with adult psychiatry is the performance of longitudinal studies, which are best suited to clarifying the developmental psychopathological aspect of mental illnesses and to bridging the gap between adult psychiatry and child and adolescent psychiatry. There is still much to do in this area, although the systems of care are already quite well developed in a number of European countries.

If the constructive work already in progress in these areas can be successfully continued, child and adolescent psychiatry will be able to overcome future challenges.

7 References

Expertenkommission der Bundesregierung (1988) Empfehlungen zur Reform der Versorgung im psychiatrischen und psychotherapeutisch/psychosomatischen Bereich. Auf der Grundlage des Modellprogramms Psychiatrie der Bundesregierung. BMJFFG, Bonn

Helmchen H, Müller-Oerlinghausen B (eds) (1978) Psychiatrische Therapieforschung. Ethische und juristische Probleme. Springer, Berlin Heidelberg New York

Kanner L (1957) Child psychiatry, 3rd edn. Blackwell, Oxford

Remschmidt H (1989) Developmental psychopathology as theoretical framework for child and adolescent psychiatry. In: Schmidt MH, Remschmidt H (eds) Needs and prospects in child and adolescent psychiatry. Hogrefe and Huber, Toronto, pp 3–24

Remschmidt H (1998) Tradition und Entwicklung in der Kinder- und Jugendpsychiatrie. Zeitschrift für Kinder- und Jugendpsychiatrie 26: 34–42

Remschmidt H, Fombonne E (1999) Entwicklungspsychopathologie. Grundlagenwissenschaft für die Kinder- und Jugendpsychiatrie und Psychiatrie. Nervenarzt 70: 577–586

Remschmidt H, Mattejat F (1994) Kinder psychotischer Eltern. Mit einer Anleitung zur Beratung von Eltern mit einer psychotischen Erkrankung. Hogrefe, Göttingen

Remschmidt H, Mattejat F (1996) Die Beiträge der kinder- und jugendpsychiatrischen und entwicklungspsychologischen Forschung zur "Objektivierung" des Kindeswohlbegriffes. Prax Kinderpsychol Kinderpsychiatr 45: 266–273

**Remschmidt H, Van Engeland H (eds) (1999) Child and adolescent psychiatry in Europe. Steinkopff, Darmstadt

*Remschmidt H, Walter R (1989) Evaluation kinder- und jugendpsychiatrischer Versorgung. Analysen und Erhebungen in drei hessischen Landkreisen. Enke, Stuttgart

*Remschmidt H, Walter R (1990) Psychische Auffälligkeiten bei Schulkindern. Hogrefe, Göttingen

Schmidt M, Remschmidt H (1989) Forschung in der Kinder- und Jugendpsychiatrie. Kinder- und Jugendpsychiatrische Klinik am Zentralinstitut für seelische Gesundheit, Mannheim/Klinik für Kinder- und Jugendpsychiatrie der Philipps-Universität, Marburg

CHAPTER

4

H. Remschmidt, M.H. Schmidt

Diagnosis and Classification

Translator: S. Henn

1 Diagnosis

It is the goal of diagnostic classification in child and adolescent psychiatry to organize information about the problems of a child or adolescent, and his or her family, so that their future course can be predicted and reasonable interventions can be chosen. The underlying conception is that the current state and past history of a disorder may be used to predict its further development.

Both the affected individuals and their families are significant sources of information. Children or adolescents can give information about their own condition and, with increasing age, can inform better than the adults around them about their behavior, their relationship to the environment, the appearance of manifestations in the intermediate term, and their subjective assessment of the situation. Adults who have a significant relationship to the patient are important sources of information about the reason for consulting a child and adolescent psychiatrist at this point, about the patient's behavior in a variety of circumstances, and about the history of the disturbance over the long term, from its first appearance to the present, acute situation.

By gathering information from both sources, the clinician not only becomes aware of differing points of view, but can also recognize the dynamics of relationships within the family system. Interviewing the parents and the child separately gives the parents the chance to tell otherwise unknown facts about their relationship with each other, and to express their own concerns. On the other hand, the child can express his or her perception of the environment, and can discuss aspects of his or her own behavior that the parents do not know about. After the separate interviews have been conducted, a joint discussion serves to establish a definition of the problem, from which the goals of the patient and family are then derived. The alternative practice of conducting only a joint interview, without speaking with the child and parents separately, is forensically unjustifiable. Games, drawings, and dialogue exercises involving the parents and the child, as well as the technique of circular questioning, are useful aids to the diagnostic process.

An interview of the child alone is indispensable. Because children may have difficulty communicating their symptoms or behaviors unless specifically asked about them, the clinician should ask about at least the following list of behavioral areas:

- physical symptoms and sleep disturbances
- separation anxiety and obsessive-compulsive phenomena
- contact disturbances and social anxiety
- relationship disturbances and sexual contacts
- school discipline and attendance
- temper tantrums and aggressive behavior
- disturbances of impulse control and antisocial manifestations
- use of illegal drugs, alcohol, and medications

Protecting the privacy of the child or adolescent is just as important as protecting the parents' privacy, as long as no overriding interests (e.g. risk of progression of the disorder) dictate otherwise. Nevertheless, the child or adolescent must be made to understand his or her obligation to pass on certain information. The ideal of total impartiality is particularly inappropriate in situations in which the child would thereby suffer, be harmed, or be put in danger.

The interview usually provides information about the following:

- the reason for the consultation
- manifestations of the disorder and their history
- previous interventions
- the young person's developmental history
- the family history
- the young person's behavior and affect
- relationships within the (entire or immediate) family
- the young person's present level of functioning, i.e. not only deficits, but also competencies

Information thus gained can be supplemented by the following:

- medical and neurological examination of the child
- neurophysiological studies (when cerebral developmental delay or neurological illness is suspected)
- imaging studies (when malformations, sequelae of trauma, or degenerative illnesses are suspected)
- laboratory studies (when physical manifestations are present, or for the purpose of exclusion in differential diagnosis)
- tests of cognitive performance and neuropsychological testing (for the assessment of scholastic abilities, and when specific developmental delays are suspected)
- diagnostic assessment of the family (if disturbed interactions within the family are suspected)
- projective diagnostic assessment (if perceptual styles, attitudes, prevailing themes and conflicts cannot otherwise be assessed)

The main problem in diagnostic assessment is that of economy, i.e. of collecting only the necessary information. This is evident both in the Practice

Parameters of the American Academy of Child and Adolescent Psychiatry (AACAP 1997, 1998) and in its German equivalent, the Guidelines for the Diagnosis and Treatment of Mental Disorders in Children and Adolescents (Deutsche Gesellschaft für Kinder- und Jugendpsychiatrie und Psychotherapie 2000). The most important source of error is not the failure to perform additional tests, but incompleteness of the information obtained by interview. Complete information is indispensable for accurate diagnosis and must not be compromised by narrowing the focus too early onto a specific suspected diagnosis.

The use of standardized interviews may prevent this error, but they are too time-consuming in practice and are suitable only for diagnostic assessment in the research setting. The available standardized interviews include the Diagnostic Interview Schedule for Children (DISC), the Composite International Diagnostic Interview (CIDI), the Structural Clinical Interview, DSM-IV Version (SCID), and the Schedule for Clinical Assessment in Neuropsychiatry (SCAN) (WHO 1994, 1997).

The German interviews that have been developed for epidemiological studies do not cover all syndromes. These include the Mannheimer Elterninterview ("Mannheim Parents' Interview," MEI; Esser et al. 1989), the Diagnostik-System für Psychische Störungen im Kindes- und Jugendalter nach ICD-10 und DSM-IV ("Diagnostic System for Mental Disorders in Childhood and Adolesence According to ICD-10 and DSM-IV"; Döpfner and Lehmkuhl 1998), and the Diagnostisches Interview bei Psychischen Störungen für Kinder ("Diagnostic Interview for Mental Disorders in Children," Kinder-DIPS; Unnewehr et al. 1995). Döpfner and Lehmkuhl attempt to combine a category-based classification with a dimensional description of the disorders under study, while Unnewehr's interview focuses on integrating clinically relevant information in treatment plans.

Questionnaires such as the Child Behavior Checklist of Achenbach and Edelbrock (1983) do not meet this need because they are based on the use of a large number of equally weighted questions which, when correctly answered without follow-up questioning, can yield information about the child's general symptoms but are inadequate for the provision of a specific diagnosis. Questionnaires have the additional disadvantage that the answers of mentally ill and normal persons may be systematically biased in opposite directions, so that unaffected persons' scores may overestimate the degree of disturbance in a particular dimension, while affected persons' scores may underestimate it (see also the comments on dimensional vs. category-based assessment in Sect. 2).

Once the assessment is complete, information about the following areas is usually available:

- current manifestations and their history
- the problems affecting the child and family, and the reason for the consultation
- competencies, coping behavior and defense mechanisms
- relevant earlier physical and mental illness of the patient and family members
- developmental level attained before onset of the disorder
- quality of relationships
- current or chronic stresses in present life circumstances
- factors maintaining the disorder

Such information is generally used in the final provision of a diagnosis according to a multiaxial diagnostic system (see Sect. 2). In this process, all abnormalities that might be the principal manifestations of a mental disorder must be taken into account, and comorbid disturbances must be listed in order of importance. Individual manifestations should be evaluated according to their significance for developmental psychopathology.

Once one or more diagnoses have been made in this way, a decision follows regarding the need for intervention. This decision is based not only on the nature of the disorders themselves, but also on the wishes of the family. Ethical problems arise when there are fundamental disagreements, e.g. when the parents do not think a serious disorder needs treatment. If a disorder cannot be treated, the parents should be so informed. The most important factor in the decision regarding treatment is the current level of functioning, i.e. the child's ability to perform age-typical developmental tasks.

The hierarchy of necessary interventions is then established: for example, syndromes that fundamentally inhibit development, such as attention disorders, require specific treatment before further interventions demanding the child's cooperation can be initiated. Once the goal of treatment has been agreed upon, information is provided about possible side effects (not only of pharmacotherapy), and the treatment setting is decided upon. A consensus among all involved individuals regarding the goals of therapy is required, except in emergency situations such as suicidal behavior, excited states, intoxications, dangerous conditions in eating disorders, intensification of separation anxiety to the point of a crisis, or physical or sexual abuse.

The next step is further diagnostic assessment with relevance to treatment, i.e. the establishment of a detailed treatment plan specific to the particular disorder (see Chap. 6, Part 1 of this volume).

2 Classification

The goal of classification is the definition of disorders and problem situations in children to facilitate both clinical treatment and scientific exchange. The wide variety of clinical presentations is thereby reduced to a manageable number of symptom complexes. The criteria for the development of such symptom complexes include the symptom profile, the history of the present illness (not necessarily connected to its etiology), its expected course, and its responsiveness to treatment. Developmental psychopathology and research on disease course and therapy incorporating these criteria are a more important element of modern classification systems than are concepts of the etiology and pathogenesis of mental disorders.

It is therefore of prime importance that classification should be oriented toward developmental aspects. It must be reliable, objective, and not overly time-consuming. Logical consistency and validity – above all, predictive validity – are essential. Reliability is not just a matter of providing a correct description, but also of the proper application of prescribed criteria; the latter is achieved not through the accumulation of professional experience but by quality-assurance measures such as consensus conferences, quality management, and the like. Objectivity, too, is enhanced by operationalized criteria. It is diminished by the use of polythetic categories, i.e. the provision of lists of features only some of which must be present, with the result that the same diagnosis may apply to different constellations of symptoms. The problem of economical use of time in classification arises because information that is easy to obtain may not be the information that is essential for predictive validity.

Logical consistency calls for the mutual exclusiveness of diagnostic classes, which should therefore ideally have a comparable degree of differentiation. With reference to predetermined standard diagnoses, Blanz and Schmidt (1994) found a degree of agreement of 50% for three- and four-digit diagnostic codes in child psychiatry, and 58% when alternative diagnoses were also considered. The corresponding figures for three-digit codes alone were 61% and 68%, respectively. With respect to the diagnosis chosen by the largest number of raters (regardless of accuracy), the rate of agreement for four-digit codes rose to 60%, and, for three-digit codes, to 71%. Nearly perfect agreement was achieved with monosymptomatic disorders, and the least agreement was found with those syndromes that are inadequately operationalized for children and adolescents, e.g. somatization disorders or depressive syndromes. This demonstrates the effect of the following on the quality of classification: the level of information in assessment, the current state of nosological research, and the logical consistency of assessment systems. There has been less investigation of the degree to which the assessors' conceptions of developmental pathology affect decisions in classification.

Disorders predominantly requiring dimensional assessment cause greater difficulties in classification (see below). The validity of the construct is more frequently referred to in these cases. However, construct validity may be less relevant to clinical decisions regarding the course of the disorder and its responsiveness to treatment. The importance of classification criteria that provide predictive validity is obvious. In this connection, the reader is referred to the results obtained by Moffitt (1993) on the onset of antisocial symptoms, by Henn et al. (1980) on the role of social ties for the course of antisocial disorders, and by Mandel (1997) on the significance of comorbid antisocial behavior with hyperkinetic disorders; for a summary of work about the course of shyness and social hypersensitivity with regard to its later significance for social phobia, see Beidel and Turner (1998).

Multidimensionality of Classification Systems

The determination of the pathological clinical manifestations, personality traits, and environmental circumstances that are ultimately relevant to therapy in child and adolescent psychiatry has led to the development of multidimensional classification systems. Such systems take also account of the random simultaneous occurrence of features that may not be causally related, and thereby enable statistical correlation between them. Multiaxial classification has significantly influenced research in child and adolescent psychiatry since the late 1970s. Rutter et al. first proposed such systems in 1969.

The two multiaxial systems currently in use are ICD-10 and DSM-IV. The adapted version of the CD-10 (WHO 1996) prescribes six axes:

Axis I: clinical psychiatric syndromes
Axis II: specific disorders of psychological development
Axis III: intellectual level
Axis IV: medical conditions
Axis V: associated abnormal psychosocial situations
Axis VI: global assessment of psychosocial disability

Axis I corresponds to Chap. F, Sect. V of ICD-10, except for codes 80–83, which correspond to Axis II, and Sect. F7, which corresponds to Axis III. Axis IV corresponds to other chapters of the ICD-10. In each case, only current diagnoses are taken into account. Axis V codifies current abnormal psychosocial situations, i.e. those present in the past six months, and does not consider earlier problems. The global assessment of psychosocial disability made on Axis VI refers

to the patients' psychosocial functioning at the time of clinical examination. It should include an assessment of their behavior within the family, with peers, at school, and in their spare time, over the course of the last three months before the examination (for chronic or subacute disorders) or the last few days or weeks (for acute disorders).

The multiaxial classification of adult mental disorders (Janca et al. 1997), which independently surveys clinical diagnoses, impairments, and contextual factors, has a narrower basis. The predictive validity of this method still awaits empirical confirmation in the field. In addition to clinical diagnostic guidelines, ICD-10 has separate research criteria for the disorders described in Chap. F. These are more narrowly operationalized than the clinical guidelines and are intended to increase the homogeneity of the disorders dealt with in the scientific studies. This goal will certainly be achieved, but further attention will have to be paid to the applicability of the results obtained in such studies to everyday clinical practice.

The Diagnostic and Statistical Manual of the American Psychiatric Association (DSM-IV) provides a system of five axes for children and adults:

Axis I: clinical disorders and other conditions that may be a focus of clinical attention
Axis II: personality disorders and mental retardation
Axis III: general medical conditions
Axis IV: psychosocial and environmental problems
Axis V: global assessment of functioning

The Axis I codes thus also include additional conditions that cannot be attributed to any individual mental illness but nevertheless occasion a need for observation or treatment. Some of these conditions are classified on Axis IV or V of the ICD-10. Axis II of DSM-IV is for personality disorders and mental retardation. While the multiaxial diagnostic system according to ICD-10 (MAS/ICD-10) places personality disorders on Axis I, DSM-IV also allows the classification on Axis II of specific personality traits or defense mechanisms that do not cross the threshold of personality disorders. Hierarchically ordered multiple diagnoses are allowed on Axes I and II. Axis III codes correspond to the Axis IV codes of MAS/ICD-10.

The degree of severity of psychosocial stress factors in the year prior to examination, as coded on Axis IV of DSM-IV, relates only to pathogenetically relevant circumstances and to the predominance of acute events or ongoing life circumstances. The degree of severity is numerically coded in Axis IV of DSM-IV, while Axis V of MAS/ICD-10 assigns abnormal psychosocial situations to nine major categories and 31 subcategories. Axis V assesses the psychosocial functional level – above all, psychosocial and professional capability, both at the time of assessment and before the onset of illness – on a quantitative scale analogous to Axis VI of MAS/ICD-10.

3 Current Problems

Earlier case definitions permitted the use of symptom formation or performance limitation as criteria for the diagnosis of mental disorders. Yet even severe disorders, such as a severe transient tic disorder, need not permanently impair a child's functioning in everyday life, while conversely even a moderately severe social phobia can affect everyday behavior across a broad front. Angold et al. (1999) have shown that 9.4% of children who came to clinical attention but could not be given a category-based diagnosis were significantly impaired in their everyday functioning. Subthreshold manifestations also appear to favor the emergence of more overt, categorically classifiable disorders 5–7 years later (Costello et al. 1999).

The need to consider subdiagnostic symptom constellations and comorbid symptoms has stimulated discussion of the advantages of dimensional diagnosis. ICD-10 differs from DSM-IV (among other ways) in allowing combined diagnoses, e.g. hyperactivity with conduct disorder (F90.1). It is known from clinical observations that the combination of these disorders has a significantly less favorable course than either of them alone (e.g. Mandel 1997). The inclusion of combined diagnoses is based on the fact that certain disorders appear in combination at greater than random frequency. Caron and Rutter (1991) have pointed out a number of mechanisms by which this might occur: two disorders might have exactly the same risk factors; they might share some risk factors and also have other risk factors that are not shared; even if they have no common risk factors at all, the risk factors of one disorder might have a more than random association with the risk factors of the other; or one disorder might be a consequence of the other.

The consistent consideration of comorbid disorders in the additive procedure of the DSM-IV leads to an increase in the number of diagnoses without implications for treatment plans. The use of factor-analytic techniques in dimensional diagnosis, e.g. in the construction of internalized and externalized symptom groups by Achenbach and Edelbrock (1983), has also been used as a search strategy and has led to interesting speculations, such as that of Rasmussen and Eisen (1992) regarding a syndrome group surrounding obsessive–compulsive disorder, including dysmorphophobia, hypochondriasis, paranoid jealousy, Gilles de la Tourette syndrome, trichotillomania, disorders of impulse control, and eating disorders, among others.

Further examples, such as the obsessive-compulsive spectrum described by Hollander (1998) or the bipolar spectrum described by Cassano et al. (1999), are partly the consequence of etiological considerations and partly hypotheses that are useful for research on pathogenesis.

The diagnostic classification of some disorders takes account of their course, as in schizophrenic or bipolar affective disorders, or of the need for various interventions, primarily in chronic, progressive disorders. Consideration of stages of disease is clearly less pronounced for illnesses primarily regarded as psychogenic. The late manifestations of such illnesses – e.g., the severe physical manifestations of anorexia nervosa – are simply seen as complications of the psychogenic process and are expected to regress with successful psychotherapy. Such illnesses are regarded as psychogenic in all stages, and thus reversible. The fact that certain symptom complexes become autonomous is not taken into account. Yet it is precisely with illnesses such as these that stage-specific observation could be important for the planning of interventions. The appearance of specific stages of a disorder during development may also represent a typology of developmental psychopathology, e.g. the appearance of hyperactivity before the emergence of many chronic motor tic disorders, which, with increasing age, may develop into obsessive-compulsive disorders.

The categories of Axis V in the multiaxial classification urgently need to be differentiated and operationalized, so that they may become more reliable (Van Goor-Lambo et al. 1990). The desire to classify protective factors at the same time as risk factors is based on the common misunderstanding that they are merely the positive counterparts of risk factors; this cannot be achieved (Laucht et al 1997, see also Chap. 5, Part 1 of this volume). Protective factors should be regarded as such only if they alter the probability of manifestations of illness when risk factors are present. Similar considerations apply to the inclusion of coping and defense mechanisms in classification.

Because the diagnoses for young children are relatively poorly differentiated, the National Center for Clinical Infant Programs has proposed a classification for 0- to 3-year-olds (Zero-to-three 1990) which simultaneously codes mental disorders, the presence and nature of bonding disorders, disorders of physical development, psychosocial stress factors, and the functional level of emotional development. Unfortunately, this classification differs from the existing multiaxial systems, and is thus unlikely to close the gap between infant psychiatry and child psychiatry. An assessment of bonding might be introduced into the existing diagnostic scheme on Axis V, as long as the coding of psychosocial risk factors were changed so that it could be based less on external circumstances than on underlying mechanisms.

The ICD-10 does not pay due consideration to developmental aspects in the diagnosis of mental disorders of children and adolescents. Not all disorders appearing in childhood and adolescence are listed in the chapters allocated to them in the ICD-10 (Chapters F7–F9); for example, age-typical parasomnias, anorexia nervosa, gender-identity disorders of childhood, and similar disorders are missing. Yet their mere inclusion would still not allow a classification oriented toward developmental psychopathology. Therefore, for these chapters on child and adolescent psychiatry, we have proposed a classification based on the course of these disorders, which distinguishes between behavioral variants and stress reactions, early-onset disorders with persistent course, transient developmentally dependent disorders, disorders with age-typical onset, age-specific interaction disorders, and early-onset adult-type disorders (see Chap. 7, Part 1 of this volume).

4 References

*Achenbach TM, Edelbrock C (1983) Manual for the Child Behavior Checklist and Revised Child Behavior Profile. University of Vermont, Department of Psychiatry, Burlington

*American Academy of Child and Adolescent Psychiatry (1997) Practice parameters. J Am Acad Child Adolesc Psychiatry 36: 1S–202S

*American Academy of Child and Adolescent Psychiatry (1998) Practice parameters. J Am Acad Child Adolesc Psychiatry 37: 1S–89S

Angold A, Costello EJ, Farmer EMZ, Burns BJ, Erkanli A (1999) Impaired but undiagnosed. J Am Acad Child Adolesc Psychiatry 38(2): 129–137

Beidel DC, Turner SM (1998) Shy children, phobic adults. Nature and treatment of social phobia. American Psychiatric Association, Washington

Blanz B, Schmidt MH (1994) Ergebnisse der Mannheimer ICD-10 Feldstudie. In: Dilling H, Schulte-Markwort E, Freyberger J (eds) Von der ICD-9 zur ICD-10. Huber, Bern, pp 231–239

**Caron C, Rutter M (1991) Comorbidity in child psychopathology. Concepts, issues and research strategies. J Child Psychol Psychiatry 32: 1036–1080

Cassano GB, Dell'Osso L, Frank E, Miniati M, Fagiolini A, Shear K, Pini S, Maser J (1999) The bipolar spectrum: a clinical reality in search of diagnostic criteria and an assessment methodology. J Affect Disord 54(3): 319–328

Costello EJ, Angold A, Keeler GP (1999) Adolescent outcomes of childhood disorders: the consequences of severity and impairment. J Am Acad Child Adolesc Psychiatry 38(2): 121–128

Deutsche Gesellschaft für Kinder- und Jugendpsychiatrie und Psychotherapie et al. (eds) (2000) Leitlinien zu Diagnostik und Therapie von psychischen Störungen im Säuglings-, Kindes- und Jugendalter. Deutscher Ärzte-Verlag, Köln

Döpfner M, Lehmkuhl G (1998) Diagnostik-System für Psychische Störungen im Kindes- und Jugendalter nach ICD-10 und DSM-IV. Huber, Bern

Esser G, Blanz B, Geisel B, Laucht M (1989) Mannheimer Elterninterview – Strukturiertes Interview zur Erfassung von kinderpsychiatrischen Auffälligkeiten. Beltz, Weinheim

Henn FA, Bardwell R, Jenkins RL (1980) Juvenile delinquents revisited. Adult criminal activity. Arch Gen Psychiatry 37: 1160–1163

Hollander E (1998) Treatment of obsessive–compulsive spectrum disorders with SSRIs. Br J Psychiatry Suppl 35: 7–12

Janca A et al. (1997) Multiaxial presentation of the ICD-10 for use in adult psychiatry. World Health Organization, Cambridge University Press, Cambridge

Laucht M, Esser G, Schmidt MH (1997) Developmental outcome of infants born with biological and psychosocial risks. J Child Psychol Psychiatry 38: 843–853

Mandel HP (1997) Conduct disorder and underachievement. Risk factors, assessment, treatment, and prevention. Wiley, New York

Moffitt TE (1993) Adolescence-limited and life-course-persistent, antisocial behavior: a developmental taxonomy. Psychol Rev 100: 674–701

Rasmussen SA, Eisen JL (1992) The epidemiology and differential diagnosis of obsessive compulsive disorder. J Clin Psychiatry 53[Suppl]: 4–10

Rutter M, Shaffer D, Shepherd M (1969) A tri-axial classification of mental disorders in childhood. J Child Psychol Psychiatry 10: 41–61

Unnewehr S, Schneider S, Margraf J (eds) (1995) Kinder-DIPS. Diagnostisches Interview bei Psychischen Störungen für Kinder. Springer, Berlin Heidelberg New York

Van Goor-Lambo G, Orley J, Poustka F, Rutter M (1990) Classification of abnormal psychosocial situations: preliminary report of a revision of a WHO scheme. J Child Psychol Psychiatry 31: 229–241

WHO (1994) Schedule for Clinical Assessment in Neuropsychiatry (SCAN). World Health Organization, Geneva

*WHO (1996) Multiaxial classification of child and adolescent psychiatric disorders: the ICD-10 classification of mental and behavioural disorders in children and adolescents. Cambridge University Press, Cambridge

*WHO (1997) Composite International Diagnostic Interview – version 2.1. World Health Organization, Geneva

Zero-to-three, National Center for Clinical Infant Programs (1990) 0–3, diagnostic classification of mental health and developmental disorders of infancy and early childhood. National Center for Clinical Infant Programs, Arlington

CHAPTER

5

H. Remschmidt, M.H. Schmidt

Epidemiology and Pathogenesis

Translator: S. Henn

1 Age-Specific Problems of Epidemiological Research

If knowledge of the frequency, causation, and maintaining factors of mental illnesses is to be used as the basis of their treatment and prevention, then incidence and prevalence figures are of significance. The sources of information used to derive these figures are important. Parents give information about psychopathological problems and their minor children's need for treatment. The life circumstances of these children are also relevant to the causation, onset, course, and treatment of these disorders. Because children have an immature central nervous system (i.e. an immature mental apparatus) and are therefore highly dependent on adults, psychiatric disorders in children can often not be conceived of except in relation to the environment.

Sources of information about psychopathological problems in children and adolescents include the young people themselves and their parents and teachers. The younger the child, the more important the information provided by adults. In general, 8-year-olds can already be asked about their feelings, while adults can usually give better information about outwardly observable behavior. Adolescents are also sources of information about their own behavior, because their delinquency, substance abuse, or sexual activity may be unknown to their parents. Thus descriptive epidemiology makes use of information from various sources, not all of which are necessarily in agreement.

Breton et al. (1999), in the Quebec Child Mental Health Survey, studied the degree of agreement of information obtained from different informants. They found that there was agreement between information obtained from parents and that obtained from children with respect to an elevated frequency of hyperactivity in boys and younger children, and also agreement of information derived from parents, children, and teachers with respect to oppositional disorders and conduct disorders, including a higher rate for boys than girls. For anxiety and depressive disorders, however, no such agreement of information across informants could be found.

In studies in which both healthy and ill individuals are investigated, their answers to questions about psychiatric symptoms differ. Healthy individuals consider even mild symptoms to be serious, while ill individuals are likely to mention only their more serious symptoms.

Because a large amount of data must be obtained, epidemiological studies are often carried out in two stages. After a screening process, all abnormal probands and some of the normal probands are tested individually. Prevalence figures are higher when they are based only on reported symptoms, and when their severity and significance are not assessed by interview, than when they are determined from category-based diagnoses.

The rate of refusal to participate in epidemiologic studies of minors is also age-specific, because parents decide whether young children will participate, and the factors motivating adolescents to refuse to participate are also age-dependent. Good epidemiological studies therefore include information about the refusal rate and about the comparability of those refusing with those participating in the study.

Standardized interviews offer good conditions for obtaining precise incidence and prevalence figures. Open interviews are of somewhat lower quality, questionnaires still lower. Observation of behavior can, of course, cover only a limited repertoire of problems, but it is helpful for specific questions, especially the ascertainment of common behavioral problems. The use of videos and computer-assisted video evaluation allows the gathering of epidemiological data about interactive behavior, for example. Video techniques are almost the only way to assess this adequately, because the simultaneous observation of two partners and their interaction in a natural context cannot otherwise be documented quickly enough.

Questionnaires have the special disadvantage that the respondent's understanding of the question cannot be checked. If the number of questions is increased to be on the safe side, overlapping information results, which raises the cost of information processing. Because answers to questionnaires usually cannot be directly encoded as diagnoses, coding is performed through an intermediate step in which symptoms are linked to syndromes by means of factor analysis. Answers to questionnaires are clinically relevant in the transition zone between normal and pathological behavior, but they cannot be used reliably to determine the presence of illness in individual cases.

The assignment of a cutoff value causes difficulties when symptom-based case definitions are used, because monosymptomatic disorders may significantly impair age-typical functioning and thus deserve to be designated as illnesses. Moreover, the number of symptoms does not increase linearly with the degree of pathology, and the transition zone between pathological and non-pathological states has no clearly defined boundaries. Requiring a larger number of symptoms to be present for a diagnosis to be assigned enlarges this transition zone but cannot solve the problem at its boundaries. Cutoff values are often defined with reference to symptom severity in a sample of patients presenting for psychiatric treatment. This is justifiable as long as the presence of major selection effects can be excluded.

Category-based case definitions in epidemiology have the advantage that clinical threshold values can be better defined and operationalization can be stipulated in advance. Prevalence figures then depend on the narrowness or breadth of the category-based definitions, i.e. on whether a classification system allows combined diagnoses, or whether the required listing of comorbid diagnoses leads to higher numbers of symptoms being counted.

Expert-defined disorder entities have no a priori implications regarding the degree of impairment of age-specific functions, even though many category-based definitions include impairment of everyday functioning among the diagnostic criteria. Combining prevalence with degree of impairment offers a way out of this dilemma and is preferable to requiring further information about the need for treatment.

Cross-sectional studies are not highly suitable for the determination of reliable prevalence figures, even if six-month prevalence is determined, as is frequently done. Longitudinal studies in which data are collected at three or more different times yield more reliable information and are useful, above all, for distinguishing normal behavioral variants from clinically relevant deviations. Various questions concerning developmental psychopathology can be answered only in longitudinal studies, including the question of the stability of psychiatric disorders, i.e. whether a decline in symptoms corresponds to a real decrease in psychopathology or only to a change in symptoms; the significance of an expanding spectrum of symptoms, which is a predictor for substance addiction among patients with antisocial disorders, for example (Robins and Price 1991); and the age-specific appearance of certain symptoms.

The random samples used in such studies should be representative and should ideally be investigated in two stages, as described above. Enriched random samples are advisable for the study of rare disorders or for the study of the context of mental disorders. Samples of patients presenting for treatment are suitable, if at all, then only for determining the incidence and prevalence of very rare disorders, on the assumption that most patients with these disorders will present for treatment.

2 Descriptive Epidemiology

Table 1 contains prevalence data for disorders in child and adolescent psychiatry gathered on the basis of ICD-9, DSM-III, or DSM-III-R diagnoses. Mental retardation may overlap with these diagnoses, and pure mental retardation is not considered. The same applies to specific developmental disorders.

Two studies reporting extreme values were excluded from the table. The study of Nylander et al. (1989), involving 4- to 6-year-old urban children in Sweden, found DSM-III diagnoses in 22% of boys and 37% of girls. On the other hand, the French study of Fombonne (1994) found ICD-9 diagnoses in only 6% of 217 children aged 6 to 11. No studies based on questionnaires are mentioned. The data yield a relatively uniform picture with somewhat higher prevalence among adolescents and somewhat lower prevalence among children; the figures generally range from 12% to 18%.

Psychiatric disorders in children and adolescents show comorbidity with one or several other psychiatric disorders in some of those affected, especially in boys. Comorbidity rates can be artificially increased by extensive splitting of disease categories, so that disorders are classified which are actually components of other disorders. An increase in frequency can also result from different disorders having overlapping symptom spectra, or from chance associations. Real (i.e. systematic or greater than chance) comorbidity arises under defined circumstances (Caron and Rutter 1991; see also Chapter 4, Part 1 of this volume).

In a representative sample, Naab (1995) found two psychiatric disorders in 25% of probands and three in 5%. The comorbidity was higher in a sample of patients presenting for psychiatric treatment. Among all patients with psychiatric disorders in the latter sample, 23% were given two diagnoses, and 19% were given three or more. Bird et al. (1988), in the Puerto Rico study, found more than one diagnosis in 51% of those characterized as psychiatrically ill according to DSM-III. This was largely due to the combination of conduct disorders and hyperactivity; comorbidity was thus more frequently seen in boys and male adolescents.

Comorbidities frequently seen clinically and in representative random samples of children and adolescents are: the combination of depressive disorders with conduct disorders (frequency ca. one third); conduct disorders (usually of the oppositional type, or with absent social relationships) with a hyperactive syndrome (ca. 40%); depressive disorders with anorexia nervosa (ca. 40%); and anxiety or other emotional disorders with hyperactivity (ca. 8%). These comorbidities should not be confused with expansion of symptoms or with symptom changes in accordance with the rules of developmental psychopathology.

In children and adolescents as in adults, the symptom spectrum in anxiety disorders or obsessive–compulsive disorders may expand toward depression, anorexia may change into bulimia, and hyperactivity may supervene in cases of tic disorders, whose symptoms may also expand in the direction of

Table 1. Prevalence data for child and adolescent psychiatric disorders

Author(s)	Country	Sample size (*n*)	Age (years)	Prevalence (%)
Anderson et al. (1987)	New Zealand	792	11	17.6
Esser and Schmidt (1987)	Germany/large city	216	8	16.2
Kashani et al. (1987)	USA/Columbia	150	14–16	18.7
Offord et al. (1987)	Canada/large city	1648	4–16	19.6
Offord et al. (1987)	Canada/provincial	1031	4–16	14.9
Bird et al. (1988)	Puerto Rico	386	4–16	18
Costello et al. (1988)	USA/city	300	7–11	22
Jeffers and Fitzgerald (1991)	Ireland/city	190	10–11	25
Fergusson et al. (1993)	New Zealand/ provincial	986	15	27
Döpfner et al. (1996)	Germany/state	1622	4–18	22.3
Lavigne et al. (1996)	USA/city		2–5	11 (m), 7 (f)
Shaffer et al. (1996)	USA/Connecticut/ Georgia/New York/ Puerto Rico	1285	9–17	10.2
Verhulst et al. (1997)	The Netherlands	780	13–18	6 (according to parental data), 5 (according to adolescent data), 8 (according to parent or child data), 2 (according to parent and child data)
Lehmkuhl et al. (1998)	Germany	1030	4–10	13.1–28.3
Breton et al. (1999)	Quebec/city/ provincial	2400	6–14	19.9 (according to child data), 15.8 (according to parent data)

obsessive–compulsive symptoms. Epidemiologic studies have shown that hyperactivity can expand to conduct disorders (independently of the comorbid combination of these two types of disorder), and that conduct disorders can expand to substance abuse (Robins and Price 1991) or become transformed into antisocial personality disorders. In longitudinal epidemiologic studies, symptoms beginning in early school age show a stability of 12%–13% into the middle of the third decade and a five-year stability of approximately 50%.

The less favorable course of the combination of hyperactivity with conduct disorder, as compared to either of the disorders alone, has already been mentioned. Russo and Beidel (1994) found a less pronounced response to stimulants in hyperactive children with comorbid anxiety disorders. According to Henn et al. (1980), children with conduct disorders and without social attachments, i.e. those with earlier relationship disorders, show a course with more delinquency than children with conduct disorders whose relationships are normal for their age.

3 Analytical Epidemiology and Pathogenesis

Psychiatric disorders in school age are at least twice as frequent in boys than in girls. Their overall rates become similar during adolescence, though the rates of individual disorders are still quite different. From the earliest ages onward, boys have more expansive disorders, i.e. disorders with hyperkinetic, oppositional, or aggressive behavior. They also have more developmentally dependent disorders, e.g. speech disorders, specific difficulties of reading and writing, enuresis, and encopresis. Autism is more frequently seen in boys. The incidence of emotional disorders is about the same in children of both genders. Only anxiety and somatization disorders are somewhat more common in girls.

Anxiety-depressive disorders increase in adolescence among girls. In early adolescence, girls temporarily show high rates of aggressive-antisocial behavior, similar to those of boys; the probability that these characteristics will regress is, however, far higher in girls than in boys. Over the course of adolescence, the

rate of substance dependency in boys increases in parallel to antisocial behavior. Clear shifts occur during adolescence in girls toward eating disorders, and in boys toward completed suicides.

Because of the gender differences described, epidemiologic studies of prevalence and incidence without reference to gender are not very helpful. Despite the differences listed, psychiatric disorders are generally more severe in boys and male adolescents than in girls and female adolescents.

Many studies have shown a slight rise in the overall prevalence of psychiatric disorders in children of primary school age, with a second discrete peak seen in many studies at the onset of adolescence. For a long time, the first peak was explained as an expression of increasing levels of activity and interest in daily events at the beginning of school, but the cause of the increase of psychiatric disorders at this age is now considered more likely to be a differentiation of emotional disorders corresponding to the typical, delayed differentiation of the emotions even in the normal context. The second peak is most likely explained by the transient increase in conduct disorders in girls at the onset of adolescence.

Long-term stability of psychiatric disorders is seen primarily in disorders with extroverted manifestations in boys, particularly in connection with earlier hyperactivity. Disorders with introverted manifestations show far less stability but are of long-term importance for the appearance of significant mental disorders in early adulthood, even if the severity of symptoms lies below the diagnostic threshold.

Developmental age mirrors brain maturity. As it increases, the child becomes less dependent on others. Consequently, interaction disorders are more frequent in early childhood and fall off with increasing autonomy. Functions that develop rapidly are particularly susceptible to disturbances. Many physical functions that develop at an early age fall into this category. The higher frequency of physical symptoms at this age may also be an expression of the immaturity of physical systems; stereotypic movements or motor tics, for example, may express an as yet inadequate capability for motor inhibition. As development proceeds, the environment places demands on functions that mature in predictable fashion. Developmental delays affecting such functions, e.g. affect control, easily become manifest as disorders, e.g. the delayed regression of physiological separation anxiety.

As the child reaches school age, there is more reliable differentiation between reality and fantasy and an increase in intrapsychic processing ability. In early adolescence, with its new developmental tasks, new fields are opened for possible mental disorders. The fact that some disorders regularly appear at a later age, e.g. depressive disorders with somatic syndrome from age 16 onward, makes it seem likely that hormonal changes significantly influence the pathogenesis of mental illnesses in adolescence.

According to Rutter (1983), physical illness doubles the risk of appearance of psychiatric disorders, and diseases of the brain triple it. Brain diseases accompanied by seizures present an additional complication. Scar formation in the central nervous system produces the highest risk of cerebral impairment. The significance of organic cerebral impairment in the earliest stages of life is certainly related to the historical period in which it occurred: in previous years, when the care of premature infants was rudimentary compared to today, perinatal brain damage and premature birth were more important risk factors for later behavioral abnormalities (e.g. Pasamanick et al. 1956). The study of increases in motor, cognitive, and behavioral impairment is now focused on premature infants with very low birth weight (Esser et al. 1995; Hall et al. 1995). These organic risks are seen mostly in motor and cognitive functions and become less significant with increasing age.

For a long time, the interaction of biological risks with unfavorable social conditions was overlooked in research, but bivariant research designs now show that psychosocial deficits influence cognitive and, even more, social and emotional development (Laucht et al. 1996, 1997). In addition, the studies carried out by Field et al. (1981) have repeatedly demonstrated that the quality of interaction between parents and children favors or inhibits the pathogenesis of mental disorders.

Models for the pathogenesis of psychiatric disorders are generally hypothetical and imprecise, because the relationship of individual developmental factors to one another cannot be varied experimentally. It was hoped that the increase of knowledge in genetics and molecular biology would allow better predictions to be made, but this has not yet been realized. There are probably genetic contributions to the mental problems of children and adolescents with pronounced mental retardation, reading and writing deficits, autism and hyperactivity, Gilles de la Tourette syndrome, and antisocial disorders. The evidence for this is derived mainly from formal genetic studies (e.g. Edelbrock et al. 1995).

Because most disorders are of multifactorial origin, genetic characteristics mainly assume the role of a predisposition or increased vulnerability to illness. It is unclear under what conditions the corresponding predispositions lead to manifest symptoms. The idea that chronic or acute stress exerts a particular effect on a genetically predisposed individual is certainly too simple. Instead, we must assume that early stressful experiences can change the readiness of an organism to react. Accordingly, research has also gradually

shifted away from very remote risk factors and toward central processing mechanisms.

Urban–rural differences and high urbanization rates (Rahim and Cederblad 1984), migration, social class, and parental level of education are thus regarded as characteristics accounting for only a small part of the variance. Characteristics linked to interactions within the family, such as family disharmony, cramped living conditions, mental illnesses in family members, and criminality of a parent are quantitatively more significant (Blanz et al. 1991). The model of cumulative risk proposed in the classic study of Rutter and Quinton (1977) has been repeatedly confirmed. More of the variance is accounted for when interactions within the family are directly included in the calculations (e.g. Esser et al. 1993).

Since the 1960s, research has been based on the assumption that the concurrence of predispositions, stress factors, and unfavorable interactions does not inevitably lead to psychiatric disorders, and that there is a group of children in critical life circumstances who remain unaffected in spite of such conditions. The size of this group varies, depending on the study cited, from 10% to 25%. Such resilience is attributed to the existence of protective factors, which have been increasingly studied since the 1980s. There are essentially three groups of variables:

1. Personality features such as autonomy, positive social orientation, internal awareness of control, and differentiated linguistic abilities
2. Family characteristics such as a cohesive family, warmth, and absence of quarrels
3. Extrafamilial features such as support systems, significant individuals outside the family to whom the child can relate, and close friends.

The concept is rightly criticized in that protective factors must not be considered the same thing as generally beneficial characteristics or circumstances, or the absence of risk factors. Protective factors are rather those that exert a protective effect where combinations of risk factors exist, i.e. they are required to have been present before the appearance of these risk factors if their effect is to be assessed statistically.

Table 2. Prevalence rates of individual mental disorders in children and adolescents

Kind of disorder	Reference	Probands	Prevalence rates (%)
Hyperkinetic disorder	Costello et al. (1988)	7- to 11-year-olds	2%–10% with a tendency to lower values, maximum in primary school age
	Lavigne et al. (1996)	2- to 5-year-olds	
Behavioral disorders	Lavigne et al. (1996)	2- to 5-year-olds	2%–10%
	Cohen et al. (1993)	10- to 20-year-olds	10%–16% (m), 3.8%–7.1% (f)
Oppositional defiant behavior	Cohen et al. (1993)	10- to 13-year-olds	14.2% (m), 10.4% (f)
Depression	Birmaher et al. (1996)	Children Children of depressed parents Adolescents	0.4%–2.5% 0.4%–8.3% 15%–45% (lifetime)
Anxiety	Lavigne et al. (1996)	2- to 5-year-olds and 10- to 13-year-olds	0.7%–12.5%
Separation anxiety	Cohen et al. (1993)	10- to 13-year-olds	11.4% (m), 13.1% (f)
Tic and stereotypic disorders	Esser and Schmidt (1987)	8-year-olds	4.6% (m), 9.3% (f)
Gilles de la Tourette syndrome	Robertson (1994)	Children	0.3%–1.3%
Enuresis	von Gontard (1998)	7-year-olds	10%
Bulimia nervosa	Ledoux et al. (1991)	Adolescents	0.3%–7%
Autism (Kanner)	Rapin (1997), Steffenburg and Gillberg (1986)	Children up to 10 years	0.04%–0.1%
Autism (Asperger)	Ehlers and Gillberg (1993)	7- to 16-year-olds	0.25%–0.35%
Schizophrenia	Remschmidt (1992)	14- to 20-year-olds	0.09%

m, boys or male adolescents; f, girls or female adolescents.

The discovery of protective factors affords the hope that primary preventive measures can be developed. Yet these factors obviously exert their effects in the context of interpersonal interactions, and the mechanisms by which they do so have not yet been well studied. Observations of mother–child interactions, interactions between friends, the formation of relationships between the child and adults, and similar processes might be worthwhile approaches. Such mechanisms are probably still relatively poorly defined, however, and research continues to focus on the question of early development of neurochemical or neuroendocrine mechanisms for coping with various stresses.

This would open up a field in the epidemiology of biological markers that has been understood to date only in very rough outlines, e.g. with the search for neurological "soft signs," particular features of the electroencephalogram (EEG), and morphological abnormalities of the central nervous system. Biological markers might be a starting point for experimental or quasi-experimental studies and, with the inclusion of functional imaging, might be important for research on pathogenesis. Most of the research in the field is thus moving away from the theory of direct traumatization and toward hypotheses of the long-term effects of congenital or acquired mechanisms on behavior. This trend is also seen in the decreasing importance attached to life events as causes of mental disorders: research in child psychiatry now tends to regard life events as consequences, rather than causes, of abnormalities in children and/or their families. Constructs such as temperament have accordingly acquired new importance.

Table 2 provides information from various sources on the prevalence of individual psychiatric disorders. The different samples used to collect these data, and the possibility that they are based on differing case definitions, must be taken into account in their interpretation.

4 References

Anderson J, Williams S, McGee R, Silva P (1987) DSM-III disorders in preadolescent children. Arch Gen Psychiatry 44: 69–76

Bird H, Canino G, Rubio-Stipec M, Gould MS, Ribera J, Sesman M, Woodbury M, Huertas-Goldman S, Pagan A, Sanchez-Lacay A, Moscoso M (1988) Estimates of the prevalence of childhood maladjustment in a community survey in Puerto Rico. Arch Gen Psychiatry 45: 1120–1126

Birmaher B, Ryan ND, Williamson DE, Brent DA, Kaufman J, Dahl RE, Perel J, Nelson B (1996) Childhood and adolescent depression: a review of the past 10 years. Part I. J Am Acad Child Adolesc Psychiatry 35(11): 1427–1439

Blanz B, Schmidt MH, Esser G (1991) Familial adversities and child psychiatric disorders. J Child Psychol Psychiatry 32(6): 939–950

Breton JJ, Bergeron L, Valla JP, Berthiaume C, Gaudet N, Lambert J, St-Georges M, Houde L, Lépine S (1999) Quebec Child Mental Health Survey: prevalence of DSM-III-R mental health disorders. J Child Psychol Psychiatry 40: 375–384

Caron C, Rutter M (1991) Comorbidity in child psychopathology. Concepts, issues and research strategies. J Child Psychol Psychiatry 32: 1036–1080

Cohen P, Cohen J, Kasen S, Noemi Velez C, Hartmark C, Johnson J, Rojas M, Brook J, Streuning EL (1993) An epidemiological study of disorders in late childhood and adolescence. I. Age- and gender-specific prevalence. J Child Psychol Psychiatry 34: 851–867

*Costello EJ, Costello AJ, Edelbrock C, Burns B, Dulcan MK, Brent D, Janizewski S (1988) Psychiatric disorders in pediatric primary care: prevalence and risk factors. Arch Gen Psychiatry 45: 1107–1116

Döpfner M, Schmeck K, Poustka F, Berner W, Lehmkuhl G, Verhulst F (1996) Verhaltensauffälligkeiten von Kindern und Jugendlichen in Deutschland, den Niederlanden und den USA. Nervenarzt 67: 960–967

Edelbrock C, Rende R, Plomin R, Thompson LA (1995) A twin study of competence and problem behavior in childhood and early adolescence. J Child Psychol Psychiatry 36: 775–785

Ehlers S, Gillberg C (1993) The epidemiology of Asperger syndrome. A total population study. J Child Psychol Psychiatry 34(8): 1327–1350

*Esser G, Schmidt MH (1987) Epidemiologie und Verlauf kinderpsychiatrischer Störungen im Schulalter – Ergebnisse einer Längsschnittstudie. Nervenheilkunde 6: 27–35

Esser G, Dinter R, Jörg M, Villalba P, Laucht M, Schmidt MH (1993) Bedeutung und Determinanten der frühen Mutter-Kind-Beziehung. Z Psychosom Med 39: 246–264

Esser G, Laucht M, Schmidt MH (1995) Der Einfluß von Risikofaktoren und der Mutter-Kind-Interaktion im Säuglingsalter auf die seelische Gesundheit des Vorschulkindes. Kindheit Entwicklung 4: 33–42

Fergusson DM, Horwood LJ, Lynskey MT (1993) Prevalence and comorbidity of DSM-III-R diagnoses in a birth cohort of 15-year-olds. J Am Acad Child Adolesc Psychiatry 32: 1127–1134

Field TM, Dempsey JR, Shuman HH (1981) Developmental follow-up of pre- and postterm infants. In: Friedman SL, Sigman M (eds) Preterm birth and psychological development. Academic, New York, pp 3–15

*Fombonne E (1994) The Chartres Study. I. Prevalence of psychiatric disorders among French school-age children. Br J Psychiatry 164: 69–79

Hall A, McLeod A, Counsell C, Thomson L, Mutch L (1995) School attainment, cognitive ability and motor function in a total Scottish very-low-birthweight population at eight years: a controlled study. Dev Med Child Neurol 37: 1037–1050

Henn FA, Bardwell R, Jenkins RL (1980) Juvenile delinquents revisited. Adult criminal activity. Arch Gen Psychiatry 37: 1160–1163

Jeffers A, Fitzgerald M (1991) Irish families under stress, vol 2. Eastern Health Board, Dublin

Kashani JH, Beck NC, Hoeper EW, Fallahi C, Corcoran CM, McAllister JA, Rosenberg TK, Reid JC (1987) Psychiatric

disorders in a community sample of adolescents. Am J Psychiatry 144: 584–589

Laucht M, Esser G, Schmidt MH, Ihle W, Marcus A, Stöhr R-M, Weindrich D (1996) Viereinhalb Jahre danach: Mannheimer Risikokinder im Vorschulalter. Z Kinder Jugendpsychiatr 24: 67–81

Laucht M, Esser G, Schmidt MH (1997) Developmental outcome of infants born with biological and psychosocial risks. J Child Psychol Psychiatry 38: 843–853

Lavigne JV, Gibbons RD, Christoffel KK, Arend R, Rosenbaum D, Binns H, Dawson N, Sobel H, Isaacs C (1996) Prevalence rates and correlates of psychiatric disorders among preschool children. J Am Acad Child Adolesc Psychiatry 35: 204–214

Ledoux S, Choquet M, Flament M (1991) Eating disorders in an unselected French population. Int J Eat Disord 10: 81–89

Lehmkuhl G, Döpfner M, Pluck J, Berner W, Fegert JM, Huss M, Lenz K, Schmeck K, Lehmkuhl U, Poustka F (1998) Häufigkeit psychischer Auffälligkeiten und somatischer Beschwerden bei vier- bis zehnjährigen Kindern in Deutschland im Urteil der Eltern – ein Vergleich normorientierter und kriterienorientierter Modelle. Z Kinder Jugendpsychiatr Psychother 26: 83–96

Naab S (1995) Komorbidität kinder- und jugendpsychiatrischer Störungen. Ein Vergleich des Ausmaßes von Überschneidungen kinder- und jugendpsychiatrischer Störungen innerhalb einer Bevölkerungsstichprobe und einer Inanspruchnahmepopulation von Kindern und Jugendlichen. Dissertation, Heidelberg University

Nylander I, Rydelius PA, Nordberg L, Aurelius G, Zetterström R (1989) Infant health and development in relation to the family situation. A review of a longitudinal prospective study in a new Stockholm suburb. Acta Paediatr Scand 78: 1–10

Offord DR, Boyle MH, Szatmari P, Rae-Grant NI, Links PS, Cadman DT, Byles JA, Crawford JW, Blum HM, Byrne C, Thomas A, Woodward CA (1987) Ontario child health study. II. Six-month prevalence of disorder and rates of service utilisation. Arch Gen Psychiatry 44: 832–836

Pasamanick B, Rogers ME, Lilienfeld AM (1956) Pregnancy experience and the development of behavior disorder in children. Am J Psychiatry 112: 613–618

Rahim SJA, Cederblad M (1984) Effects of rapid urbanisation on child behaviour and health in a part of Khartoum, Sudan. J Child Psychol Psychiatry 25: 629–641

Rapin I (1997) Autism. N Engl J Med 337(2): 97–104

Remschmidt H (1992) Psychiatrie der Adoleszenz. Thieme, Stuttgart

Robertson MM (1994) Annotation: Gilles de la Tourette syndrome – an update. J Child Psychol Psychiatry 35: 597–611

**Robins LN, Price RK (1991) Adult disorders predicted by childhood conduct problems: results from the NIMH Epidemiologic Catchment Area project. Psychiatry 54: 116–132

Russo MF, Beidel DC (1994) Comorbidity of childhood anxiety and externalizing disorders: prevalence, associated characteristics, and validation issues. Clin Psychol Rev 14: 199–221

Rutter M (ed) (1983) Developmental psychopathology. Guilford, New York

Rutter M, Quinton D (1977) Psychiatric disorder – ecological factors and concepts of causation. In: McGurk H (ed) Ecological factors in human development. North-Holland, Amsterdam, pp 173–187

Shaffer D, Fisher P, Dulcan MK, Davies M, Piacentini J, Schwab-Stone ME, Lahey BB, Bourdon K, Jensen PS, Bird HR, Canino G, Regier DA (1996) The NIMH Diagnostic Interview Schedule for Children Version 2.3 (DISC-2.3): description, acceptability, prevalence rates, and performance in the MECA Study. Methods for the Epidemiology of Child and Adolescent Mental Disorders Study. J Am Acad Child Adolesc Psychiatry 35: 865–877

Steffenburg S, Gillberg C (1986) Autism and autistic-like conditions in Swedish rural and urban areas: a population study. Br J Psychiatry 149: 81–87

*Verhulst FC, van der Ende J, Ferdinand RF, Kasius MC (1997) The prevalence of DSM-III-R diagnoses in a national sample of Dutch adolescents. Arch Gen Psychiatry 54: 329–336

von Gontard A (1998) Annotation: day and night wetting in children – a paediatric and child psychiatric perspective. J Child Psychol Psychiatry 39: 439–451

CHAPTER

6

H. Remschmidt, M.H. Schmidt

Therapy of Children and Adolescents

Translator: S. Henn

1 Introduction

In this chapter, the most important methods of treating mental disorders in children and adolescents are briefly outlined. Their specific application to individual illnesses is described in Chap. 7 (Part 1 of this volume). Current reviews of psychotherapeutic methods can be found in Prout and Brown (1999) and Remschmidt (1997), and current reviews of psychopharmacological therapy can be found in Nissen et al. (1998), Rosenberg et al. (1994), and Kutcher (1997).

The types of treatment used in mental disorders may be classified in several different ways, including the following:

- by method (e.g. psychopharmacological therapy, functional therapy, behavior therapy, psychoanalysis)
- by setting (e.g. individual-centered therapy, family therapy, outpatient treatment, day clinic, or inpatient treatment)
- by the disorders to be treated (e.g. anxiety syndromes, obsessive–compulsive syndromes, hyperkinetic syndrome)

Which method will be used in which setting should ideally be determined on the basis of empirically demonstrated effectiveness.

2 Psychotherapy

Individual-centered treatment methods include all those that are carried out selectively in the individual patient. These include psychotherapeutic methods which are derived from very different theoretical schools of thought, such as psychoanalytically oriented therapy, behavior therapy, functional therapy (e.g. perception training, psychomotor training), creative methods (e.g. catathymic picture experience, music therapy), and cognitive therapy. Individual-centered treatment methods are indicated for practically every patient; the large number of indications cannot be discussed in detail in this brief introduction to the subject.

A *psychodynamic method* is suitable particularly in those cases in which the patient can cooperate adequately and the symptoms are diffuse rather than circumscribed (e.g. crises of individuation, anxiety neuroses). Age is also important. In children of preschool and primary school age, psychoanalytic (more specifically, psychodynamic) treatment must make use of game-playing (Dührssen 1980), while in adolescents the usual verbal method may be used (Seiffge-Krenke 1986), as long as due consideration is taken of the patient's stage of life (Remschmidt and Quaschner 1997).

Behavior therapy is used for habitual symptoms, i.e. those which have arisen on the basis of abnormal habit formation, and for circumscribed disorders. Examples are enuresis, habitual behaviors (e.g. nail biting, hair pulling, excessive restlessness, tics), phobias, some psychosomatic illnesses (e.g. psychogenic eating disorders, including anorexia nervosa), and a large number of disorders in which symptoms can be treated with behavior modification without curing the underlying illness (e.g. autistic syndromes, oligophrenia, schizophrenia) (Steinhausen and von Asten 1993). Behavior therapy has developed considerably in recent years and has incorporated cybernetic models and systemic approaches into its techniques (Margraf 2000). Cognitive behavior therapy now has a special status, as it is readily applicable in children, e.g. for depressive disorders (Graham 1998).

There are also many indications for *functional therapies*. These are mainly used to treat circumscribed deficits (e.g. dyslexia, arithmetical disorder, perceptual and concentration disorders) and to remediate developmental deficits or delays (e.g. in motor and language development).

Creative treatment methods are used whenever a direct approach with verbal psychotherapy is difficult or impossible because of the patient's age and developmental stage, or because of the disorder itself.

Cognitive approaches to treatment (insight therapies) are indicated mainly in neurotic and psychoreactive disorders. They are now being increasingly applied, particularly in depression. In addition, progress has been made in evaluating these approaches (Graham 1998).

In very broad terms, *family-centered treatment* includes family counseling (parent counseling), psychodynamically oriented family therapies, behavior-oriented family therapies, child-centered family therapy, and various "home treatment" methods (treatment in the patient's own environment). The indication for a family-centered method must be carefully established by evaluating both the disorder and the patient's overall situation (Mattejat 1997; Remschmidt and Mattejat 1998).

Group-centered treatment methods include open group psychotherapies (analytic or non-analytic methods), goal-directed group therapies (e.g. assertiveness training, group therapy with children who have contact difficulties or are aggressive), autogenic training in groups, group play therapy, and parent groups with various goals. Group therapies have long been established in our clinical practice. They have proved effective in the following areas: open group therapy for adolescents (with various disorders, particularly with

identity crises) and goal-directed group therapies in adolescents with contact difficulties as well as aggressive and uncontrollable children (Niebergall 1997; Remschmidt 1992).

3 Psychopharmacological Treatment

Psychoses in adults and adolescents are still the domain of psychopharmacological treatment (see also Chap. 6, Part 1 of this volume, and Chaps. 13 and 23, Vol. 3, Part 1). Other important indications are the following: hyperkinetic syndrome, obsessive–compulsive disorders, Gilles de la Tourette syndrome, anxiety disorders, and depression (Kutcher 1997; Nissen et al. 1998). Pharmacodynamic mechanisms in adults do not necessarily apply to children because of differences between children and adults regarding the spontaneous course of mental disorders, developmental stages, and environmental factors.

The use of psychotropic drugs in children is often considered in the treatment of conditions that are not truly medical diagnoses. For example, it is pointless to give medication to a child who is having difficulty at school if diagnostic evaluation (which is, sadly, often not performed) reveals that the problem is due to excessive demands placed on the child by overambitious parents. The guiding principle in the choice of a treatment plan remains that the prescription of medication for the disorders discussed here cannot replace or obviate the need for discussion, psychotherapy, or special education. There is no pharmacological treatment which, when given in a particular dose, can eliminate a mental disorder.

Only general principles can be stated for the dosage and duration of treatment with psychotropic substances. These can be summarized as follows:

- The indication should be substantiated by meticulous assessment leading to the provision of a diagnosis (e.g. hyperkinetic syndrome, schizophrenic psychosis, tic, Gilles de la Tourette syndrome).
- In general, treatment should begin at a low dose, which is then slowly raised.
- Treatment should be given at the optimal dose for at least several weeks, because a treatment that is too short will often fail.
- Abrupt discontinuation is to be avoided.
- Once symptoms improve, a maintenance dose should be given, or not given, depending on the clinical situation.
- The method of administration should be determined by pharmacokinetic and pharmacodynamic considerations and not by rigid preset schemes.
- It is advisable for physicians to use a few psychotropic drugs with which they are highly familiar, rather than prescribe a large number of drugs with which they have insufficient personal experience.

3.1 Neuroleptics

Neuroleptics are a group of chemically and pharmacologically distinct substances which, despite their structural differences, share a number of effects:

- Reduction of physical (CNS) and mental tension
- Psychomotor sedation
- Dampening of excitement and aggression
- Antipsychotic/antischizophrenic effect
- Neurological and autonomic side effects
- Predominantly subcortical mechanism of action

Neuroleptics also alter autonomic function, causing undesired side effects such as hypotension, tachycardia, and dry mouth. Furthermore, most neuroleptics have antiemetic, antipruritic, antipyretic, analgesic, and sometimes appetite-stimulating effects.

The mechanism of action of neuroleptics is now explained mainly as the blockade of dopamine receptors. The effectiveness of the neurotransmitter dopamine is reduced by blockade of the dopamine receptor. It is suspected that blockade of the dopamine receptors leads to a compensatory increase of catecholamine synthesis. Yet dopamine receptor blockade cannot be the sole mechanism of action of neuroleptics, as they also have antiserotonergic, antihistaminic, anticholinergic, and antiadrenergic effects, which are particularly prominent in the new, atypical neuroleptics and determine both their activity profiles and their side effects.

The neuroleptics most frequently used in child and adolescent psychiatry can be classified into six groups:

1. Phenothiazine derivatives: levomepromazine (e.g. Neurocil), promethyzine (e.g. Atosil), thioridazine (e.g. Mellaril), perazine (e.g. Taxilan), trifluoperazine (e.g. Jatroneural), fluphenazine (e.g. Dapotum, Lyogen, Omca), perphenazine (e.g. Decentan).
2 Thioxanthene derivatives: chlorprothixene (e.g. Truxal), flupentixol (e.g. Fluanxol).
3. Butyrophenones: haloperidol (e.g. Haldol), trifluperidol (e.g. Triperidol), benperidol (e.g. Glianimon), bromperidol (e.g. Impromen).
4. Diphenylbutylpiperidines: pimozide (e.g. Orap), fluspirilene (e.g. Imap).
5. Benzamides: sulpiride (e.g. Dogmatil), tiapride (e.g. Tiapridex).
6. Atypical neuroleptics: clozapine (e.g. Leponex), olanzapine (Zyprexa), risperidone (Risperdal).

Indications and Side Effects

The indications for neuroleptic treatment in children and adolescents, as in adults, are based on "target symptoms." In acute psychotic conditions with predominantly productive symptoms, the butyrophenone derivatives (especially haloperidol and benperidol) have been found to be effective, as have the phenothiazine derivatives perazine, fluphenazine, perphenazine, and chlorprothixene. If the psychotic condition is accompanied by marked restlessness, then levomepromazine is indicated for its calming effect.

Clozapine has also proved to be effective in both acute and chronic psychotic conditions. Its advantage, in addition to the general absence of extrapyramidal motor side effects, is its positive effect on psychotic symptoms. Because of side effects on the hematopoietic system (most importantly agranulocytosis), the preparation has been taken off the market. However, it can be used as long as special precautions are taken (regular and repeated determination of the complete blood count, ruling out of epileptiform activity by EEG, and testing of liver and kidney function), and the patients and their parents are informed of possible side effects.

If a lack of drive, negativism, autistic behavior, inhibition, and withdrawal are prominent, then the use of sulpiride is advisable. Success has also been reported with haloperidol and fluphenazine.

Depot neuroleptics are indicated for chronic schizophrenic psychoses. Haloperidol decanoate and fluphenazine decanoate have been used successfully. These medications are given intramuscularly, and their effect lasts from 1 to 4 weeks, depending on the substance. These medications are only rarely used in children and will not be further discussed here.

For psychotic and nonpsychotic restlessness and excitement, neuroleptics with a strong hypnotic effect, such as levomepromazine, are recommended.

A number of neuroleptics have a good or very good effect on tics and Gilles de la Tourette syndrome. Haloperidol, pimozide, and the benzamide derivative tiapride have been used with success.

If hypermotor phenomena are prominent following disorders of brain function in early childhood, thioridazine and the butyrophenone derivative pipamperone can be used successfully. It has been shown that thioridazine in lower doses has a positive effect on concentration and motor ability, which decreases at higher doses.

The side effects of neuroleptic treatment can be summarized as follows:

- Extrapyramidal motor symptoms (EPS): pharmacogenic parkinsonism, acute dyskinesia ("early dyskinesia"), akathisia, tardive dyskinesia ("late dyskinesia")
- Autonomic symptoms: dry mouth, tachycardia, accommodation disorders, constipation, micturition disorders, decreased sweating, increased intraocular pressure (glaucoma)
- Cardiovascular disturbances: hypotension, ECG abnormalities (prolonged QT interval, T-wave abnormalities), arrhythmias
- Mental symptoms: fatigue, pharmacogenic depression
- Endocrine disturbances: galactorrhea, gynecomastia, menstrual disturbances, sexual disturbances
- Neurological side effects: epileptic seizures, delirium
- Hematopoietic effects: transient leukocytosis, eosinophilia, lymphocytosis, agranulocytosis
- Dermatological disturbances: cutaneous allergic reactions, light sensitization
- Hepatic effects: (transient) elevation of transaminases, cholestatic jaundice, liver failure
- Ophthalmologic disturbances: clouding of the lens and cornea, retinal pigmental changes
- Thermoregulatory disturbances: temporary mild fever ("drug fever"), high fever as a component of neuroleptic malignant syndrome
- Metabolic disturbances: decreased glucose tolerance, increased appetite

It should be pointed out that some of the side effects listed are very rare. For example, agranulocytosis appears in only 0.1% of patients and is reversible if diagnosed early enough. This makes it even more important for the physician to be well-informed about side effects.

3.2 Antidepressants

Antidepressants are used more sparingly in children and adolescents than in adults. Their indications have widened recently because of new findings (for a review, see Remschmidt and Schulz 1995).

Antidepressant substances can be divided into five classes:

1. Tricyclic antidepressants
2. Tetracyclic antidepressants
3. Monoamine oxidase inhibitors (MAOI)
4. Selective serotonin reuptake inhibitors
5. Lithium salts

Lithium salts are a group of agents that cannot be designated as antidepressants in the narrower sense. However, they will be discussed here because they have an antidepressant effect.

Antidepressants (thymoleptics) include various substances of different chemical structures that share the

effects of brightening mood, increasing motivation, and relieving anxiety. The effectiveness profiles of individual antidepressants are described in reference to these three components. The following types of effectiveness profile are found (Kielholz 1971):

- Imipramine type: antidepressant and mild psychomotor activating effect
- Amitriptyline type: antidepressant and calming effects
- Desipramine type: antidepressant and marked psychomotor activating effect

Tricyclic antidepressants inhibit the reuptake transport of norepinephrine and other transmitters into neurons. Thus, increased levels of these amines are available at the receptor. Tricyclic antidepressants and MAOI lead to the same end result through different mechanisms of action. Their antidepressant effect is also based on this. The individual tricyclic antidepressants differ in the extent to which they inhibit the synaptic reuptake of norepinephrine and serotonin.

A similar mechanism has been described for tetracyclic antidepressants as for the tricyclics. These, too, inhibit norepinephrine reuptake into neurons.

Tricyclic antidepressants are not effective in childhood depression as demonstrated by controlled studies (Remschmidt and Schulz 1995). Their main indications are other fields of treatment and the period of adolescence and adulthood.

Indications

Although the main indication for antidepressants is endogenous depression, the use of these substances is also indicated as a supportive measure in other forms of depression. The indications for the use of antidepressants are based on so-called target symptoms, which are included in the three-component scheme proposed by Kielholz (1971). The choice of medication is guided by the symptoms and the desired therapeutic effect. If relief of depression and strong psychomotor activation are the objective, desipramine-type antidepressants are given; if relief of depression and mild psychomotor activation are the objective, imipramine-type antidepressants are given; and if relief of depression and a calming action are the objective, amitriptyline-type antidepressants are given.

Antidepressants are generally given orally. The dose is increased gradually so that the full dose is reached in 3 to 5 days. Most antidepressants have relatively long half-lives, and can thus be given twice instead of three times daily. The onset of effectiveness of most antidepressants is within 1 to 2 weeks of the start of treatment. Frequent errors in their administration are inadequate dosing and premature discontinuation. These medications should always be discontinued with a gradual taper.

Drug treatment of obsessive–compulsive syndromes is based essentially on the use of antidepressants. Clomipramine has proved to be especially useful. It is a chlorinated imipramine which preferentially inhibits serotonin reuptake. A dosage of 50–75 mg/day is usually aimed at in children under 14 years of age, and 75–100 mg/day in those over 14 years of age. The therapeutic plasma level has been reported to be 50–140 ng/ml. Fluoxetine, a new selective serotonin reuptake inhibitor, also provides effective treatment, especially in adolescents; this substance has a major activating effect and the dose should therefore be increased cautiously and gradually. As there has been less experience with fluoxetine in children and adolescents, drug treatment of obsessive–compulsive syndromes is based primarily on the use of clomipramine. If the response to tricyclic antidepressants is inadequate, the administration of fluoxetine can be considered.

Antidepressants are also useful in the treatment of school phobia. It should be kept in mind, however, that psychotherapy, rather than drug treatment, is the main component of therapy for this disorder. The administration of tricyclic antidepressants has proved to be useful for patients with this disorder who are depressed.

Numerous observations, as well as meticulous double-blind studies, are available concerning the effectiveness of tricyclic antidepressants (especially imipramine) in enuresis, about which there is no longer any doubt.

Although the pharmacological therapy of hyperactivity consists mainly of the administration of stimulants (e.g. methylphenidate), success has also been achieved with tricyclic and tetracyclic antidepressants. Again, most experience has been with imipramine. Comparative studies show, however, that methylphenidate is superior to imipramine.

In somnambulism (sleepwalking), too, the administration of tri- and tetracyclic antidepressants has proved successful. The greatest amount of data is available for imipramine. The effect is attributed to a reduction of depth of sleep and a shortening of rapid eye movement (REM) sleep. As for enuresis, a one-time evening dose of the antidepressant is administered, e.g. 10–20 mg imipramine.

Contraindications and Side Effects

Previous treatment with MAOI is an absolute contraindication. If tri- or tetracyclic antidepressants are to be used after such treatment, then there should be an interval of at least 14 days in between.

Particular attention should be paid to cardiotoxicity and the reduction of the seizure threshold caused by tricyclic antidepressants. It is advisable to obtain an ECG before initiating their use in order to identify any previously existing arrhythmias, which present a contraindication. An EEG should also be obtained.

The most important undesired side effects of the tri- and tetracyclic antidepressants are the following:

- anticholinergic side effects: dryness of mucous membranes, mydriasis, sweating, tachycardia, accommodation disturbances, fatigue, somnolence
- cardiac arrhythmias and ECG changes: frequent rhythm disturbances (up to absolute arrhythmia), flattened T-waves, hypotension
- Micturition disorders (up to urinary retention)
- Hematopoietic abnormalities: leukopenia, eosinophilia, agranulocytosis
- Hepatic disturbances: jaundice, elevated transaminases and alkaline phosphatase
- Lowering of the seizure threshold, possibly leading to epileptic seizures
- Cutaneous allergic reactions have been seen, as well as gynecomastia, in some cases

These side effects most frequently appear in the first few weeks after treatment is begun, so especially careful observation of patients is required during this period.

MAOI are infrequently used in children and adolescents, although a new generation of these compounds with fewer side effects is currently under development. MAOI interact with various substances, particularly tyramine, which is present in various cheeses and many other foods. The inhibition of tyramine degradation can lead to major fluctuations of blood pressure (usually hypertension). Furthermore, there are interactions with various other foods. Thus the use of MAOI has been restricted. For the new generation of so-called reversible MAOI, the side effects listed here do not apply. Experience with these substances in child and adolescent psychiatry is still limited at present.

3.3 Lithium Salts

The major area of indication for lithium salts is still manic-depressive illnesses. These are very rare in children, but become more common in adolescence. Lithium is given prophylactically. It has been shown that lithium can prevent the reappearance of depressive or manic episodes, or at least postpone them. The effect is documented for both uni- and bipolar endogenous phasic illnesses. Lithium salts have also been used to treat various other mental disorders in children and adolescents, although there have been no large systematic studies. Positive results have also been described in aggressive behavior disorders with a phasic pattern and in impulse control disorders, for example. In such cases, however, an affective component should be also present if lithium salts are to be given.

The dose is gradually increased from 6 to 18 mmol/l. 7 days after the initiation of treatment, the serum lithium level should be determined 12 hours after ingestion of the last tablet. The dose is then increased until the blood level reaches approximately 0.8 mmol/l (maximum tolerated level, 0.8–1.2 mmol/l). Slow-acting preparations which need only be given twice per day are recommended (e.g. Hypnorex retard, Quilinum retard).

In addition to the obligatory provision of information to parents and patients, an ECG and an EEG should be performed in each case and a complete blood count should be obtained before lithium treatment is begun. Thyroid and liver function tests should also be performed. Because of the danger of goiter formation, the neck circumference should be measured and the dermatological examination documented.

Furthermore, tremor, weight gain, polyuria, nausea, vomiting, and diarrhea can occur. Cardiotoxic effects are rare. EEG occasionally reveals paroxysmal dysrhythmic disorders, more rarely seizures. Careful monitoring of medication to ensure that the lithium level remains in the therapeutic range is the best way to prevent these side effects.

3.4 Tranquilizers

Tranquilizers are substances of varying chemical structures that possess a sedating but not hypnotic effect and that have neither antipsychotic nor antidepressant effects at higher doses. They also act specifically against anxiety.

The following classes of tranquilizers are used in child and adolescent psychiatry:

- Meprobamate
- Diazepines (e.g. Valium)
- Lorazepam (e.g. Tavor)

Great care should be taken when prescribing tranquilizers to children and adolescents, not least because of the danger of dependency. Longer-term drug use (over many months) is justified only in exceptional cases.

As for indications, tranquilizers can be used, albeit temporarily, in almost all conditions accompanied by anxiety, tension, contact inhibitions, sleep disturbances, and compulsive symptoms. A particularly important indication is present in certain types of epilepsy, where diazepam and mogadan in particular are used (e.g. for myoclonic seizures).

3.5 Stimulants

Stimulants have a central stimulating effect and can lead to dependency in adults, while they can have a

positive influence on attention and motor functions in children. Recent reviews can be found in Kutcher (1997) and Nissen et al. (1998). The following stimulants are in use:

- Methylphenidate (e.g. Ritalin)
- Dextroamphetamie
- Pemoline

The chemical structure of the amphetamines bears a striking similarity to that of the neurotransmitters dopamine and norepinephrine. Methylphenidate and amphetamine affect catecholamine release through different mechanisms: methylphenidate causes dopamine release from reserpine-sensitive granules, while dextroamphetamine activates the release of newly synthesized cytoplasmic amines (reserpine-resistant). Both mechanisms of increased catecholamine release cause these neurotransmitters to flood the synaptic cleft and interact with postsynaptic receptors and presynaptic autoreceptors. In addition, amphetamines cause a dose-dependent blockade of the reuptake of dopamine, norepinephrine, and serotonin into nerve terminals. This mechanism, similar to the effect of tricyclic antidepressants on norepinephrine and serotonin reuptake, leads to increased concentration of the neurotransmitter in the synaptic cleft.

Stimulants are used in children almost exclusively to treat hyperactivity. The doses used are shown in Table 1.

In recent years, hyperactivity has been treated mainly with methylphenidate (e.g. Ritalin). Because of its short half-life, it must be administered twice daily (optimally in the morning and at midday). Methylphenidate should not be given in the evening, as it can adversely affect falling asleep or the quality of sleep. The medication should be temporarily discontinued on weekends and during vacations. This measure prevents the appearance of growth disorders.

The following side effects are known:

- Loss of appetite
- Sleep disturbances
- Tachycardia
- Lowering of the seizure threshold (methylphenidate is thus contraindicated in the presence of an increased seizure tendency)

Table 1. Range of doses of stimulants in school children

	Total dose (mg/kg body weight)	Individual doses (*n*)
Methylphenidate	0.20–1.0	1–3
Dextroamphetamine	0.15–0.50	1–3
Pemoline	0.50–2.0	1
Deanol	1.0–3.0	1–3

Various studies have shown that treatment with methylphenidate also leads to markedly better mother–child interaction and an improved teacher–pupil relationship. Because the children can concentrate better, they are also better able to meet scholastic demands. They are markedly less disruptive than when they were without medication. It should be pointed out that only about 60% of hyperactive children are responders; responses can be seen within a few days. If there is no response to medication, further administration is pointless. It is not yet clear what determines whether an individual will respond. Stimulants should be used only for clearly defined and unambiguously diagnosed hyperactivity. They are not a suitable remedy for difficulties at school or unspecified behavior problems. The dependency often feared does not occur with hyperactivity.

Stimulants are subject to regulations governing the prescription of narcotics and, in many countries, special prescription forms must be used.

4 Treatment Settings

The treatment setting (modality of treatment) is the framework in which treatment can be most successfully carried out. It must be decided whether to provide outpatient or inpatient treatment, day treatment, or home treatment. All of the methods mentioned above can be carried out in a number of different settings. The choice of setting should take two factors into account:

1. Empirical knowledge of the effectiveness of individual treatment methods in each setting
2. The possibility of establishing an adequate therapeutic relationship with the child and family

Outpatient therapy is most frequently performed, and there are scarcely any individual illnesses in which it is contraindicated. It is indicated both in predominantly organic illnesses and in psychogenic ones.

The usefulness of outpatient treatment is limited under the following circumstances:

- Suicidal behavior or other forms of self-endangerment
- Inability to provide effective treatment for very severe or chronified disorders
- Extreme family pathology, because of which a separation of the parents (or other significant adults) from the child is advisable

Practically all treatment methods that have been confirmed to be effective can be used in an outpatient setting.

The decision to provide inpatient treatment is generally guided by the following considerations:

- Severe and/or chronified illness
- Danger to self or others

Some treatment measures are possible only in specialized institutions. The most important types of institution for treating mental disorders in children and adolescents are the following:

- Outpatient care: child and adolescent psychiatrists and psychotherapists in private practice, institutional ambulatory care centers and outpatient clinics, child and adolescent psychiatric services, mobile child and adolescent psychiatric services with treatment responsibilities, child guidance and family counseling centers, early development centers, pediatric social welfare centers
- Partial hospitalization: day clinics for mentally ill and handicapped children and adolescents, with the option of overnight admission
- Inpatient care: university departments of child and adolescent psychiatry, government hospitals or hospital departments for child and adolescent psychiatry, departments of child and adolescent psychiatry in general or pediatric hospitals
- Complementary rehabilitative care: rehabilitation facilities for special groups of patients (e.g. children with severe traumatic brain injury or refractory epilepsy), halfway houses, group homes, homes of various other types

Not every type of institution listed above is present in every geographic region. Sometimes, the appropriate institution is far from the patient's home. Ideally, the choice of a particular form of therapy should be based on empirical evidence of its effectiveness to treat the disorder in question. Unfortunately, such evidence is still lacking in many cases.

It is often difficult in practice to base the choice of treatment on empirical data. Nonetheless, the therapist must consider whether the prospective treatment method is appropriate for the patient's age and developmental stage.

5 Legal Aspects

According to German law, children have legal capacity "upon completion of birth" (Sect. 1 of the German Civil Code – *Bürgerliches Gesetzbuch,* BGB); for example, they are able to take legal action (obviously represented by their parents or another adult). Until children reach majority, they are legally under the care of their parents (Sect. 1626 BGB). Before the age of 18, children and adolescents acquire partial legal rights in a series of steps, through which they gradually attain legal majority. The most important milestones are at the end of the 6th year (obligation to attend school), the 14th year (criminal responsibility), and the 18th year (full majority). Until age 7, children are legally incompetent (Sect. 104 BGB) and are unable to be delinquent, i.e. they cannot be held responsible for damages inflicted on another person. Between 7 and 14 years of age, they have limited competency (Sect. 106 BGB) and limited (civil, not criminal) delinquency (Sect. 828 (1) BGB). Adolescence begins at age 14, as does criminal responsibility.

In Germany, health insurance companies are responsible for treatment and rehabilitation, and children are usually insured along with their parents. There are a number of therapeutic measures, however, that are covered under other relevant laws. The most important laws are the Juvenile Court Act (*Jugendgerichtsgesetz*, JGG) and the Federal Social Welfare Act (*Bundessozialhilfgesetz*, BSHG). In addition, matters of family law (e.g. laws concerning parental care, divorce law, adoption law) are regulated in the German Civil Code.

The German Child and Adolescent Assistance Act (*Kinder- und Jugendhilfegesetz*, KJHG) lists the following youth services that may be used in addition to therapeutic measures:

- Assistance with child-rearing (Sect. 27 KJHG): The individual(s) having care and custody of the child are entitled to assistance with child-rearing "when a suitable upbringing for the welfare of the child or adolescent is not guaranteed, and assistance is appropriate and necessary for his or her development."
- Child guidance (Sect. 28): "Child guidance clinics and other counseling services and facilities are designed to support children, adolescents, parents, and other guardians in clarifying and coping with individual and family-related problems and the underlying factors that cause them, in the resolution of child-raising issues, and in separation and divorce."
- Social welfare group work (Sect. 29): This is designed to help primarily older children and adolescents "overcome developmental difficulties and behavioral problems."
- Educational supervisors, welfare helpers (Sect. 30): These persons help children and adolescents cope with developmental problems.
- Social and educational family assistance (Sect. 31): These measures serve to support families that are unable to perform their educational tasks adequately.

- Child-rearing in a day group (Sect. 32): This measure is intended to promote the development of children and adolescents through social learning in a group.
- Full-time foster care (Sect. 33): This generally involves the placement of children and adolescents in foster families.
- Care in a children's home, other forms of supervised accommodation (Sect. 34): For children who cannot be adequately helped by the measures already mentioned, care in a children's home or other supervised accommodation can be arranged.
- Intensive, individual socio-educational care (Sect. 35): This measure is essentially for adolescents who need particularly intensive support in order to be able to lead responsible lives.
- Temporary care of children and adolescents (Sect. 42): This is the temporary accommodation of a child or adolescent by an appropriate person, facility, or other supervised living arrangement.
- Options for participation (Sect. 36): Options for participation are addressed in Section 36 of the KJHG, which stipulates that an assistance plan should be drawn up. With regard to participation in the choice of measures, the law states that the guardians and the child or adolescent are to be counseled before youth services are instituted. Such counseling may be within the competence of the physician, particularly when social youth services are to be combined with medical treatment.

The Federal Social Welfare Act (BSHG) concerns "social assistance for living arrangements and support in special situations." The task of social assistance is to provide a wide variety of social services enabling the recipients to lead well-organized lives. In the context of therapy and rehabilitation, assistance with integration for the handicapped (Sect. 39 BSHG) plays a major role. Its objective is "to prevent an impending handicap or to eliminate or alleviate the consequences of an existing handicap, and thereby enable or facilitate the participation of the handicapped person in community life."

The Juvenile Court Act (JGG) applies to adolescents (aged 14–18) being prosecuted under criminal law. Sect. 10 section 2 JGG concerns reformatory treatment, which can be imposed by the court and is not restricted with respect to the measures to be taken. In other words, practically all treatment methods may be recommended, as long as they appear likely succeed. Reformatory treatment may also take the place of punishment. The expert in adolescent psychiatry faces important tasks in this setting.

Matters of family law are regulated by the German Civil Code (BGB). It is beyond the scope of this discussion to deal with these issues individually. Essentially, the following issues are involved:

- Regulation of custody when parents divorce (Sect. 1671, 1672, 1680, 1681 BGB)
- Regulation of visiting rights (Sect. 1634 BGB)
- Procedure when the child's welfare is endangered (Sect. 1666 BGB)
- Adoption issues (Sect. 1741-1772 BGB)

6 References

Dührssen A (1980) Psychotherapie bei Kindern und Jugendlichen, 6th edn. Vandenhoeck and Ruprecht, Göttingen

**Graham P (ed) (1998) Cognitive behaviour therapy for children and families. Cambridge University Press, Cambridge

Kielholz P (1971) Diagnose und Therapie der Depressionen für den Praktiker, 3rd edn. Lehmanns, Munich

*Kutcher SP (1997) Child and adolescent psychopharmacology. Saunders, Philadelphia

*Margraf J (ed) (2000) Lehrbuch der Verhaltenstherapie, vols 1 and 2, 2nd edn. Springer, Berlin Heidelberg New York

Mattejat F (1997) Familientherapie. In: Remschmidt H (ed) Psychotherapie im Kindes- und Jugendalter. Thieme, Stuttgart, pp 148–174

Niebergall G (1997) Gruppentherapie, Psychodrama und Rollenspiel. In: Remschmidt H (ed) Psychotherapie im Kindes- und Jugendalter. Thieme, Stuttgart, pp 134–147

Nissen G, Fritze J, Trott GE (1998) Psychopharmaka im Kindes- und Jugendalter. Fischer, Stuttgart

Prout HT, Brown DT (eds) (1999) Counseling and psychotherapy with children and adolescents. Wiley, New York

*Remschmidt H (1992) Psychiatrie der Adoleszenz. Thieme, Stuttgart

*Remschmidt H (ed) (1997) Psychotherapie im Kindes- und Jugendalter. Thieme, Stuttgart

Remschmidt H (ed) (2000) Kinder-und Jugendpsychiatrie. Eine praktische Einführung. Thieme, Stuttgart

Remschmidt H, Mattejat F (1998) Familiendiagnostisches Lesebuch. Enke, Stuttgart

Remschmidt H, Quaschner K (1997) Tiefenpsychologisch fundierte Psychotherapie. In: Remschmidt (ed) Psychotherapie im Kindes- und Jugendalter. Thieme, Stuttgart, pp 80–91

Remschmidt H, Schulz E (1995) Psychopharmacology of depressive states in childhood and adolescence. In: Goodyer IM (ed) The depressed child and adolescent. Developmental and clinical perspectives. Cambridge University Press, Cambridge, pp 253–279

Rosenberg DR, Holttum J, Gershon S (eds) (1994) Textbook of pharmacotherapy for child and adolescent psychiatric disorders. Brunner and Mazel, New York

Seiffge-Krenke I (1986) Psychoanalytische Therapie Jugendlicher. Kohlhammer, Stuttgart

*Steinhausen HC, von Aster M (eds) (1993) Handbuch Verhaltenstherapie und Verhaltensmedizin bei Kindern und Jugendlichen. Beltz, Weinheim

CHAPTER

7

H. Remschmidt, M.H. Schmidt

Disorders in Child and Adolescent Psychiatry

Translators: E. Taub, S. Henn

1 Introduction

We use the term "disorders" in this chapter in accordance with the international trend, reflected in both of the current systems of classification (ICD-10 and DSM-IV), toward using this term and avoiding the term "disease." It remains to be seen how long this change will last. The justification for this change in the current classification systems is that the concept of a disorder carries fewer etiological and theoretical implications than that of a disease. The decision not to speak of diseases is not intended to be final; it merely reflects the fact that current knowledge is too limited to allow an etiologically-based classification (Remschmidt 1988).

Until this is possible, there will remain several different ways to describe concepts of diseases or disorders in child and adolescent psychiatry:

1. Definition of different levels on which concepts of disorders can be placed, in the form of constructs. This was the approach of Häfner (1983).
2. Definition of relatively broad areas in which pathological organic changes, behavior, and experience occur. Such a suggestion was made by one of the pioneers of European child and adolescent psychiatry, Moritz Tramer (1942), who classified diseases into 4 groups according to the prevailing state of knowledge at that time: (1) somatic disorders and illnesses, (2) somatopsychic disorders and illnesses, (3) psychosomatic disorders and illnesses, and (4) mental disorders and illnesses. It is of interest that Tramer distinguished between disorders and illnesses, thus anticipating current concepts to some extent, although he defined the concept of disorder less precisely than that of an illness.
3. Orientation toward developmental aspects and considerations of disease course. We have adopted this point of view in our own scheme of classification in this chapter, in that we do not regard "disorder" and "disease" as mutually exclusive entities but rather view disorders as the preliminary stage of diseases. Thus, following Häfner (1983), we define a disorder in child and adolescent psychiatry as "a condition of involuntarily disturbed life functions, which has a temporal dimension in terms of its onset, course, and (sometimes) termination, and which definitively prevents a child or adolescent from actively participating in, and mastering, age-appropriate activities" (Remschmidt 1988, p. 146).

In view of the considerable improvement of our understanding of the development and course of disorders and illnesses in child and adolescent psychiatry over the last two decades, we believe we are justified in using the last-described point of view as the basis of a classification of these disorders and illnesses. In consideration of developmental aspects and of disease course, we classify them into six types:

1. *Behavioral abnormalities* are deviations from normal functioning in various areas over the course of development (e.g. autonomic functions of attachment behavior, language development, social development), which are not yet severe enough to be designated as disorders or illnesses, but whose detection is nonetheless important, so that preventive measures can be taken and parents and significant others can be appropriately counseled.
2. *Early-onset disorders with lasting developmental impairment* include a number of etiologically diverse disorders that begin in the early and middle years of childhood and result in lifelong impairment, or at least have demonstrable effects lasting into the middle years of adulthood (up to approximately age 30).
3. The characteristic feature of *developmental disorders* is their dependence on the developmental and maturational processes of the early and middle years of childhood, approximately until puberty. As maturation progresses, they become significantly less common.
4. *Disorders of age-specific onset* are typified by their onset in a specific age period. The typical course of disorders of this type is different for each disorder.
5. The hallmark of the *developmentally dependent interaction disorders* is that they occur in the context of interpersonal interactions that are specific to certain age periods and levels of development.
6. *Early-onset adult-type disorders* are disorders that are typically found in adults but often begin in adolescence, generally continuing into adulthood.

Four of the six types of course are depicted schematically in Fig. 1. Behavioral abnormalities and developmentally dependent interaction disorders are difficult to depict in this way, and have been left out of the figure.

In what follows, we classify mental disorders in childhood and adolescence according to these 6 types of course. A comprehensive review of the subject is not possible, because of space limitations. Our discussion must, therefore, remain on an illustrative level. We hope that it will serve to orient the reader in the field and provide the key to a deeper understanding.

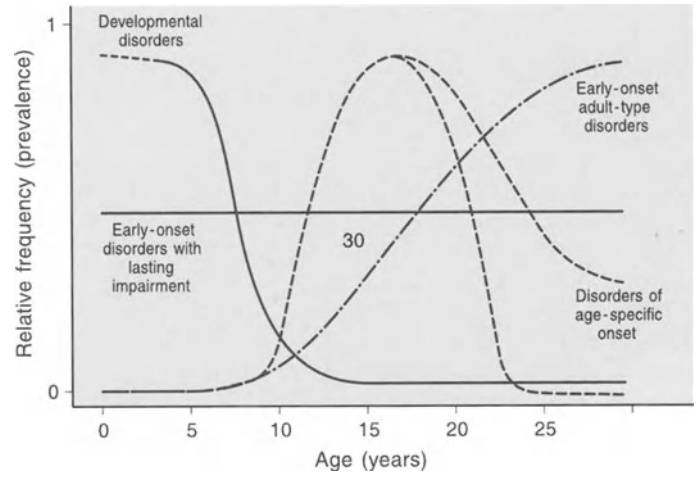


Fig. 1. Types of course of mental disorders in childhood and adolescence

2
Behavioral Abnormalities and Stress Reactions

Changes in physical and mental function and behavior that occur together at certain times in the life history of the individual constitute a process retrospectively designated as development. Three major types of factors participate in development: maturational processes, adjustment processes, and control by the individual.

Many developmental processes are set in motion by maturation and then come increasingly under the influence of learning and adjustment processes or of individual control. Genetically based processes of maturation are modified by the biological environment, and, in particular, by somatic disorders acquired at an early age; limitations of central nervous function play a major role in these processes.

Adjustment processes are set in motion mainly by social learning: the environment generally places demands on the individual, who must then adjust his or her behavior accordingly. Repeated accommodation results in changes of behavior. The demands of everyday life may be perceived as stressors of variable intensity (physical illness, separation, loss of significant others, scholastic demands).

Thus, even at an early age, children develop coping mechanisms whose biological correlates have as yet been the subject of little research (e.g. Lundberg 1986). There is no doubt that these mechanisms are influenced, in part, by the intensity of the stressful experience. Behavior under the individual's own control is easily observed even in 2-year-olds: children behave differently in different contexts.

Individual developmental factors interact with each other in a manner that is not merely additive. Thus, the relative importance of each will be of lesser concern to us here than the interactions between them and the mechanisms underlying them. Biological processes are heavily involved and determine, among other things, the range of possible reactions in the framework of learning processes. Life experiences make a further contribution to coping styles and to the development of defense mechanisms. The individual creates his or her own personal environment through variable behavior in varying contexts (Bouchard and McGue 1990). The greater importance of the individual environment, which a child does not share with his or her siblings, as compared to the common environment is one of the reasons why genetic influences become increasingly manifest over the course of development. The activity of the individual is a more important cause of diversity in behavior than the variability of the environment. Maximal diversity arises in functions that develop rapidly, because some children develop these functions earlier, and others later.

Genetic predispositions, differing opportunities and demands in the social environment, active behavior, and varying speeds of development create variations in behavior whose extremes lie near, or may overlap with, pathological behavioral abnormalities. The differential diagnosis of such variations from incipient psychiatric disorders may be difficult. The experienced clinician relies more on qualitative than on quantitative abnormalities for diagnostic decision-making.

Typically, however, behavioral variants are characterized by a temporal lead or delay compared with normal behavior, or are extremely deviant in their intensity. They are not, in themselves, a sign that mental disorders are present, but they do constitute a risk factor for them: speech dysfluency of physiological origin, for example, may appear to other persons to be a manifestation of deficient self-control, and an attempt may be made to correct it forcibly, with the

result that overt stuttering develops. Girls whose menses begin early have an elevated risk of contact with older male adolescents, which, in turn, elevates their risk of consuming hashish at a young age (Magnusson et al. 1986).

2.1 Motor Development, Autonomic Functions, and Sexual Development

The assessment of the psychomotor development of prematurely born children must be corrected for gestational age. In the third year of life, the motor overactivity characteristic of that age is so common that it may be confused with the manifestations of a hyperkinetic disorder. Motor coordination is well enough developed in the fourth year of life that the child becomes able to participate in games of skill. Nonetheless, motor function is not yet under complete control: complicated, selective movements are often accompanied by involuntary accessory movements even into early school age. Persistent ambidexterity is considered a sign of late maturation.

Sex differences are apparent in the motor development of school-age children. Strength and skill are important sources of social recognition by peers for children of both sexes – for girl aged 11 years and older, personal appearance is another important source. There is a wide spectrum of motor control of speech: stammering is found in one quarter of all five-year-olds and must be diagnostically differentiated from auditory disturbances and sound agnosia. Speech dysfluency is common in the third and fourth years of life, more so in boys than in girls. Motor development determines the pace of acquisition of drawing and writing skills until age 9; in later school years, the speed of writing is determined mainly by temperamental factors, and, after puberty, by the influence of practice. Stereotyped motor patterns may still reflect immature inhibitory processes in the fourth and fifth years of life.

Young children vary widely not only in the duration of sleep, but also in their behavior while falling asleep. At least one third of four-year-olds display stereotypic movements before falling asleep, need transitional objects, and climb into their parents' beds when they wake up at night (Klackenberg 1987). Resistance to certain foods, or to being fed, are more likely to be an expression of a mother-child conflict than of early eating disorders. Vomiting is not uncommonly a habituated reaction to excitement in irritable infants.

Overweight is found in 13.5% of 5-year-olds, obesity in up to 2.5%. Overweight 5-year-olds are likely to have been overweight babies, although only a small fraction of overweight babies go on to become obese children. Eating behavior does not become stable until school age (Woolston 1991). The only behavioral variants that are considered to occur in the first 6 months of life are excessive crying, disturbances of falling back asleep, and/or isolated feeding disorders. Bladder and bowel control are usually achieved by the age of 3 years, but the range of normal variation is so wide that the diagnosis of enuresis cannot be made before the age of 5 years. A transient reappearance of enuresis under stress may represent an adjustment reaction (see below).

The acceleration of longitudinal growth that marks the onset of puberty occurs in girls aged 10 to 11 and in boys aged 12 to 13. The high point of puberty occurs approximately 2 years later in each sex, with a corresponding time difference between the sexes; the tempo of puberty, however, is subject to variation. Sexual fantasies and masturbation appear in boys at about the age of 13, in girls at the age of 15 on average. By the age of 18, 50% of male and 60% of female adolescents have experienced coitus; this implies a shift toward earlier sexual experience among girls.

2.2 Cognitive Functions and Language

Practice in differential perception is important for the development of intellectual function. The preference for color that prevails until the 6th or 7th year of life is increasingly replaced by a preference for form. In the seventh year, stability of spatial location is achieved, both in perception and in reproduction; mirror-writing still occurs in the first grade, however, especially among left-handers. The concept of number develops in the fourth year, while spatial orientation remains dependent on the child's own position until the early school years, i.e. until reversible cognitive processes come into use. As for the concept of time, an idea of the future develops in the fifth year, before the differentiation of the past; a temporal continuum comes about only in school age. Perceptual illusions are a common symptom of fever until early childhood.

The attention span lengthens between the ages of 5 and 7 years but remains highly dependent on motivation and context. Individual reaction times become shorter until age 13. Simultaneous memory begins in the third and fourth years, while ordered reproduction can be performed only in school age; it is supported by language development. The beginning of school age is accompanied by a massive expansion of short- and long-term memory. Model learning is possible as early as the first year. School education leads to a verbal orientation of learning, at first as multiple discrimination, and later in the form of concept and rule learning, and of problem solving.

Remnants of the magical world view of the preschooler are still recognizable as late as the seventh year; corresponding rituals must sometimes be differentiated from early obsessive-compulsive phenomena. Conceptual thinking is possible after the fourth year, and simple, reversible thought processes are present in the seventh year, while, at the same time, the progressive building up of concepts is a stimulus to causal thinking. The level of intelligence and life experiences exert an important influence on the acquisition of wide-ranging cognitive abilities. A major portion of cognitive variance is already accounted for by the eighth year of life. From the tenth year onward, there is a relative stability of intellectual performance until adolescence; before the tenth year, measurable intellectual performance is more strongly milieu-dependent.

Nonverbal functioning is more suitable than language-bound intelligence for use as a criterion to distinguish a temporary developmental lead from special intellectual talent, because the frequency of vocal expression is correlated with social class even before the sixth year of life. Passive vocabulary develops earlier than active, and language comprehension earlier than language production. Language develops earlier in girls, and with rarer disturbances. Grammatical correctness is largely achieved in the fourth year, and the vocabulary trebles in size while the child is in kindergarten. From the fourth year onward, language serves to communicate thought processes in a manner independent of visual context. Reading may normally be acquired as late as age 8. General intelligence is more closely correlated with language comprehension than with language production. By the end of the primary school years, an understanding of double meanings is acquired, and children can use language to express how they feel.

2.3 Emotions and Motivation

In terms of mental development, displeasure reactions appear earlier than pleasure reactions, and they also become differentiated earlier. The differentiation of the emotions depends at first on autonomic processes, later on imitative processes. Emotional attachment to significant others favors exploration, which occurs independently of significant others in the third and fourth years of life.

From the end of the first year onward, the child reacts sensitively to separation, because the concept of time is still lacking. As perception becomes differentiated, the emergence of anxieties is favored, first by conditioning, and later by model learning. Improvements in cognitive structures also enable the child to have false ideas of danger, which may be reinforced verbally; conversely, the infant's fears can be neutralized by the presence of its mother. Separation requires practice – for example, in the context of going to sleep.

Defiant reactions in the second year of life represent the young child's active attempts to deal with the environment. Because affect regulation is not yet fully developed, these reactions may escalate to the level of temper tantrums. Long, continued crying may turn into respiratory affective convulsions with cyanosis, inspiratory pauses, and ensuing disorientation. By the age of eight at the latest, children can say how they feel more accurately than their parents can. Anxious affect develops earlier than depressive affect, though the former is often inappropriate to the real-life situation, in which no threat is actually present. The greater timidity of girls decreases continuously from the age of 10 onward. As children approach adolescence, they discuss their feelings more readily with their peers, girls more than boys; at later ages, such discussions also take place in groups of mixed sex.

Problems with poor school performance become less common with adolescence, but they may emerge once again in a manner related to social class. Uncertainties about the future and about self-esteem are not the main cause of death wishes and suicidal fantasies, which occur at least once in one fourth to one third of all adolescents. The correlation between such fantasies and suicidal thoughts in the narrower sense is rather low.

Curiosity is one of the child's earliest observable motivations. Motivation to perform, in the sense of competitive behavior, presupposes the cognitive ability to distinguish measurable effects of one's actions. Orientation toward success and avoidance of failure are recognizable as early as the fifth year, and children place demands on themselves from the fourth year onward. The five-year-old is aware of his or her own failure, and also of the different reactions that success and failure generate. The motivation to perform develops in a class- and milieu-dependent fashion, and is optimal when the external expectation always lies just below the limit of what the child can achieve. Even before school age, it becomes unnecessary to reward good performance immediately, but performance-oriented games come to be preferred around this time. When the magical world view disappears, children want to influence their own success. The developmental task of the primary-school child is to replace person-oriented with object-oriented performance motivation. Motivation can be maintained by an attitude that is as demanding as it is supportive. Lasting motivation to perform can protect the individual from the difficulties of orientation typical of adolescence and eases the fulfillment of educational demands.

2.4 Contacts, Conduct, and Self-Image

Shy behavior in school age may reflect either a lack of social competence, or social hypersensitivity (see Sect. 4). Both shy and socially hypersensitive children generally have no difficulty with attachment and therefore want contact with others, even if they want to determine its extent and quality themselves. Normal attachment behavior may be observed in grief reactions to separation (Ainsworth 1978). Securely attached children have an egocentric perspective that recedes only in the sixth or seventh year, in favor of an interactive perspective. When the "difficult" age is past, limits to self-determination are experienced and internalized in this way.

The boundaries separating the self from other persons and things, which are perceived in a merely physical sense in early childhood, come to be perceived mentally during early adolescence. A concept of self arises out of the uniformity and continuity of the individual's own experience, while the adoption of a perspective, something already learned in childhood, now becomes emotionally deepened. Conceptions of self are differentiated to a varying extent; Evans et al. (1994) pointed out the lesser degree of differentiation in adolescents with mental disorders. Self-evaluation begins in the primary school years. A high degree of experienced agreement between self-image and others' image of oneself strengthens self-esteem and leads to self-acceptance (Robson 1988). A high degree of self-esteem enables a more distanced perception and evaluation of the behavior of others, because adolescents with high self-esteem do not have to mobilize defense mechanisms to the same extent.

In a study of defense mechanisms (Perry and Cooper 1989), a distinction was drawn between mature and immature mechanisms. In the context of social development, self-acceptance obviates the need for immature defense mechanisms and thereby permits the establishment of mature moral judgments (Kohlberg 1997), which respect the interests of others and accord with overarching principles of behavior.

The clinically relevant stress reactions described below may be understood as reflecting an overloading of usual coping strategies. The highly affective tone of the acute stress reaction reflects a failure of affect control; the re-experiencing of traumatic situations in post-traumatic stress disorder reflects a failure of reality control; and the temporary inadequacy of coping mechanisms seen in adjustment disorder reflect the excessive stress placed on them.

2.5 Acute Stress Reactions

Even when extraordinary stresses do not lead to the appearance of a disorder corresponding to the particular vulnerability of the individual, they can still have a directly traumatizing effect. Successful coping with the stressful situation does not take place immediately. The child or adolescent is temporarily limited in his or her ability to go about the activities of daily living. The most important factor is the severity of the transient, triggering event, which is generally of a negative character for the involved person, or for another person close to him or her. Such events include natural disasters, accidents, wartime experiences, crimes, and other experiences of loss. Physical exhaustion or chronic illnesses heighten the risk of an acute stress reaction.

The severity of the stress reaction depends both on the stressful event itself and on the individual's coping mechanisms, as becomes clear when one considers the individual nature of these reactions. They generally begin with a diminution of consciousness or attention, which results in stupor or agitation. Because the individual is disoriented, stimulus processing is not possible, or takes the form of hyperactive reactions. Possible affective reactions include anxiety, despair, and depression, as well as anger and withdrawal. The manifestations usually fluctuate, and accompanying autonomic anxiety reactions (flushing, tachycardia) are common. Intervention is necessary only when self-injury or suicidal actions are likely to occur.

2.6 Post-traumatic Stress Disorders

Post-traumatic stress disorders appear after a long delay, or gradually over a long period, with latencies of weeks to months, or rarely even longer. For persons under 18 years of age, the lifetime prevalence of such disorders is said to be as high as 6% (Giaconia et al. 1995). Older children are more severely affected than younger ones, and, from pre-adolescence onward, girls are more commonly affected than boys. Increased autonomic excitability, anxiety disorders, and anxious or obsessive-compulsive personality structures favor the development of post-traumatic stress disorders, as does repeated traumatization (Garrison et al. 1993). Successful coping with earlier trauma does not seem to have a protective effect. Close family ties, however, and other compensatory mechanisms are considered to be protective, as is crisis intervention immediately after the stressful event.

As in adult patients (Yule 1992), this disorder is characterized by repeated recollection of the inciting event, or by a re-experiencing of the stressful emotions associated with it, in response to key stimuli. The situation of recollecting the event, and analogous constellations, are avoided in a phobic manner in order to prevent the occurrence of the accompanying vegetative phenomena, which may include excessive alertness, fearfulness, difficulty concentrating, irritability, hyperactivity, and sleep disturbances. Depressive and dissociative reactions may also occur. The intensity of recollection is independent of the immediate reactions.

Manifestations of the disorder among children, in particular, include fear of the dark and separation anxiety (even when the child sleeps in the parents' bed), failure in school, irritability towards peers and adults, regressive manifestations, difficulty talking about the event with trusted persons (for fear of making them sad), fear of threatening situations for significant others, and, sometimes, feelings of guilt at having survived the event. Information provided by the children themselves is of greater diagnostic relevance than that obtained from their parents. It is typical for the recollections to be experienced as forcing themselves upon the sufferer. The diagnosis is made only if the manifestations of the disorder arise within 6 months of the event. Panic reactions may also occur.

In most persons suffering from post-traumatic stress disorder, the manifestations become milder over time; a transition to a chronic personality change, as occurs after repeated, severe traumatization (for example), is rare. Post-traumatic stress disorder among children is more likely to become chronified if the patient is a girl, if the underlying trauma involved personal threat or experiences of interpersonal violence, or if the coping behavior of significant others is inadequate. Its treatment consists of repeated discussion or reenactment of the stressful event. Resistance preventing psychotherapeutic access to the traumatic event is only transiently operative. A feeling of control over the event, and the anxiety and excitement associated with it, is obtained by talking about it in an anxiety-free atmosphere.

Severely traumatized children are, therefore, treated individually. When the traumatic experience was a shared one, discussion in a group of affected persons is helpful. Treatment in a family-therapeutic setting is reasonable when several members of the family have been traumatized, or when a high degree of family resistance prevents treatment of the child alone. Trauma interviews (Pynoos and Eth 1986) may be helpful in reviving the event. Children are given propranolol in some cases for pharmacological support, while older adolescents may be given fluoxetine, which has been found useful in adults. A purely cognitive approach to the event, without reliving the associated affect, is not an effective form of treatment.

2.7
Adjustment Disorders

Adjustment disorders generally arise not because of dangers inherent in the present situation, but because of lasting, negative changes in life circumstances. Individual predisposition, or the material and social environment, may also be partially responsible. Manifestations of these disorders include depressive behavior, which may be mixed with anxiety, worry, or tension; in children, regressive phenomena such as the reappearance of bedwetting or a reversion to baby talk; and, in adolescents, aggressive or antisocial behavior. One or more of these manifestations generally appear within one month of the event and last as long as six months; depressive reactions may last even longer. In the Puerto Rico Study (Bird et al. 1988), adjustment disorders had a frequency of 4.2% and accounted for one quarter of overall morbidity. Their treatment is based on their most prominent manifestations in the individual patient, with the application of cognitive restructuring. Their prognosis is generally as favorable as that of stress reactions requiring no treatment.

3
Early-Onset Disorders with Lasting Impairment

3.1
Mental Retardation

Mental retardation (ICD codes F70-F79) is a congenital or acquired limitation of cognitive performance. Its nature and intensity are highly variable. The contrasting concept of dementia refers to a deterioration of mental functioning, i.e. it presupposes that a high intellectual level was present until a particular event occurred (e.g. encephalitis or epilepsy).

These disorders are to be distinguished from disorders of impaired intellectual capacity in children resulting from extreme neglect and lack of encouragement, which are designated as mental hospitalism, or deprivation syndrome.

One may assume that mental retardation of varying degrees of severity is present in ca. 9–10% of the general population. The more severe the retardation, the more rarely it is found. Severe and very severe mental retardation are generally caused by diseases or injuries of the brain, malformation syndromes, inherited metabolic anomalies or chromosomal aberrations. In the statistical distribution of the intelligence quo-

tient (IQ), the prevalence rates at the lower end of the distribution do not correspond to the values that would be expected on a Gaussian curve; the lower grades of intelligence are significantly over-represented. This is explained by the fact that mental retardation may be the result of a combination of genetic and organic causes (Zigler 1967). The mean IQ of patients with organic mental retardation is approximately 35.

The most important classifications of mental retardation and dementia are based on severity, the causes of the intellectual impairment, and the opportunity for improvement (reviewed in King et al. 1997; Neuhäuser and Steinhausen 1998; State et al. 1997).

Varying levels of intellectual capacity can be measured with the intelligence quotient, which deals with the entire spectrum of intellectual capacity, and not just with retardation. These varying levels are listed in Table 1 according to the multiaxial classification scheme (MAS) for mental disorders of children and adolescents according to the ICD-10 (Remschmidt and Schmidt 1994). Brief descriptions of the most important types of mental retardation are provided.

In the MAS, mental retardation is assigned to a separate, third axis of classification independent of psychiatric illnesses and specific developmental deficiencies. The DSM-IV also allows a multiaxial classification, in which intellectual impairments are assigned to the second axis.

Because classification by IQ is inadequately descriptive, mental retardation is also broken down into two types that are distinguished by differing ability to learn in school:

- learning difficulty (IQ ca. 50–80), and
- mental handicap (IQ ca. 30–55).

The classification of mental retardation by etiology is problematic, in that the etiology can be determined in fewer than 30% of cases. Furthermore, the elucidation of the etiology depends on the severity of the mental retardation. More severe cases tend to be associated with organic insults, malformations, metabolic disorders, and other somatic abnormalities. An extensive review is found in Neuhäuser and Steinhausen (1998).

The diagnostic evaluation of mental retardation and dementia has the following objectives:

- Discovery of the cause of the disorder. A precise determination of the etiology can lead to specific treatment recommendations.
- A differentiated description of the pattern of the disorder, which is helpful in the formulation of an individual treatment plan and in predicting the prognosis. A thorough psychological assessment is used to meet this objective. Treatment measures are accordingly based, not just on global IQ scores, but on the individual patient's profile of abilities and skills.
- Acquisition of important information for genetic family counseling. Genetic counseling is provided to the children themselves when they reach reproductive age, and to their parents and siblings.

Mental retardation must initially be differentiated from dementia. The most important distinguishing feature of dementia is a measurable decline in the level of intellectual functioning, which may have very diverse causes. Its origin may lie in the progression of an underlying illness, such as Rett syndrome or Heller's infantile dementia, the etiology of which is still unknown. Mental retardation must also be differentiated from autistic disorders, epilepsy, and Landau-Kleffner syndrome (acquired aphasia with epilepsy).

Because the treatment in most cases must remain at the symptomatic, rather than the causal, level, the treatment plan must be directed toward the areas that have the greatest potential for improvement and are associated with the greatest impairment in the patients' everyday lives. Therapeutic interventions generally cover the following areas:

- *Functional therapy* is indicated when there are motor disturbances or special deficits.
- *Behavioral therapy* is performed to reduce or eliminate accompanying behavioral abnormalities (e.g. self-destructive behavior). It is also used successfully to further the goal of personal independence (e.g. in dressing, eating, shopping, self-assertion).
- *Pharmacotherapy* may be used to treat transient, severe behavioral abnormalities, or to treat marked hyperactivity and the associated attention disorders.
- *Counseling of parents and the social environment* is beneficial for children and adolescents in many ways. Its most important objective is to help persons in the social environment achieve a realistic attitude toward the intellectually handicapped child or adolescent and learn how to deal with him or her appropriately.
- *Support* in school, in vocational training, and in obtaining a job is of central importance.

With respect to the components of the rehabilitation of the mentally retarded, it is useful to retain the distinction drawn in the international classification scheme of the WHO between "impairment" (the mental and physical problems themselves), "disability" (the resulting functional limitations), and "handicap" (the negative social consequences). As the impairment is usually fixed, rehabilitation measures are generally oriented toward the remediation of disability and handicap. A series of graded interventions are provided by institutions specializing in the rehabilitation of

Table 1. Variations in intellectual capacity (MAS – ICD-10; Remschmidt and Schmidt 1994; WHO 1996)

Level of intelligence (IQ)	Verbal Classification	Brief Characterization
Very high intelligence (IQ > 129)	Intelligence far above average	
High intelligence (IQ 115–129)	Intelligence above average	
Normal range (IQ 85–114)	Average intelligence	
Low intelligence (IQ 70–84)	Below average intelligence, borderline mental subnormality	These children have mildly impaired intellectual capacity but are able to function independently in everyday life and often complete primary and secondary school.
Mild mental retardation (F70) (IQ 50–69) (in adults, intellectual age of 9 to <12 years)	Mild mental subnormality	Children with intellectual capacity in this range generally attend a special school for the learning-impaired. Their practical intelligence is usually better than their theoretical intelligence. They can carry out concrete cognitive operations and acquire the basic elements of scholastic skills.
Moderate mental retardation (F71) (IQ 35–49) (in adults, intellectual age of 6 to <9 years)	Imbecility	Children with intellectual capacity in this range can usually be taught appropriately only in special schools for the practically educable mentally handicapped. Their mental/emotional development is far behind that of non-handicapped children. They cannot acquire scholastic skills.
Severe mental retardation (F72) (IQ 20–34) (in adults, intellectual age of 3 to <6 years)	Severe oligophrenia	Some of these children can still be taught in special schools for the practically educable, but this presents considerable difficulties for many children, not least because of additional impairments that are often present (e.g. paralysis, deformities).
Profound mental retardation (F73) (IQ < 20) (in adults, intellectual age of <3 years)	Idiocy	These children usually have an extremely limited learning ability. In many cases, they cannot walk, eat independently, or speak. They tend to display stereotypic movements and primitive reactions. The developmental age they can achieve corresponds approximately to that of 18-month-old infants. Their intellectual functions are restricted to the sensorimotor level. They cannot make their learning more efficient by means of language.

the severely retarded with the goal of integrating them in the workplace and in society. Further information (in German) is found in the guide *Rehabilitation Behinderter* ("Rehabilitation of the Handicapped") published in 1994 by the German Federal Task Force on Rehabilitation.

Prenatal diagnosis and genetic counseling are means of primary prevention of mental retardation, as are dietary measures for the prevention of severe metabolic disturbances (e.g. in phenylketonuria or hypothyroidism). Continuous and meticulous prenatal care is also of obvious importance, both to prevent injury to the developing fetus and to provide early warning of possible complications, so that suitable plans can be made for the remainder of pregnancy and delivery.

The course and prognosis of mental retardation depend on its type and severity, on possible progression of the underlying disorder, and on the kind of support provided. Recent years have witnessed a change of attitude favoring greater integration of the mentally retarded into the community. This change and the increased availability of workshops for the handicapped and other "protected" work environments has led to a major improvement in the quality of life of mentally retarded young people.

3.2 Autism

3.2.1 Early Childhood Autism

The most prominent manifestation of early childhood autism (ICD-10 classification, F84.0) is a major disturbance of interpersonal contact whose first signs appear in the first months of life, although they are still difficult to recognize at this time. The children are delayed in their emotional and motor development, fail to make eye contact, and treat significant others with resistance and rejection whenever they try to make

contact. Further manifestations include an anxious clinging to the familiar (fear of change), various language abnormalities (delayed language development, echolalia, pronoun reversal, ungrammatical speech), and often abnormalities of the voice and of speech melody, as well as a number of other behavioral abnormalities, commonly including obsessive–compulsive behavior, self-destructiveness, aggressive and impulsive behavior, and a lack of fear in the presence of real danger. A comprehensive review is given by Cohen and Volkmar (1997).

Studies of unselected samples have revealed that the prevalence of early childhood autism is approximately 4–5 per 10,000 children and adolescents, and that boys are more commonly affected than girls. The sex ratio is in the range of 2:1 to 3:1.

Over the course of development, the manifestations of the disorder change in a somewhat predictable fashion. The predominant manifestations in early childhood, such as sensitivity to noise, paroxysms of fear, psychomotor unrest, sleep disturbances and a tendency to touch people or objects, steadily decline in intensity (Weber 1970, 1985).

Factors thought to play a role in the pathogenesis of early childhood autism include the following:

1. The findings of family studies and twin studies imply that *genetic factors* play an important role. Family studies on the frequency of early childhood autism among the siblings of autistic probands have revealed that they have an approximately 3% risk of developing the disorder, which is 60 to 100 times higher than the risk in the general population. Twin studies have shown that the concordance rate for monozygotic twins is far higher than that for dizygotic twins.
2. There have been many observations of *structural and functional abnormalities of the brain* in children with early childhood autism. These include a number of neurobiological abnormalities (e.g. disturbances of the sleep rhythm, eating disorders, abnormal crying, elimination disorders, excessive excitability) that are not found in normal children. Moreover, approximately 30% of all children with early childhood autism later develop epileptic seizures. Early childhood autism is associated with a number of other neurological syndromes more often than would be predicted by chance. Structural anomalies of the brain and abnormalities of cerebral metabolism have also been reported in a fairly large number of cases.
3. The hypothesis of a *disturbance of cognitive and emotional development* is based on the observation of Leo Kanner (1943) that children with early childhood autism have an "innate inability to form the usual, biologically provided affective contact with other people" (ibid., p. 250). This basal deficit makes such children unable to achieve further normal progress in development. They thus fail to learn at an early age, as normal children do, that other persons, too, have affective relationships, ideas and thoughts that one can sometimes infer from their behavior, and into which one can even, at times, project oneself.

The diagnosis of early childhood autism is made on the basis of the history and clinical observation. Aids to diagnosis exist in the form of scales with diagnostic criteria.

The differential diagnosis of early childhood autism from childhood schizophrenia is of great practical and clinical importance. The disorder must first be differentiated from autistic personality disorder (Asperger syndrome). It must then be differentiated from Rett syndrome, which is characterized by a progressive decline resulting in dementia. The possibility of sensory deficits and oligophrenia must also be considered. When speech peculiarities are present, audimutism (developmental dysphasia or aphasia) must also be ruled out.

The guiding principle of therapy, intervention, and rehabilitation in children with early childhood autism, as experience has amply confirmed, is that strongly behaviorally oriented, direct and structured treatment methods are more likely to succeed than "laissez-faire" methods that leave the patients to their own devices.

Every treatment plan must be based on a consideration of the developmental profile of the individual autistic child and be "custom-tailored" to it, to a certain extent. From this point of view, the following measures have been found useful and productive:

1. *Early intervention* is indicated in all autistic syndromes. Autistic syndromes should be diagnosed early, and development-promoting measures should then be instituted, with the aim of positively influencing the development of autistic children, in the sense of making their behavior more like that of healthy children.
2. Many *behavior-therapeutic methods* are used, ranging from operant conditioning and the use of reinforcements and aversive stimuli to "prompting" (the provision of specific help), "shaping" (i.e. of behavior) and "fading" (a gradual withdrawal of prompting).
3. A number of different *body-based treatments* may be used; these are a very heterogeneous group of methods whose theoretical justification and empirical confirmation remain controversial (for a review, see Weber and Remschmidt 1997):
   - *Developmental stimulation (after Delacato)* is intended to bring the children up to pace with missed developmental steps and bears the danger that the extent and intensity of the treatment will

overtax the children, their parents, and other family members.
- *Integrative physical therapy* is intended to promote the child's overall personality through the use of the body. It emphasizes the importance of emotional experience (communication of warmth, security, and understanding).
- *Holding therapy* is based on the idea that autistic children lack primal trust and tries to reestablish it by body contact (holding), even when the children resist (Tinbergen and Tinbergen 1984). Despite several encouraging case reports, the effectiveness of the method cannot yet be regarded as proved.
- *Facilitated communication* is a therapeutic measure that is designed to help people with severe disturbances of communication express themselves without speaking. This is done with the aid of letter tables or computers. The therapist can support the writing hand, or the arm. Most studies performed to date have failed to demonstrate the efficacy of facilitated communication (Howlin 1997).

4. We may use the term crisis to designate a situation that lies outside the range of the experience and behavior of the individual up to that point in time and that cannot be resolved by any available strategy. *Crisis intervention* is the combined use of all suitable measures to end the crisis, or at least eliminate the acute danger. Crisis intervention measures must be applied quickly and specifically and are distinct from long-term methods of treatment and rehabilitation (Weber and Remschmidt 1997).

The most important indicators of prognosis are the state of language development and the level of intelligence in the fifth and sixth years of life. Autistic children who have developed language relatively well up to this point and are of relatively high intelligence (IQ > 80) have a more favorable prognosis, as long as further complications, such as epileptic seizures or other neurological problems, do not arise.

3.2.2 Atypical Autism

In children with atypical autism (ICD-10 classification: F84.1), some, but not all, of the characteristic features of early childhood autism are observed. The disorder is frequently not present from birth, but rather develops after an organic illness in the first few years of life (e.g. encephalitis). Most of these children are mentally retarded and display so-called autistic features, i.e. features also found in early childhood autism. These include: delayed language development, echolalia, delayed motor development, stereotypic movements, (often) attention disorders, and, to a variable extent, disturbances of social communication. Many of these children also have neurological abnormalities, which are often attributable to a concurrent neurological illness or to the residua of a past neurological illness. This subject is reviewed by Cohen and Volkmar (1997).

The treatment consists of two components: treatment of the underlying illness, when there is one, and modification of autistic behavior. If the underlying illness is progressive, as in tuberous sclerosis or other chronic progressive illnesses, effective treatment is scarcely possible. The behavioral-therapeutic approach to the autistic manifestations must also take the underlying illness into account. If additional behavioral abnormalities are present, neuroleptic medication may be required.

The prognosis is a function of the underlying illness, the severity of the intellectual impairment, and the severity of the autistic manifestations. If a progressive organic disease of the brain is present, the prognosis is unfavorable; in cases where a disease of the brain has reached a stable state (e.g. in the aftermath of encephalitis), the prognosis depends on the severity of the intellectual impairment.

3.2.3 Asperger Syndrome

Asperger syndrome (ICD-10 classification F84.5) can be regarded as an extreme variant of the schizoid personality structure. Children with Asperger syndrome have a less pronounced disturbance of interpersonal relationships than children with early childhood autism, and they are often of normal or above average intelligence. They come to attention later than children with Kanner syndrome, typically only when special demands are placed on their ability to integrate themselves into social groups, e.g. when beginning kindergarten or (at latest) primary school. The severity of the disorder is highly variable. A review is provided by Weber and Remschmidt (1997).

Language development occurs early. The children often begin to speak before they can walk freely and acquire a versatile language with a large vocabulary and neologisms. Their language is disturbed in its communicative function in a different way from that of children with Kanner syndrome; they talk whenever they want without adapting to their audience (spontaneous talk), and they often talk to themselves. They never show the characteristic disturbances of preverbal and verbal communication found in children with early childhood autism, but they do have comparable abnormalities of the speaking voice.

They are capable of original thinking and have good logical and abstracting abilities. They often have excessively intense, narrowly defined, and unusual

special interests. They sometimes have good lexical knowledge of specific areas, but they mostly just store up knowledge without organizing it into larger contexts. Despite their high intelligence, they often do badly at school, because (like children with early childhood autism) they have a marked disturbance of attention arising not from external, but from internal distraction, i.e. they are preoccupied with themselves. Their physical clumsiness is evident from an early age, and dyspraxic disturbances are also frequently observed.

Children with Asperger syndrome can cope with other people or social situations only to a limited extent. They are inconsiderate in the furthering of their own desires, they are often happy about others' distress, they have no feeling for personal distance, and they lack a sense of humor.

The diagnosis is made on the basis of history-taking, examination, and observation of behavior. Personal observations made outside the examining room are especially helpful. A meticulous psychological examination with consideration of neuropsychological aspects, cognitive functions and personality is very important for diagnostic classification. Problems of differential diagnosis arise with respect to the following disorders: early childhood autism, personality disorders secondary to organic brain insults, and specific personality disorders.

In the broader genetic circle of children of Asperger syndrome, some investigators have found an increased number of persons with similar or identical abnormalities of personality. An inherited factor is thus thought to be involved in the etiology of autistic psychopathy. Autistic personality disorder is explained as a disintegration of the intellectual and emotional areas of the personality, or as a disturbance of intuitive ability. Harmful substances in the environment and early organic brain insults are further (co-)factors in etiology.

The presumption that Asperger syndrome has a genetic basis does not preclude the possibility that it might be treated effectively. The most important persons in the lives of these children (parents and others) should be counseled as early as possible, so that further progression can be prevented to the greatest possible extent. Adequate counseling requires a thorough acquaintance with the personalities of children in general and of the specific child (patient) in particular, including his or her overall state of maturity. If secondary neuroses are present, psychotherapy with involvement of the parents may be necessary. Sensorimotor exercise therapy should be performed early to treat motor disturbances.

The prognosis is usually favorable. The older these children become, the better their social integration proceeds, because of their increasing intellectual maturity, their rational adjustment to people and situations, and the fact that communication on a rational level is easier among adults; some of these children also achieve an improvement in social maturity, after a delay. Most adults with this personality disorder become well integrated in an occupation, often a demanding one in which they perform at an above average level. Their interpersonal relationships remain difficult, however, particularly with their closest relatives. Many adults with severe Asperger syndrome remain aloof loners and adopt "outsider"-type occupations, which they change frequently.

3.3 Rett Syndrome

This syndrome has the ICD-10 classification F84.2, in the broader category of profound developmental disorders. It is classified analogously in DSM-IV.

This syndrome was originally described by the Austrian child psychiatrist Andreas Rett (1966). It occurs predominantly in girls and first appears between the 7th and 24th months of life, after previously normal (or nearly normal) development. Its characteristic manifestations are:

- total loss of goal-oriented use of the hands, accompanied by frequent episodes of hyperventilation;
- total or partial loss of language;
- regressive social development and play development in the first years of life;
- slowing of head growth, with severe and progressive intellectual impairment;
- peculiar, stereotypic "winding" movements of the hands;
- neurological manifestations such as truncal ataxia and apraxia, often in association with scoliosis and kyphoscoliosis;
- tooth-grinding; and
- in approximately 50% of patients, epileptic seizures after the 18th year of life.

The full clinical picture of the disorder has been observed to date only in girls. Its prevalence in the female population up to age 17 is approximately 1 in 10,000.

The cause of this syndrome is not yet known. Family studies and observations of a small number of twin pairs have led to the conclusion that the cause is, at least in part, genetic. Cases of this syndrome in combination with other disorders suggest that what is inherited is not the classical phenotype of Rett syndrome itself, but rather a particular pattern of neuropsychiatric impairment that progresses until adolescence. Morphological and histological studies have shown that the cerebral development of patients with Rett syndrome is abnormal in a way that suggests a genetically transmitted neurodegenerative disorder.

The diagnosis is made on the basis of the clinical manifestations and can be assigned fairly securely in the first 3 to 5 years of life, as long as the diagnostician is acquainted with the syndrome. Its diagnostic differentiation from other neurodegenerative disorders requires a meticulous neurological examination, EEG, and computed tomography of the brain, as well as laboratory studies to rule out disorders of amino acid, lipid, and polysaccharide metabolism. The EEG during stage II sleep often reveals spike-discharges in the central and centroparietal regions.

The differential diagnosis of Rett syndrome first requires its differentiation from early childhood autism. A number of metabolic disorders must also be ruled out, as well as epilepsy as an independent disorder.

Treatment directed at the cause of the disorder does not yet exist, nor is there any way of arresting its progression. Emphasis is thus placed on symptomatic measures intended to achieve the following goals:

1. Education of parents and other involved persons about the nature of the illness.
2. Facilitation of nonverbal communication.
3. Medical treatment of epileptic seizures (Uldall et al. 1993).
4. Medical treatment of behavioral abnormalities (Zappella et al. 1990).

The prognosis is unfavorable, as the disorder is characterized by a steady intellectual decline and progression of the neurological abnormalities until puberty. It seems to remain relatively stable thereafter.

3.4 Disintegrative Disorders

The disintegrative disorders are also classified among the profound developmental disorders, both in ICD-10 (F48.3) and in DSM-IV.

The prototype of childhood disintegrative disorders is Heller's infantile dementia, described by Theodor Heller in 1908. All of these disorders are characterized by a loss of previously acquired abilities and skills after a period of normal development. Their most conspicuous manifestations are the loss, or progressive decline, of language and of intellectual, social, and communicative skills.

The age at which manifestations usually appear is between the 2nd and 4th years of life. The disorder usually begins insidiously. The children become mildly irritable and withdrawn, can no longer make themselves understood through speech, have disturbances of memory and perception, are anxious or aggressive, can no longer cope with social situations, often lose already acquired bladder and bowel control, and develop stereotypic movements, until the fully developed picture of a dementia emerges. The behavioral withdrawal and impaired communication skills sometimes create the impression of autism. It is noteworthy that the facial features of these patients do not become coarsened, despite their often severe dementia; this phenomenon is known among German-speaking child psychiatrists as the *Prinzengesicht* ("princely face").

This is a very rare disorder affecting an estimated 10 per million children. According to Gilberg (1995), children with disintegrative psychoses (Heller syndrome) make up about 1% of all patients in the spectrum of autistic disorders. The sex ratio of affected boys to girls is 8:1.

The etiology of this disorder remains unknown. An organic cause in the brain is presumed because of the characteristic course (onset of the dementing process between the 2nd and 4th years of life). A study of 76 cases by Volkmar (1992) revealed that the illness begins between the 2nd and 3rd year in 40% of cases, and between the 2nd and 4th year in 80%. Further evidence for an organic etiology is provided by the frequently observed EEG abnormalities and the presence of seizures in just over 50% of patients.

The diagnosis is usually made on the basis of the characteristic clinical manifestations. This disorder must be distinguished from early childhood autism and Rett syndrome.

There is no etiologically directed treatment to date. The treatment consists of measures that make it easier for the patient's family and others to deal with the disorder, and that protect the patient from further injury.

The prognosis of this disorder is unfavorable. Its inexorable downward course cannot be influenced by any currently available therapy.

3.5 Expressive Language Disorder

In ICD-10, expressive language disorder (F80.1) is listed among the specific developmental disorders of speech and language (F80), as is receptive language disorder. In DSM-IV, both disorders are subsumed under the designation "communication disorder."

Children with expressive language disorder display a discrepancy between their impaired ability to use spoken language and their otherwise normal, or nearly normal, intellectual ability. Language comprehension is normal in this disorder (unlike receptive language disorder). The disorder becomes apparent during language acquisition when the children remain unable to produce individual words or sounds, which results not only in a temporal delay, but also in a qualitative abnormality of language development. This qualitative abnormality is already manifest in the first and second

years of life when the child's first words are unintelligible, and continues as an inability to produce simple two-word sentences at the age of 3, marked uncertainty over the choice of words used to designate an object or to describe a situation, poverty of vocabulary, and abnormalities of sentence length, syntax, and grammar (Bishop 1994; Remschmidt, in press).

The following criteria are obligatory for the diagnosis:

1. The expressive speech disturbance must be significantly worse than the child's overall cognitive level.
2. Language comprehension must be within the normal range for age.
3. Nonverbal communication by means of facial expression and gestures must not be disturbed, as reflected above all by the child's attempts to compensate for his limited linguistic ability through nonverbal communicative gestures.

As in receptive language disorder, the language abnormalities of expressive language disorder are often accompanied by additional emotional and behavioral disturbances that are not specific to the disorder, but rather belong to the individual child's repertoire of reaction patterns. Aggressive and defiant behavior, depressive reactions and social withdrawal, attention disorders and hyperactivity are often seen.

These disorders must be distinguished from other disorders that are associated with expressive language disorder at a rate higher than chance would predict, including dyslexia and early-onset disorders of articulation.

3.6 Receptive Language Disorder

Receptive language disorder (ICD-10 classification: F80.2) is classified in both major systems in a manner analogous to expressive language disorder.

Children with receptive language disorder are characterized by a significant discrepancy between their language comprehension and their cognitive level. Parents become aware of the disorder when the child fails to react to simple vocal expressions or to commands expressed in language, despite normal, or nearly normal, overall development. The deficient comprehension of language naturally results in a failure of normal language development, so that expressive language is nearly always impaired as well. This manifests itself in disturbances of word and sound production.

As early as the first year of life, it becomes apparent that these children, once they emerge from the normal babbling phase, produce only more or less unarticulated sounds and cannot understand their parents' onomatopoeic expressions or simple verbal commands. In the second year, they are unable to name objects or follow simple commands, unless these are underscored by gestures or by corresponding eye contact. In the third and fourth years, they cannot adequately understand the grammatical structures of language (questions, negations, comparisons), nor can they understand the facial and gestural expressions accompanying speech (Bishop 1994; Remschmidt, in press).

Because these children nearly inevitably have difficulty in understanding and being understood, a number of behavioral abnormalities arise, including hyperactivity, attention disturbances, aggressive behavior, fearfulness, isolation, and rejection by peers.

The cause of this disorder usually lies in an organic brain disturbance that must be presumed in most cases and can only rarely be identified. In many cases, a number of other organic abnormalities accompany the disorder.

The diagnosis and differential diagnosis of receptive language disorder are made by means of a meticulous history and a comprehensive neurological and neuropsychological examination. Hearing must be tested next, as well as language comprehension and, where applicable, the auditory discriminating capacity for speech-related sounds. The performance of an EEG serves either to rule out epilepsy or to give evidence for hypersynchronous activity, and neuroimaging techniques (computed tomography, magnetic resonance imaging) can be used to detect structural brain anomalies. Particular attention is paid to a careful assessment of language comprehension and of language production, which is usually poorly differentiated. The prognosis depends largely on the level of language development.

The differential diagnosis of receptive language disorder includes early childhood autism, which can be excluded relatively easily if its other characteristic symptoms are absent. A further entity to be excluded is Landau-Kleffner syndrome (acquired aphasia with epilepsy).

Because no treatment directed at the etiology of the disorder is currently available, therapy consists of the following symptomatic measures:

1. Parents and other persons in the child's environment should be thoroughly informed about the nature of the disorder and the fact that secondary behavioral abnormalities often arise as a result of the impairment of verbal communicating ability.
2. If the child is cooperative or can be taught to cooperate, exercises can be used to extend his or her repertoire of sounds and words, in small increments.

3. Opportunities should be created for the stimulation of speech and the promotion of nonverbal communication, if verbal communication is massively impaired.
4. In speech therapy, the emphasis is no longer placed on highly structured exercises and rigidly applied treatment programs based on imitation, because it has been found that children do not remain motivated to continue with such exercises in the intermediate term, and that the patterns of communication acquired and practiced during the exercises are difficult to transfer to everyday, real-life situations. Thus, attempts are increasingly being made to incorporate elements of verbal communication, and practice of these elements, into normal play situations.
5. The introduction of sign language is recommended only if the child has a severe hearing impairment and can therefore not be expected to achieve adequate speech comprehension and production.

In half of all cases, the outcome is unfavorable, with persistent disturbance of language comprehension and inadequate acquisition of expressive language. The secondary consequences for the affected persons are a lack of satisfactory friendships and partnerships, and increasing social non-integration and isolation.

3.7 Landau-Kleffner Syndrome

Landau-Kleffner syndrome is listed in the ICD-10 among the specific developmental disorders of speech and language (classification: F80.3).

This illness consists of a predominantly receptive disturbance of language that first manifests itself at the age of 3 to 6 years along with the simultaneous appearance of epileptic seizures or of hypersynchronous EEG activity (Aicardi 1992). It may begin suddenly or insidiously and is at first characterized by a marked disturbance of language comprehension, which is usually followed by a disturbance of verbal expression.

It is understandable that children suddenly or gradually confronted with a deterioration of their comprehension of language and of their ability to speak react with emotional and behavioral abnormalities. These may appear in the form of depressive reactions, extreme fearfulness, or aggressive or defiant behavior, and create further problems for treatment.

The prognosis of this illness is unfavorable and its major manifestations persist despite the occurrence of repeated, but transient, episodes of improvement. The earlier the disorder arises, the less favorable its course (Bishop 1985). Thus, children in whom Landau-Kleffner syndrome becomes apparent before the fifth year of life have considerable problems with language comprehension lasting into adulthood, despite intensive efforts to improve the situation through therapy. It has also been found that anticonvulsive drug therapy, though it may prevent seizures and improve the EEG pattern, does not lead to a significant improvement of language functioning (Deonna et al. 1995).

The cause of the disorder is currently unknown. An organic etiology is assumed because of the characteristic pattern of manifestations and the accompanying abnormal EEG activity. The disorder has no known genetic basis.

The diagnosis is usually made on the basis of the typical history, the clinical manifestations, and the EEG findings. History-taking often reveals that the decline of language expression was preceded by a phase of abnormal behavior reflecting a disturbance of language comprehension. Many children with Landau-Kleffner syndrome may be thought to be deaf once the full clinical picture develops; for this reason, a hearing examination is urgently indicated.

It may be difficult to distinguish Landau-Kleffner syndrome from the disintegrative disorders of childhood, because a phase of normal development precedes the onset of both disorders. The differential diagnosis is established on the basis of the epileptic seizures and the characteristic bitemporal EEG abnormalities seen in Landau-Kleffner syndrome. Early childhood autism differs from Landau-Kleffner syndrome in that it is present from the earliest stages onward and has characteristic manifestations that are not found in Landau-Kleffner syndrome.

This illness, too, has no etiologically directed treatment, though the epileptic seizures may be alleviated by appropriate medication. Parents and other persons in the child's environment must be educated about the nature of the disorder so that they may better understand not just the underlying disease, but also the secondary behavioral abnormalities that result from it.

The epileptic seizures require treatment; carbamazepine in age-appropriate doses has been found to be useful. Such treatment generally does not improve language function, nor does it eliminate abnormal EEG activity. There is as yet no conclusive evidence in favor of corticosteroid treatment, but a trial of such treatment seems to be indicated in the early stage of the disease (Lerman et al. 1991).

3.8 Specific Disorders of Reading and Spelling (Dyslexia)

The specific disorders of reading and spelling (dyslexia) are classified in the ICD-10 among the specific developmental disorders of scholastic skills (F81). This heading covers the following disorders: specific reading disorder (F81.0), specific spelling disorder (F81.1), specific disorder of arithmetical skills (F81.2), mixed disorder of scholastic skills (F81.3), other developmental disorders of scholastic skills (F81.8) and unspecified developmental disorder of scholastic skills (F81.9).

The hallmark of specific reading disorder is a major, specific impairment of the development of reading ability that cannot be attributed to mental retardation, a sensory deficit, or the consequences of an illness. Reading difficulties are often the result of developmental disorders of speech and language. Reading disorder is often accompanied by spelling disorder, which is why the two disorders are commonly designated together as reading/spelling disorder (dyslexia). At later ages (e.g. in adolescence and adulthood), the spelling deficits are generally more severe than the reading deficits. Specific diagnostic criteria include the following:

- reading performance significantly below the norm for age (e.g. 1.5 to 2 standard deviations below the norm), which cannot be attributed to mental retardation,
- spelling performance that is below the norm to a comparable degree,
- the occurrence of developmental speech disorders in preschool age,
- an IQ over 70 and
- normal school experience.

The impairment in the development of reading skills manifests itself mainly in the following ways: the omission, substitution, or jumbling of words, low reading speed, difficulty starting to read aloud (long pauses or losing one's place in the text), incorrect parsing, exchanging of words and deficient reading comprehension, and an inability to reproduce verbally what has been read, or to draw conclusions or instructions for action from it.

The clinical picture of spelling disorder involves various types of errors that are not specific to this disorder, e.g.: the turning around and omission of letters, introduction of letters in the wrong place, and incorrect ordering of letters. The errors are inconstant, i.e. the same word is written with different mistakes at different times (Maughan and Yule 1994; Remschmidt, in press; Warnke and Roth 2000).

Children with reading/spelling disorder characteristically have a number of other disturbances: abnormal language development can be demonstrated in roughly 60%, one third have attention deficits and hyperactivity, and 5% to 10% have visual or visuomotor abnormalities. Secondary problems resulting from reading/spelling disorder include disturbances of learning and performance behavior, emotional disturbances, impaired concentration and hyperactivity, psychosomatic symptoms (headache, stomachache, and nausea, particularly before tests in school), depressive mood abnormalities and conduct disorders. The disorder is significantly more common in boys than in girls (sex ratio approximately 2:1). Its prevalence is approximately 6% to 7% in the second and third grades and 8% in a representative population of schoolchildren up to 18 years of age (Remschmidt and Walter 1990a,b).

Although many studies have been performed on the question, the etiology of reading/spelling disorder remains unknown. A number of hypotheses have been advanced (cf. Warnke and Roth 2000) that are not mutually exclusive, each of which can be supported by a certain amount of evidence:

1. The hypothesis of a genetic etiology is based on the increased frequency of reading/spelling disorder in certain families, and on the findings of twin studies. Both types of findings imply the existence of an important genetic factor. The results of more recent studies suggest that reading/spelling disorder and its component disorders may be traced to abnormalities in a single gene. There is already a replicated finding that reading/spelling disorder may be linked to a particular chromosome.
2. The hypothesis of a disturbance of information processing is based on findings concerning the processing of both verbal and visual information. As for the former, it is known that as many as 60% of children with reading/spelling disorder suffer from disturbances of language development that involve various aspects of language acquisition (deficient phonological awareness, impaired auditory discrimination, difficulty in the storage and reproduction of verbal material). As for the latter, children with reading/spelling disorder cannot process visual information adequately, as normal children can. They need more time to perform the same tasks, and their performance often differs qualitatively as well, as can be demonstrated by differences in evoked potentials. These findings may be interpreted as implying a disturbance of the temporal dimension of information processing.
3. Hypotheses of abnormalities in brain structure and function are based on the following arguments: A number of studies have shown that the planum temporale, an important structure in human language development, is of comparable size in the two

hemispheres in persons with reading/spelling disorder, though it is larger on the left side in the great majority of normal persons. One may, therefore, conclude that the normal cerebral asymmetry fails to develop to an adequate extent in persons with dyslexia. Furthermore, autopsy studies have revealed histological abnormalities and abnormalities of vascular supply in the brains of patients with dyslexia. Studies employing evoked potentials have revealed both temporal and qualitative abnormalities of information processing in the brain of dyslexics. Finally, some children with reading/spelling disorder also have significant disturbances of attention, which mainly relate to selective, i.e. directed, attention, rather than to long-range attention.

The disorder is diagnosed on the basis of meticulous history-taking, clinical examination, and psychological testing. During history-taking, particular attention must be paid to language development in the preschool years and to the possible occurrence of reading or spelling difficulties in other members of the family. The clinical psychiatric and neurological examination is directed toward additional manifestations and neurological findings. Clinical examination should also include auditory and visual testing, and testing of other sensory modalities. Finally, psychological testing is performed in order to assess reading and spelling ability in a standardized fashion, as well as calculating ability (important for differential diagnosis).

Reading/spelling disorder must be differentiated from the following other disorders: a disturbance of reading and spelling secondary to low intelligence; acquired dyslexia or alexia; an acquired disturbance or reading and spelling resulting from an emotional disorder; and specific spelling disorder without impairment of reading (F81.1).

The treatment of reading/spelling disorder consists of three types of intervention (Warnke 1987; Warnke and Niebergall 1997):

- functional therapy of reading and spelling,
- helping the child cope with the disorder mentally and emotionally, and
- treatment of secondary mental manifestations with inclusion of family members and other persons in the child's environment, possibly with support from the child's school.

The individual German federal states formulated specific guidelines and directives for assistance to children with reading and spelling disorders in response to a recommendation issued on 20 April 1978 by the conference of German ministers of education and cultural affairs. These guidelines and directives differ in detail, but all share the following elements (Warnke and Roth 2000):

- All children with reading/spelling disorder have a right to scholastic assistance, e.g. through special classes or individual attention.
- Spelling performance must not play a decisive role in promotion or non-promotion to the next grade, in transfer to a special school, or in the transition from primary to secondary school.
- Spelling mistakes must not be considered in the grading of written assignments, such as essays, that are not specifically intended as spelling tests.

Most of the German states also allow the option of not grading spelling performance.

There is no medication that has a specific influence on reading/spelling disorder, but medications are occasionally useful for the treatment of additional and secondary disturbances. Antidepressants, for example, have been found useful for the treatment of accompanying depressive manifestations. The use of stimulants can be considered when dyslexia is coupled with a hyperkinetic syndrome. Treatment with nootropic agents is controversial. A number of findings suggest that children treated with these substances read more rapidly and accurately and have a better verbal memory performance (Wilsher 1986).

An uncompensated reading/spelling disorder is a major risk factor for additional mental disturbances, antisocial behavior, and possible failure of vocational development. Various studies have found that 25% to 75% of juvenile delinquents suffer from specific learning disorders, most commonly dyslexia. Approximately 30% of children with reading/spelling disorder manifest behavioral abnormalities (Rutter et al. 1970, 1976); the corresponding figure for adolescents is 50% (Korhonen 1984). The prevalence of reading/spelling disorder among prison inmates is also notably high (Weinschenk 1965).

Only 20% to 25% of children with reading/spelling disorder achieve age-appropriate spelling performance during the primary school years, and only 4% of children with severe dyslexia achieve normal spelling performance (Watson et al. 1982).

3.9 Specific Disorder of Arithmetical Skills

This disorder, too, is classified in ICD-10 under the specific developmental disorders of scholastic skills (ICD classification: F81.2).

The specific disorder of arithmetical skills – which we will refer to here as arithmetical disorder – is characterized by a marked impairment of arithmetical ability lying well below the child's general intelligence level. A discrepancy of at least 1 or 1.5 standard deviations, or of 12 *t*-value points, is generally required for the diagnosis. The disorder becomes

apparent in the early school years (no later than the fifth school year) and is often found in association with developmental disorders of speech or language, of motor function, or of visuomotor function. These children can read and spell normally, and their inability to calculate is a more or less isolated weakness within the overall performance profile. These children manifest additional social or emotional behavioral abnormalities, which may often be secondary consequences of the arithmetical disorder. The ICD-10 diagnostic guidelines require that impairment of arithmetical skills should not be attributable to lack of training, visual or auditory deficits, or neurological illnesses.

The manifestations of arithmetical disorder are variable, because different components of calculating ability can be affected to differing extents. There may be disturbances of spatial representation and spatial thinking, a lack of understanding of mathematical terms or signs, difficulty in arranging numbers in the right order, a diminished capacity for abstraction, etc. (Remschmidt, in press). In everyday life at school, the disorder manifests itself in deficient or unsatisfactory performance on arithmetical tasks, while performance in other areas is at a normal level.

The prevalence of arithmetical disorder not associated with other developmental and learning disorders is estimated at 1% among children. If one includes cases that are associated with disorders of these types, its prevalence lies between 4% and 6%. From the educational point of view, approximately 15% of schoolchildren require special help with arithmetic (Warnke, in press). The disorder seems to be more common in girls. These epidemiologic data should be used with caution, however, because studies of unselected population samples are not yet available.

The etiology of arithmetical disorder (not related to lack of training or to neurological disease) remains essentially unknown. Its common association with other developmental disorders implies that various cerebral processes become impaired in the course of development. It is still unclear to what extent genetic factors may be involved.

The diagnosis is made on the basis of the history, clinical examination, psychological testing and, where appropriate, neuropsychological assessment. Intelligence testing is mandatory, as is the application of age-specific tests of arithmetic and spelling. Furthermore, language development, motor development, visuomotor function and attention should be specifically tested in every case. Finally, emotional and other problems should be explored, and the family background should be evaluated, as part of the diagnostic process.

As for differential diagnosis, arithmetical disorder must be distinguished from difficulty with arithmetic in the context of mental retardation, neurological disease, lack of training, neglect, and other psychiatric disorders (e.g. obsessive-compulsive disorder, depression, anxiety disorders). Difficulty with arithmetic may occur as a reaction to experiences of failure in arithmetic that then engender severe performance anxiety, so that the difficulty perpetuates itself.

The appropriate treatment consists of functional therapies performed on an outpatient basis and based on a meticulous analysis of the deficient components of the calculating process. Three specific areas are addressed:

- preconditions for adequate performance of arithmetical tasks: spatial representational ability, the concept of quantity, the capacity for abstraction and other basic underlying abilities;
- the type of error made while calculating (e.g. reading errors, orthographic errors, errors when writing numbers); and
- calculating strategies, i.e. the systematic application of basic abilities so that the correct answer is obtained. More complex problems can be solved in various different ways, which should first be illustrated individually with concrete examples, and then practiced in abstract form.

The child's motivation to undergo functional therapy is extremely important, as success is achieved slowly and in small steps.

3.10 Hyperkinetic Conduct Disorder

Hyperkinetic conduct disorder (ICD-10 classification: F90.1) is coded in ICD-10 as a subcategory of the hyperkinetic disorders (F90). In DSM-IV, various subcategories of hyperkinetic disorder are coded under 314.xx (attention deficit hyperactivity disorder), and – in contrast to ICD-10 – there is no mixed category for the combination of hyperkinetic disorder with a conduct disorder.

Hyperkinetic conduct disorder is characterized by the cardinal manifestations of hyperkinetic disorder – attention deficit, motor hyperactivity and, in some cases, marked impulsiveness – and additionally by the antisocial, aggressive, and rebellious behavior typical of conduct disorders.

The attention deficit manifests itself in the premature termination of tasks and activities. The children appear to lose interest in tasks easily because they are distracted by other stimuli. Motor hyperactivity involves not only a compelling inner drive toward movement, but also an excessive restlessness that is particularly evident in situations demanding relative tranquillity. This feature is found above all in structured situations requiring a large measure of self-control.

Impulsiveness, i.e. an inclination toward excessively rapid and poorly thought out activity, appears in everyday life and particularly in situations in which performance is demanded. In school and other performance situations, these children have an "impulsive working style," i.e. they have trouble waiting their turn, and they interrupt others and shout out their responses without fully answering the question asked. Impulsiveness is not only the manifestation that appears regularly in the greatest number of situations, but also the one that persists for the longest time over the course of development (as these children grow older). Not least, the impulsive behavior of these children, in combination with their antisocial, aggressive, and rebellious behavior, often leads to difficulties and rejection in the social sphere.

The available epidemiological data concerning the frequency of the overall group of the hyperkinetic disorders are highly divergent. All in all, it is thought that some 3% of school-age children suffer from a disorder of this type. Boys are significantly overrepresented; the ratio of affected boys to girls is approximately 3:1, in clinical studies as high as 6:1 or 9:1.

No single cause of hyperkinetic disorder has been discovered, though many possible etiological factors have been discussed. Constitutional factors, i.e. genetic predispositions, seem to play a decisive role. On the other hand, the severity and long-term course of the disorder and the nature of the accompanying manifestations seem to depend very strongly on environmental influences. This is particularly true of hyperkinetic children with conduct disorder.

The task of diagnosis presents considerable difficulties. The main reason for this is that the manifestations of the disorder are many and varied; a further reason is that they are situation-dependent and thus subject to fluctuation. The developmental dimension must also be taken into account in the diagnostic process, because the normal degree of motor activity varies in close relation to a child's age and stage of development. Not least, normative conceptions also play a role in the labeling of a child as "abnormal." Because of all of these problems, the diagnostic process should be carried out as thoroughly and comprehensively as possible, with attention to all of the relevant dimensions.

As far as differential diagnosis is concerned, it must be borne in mind that abnormalities on the basis of comorbid disorders may be found. In particular, the differentiation of hyperkinetic conduct disorder from the conduct disorders proper presents major difficulties. The creation of a special category for the area of overlap between hyperkinetic disorders and conduct disorders is a product of this state of affairs. Aside from conduct disorders, other possible comorbid disorders include affective disorders, anxiety disorders and specific learning disorders.

The many and diverse therapeutic approaches that have been tried are an indication that the treatment of this disorder must be multidimensional and multimodal. There is no single therapeutic measure that is effective by itself in the treatment of hyperkinetic disorders. This fact has led to the formulation of treatment programs that contain a variety of different elements and are flexible enough to be tailored to the individual needs of each patient.

Pharmacotherapy may play an important role, depending on the severity of the disorder. The use of stimulants (methylphenhydate, dextroamphetamine) is the major form of drug treatment. Antidepressants (e.g. imipramine) and neuroleptics have also been tried. It has been demonstrated that treatment with stimulants can alleviate the specific hyperkinetic manifestations (e.g. the attention deficit) but cannot eliminate the conduct disturbance. Stimulants are ineffective in 10% to 15% of hyperkinetic children and lose their effect after 6 months in a further 20%. In approximately 40% of children, the treatment can be discontinued after 1 or 2 years without return of manifestations to any marked degree. Pharmacotherapy can never be used as the sole form of treatment. It meets the objective of enabling these children to cope with an important stage of their development, but does not eliminate the disorder or affect its underlying causes.

Pharmacotherapy is often combined with behavior therapy. Operant techniques have proved useful not only in the form of programs and written instructions, but also as aids to orientation in everyday activities and in parent counseling. Self-instruction training is an attempt to improve the capacity for self-control by systematic verbalization and practice of instructions for dealing with problem situations. Social competency training is intended to make the child behave toward others in a better adapted and more socially acceptable way, and thereby stabilize the child's emotional well-being.

Emotional stabilization is further promoted by functional therapies that alleviate deficits of learning and performance and thereby increase self-esteem.

Finally, working together with the child's parents is an important part of treatment. This is not merely a matter of educating the parents thoroughly about the disorder. The parents' perception of the child, which is usually very strongly fixated on his or her inappropriate behavior, must also be changed.

Contrary to the previously widespread notion that children "grow out of" hyperkinetic disorders when they reach puberty, it is now known that at least three quarters of affected persons go on to have further difficulties in school and vocational training, in the family, and in general social adjustment. This state of affairs continues even into adulthood. At least 60% of affected young adults still have manifestations of the

disorder. Particularly at risk are those with low intelligence, low socioeconomic status, and a high degree of aggressiveness. Limited social contact with peers, emotional instability and, above all, psychopathological abnormalities in parents also confer a poor prognosis. The affected individuals are especially at risk for addiction and delinquency and show higher rates of these problems than other persons in comparable age groups.

4 Developmental Disorders

As explained above in Sect. 2, behavioral changes occurring over the course of time, including developmental processes, are partly under biological control and are thus normally subject to wide variation. Observation is indicated when a child's developmental progress is at the border of the normal range. Conditions lying outside this range are considered pathological. This remains true even when it is certain that such partial developmental delays will be fully compensated in the intermediate term without external support. Such delays can transiently impair the interplay of the affected function with other developmentally relevant functions, or they can cause other persons to perceive the child or adolescent so differently that they no longer place age-appropriate demands on him or her.

Developmental disorders are quantitative or qualitative deviations from behavior that is expected in normal development. They most often resolve spontaneously in the intermediate term, but they may nevertheless permanently affect the overall development of an individual, either directly or indirectly (i.e. through the behavior of other persons).

The pathogenetic basis of developmental disorders is usually biological. Familial clustering of certain developmental disorders and their more common occurrence in boys lead one to suspect a genetic component. Delays of cerebral maturity are also a possible pathogenetic mechanism and are mainly attributable to impairments or illnesses of the immature central nervous system. Developmental disorders cannot manifest themselves before the expected developmental appearance of the function in question; thus, their onset is usually tightly coupled to the time of appearance of rapidly developing functions.

Functions of this type are especially susceptible to outside influence and also especially dependent on support from the environment. Their development also depends on the individual's motivation to learn. Lack of support, disturbing external influences, or conflicting motivation can promote the emergence of a developmental disorder. The determination of pathogenesis is generally difficult in the individual case. Medicolegal questions about the origin of bedwetting or of an attention disorder, for example, cannot be definitively answered; often, one can only say that there is no single underlying cause.

The course of developmental disorders is usually favorable. Some, however, persist because a partial developmental delay becomes fixed and autonomous, while others are prolonged because the disorder creates a gain of some type for the individual. The risk that the disorder will become fixed is higher if it includes behavior that deviates from the norm not just quantitatively, but also qualitatively. In a small number of cases, persistence of a "developmental" disorder is actually due to misdiagnosis of a condition more correctly interpreted as an age-specific disorder with an open prognosis or a tendency to persist (see Sect. 5).

4.1 Disorders of Motor Development and Articulation

In the specific developmental disorder of motor function and in specific speech articulation disorder (ICD-10 classifications: F82 and F80.0, respectively), the developmental level of coarse and fine motor function, or of articulation, deviates from the attained developmental level of cognitive function. The corresponding perceptual functions are always involved as well, e.g. kinesthetic sense or auditory discrimination. The major manifestation of the motor disorder is a slowed performance of complex tasks requiring a high degree of coordination, such as climbing stairs with one foot after the other, tying knots, balancing, riding a bicycle, eating with knife and fork, writing and drawing. Often, too much or too little force is applied for the proper performance of the task. In the articulation disorder, sounds are not acquired at the usual times or in the usual order. Omissions, distortions, and substitutions occur which may make the patient's speech difficult to understand. The two disorders frequently occur together.

Boys are affected more often than girls. The six-month prevalence of these disorders in primary-school age has been reported as 1.5% for the motor disorder and 5% for the articulation disorder; the prevalence is higher in younger age groups.

The fact that these disorders occur with familial clustering implies that they have a genetic component, while the fact that they improve with increasing age implies the involvement of maturational processes. Organic brain disorders seem to make these disorders more likely to occur. Practice seems to play no role in the motor deficit, as practice can speed up motor function only slightly. Articulation disorders are more likely to occur when the child learns to speak from bad models in the environment, or when there are distur-

bances of auditory discrimination and sequential stimulus processing.

Neurological illnesses, sensory deficits, mental retardation and lack of external support must be excluded in the differential diagnosis. The diagnosis of disorders of motor development and articulation requires that they manifest themselves at the time of the expected appearance of the function in question. "Soft" neurological signs and delayed maturation of background EEG activity are common, but not obligatory for the diagnosis. The diagnosis is supported by the occurrence of other cases in the family. Articulation disorders are not diagnosed as an independent entity when they occur as accompanying manifestations of expressive or receptive language disorders. The diagnosis is based on a major discrepancy between motor function or articulation and the child's intellectual level.

Functional therapy is the usual form of treatment; its effectiveness is better documented for articulation disorders than for disorders of motor development. The prognosis for spontaneous improvement is much better than that of the specific developmental disorders of speech or written language (Esser 1991). Comparative longitudinal studies of the effect of treatment are lacking; indeed, even the "spontaneous" remissions seen in cohort studies mostly took place after functional therapy.

4.2 Stereotyped Movement Disorders and Transient Tic Disorders

Stereotyped movements (ICD-10 classification: F98.4) are nonfunctional, voluntary, repetitive, often rhythmic movements of a single type. Tics (ICD-10 classification: F95.0) are repetitive, sudden, quick, nonrhythmic movements of specific muscle groups; the most common transient tics are blinking and grimacing. Tic disorders are of age-specific onset and are accordingly discussed in Sect. 5.2. Both types of disorder usually appear in preschool age and affect approximately 3% of this age group. Children who are mentally abnormal in other ways often develop stereotyped movements with self-injurious behavior. Stereotyped movements not involving self-injurious behavior generally remit when the children reach school age (Werry et al. 1983).

Stereotyped movements occur under conditions of either decreased or increased stimulation; transient tics occur spontaneously, although they may be reinforced by stress. The most common stereotyped movements are head-tossing and body-rocking (often before falling asleep); nail-biting and thumb-sucking are also common. Pulling out hair and eyelashes is a disorder of impulse control. Stereotyped movements are often components of other illnesses (e.g. autistic syndromes) and, in such cases, are not diagnosed as a separate entity.

4.3 Pica and Rumination

Pica and rumination (ICD-10 classifications: F98.3 and F98.2) are rare eating disorders of childhood consisting of the regular eating of inedible substances (pica) or the regurgitation of swallowed food (rumination). Rumination is closely related to the stereotyped movement disorders. Pica is more common in mentally retarded or neglected children, and also in those with iron-deficiency anemia (Parry-Jones and Parry-Jones 1992).

4.4 Sleep Terrors, Sleepwalking, and Nightmares

Sleep terrors (ICD-10 classification: F51.4, also called "pavor nocturnus") are found in approximately 4% of preschool children. These children awaken during the first third of the night (sleep stage III or IV) with signs of panic, increased motor activity, high levels of autonomic discharge, and disorientation that may last several minutes. They have no recollection of the event afterwards. Lack of recollection of the event also characterizes sleepwalking (ICD-10 classification: F51.3), a disorder whose prevalence reaches a peak in the 12th year of life. Children with this disorder arise from bed in the first third of the night (sleep stage III or IV), occasionally several times per night. While they are walking around, they are poorly responsive and their behavior is difficult to influence, and they have a fixed facial expression and an empty gaze. They are briefly disoriented afterwards. In nightmares (ICD-10 classification: F51.5), vivid dreams occurring during REM-sleep in the last third of the night – or even during daytime sleep – lead to anxiety and a feeling of danger. Recollections of nightmares are vivid and plastic and arouse the fear that they will return, possibly leading to difficulty falling asleep (Clore and Hibel 1993).

Disturbances of maturation of the central nervous system are thought to be responsible for sleep terrors and sleepwalking, and special features of emotional development are thought to be responsible for nightmares. Traumatic experiences are thought to cause nightmares in only a small minority of cases.

Disturbances of maturation are characteristic and easily diagnosed from parents' descriptions. As for differential diagnosis, nocturnal seizures can be ruled out by sleep EEG, and, in case of doubt, by investigation in a sleep laboratory.

These childhood parasomnias have a favorable prognosis. Sleep terrors frequently become transformed into sleepwalking as the child enters school age. Sleepwalking rarely persists and, when it does, usually in combination with major psychopathology. Parent education is mandatory in sleepwalking to prevent injury to the children. The rare persistent forms of sleepwalking may be treated with tricyclic antidepressants. In children with nightmares, the child's fear that the bad dreams will return may require treatment.

4.5 Enuresis and Functional Urinary Incontinence

These disorders, subsumed under the heading of nonorganic enuresis (ICD-10 classification: F98.0), require further differentiation. Two forms of primary enuresis (failure of bladder control) may be distinguished: an isolated and a combined form.

In the isolated (monosymptomatic) form, large volumes of urine are voided during non-REM sleep in the first few hours after falling asleep. The affected children are not awakened by the elevated intravesical pressure and are also difficult to awaken by other means. This isolated, primary nocturnal enuresis tends to occur in families. In the combined form of the disorder, diurnal enuresis often occurs in parallel with symptoms of functional urinary incontinence, encopresis, and other mental abnormalities, and in contrast to the isolated form, bedwetting may occur several times per night. Bedwetting is a stimulus that awakens the child (von Gontard 1998). The primary form accounts for almost one third of patients seen in clinical practice, and secondary nocturnal enuresis for a little more than one quarter.

In secondary enuresis, bedwetting returns after an interval of good bladder control lasting more than 6 months. It is often associated with daytime urge incontinence, and with mental abnormalities of a mainly extroverted type; familial occurrence of enuresis is common. The six-month prevalence of nocturnal enuresis is 15% among 5-year-olds, 8% among 8-year-olds, and 3% among 13- to 15-year-olds. After puberty, boys are affected nearly exclusively; until puberty, the ratio of boys to girls is 2:1.

Pure diurnal enuresis is rare and is almost exclusively due to functional urinary incontinence (Hjälmas 1992), which occurs somewhat more commonly in boys than in girls, at a rate of 5% among 7-year-olds and 1% among adolescents. Pure urge incontinence predominantly affects girls (van Gool 1989). The frequent bladder emptyings in this disorder are of only about 60% of the total bladder capacity, yet no delay can be tolerated. Fatigue exacerbates the symptoms. Bladder-holding cannot prevent the release of small quantities of urine. Urge incontinence often occurs in combination with secondary nocturnal enuresis. A large fraction of children with functional urinary incontinence are incontinent during the day because of deliberate postponement of urination; the urinary volume is elevated, and bladder-holding is common. This disorder is more common in boys than in girls, often combined with encopresis and constipation, with incomplete bladder emptying, and with psychiatric comorbidity (also with extroverted manifestations) (von Gontard et al. 1998).

A small subgroup of children with functional urinary incontinence suffer from incoordination of the detrusor and sphincter musculature. Inadequate sphincter tone prolongs the time needed for bladder emptying and makes it incomplete. Detrusor hypertrophy and obstruction of the urinary pathway result (Hinman 1986; Olbing 1993).

A major genetic component of primary nocturnal enuresis has been identified and traced to loci on chromosomes 8, 12, 13 and 22 (Eiberg et al. 1995; von Gontard 1998; von Gontard et al. 1998). In other forms of enuresis, too, familial occurrence points to a genetic predisposition. Pathogenetic factors include a reduction of antidiuretic hormone, increased depth of sleep, and, in some cases, inconsistent toilet-training. Secondary nocturnal enuresis often begins in the context of external stress and in the framework of defiant or regressive behavior (e.g. after the birth of siblings), but it often persists after the triggering factor has been removed. Secondary mental phenomena may complicate the treatment of primary nocturnal enuresis as well as its secondary psychopathology. An important contributing factor in daytime incontinence is the activation of the detrusor muscle during the vesical filling phase, a sign of incomplete maturation. The cause of detrusor-sphincter incoordination is unknown. Voluntary postponement of urination acts in opposition to physiological bladder emptying. Recurrent urinary tract infections or asymptomatic bacteriuria are commonly associated with functional urinary incontinence.

The subtypes of the disorder are diagnosed on the basis of the family history, onset and frequency of manifestations, length of asymptomatic intervals, quality of toilet-training, other developmental disorders, current stressors, and other psychiatric manifestations. The information to be obtained includes the amount of urine voided and whether bedwetting awakens the patient at night, as well as the frequency and volume of urination by day and the possible occurrence of urge symptoms and holding maneuvers.

The physical examination is directed toward external signs of malformation of the urinary tract, while laboratory testing is performed to detect bacteria and

other signs of urinary tract infection (Hansson 1992) and sonography is performed to detect detrusor hypertrophy and residual urinary volume (Olbing 1993). Sphincter-detrusor incoordination can be detected only by a urination program and is not diagnosed in children below 5 years of age. The asymptomatic interval in secondary enuresis must be at least 6 months long. Other entities to be ruled out include dysuria, polyuria, hematuria, incomplete voiding, mechanical problems (outflow obstruction, reflux, cystitis), neurological problems (nighttime epilepsy), and, rarely, the onset of type I diabetes. If there is urinary dripping, anomalies of the urinary tract must be ruled out.

The annual rate of spontaneous remission is approximately 15% of the age-specific prevalence until late adolescence and is somewhat higher in preschool children. The course is more favorable in girls. Functional urinary incontinence has the best prognosis; the prognosis of secondary nocturnal enuresis is relatively favorable, that of the isolated and combined primary forms less so. These disorders require treatment even when the spontaneous prognosis is good, because their manifestations lead to social disadvantage and to impairment of the patients' self-esteem.

Before the proposed treatment is begun, it should be established that the child and the parents will be able to comply with it. Measures that tend to reinforce the manifestations of the disorder (fluid restriction, diapers) should be avoided. Hyperkinetic disorders should always be treated first, as should oppositional defiant disorders, while conduct disorders should be treated at least in parallel. Self-control and reinforcement techniques are adequate to control manifestations in 30% of primarily and secondarily incontinent children; these techniques require a consistent transfer of responsibility to the child and can succeed only if behavior-therapeutic rules are meticulously followed. If this form of treatment remains ineffective for 6 weeks, treatment with enuresis alarms is recommended, which, when meticulously applied, leads to improvement in 75% of patients within 8 weeks. Because the rate of nighttime urination is 40%, overlearning by means of fluid loading is advisable. If this should fail, alarm training is coupled with intensive bladder training; this has been reported to yield a success rate of at least 75% in 4 weeks, with a 30% relapse rate. A high degree of parental cooperation is required.

The administration of tricyclic antidepressants carries the risk of cardiac side effects, is associated with a relapse rate of 50%, and is thus not a treatment of first choice. Desmopressin treatment has a success rate of approximately 50%, but its relapse rate is on the same order of magnitude and there is a risk of side effects from fluid restriction. Pharmacological intervention is indicated only if there is a short-term need for successful treatment, or as a supportive measure accompanying behavior-therapeutic methods.

The various forms of functional urinary incontinence generally require treatment. Daytime incontinence due to voluntary postponement of urination can be treated successfully by consistent, controlled voiding according to a behavior-therapeutic treatment program. Detrusor-sphincter coordination can be achieved through biofeedback training; the success rate is 50%, and the rate of partial success 20%. The treatment of urge incontinence begins with the elimination of urinary tract infections, which leads to a full remission in some patients. In the second step, if needed, behavior-therapeutic principles are applied to make the patient aware of the pressure of a full bladder and to train him or her to empty the bladder completely. The third step consists of additional treatment with oxibutinin or, in older children, bladder capacity training. The success rate of this form of treatment has been reported to be 70%.

4.6 Encopresis

Encopresis (ICD-10 classification: F98.1) usually occurs by day, more rarely by night, and usually secondarily, i.e. after the acquisition of sphincter control. Repeated, voluntary or involuntary deposition of stool in varying quantities in inappropriate places occurs in 2% of 8-year-olds and 0.6% of 13-year-olds and is approximately four times more common in boys than in girls. Encopresis is isolated in 80% of cases and is due to constipation or fecal retention as a result of genetic predisposition, nutritional habits or avoidance of defecation, and sometimes as a result of depression (Christophersen and Rapoff 1983; Levine 1975).

When constipation is the underlying mechanism of encopresis, it is not perceived by the child, and its pathogenetic mechanism is unclear. In such cases, soft stool generally travels past more solid stool already located in the rectum. 15% of cases of encopresis are due to an irritable colon reacting in this way to stress or anxiety. This form of the disorder is often associated with emotional disorders. Deliberate, manipulative encopresis accounts for 5% of cases and is usually associated with oppositional defiant disorders or antisocial manifestations. The age at which toilet-training is given is irrelevant, but its quality is crucial. The secondary form of encopresis is more common than the primary form. Secondary combinations with other mental disorders are frequent, as is a secondary combination with enuresis.

Diagnostically relevant considerations include the nature and history of the manifestations, accompanying psychiatric disorders, personal maturity, and

current or chronic family stresses, as well as physical manifestations of constipation. The diagnosis is not made in children under 5 years of age.

The need for treatment arises from the social consequences of the disorder and the danger of chronic constipation. In secondary forms, the emotional or defiant-antisocial manifestations are treated at least in parallel with the behavior-therapeutic regularization of bowel function. In primary forms, the treatment of constipation is the next most important objective after the achievement of bowel control and is carried out with behavior-therapeutic measures.

4.7 Age-Specific Somatoform and Conversion Disorders

Recurrent headaches and stomachaches in childhood are often precursors of somatoform functional disturbances of autonomically controlled organ systems. Such disturbances have been reported to occur in 10% to 30% of school-age children but are found in only 1.5% of children presenting for treatment. They include dysphagia, globus hystericus, abdominal symptoms, vertigo and palpitations, and, less frequently, nausea, dysuria and hyperventilation. 40% of affected children also have dissociative symptoms such as motor or sensory disturbances, visual disturbances, and aphonia. In adolescence, non-specific pains become less common and psychogenic seizures become more common. Girls are more commonly affected than boys, and conflict-associated unconscious affects can usually be discerned. Suffering and a need for sympathy are found in place of the "belle indifférence" typical of somatoform disorders in adulthood.

The behavior of adult role models seems to play a role in pathogenesis. Sexual abuse is often said to contribute to the generation of these disorders.

The diagnostic assessment of these disorders requires both an elucidation of the patient's past history and behavioral models, and the early and thorough exclusion of somatic disorders. It is typical for patients to reject an interpretation of their symptoms as psychogenic, even though they seem to be relatively unconcerned about them. Depending on the triggering situation, the symptoms may subside in weeks or months, but they may recur in situations of chronic stress.

Younger children are treated indirectly through the influence of their parents, while school-age children are treated directly, particularly when the symptoms are acute. The somatic manifestations are accepted as such, and therapeutic use is made of their regression under placebo treatment and suggestion. Attention is paid to minimizing the patient's gain from illness and to building up an inner conviction of being able to control the symptoms. The psychogenic interpretation is brought into play gradually as cognitive maturity develops.

4.8 Attention Deficit Hyperactivity Disorder

Longitudinal studies show that remission eventually occurs in approximately 50% of cases of pure attention deficit disorder or of simple attention deficit hyperactivity disorder. The most important positive prognostic factor is the absence of concurrent oppositional defiant or aggressive-antisocial disorders, as the combination of a hyperkinetic disorder with a conduct disorder is considered to have a negative prognosis.

It is currently believed that the combined disorder is more strongly genetically determined (Silberg et al. 1996), while simple attention deficit hyperactivity disorder is associated with faulty maturation of the central nervous system. The prevalence of these disorders in the primary-school age group is approximately 3%; they occur predominantly in boys.

The diagnosis and treatment of the prognostically less favorable combined hyperkinetic disorders, which confer a significant degree of handicap, are discussed in Sect. 3.10. It is unclear whether hyperkinetic disorders have a greater tendency to overlap with tics, stuttering and specific developmental disorders when they are associated with conduct disorders.

4.9 Age-Specific Phobias

Monosymptomatic (specific) phobias are related to particular objects and situations. The commoner kinds include animal phobias (e.g. of spiders or dogs) as well as fear of enclosed spaces (claustrophobia), of large public places (agoraphobia), of the dark, and of specific situations. Monosymptomatic phobias may arise in childhood, adolescence, or early adulthood.

Phobic disorders of childhood are usually abnormally intense versions of normal age-specific anxieties, e.g. fear of animals, of the dark, or of storms; panic symptoms accompany them only rarely. Between 7% and 10% of children entering school age suffer from such disorders; boys and girls are approximately equally affected. It is unclear whether a genetic component plays a role in the process of abnormal intensification of anxieties (Beidel and Turner 1998) Avoidance behavior on the parents' part in dealing with these symptoms appears to be significant. It is of importance for the diagnosis that the phobic manifestations are no longer appropriate to the child's age or developmental stage. Their spontaneous course is

favorable, and treatment is necessary only in some cases, when the child's ability to function in everyday life is impaired. Behavior-therapeutic techniques such as prolonged exposure, systematic desensitization and model learning require that both the child and the parents are motivated to cooperate.

Social phobias are common in adolescence and early adulthood. As certain social situations – e.g. scholastic tests, eating or speaking in public, getting together with the opposite sex, and public appearance of any kind – become increasingly important in the lives of adolescents, fear and anxiety become directed toward them. Affected adolescents may fear that such critical situations will cause them to feel faint or vomit, or that they will be exposed to ridicule. Such fears are associated with the usual physiological manifestations of anxiety. In contrast to other phobias, social phobias are approximately equally common in boys and girls.

The etiology and pathogenesis of monosymptomatic (specific) phobias and social phobias are similar. There are three arguments favoring the hypothesis that constitutional and genetic factors play a major role:

1. Twin studies show that anxiety syndromes generally manifest a higher concordance in mono- than in dizygotic twins; the same is true of social anxieties (Torgersen 1979).
2. Phobic patients are more excitable in general, and less capable of habituation, than normal control subjects (Lader and Wing 1966).
3. Anxiety disorders (not only phobias) tend to occur in families, and the children of anxious parents are more likely to be anxious.

Predisposing psychological and psychosocial factors include premorbid personality traits (contact avoidance, depressive tendency, social withdrawal), family dynamics (similar behavior among family members) and conditioning processes. According to the concept of "preparedness," phobic content can be explained as the product of evolution. Phobia is always directed toward objects that were associated with danger during the evolutionary development of the species (e.g. dangerous animals, heights, sharp objects), never toward modern technological appliances (e.g. washing machines, cars, radios, or television sets). Cognitive theories of emotion (Lazarus 1966, 1981) rest on the assumption that anxiety arises from the cognitive assessment of a situation as dangerous.

The diagnosis is made on the basis of the history and clinical manifestations, which can, in part, be directly observed. To be excluded in the differential diagnosis are other anxiety syndromes and disorders of other types in which anxiety plays a major role (e.g. schizophrenic psychoses, depressive disorders).

Various behavior-therapeutic strategies are the treatment methods of first choice for both monosymptomatic and social phobias. Success has been reported with systematic desensitization and, in particular, with "flooding" (exposure and prevention of reaction). Very recently, cognitive strategies, with or without relaxation exercises, have come to the fore. Among these are various forms of training in problem-solving and self-instruction, combined with self-confidence training. Cognitive strategies seem to be inferior to the method of exposure and reaction prevention (Reinecker 1990).

Drug treatment is also available. Two substance groups are most commonly used, the antidepressants and the benzodiazepines. Patients with generalized anxiety disorders respond better to antidepressants than do patients with monosymptomatic anxiety (Zitrin et al. 1983). The appropriateness of antidepressant administration depends largely on whether depression is present as an accompanying manifestation. The use of benzodiazepines in adolescents is only recommended in severe cases because of their potential for addiction, and only in the short term (for no more than 6 weeks).

As for the course of these disorders, the phylogenetically explicable fears (animal phobias, claustrophobia, fear of heights) seem to persist longer. Social phobias, if untreated, also tend toward chronification (Agras et al. 1969). Favorable prognostic factors include the absence of a depressive mood disturbance, consistent avoidance of the anxiety-producing situation, motivation to undergo treatment, and satisfactory progress in the initial treatment sessions (Marks 1987).

4.10 Social Hypersensitivity

This disorder consists of a pathological exaggeration of age-typical worries and anxieties relating to contact with strangers or social situations that seem threatening. Its prevalence is 2.5% (although a study of adolescents alone revealed a higher value), and boys are more commonly affected than girls.

The pathogenesis of the disorder seems to lie in the intensification of age-typical avoidance of strangers by temperamental traits. The major factor is the use of ineffective social strategies by the sufferer, rather than rejection by others. Role models suffering from social anxiety, and reinforcement of avoidance behavior in the affected person's upbringing, are further contributing factors. Traumatic experiences are rare as triggering factors. Hypersensitivity may be directed toward peers, older individuals, adults, or strangers in general. Attachment behavior is undisturbed, but the social circle is frequently small because of avoidance. The lack of social experience favors accompanying symptoms of anxiety and a feeling of incompetence. In

older children and adolescents, social status is negatively affected.

The differential diagnosis requires the separation of this disorder from social phobias consisting of specific fears related to the other individual – these usually begin in late childhood – and from acute reactive disorders (see Sect. 2). The need for treatment is a function of the degree of social impairment. Unfavorable prognostic signs include the (rare) occurrence of panic symptoms, familial loading with anxiety disorders, worsening of symptoms over time, and their failure to improve with social support. A transition from social hypersensitivity to social phobia is rare, but this possibility should be borne in mind (Beidel and Turner 1998). The goals of treatment are the acquisition of competence through model learning, the correction of distorted social perceptions and the strengthening of the inner conviction of self-control. The treatment is supplemented, if necessary, with exposure and reaction prevention.

4.11 Oppositional Defiant Disorder and Socialized Conduct Disorder

The significance of oppositional disorders is unclear. Persistently disobedient and rebellious behavior with a hostile component but without aggressive or antisocial acts is often considered to be a possible preliminary phase of a conduct disorder. Oppositional disorders appear mainly in the early school years and affect boys more often than girls. Severe accompanying emotional abnormalities are rare, chronic stress and deficient education are common, and temperamental factors are considered to predispose to these disorders. A disturbance of this type can be classified among the transient age-specific disorders only if there is no accompanying hyperkinetic disorder. There are no reliable epidemiologic data for oppositional defiant disorder. The oppositional behavior is directed mainly toward authority figures in the near environment, whose demands are passively or actively disregarded. The anger of others is deliberately aroused, and others are blamed for the resulting difficulties. Provocations are frequent, and a low frustration threshold leads to frequent temper tantrums.

The emergence of conduct disorders in early adolescence is of prognostic significance; according to Moffitt (1993), approximately 85% of conduct disorders appear at age 10 or above. The affected adolescents, despite their persistently antisocial behavior, are able to make social attachments and often have good relations with their peers. Conversely, the lack of social attachments to peers, physical aggression, and the combination of an early-onset conduct disorder with a hyperkinetic disorder are frequently seen features that, when present, confer an unfavorable prognosis (Biederman et al. 1996). The prevalence of these disorders rises from 1–2% in the primary school years to approximately 6% in adolescence. Boys are more commonly and more severely affected than girls. There is a risk of substance abuse and delinquency. Lack of social correction promotes the persistence of the behavioral disturbance. Its manifestations range from fighting, bullying, vandalism, lying, stealing, outbursts of rage, and truancy to running away from home, setting fires, physical assault, and other criminal offenses. Mild accompanying emotional disorders are common; transient, severe depressive disorders also occur as comorbidity.

Unfavorable temperamental traits and faulty upbringing, usually consisting of a lack of sufficient parental control despite adequate affection, favor the development of oppositional behavior in childhood. Antisocial role models in the child's own family or (more commonly) among peers, failure at school and social rejection are clearly major predisposing factors for the development of age-specific antisocial behavior, which only manifests itself in specific contexts (e.g. not at school or in the workplace; Matt 1995). Less is known about the origin of this disorder than about its course.

The diagnosis is made on the basis of the patient's history and the exclusion of primary disturbances of cognitive performance (e.g. deficient verbal differentiation of affective experience) and hyperkinetic syndromes. The context in which the abnormal behavior arises requires careful attention, as do the extent and quality of social attachments and possible membership in groups engaging in deviant behavior. Persistence of the abnormal behavior is another important factor for the diagnosis; a single incident of delinquency does not suffice. The abnormal behavior must not be a component or consequence of another psychiatric disorder.

The oppositional defiant disorders are considered to require treatment because of the risk that they will be transformed into conduct disorders (which has not been adequately quantified to date). Conduct disorders beginning on the threshold of adolescence, accompanied by good social contact and the ability to form attachments, are considered to have a good prognosis. Such disorders generally remit during adolescence in girls, but may not remit in boys until the mid-twenties. They require treatment because of the risk of substance abuse, which increases as a function of the number of antisocial manifestations. As stated by Patterson et al. (1989), the measures that have proved effective in the treatment of oppositional defiant disorders are generally directed toward the parents' child-rearing behavior. Important factors for successful treatment include

the involvement of significant persons in the child's upbringing, the analysis of misperceptions, the development of alternative behaviors in both the child and the parents, the acquisition by the parents of effective means of control and discipline, and the maintenance of success by continuous repetition. Other measures directed toward the child's problem-solving ability are more appropriate for older children and adolescents. The treatment of substance abuse has priority. The treatment of adolescents may be focussed on specific manifestations such as truancy, stealing, and setting fires. If everyday behavior is impaired in multiple ways, therapy involving multiple systems is necessary. Field interventions improve the prospects for success (for a discussion of the current state of research on treatment effectiveness, cf. Kazdin 1997).

4.12
Crises of Sexual Maturation

This term covers emotional – often, anxious or depressive – reactions to the individual's own sexual orientation. The reaction is usually to a deviant sexual orientation, identity, or preference. These disorders occur in adolescence, when the individual first becomes aware of such deviations, and become significant because the establishment of sexual identity is a developmental task in this phase of life. The adoption of such deviant sexual roles may be more difficult than the establishment of heterosexual identity, and is often associated with subjective suffering. There are no reliable epidemiological data on the frequency of such crises.

The diagnostic difficulty lies in the differentiation of transient disturbances in the course of sexual development from permanent sexual attitudes. The former may be perceived as transitional manifestations requiring no treatment; while the latter may require counseling or psychotherapy, of which the main objective is the development of self-acceptance. These tasks can be accomplished only on the basis of meticulous history-taking with respect to psychosexual development. The therapist must also possess a thorough knowledge of the relevant aspects of developmental psychology, including behavioral variants, as well as experience in dealing with adolescents with complex problems (Rekers 1982).

5
Disorders of Age-Specific Onset

The following sections deal with various disorders whose common feature is that they begin in childhood or adolescence and continue into adulthood. Their manifestations may remain constant or may change over the course of the transition into adulthood. The latter is the case for some of the tic disorders, for example, which may continue into adulthood as obsessive–compulsive disorders, or for those eating disorders that begin as anorexia nervosa and then become transformed, as the patient grows older, either into a subclinical form of anorexia nervosa, or into bulimia nervosa.

5.1
Disorders of Speech Fluency

5.1.1 Stuttering

Stuttering is a usually situation-dependent disturbance of speech fluency. Its ICD-10 classification is F98.5.

Stuttering is found in approximately 5% of 5-year-old boys and 2% of 5-year-old girls. It may begin during the phase of language development (between the third and the sixth years), in the early school years (sixth and seventh years), or during puberty (twelfth through fourteenth years).

Three forms of stuttering are distinguished:

1. *clonic* stuttering, i.e. repetition during the initiation of speech;
2. *tonic* stuttering, i.e. blockage in the course of speaking; and
3. *tonic-clonic* or combined stuttering, in which both of the above are present.

The most important manifestation for the diagnosis is the patient's tonically or clonically distorted speech. Tonic stuttering is distinguished by clearly evident straining at the beginning of words or sentences, clonic stuttering by the repetition of sounds, syllables, or words. Often, both components are found. The disorder is easily identified as a disorder of speech fluency. Moreover, there are usually difficulties in the coordination of other functions coupled to the process of speaking (breathing, facial expression, gestures, vocalization). Finally, nearly all patients have major autonomic reactions while speaking (sweating, blushing, irregular breathing). It is an important observation that stuttering does not occur during singing in the great majority of cases; this can also be put to therapeutic use. In the differential diagnosis, stuttering must be distinguished from cluttering. Stammering, a disturbance of sound formation, is easily distinguished from stuttering, in which sound formation is normal.

There are many hypotheses concerning the origin of stuttering, which can be summarized under the following four headings:

1. Recent studies have shown that the influence of *heredity* in the stuttering syndrome is less than previously thought. A genetic component is thought to be probable in approximately 8% of cases.
2. *Organic brain abnormalities* are present in a considerable percentage of patients who stutter. These are revealed by abnormal EEG findings, which may be found in a relatively high percentage of stutterers. The hypothesis of an organic etiology is also supported by the fact that stuttering occurs in as many as 20% of patients with brain damage.
3. *Psychogenic influences* certainly play a major role in the appearance of this disorder. According to Schilling (1963), most stutterers (68%) suffer from neurotic disorders, and psychogenic influences cannot be definitively ruled out in 12%, while psychogenic influences can be definitively ruled out in only 20%.
4. According to the *multifactorial hypothesis*, several factors acting together cause the disorder. For example, a mild cerebral dysfunction may act in concert with aggravating environmental influences to produce stuttering.

Just as the hypotheses as to the etiology of stuttering are diverse, so, too, are the approaches to its treatment. These may be summarized under the following five headings. Often, several different treatment methods are used in combination:

1. *General measures* primarily concern parents and other persons interacting with the child. The basic principles include: disregard of stuttering, attentive and patient listening when the child speaks, and elimination or moderation of the circumstances that are presumed to have led to stuttering or to maintain it.
2. Many *functional treatment methods* have been used for stuttering, with variable degrees of success, including: breathing exercises, speech therapy, the use of rhythm, relaxation exercises (e.g. autogenic training), and application of the principle of delayed speech feedback.
3. *Psychoanalytic treatment* is based on the assumption that stuttering is an expression of a neurotic disorder resulting from experiences of conflict. Resolution of conflict is the goal of analytically oriented play therapy.
4. From the *behavior-therapeutic* point of view, stuttering is regarded as a learned mode of behavior that may be eliminated by treatment based on learning principles.
5. *Drug treatment* generally cannot eliminate the manifestations of stuttering in the etiological sense, but it can definitely reduce the associated anxiety and tension. The most frequently used agents are chlordiazepoxide (Librium) and butyrophenone derivatives such as haloperidol (Haldol).

As for prognosis, one third of patients become asymptomatic after prompt and intensive treatment; one third are considerably improved; and one third remain unchanged.

5.1.2 Cluttering

Cluttering is a disorder of speech fluency characterized by rushed speech, swallowing of sounds, syllables, or words, and garbling of various components of speech. It has been given the ICD-10 classification F98.6.

The process of speaking is accelerated and hurried, and the flow of speech is chaotic. The patient's speech is sometimes difficult to understand because of slurred articulation or distortion of words. In contrast to stuttering, speech improves when the clutterer pays attention to it. The disorder occurs in approximately 1–1.5% of 7-year-olds; prevalence declines with increasing age.

The disorder is mainly attributed to hereditary influences or organic brain abnormalities. The frequent familial occurrence of the disorder is evidence for the former. Evidence for the latter is provided by a high frequency of electroencephalographically detectable, though unspecific abnormalities in children and adolescents who clutter.

The manner of treatment depends on the predominant component. In pure cluttering, speech therapy is indicated, with particular emphasis on speaking slowly, directing attention to the speaking process, and beating time along with the syllables. The therapist should also make the family aware of the importance of speaking slowly.

5.2 Tic Disorders and Gilles de la Tourette Syndrome

Tics are involuntary, recurrent, circumscribed movements or vocal productions of sudden onset that are experienced by patients as irresistible but may nevertheless be suppressed at times. Gilles de la Tourette syndrome is a special form of tic disorder involving multiple motor tics and at least one vocal tic; the motor and vocal tics need not occur simultaneously. The vocal tics may involve a variety of sounds (e.g. throat clearing, coughing, grunting, or uttering obscene words or phrases).

In what follows, we will maintain the distinction between transient tics, chronic motor or vocal tics, and Gilles de la Tourette syndrome. Transient tic disorders

are very common in childhood and adolescence and are discussed in Sect. 4.

In chronic motor or vocal tic disorder, motor or vocal tics occur, not simultaneously, though possibly in temporal succession. In contrast to transient tic disorder, the manifestations persist for more than one year and usually involve multiple tics.

The syndrome originally described by Gilles de la Tourette in 1885 is characterized by multiple motor tics and phonation tics (i.e. the uttering of unarticulated sounds, or coprolalia). The disorder always begins in childhood and often continues into adulthood. Like other tic disorders, Gilles de la Tourette syndrome is 2 to 3 times more common in male children and adolescents than in females. The average age of onset is approximately 7 years. The motor tics most frequently involve the head and face, less frequently the trunk or the upper or lower limbs. The vocal tics that are always associated with the disorder may involve sounds of various types, words, coprolalia, or palilalia.

The illness usually begins as an isolated motor tic (in 50% of patients), which is most often localized to the face. Other possible initial manifestations include sticking out the tongue, sniffling, hopping, throat-clearing, stuttering, explosive production of sounds or words, and coprolalia (Shapiro and Huebner 1985). It is typical of the disorder for the initial manifestations to be replaced by others over the course of time (Hebebrand et al. 1997a,b). The severity of the disorder (i.e. the intensity of its manifestations) is highly variable over time.

Current hypotheses of the etiology of Gilles de la Tourette syndrome (for review see Leckman and Cohen 1999) are based on genetic and neurobiological findings, while environmental factors seem to play a role in determining the severity of the disorder. As for genetic causes, family studies have revealed that first-degree relatives of index patients with Gilles de la Tourette syndrome have a 5% probability of having it themselves, and a 10% probability of having a chronic tic disorder. Twin studies have revealed concordance rates of approximately 50% for monozygotic twins. An autosomal dominant inheritance pattern was postulated on the basis of family studies, but has not been confirmed to date; the disorder seems to be polygenic (Tourette Syndr. Assoc. 1999). Neurobiological hypotheses, based on the clinical features of the disorder, invoke possible functional disturbances in the basal ganglia; tics are thought to result from disturbed neural activity in subcortical regulatory circuits. The biochemical findings to date are not uniform. Recently, there has been interest in the immunological findings of Sydenham's chorea, a disorder in which children develop tics after a rheumatic infection, which may be considered a model for Gilles de la Tourette syndrome.

Items to be excluded in the differential diagnosis include other disorders producing choreiform movements (including Sydenham's chorea, whose relation to Gilles de la Tourette syndrome has recently attracted interest, as just mentioned). Myoclonus, which sometimes occurs continuously, must also be excluded; when present, it is usually due to another neurological illness, whose other manifestations simplify the differential diagnosis. A special variety of myoclonus is the type that normally occurs while the child is falling asleep. Finally, epileptic seizures and, not least, obsessive-compulsive manifestations should be excluded. The latter may be difficult to distinguish from Gilles de la Tourette syndrome, for two reasons: Obsessive-compulsive phenomena may occur as part of the syndrome (e.g. continuously repeated words, sentences, or melodies); moreover, in some 30% of cases of the syndrome, its manifestations change in the direction of an obsessive-compulsive disorder, or else an obsessive-compulsive disorder arises in addition to the already present Gilles de la Tourette syndrome (Leckman and Cohen 1999).

Tic disorders, because they are highly visible, often lead to considerable social impairment. Patients and their parents are thus highly motivated to undergo treatment. The type of treatment recommended depends on the severity of the manifestations and of the associated social impairment. Treatment usually consists of three components: thorough education as to the nature of the illness, drug therapy, and individually tailored psychotherapy.

D_2-receptor antagonists such as tiapride, haloperidol, and pimozide are the mainstays of drug therapy, while α_2-adrenergic agonists such as clonidine are used as agents of second choice. Clonidine probably affects the dopaminergic system indirectly through an alteration of norepinephrine metabolism. Drug therapy should be initiated and terminated gradually. Its goal should not necessarily be the complete elimination of tics, but rather a reduction of manifestations so that the patient can get along in everyday life (Remschmidt and Hebebrand 1993). If obsessive-compulsive features are also present, antidepressants such as clomipramine or selective serotonin reuptake inhibitors such as fluoxetine and fluvoxamine can be tried, perhaps in combination with neuroleptics. As far as psychotherapeutic treatment is concerned, behavior therapy has proved useful.

The prognosis of tic disorders is favorable, with the exception of Gilles de la Tourette syndrome. Transient tics tend to regress spontaneously: after a follow-up interval of three years, 70% of patients have major improvement or even disappearance of the tics. The manifestations persist in approximately 30% (Remschmidt and Remschmidt 1974). In Gilles de la Tourette syndrome, approximately 30% of cases have a transi-

tion to obsessive-compulsive manifestations. In the great majority of cases, the syndrome persists, with a fluctuating tendency toward improvement.

5.3 Conduct Disorders, Antisocial Disorders, and Delinquency

Conduct disorders are characterized by antisocial, aggressive, and rebellious behavior that appears repeatedly and displays a tendency to persist. In the ICD-10, four types of conduct disorders are defined:

- conduct disorder confined to the family context (F91),
- unsocialized conduct disorder (F91.1),
- socialized conduct disorder (F91.2), and
- oppositional defiant disorder (F91.3).

The first three types of conduct disorder are defined in relation to the social sphere and the relationship structure in which each occurs, while the last is a characteristic pattern of manifestations that can appear in practically any social sphere and is not bound to a specific type of relationship. Socialized conduct disorder and oppositional defiant disorder are strongly developmentally dependent and are accordingly described more extensively in Sect. 4. Conduct disorders almost always begin in childhood and early adolescence and, when severe, often undergo a transition to personality disorders in the adult.

The question must be raised whether it is truly a clinical entity that is being discussed here, because conduct disorders often must be regarded, not as psychiatric illnesses, but rather as a social failure caused by faulty education, and should accordingly be treated by educational means. In any case, the manifestations of these disorders are extremely diverse. The ICD-10 research criteria (Dilling et al. 1994) list no fewer than 23 manifestations and modes of behavior that can be found in these disorders, ranging from frequent arguing, lying, physical conflicts, persistent violation of rules set by parents, cruelty to animals, stealing, truancy, running away, and bullying to actual criminal offenses such as burglary, arson, and the use of dangerous weapons (e.g. sticks, knives, guns).

Two types of conduct disorders are traditionally distinguished: those beginning in childhood (usually before the 10th year) and those beginning in adolescence (from the 13th and 14th years onward). Conduct disorders may be classified as one of the four types listed above only if the general criteria for a conduct disorder (F91) are fulfilled. It can then be decided, on the basis of the more specific criteria for each disorder, whether the conduct disorder is confined to the family context (F91.0), whether it is associated with a lack of social relationships (F91.1) or with existing social relationships (F91.2), or whether it manifests itself primarily in oppositional and defiant behavior (F91.3).

The ICD-10 research criteria also provide a scale of severity for conduct disorders, including the three levels mild, moderate, and severe. It is difficult, and sometimes problematical, to determine the boundaries between these disorders and variations of normal behavior, which cover a wide range in childhood and adolescence. The division of conduct disorders into socialized and unsocialized, introduced by Jenkins (1968), has been found to be worth retaining, as it yields a certain degree of insight into the dynamics of these disorders; socialized conduct disorder has a better chance for improvement, and thus a more favorable prognosis, than unsocialized conduct disorder.

As in most mental disorders and illnesses, as well as most behavioral deviations, the causation of conduct disorders is thought to involve an interaction of biological and environmental factors, in which the number and type of risk factors in both areas influences both the probability that these disorders will arise, and their severity. The interrelatedness of the individual risk factors must be borne in mind, of course, as well as the individual's active attempt to deal with his or her social environment. The following etiological hypotheses have been proposed:

- *Biological hypotheses* emphasize on disorders of cerebral function, neuropsychological abnormalities, chromosomal aberrations, genetic factors, and constitutional variables. The last category also includes personality, whose precursors in childhood are summarized under the term "temperament." Twin studies and adoption studies indicate the presence of a genetic factor. Neuropsychological and neurophysiological studies have revealed that children and adolescents with severe conduct disorders have both neuropsychological deficits and a diminished arousal response. The more common occurrence of antisocial behavior in male children and adolescents is attributed to hormonal factors, as it is known that androgens not only influence brain development, but also may be responsible for aggressive behavior in male adolescents. Furthermore, the autonomic nervous system has been found to be less reactive in adolescents with conduct disorders, as revealed by a diminished reactivity of the pulse and of skin conductivity.
- *Psychogenic hypotheses* lay emphasis on the influence of early childhood development. Deprivation and unfavorable family circumstances and, in adolescents, peer influence (antisocial cliques), drug abuse, and depictions of violence in the media are considered to be important factors. Several studies have shown that the moral development of delinquents is delayed.

- *Sociological hypotheses* focus attention on the influence of unfavorable social circumstances, bad role models and family desolation.

None of the approaches listed above have yet been able to provide an adequate explanation for conduct disorder, antisocial behavior, and delinquency.

The diagnosis is made on the basis of the past history, clinical manifestations, and, for a number of disorders, direct observation. Particular attention must be paid to developmental conditions. Conduct disorders must be distinguished form personality disorders and antisocial and delinquent behavior in the context of other psychiatric illnesses (e.g. psychoses).

In accordance with the presumed multifactorial origin of conduct disorders, their treatment must be based on a number of different factors. Treatment methods are divided into three types, after Kazdin (1995), according to the major focus of the intervention: child-focused, family-focused, and community-based (see Table 2).

- *Child-focused methods* are currently based mainly on behavior-therapeutic techniques; training in problem-solving has been found to be effective. In this approach, social situations are acted out with the child and constructive outcomes of conflicts are anticipated, so that the child will behave in the actual situation as previously practiced during problem-solving training.
- *Family-focused methods* are most often used in younger children with conduct disorders. Parental training is intended to improve the parents' ability to bring up their child, while family therapy, which is less commonly performed, is intended to improve family relationships.
- *Community-based methods* include the participation of the affected child or adolescent in group activities in and out of school and the workplace, so that social behavior in peer groups can be practiced under supervision, and the implementation of youth service measures by the responsible government offices. The German Law on Child and Youth Service (*Kinder- und Jugendhilfegesetz, KJHG*) provides for various measures, ranging from assistance with child-rearing (Sect. 27), parental counseling (Sect. 28), supervision of parents (Sect. 30), socioeducational family assistance (Sect. 31), and day care (Sect. 32) to full-time care (Sect. 33), care in a children's home (Sect. 34), and intensive individual socioeducational care (Sect. 35 KJHG).

Longitudinal studies of children with conduct disorders have shown that about half of all cases remit, and half persist into adulthood. Conduct disorders beginning in the pre-school years have an unfavorable prognosis and are associated with a lack of social relationships and with impulsive and hyperactive behavior. The prognosis is further worsened by the frequent comorbidity of these disorders with alcohol and drug abuse. Some cases of severe conduct disorder are evident in adulthood as personality disorders, mostly of the dissocial type.

5.4 Eating Disorders: Anorexia Nervosa and Bulimia Nervosa

Anorexia nervosa and bulimia nervosa are common eating disorders in childhood and adolescence. The age-specific incidence of anorexia nervosa is highest at 14 years, that of bulimia nervosa at 17 to 18 years. The prevalence of anorexia nervosa has been reported to be 1%, while bulimia nervosa is more common, occurring in approximately 2% to 3% of all boys and girls. Both illnesses predominantly affect females.

As seen in Table 3, the most important feature distinguishing bulimia nervosa from anorexia nervosa is the occurrence of hunger attacks. Longitudinal studies have shown that a transition from anorexia nervosa to bulimia nervosa is common, while the reverse is extremely rare. Reviews are found in Garner and Garfinkel (1997) and in Vandereycken and Beumont (1998).

The major components of the clinical picture of anorexia nervosa are marked weight loss and refusal of food, abnormal eating behavior, constipation, and absence of menstruation. These patients try at all costs to achieve progressive weight loss or to maintain an extremely low weight. They do so not only by refusing to eat, but often also by self-induced vomiting or laxative abuse. Many patients also try to maintain a low weight by keeping the body in motion at all times (constantly pacing back and forth, extreme gymnastics). Fasting may be the exclusive means of attaining the desired weight loss (so-called restrictive form of anorexia nervosa), or additional means of reducing nutritional intake may be used, such as vomiting and laxative or diuretic abuse.

The mental abnormalities characteristic of these girls include an asthenic personality, an inclination to depressive mood disturbances, marked ambition to succeed, and, frequently, hysterical or schizoid personality traits, along with usually normal intelligence. The depressive mood disturbances are mainly a function of weight (Herpertz-Dahlmann and Remschmidt 1989). Extreme weight loss is associated with a massive change of attitude and an often entirely unrealistic body image.

The physical manifestations are all a function of weight and mostly regress after the patient's weight is normalized. These include acrocyanosis, hair loss, hematopoietic abnormalities (leukopenia, anemia), electrolyte disorders, low levels of all hormones except cortisol, osteoporosis, and cerebral pseudoatrophy,

Table 2. Antisocial behavior: therapeutic focus and processes of major classes of treatment. (After Kazdin 1987)

Treatment type	Focus	Key processes
Child focused		
Individual psychotherapy	Focus on intrapsychic bases of antisocial behavior, especially conflicts and psychological processes that were adversely affected over the course of development.	Relationship with the therapist is the primary medium through which change is achieved. Treatment provides a corrective emotional experience by providing insight and exploring new ways of behaving.
Group therapy	Processes of individual therapy, as noted above. Additional processes are reassurance, feedback, and vicarious gains by peers. Group processes, such as cohesion and leader.	Relationship with the therapist and peers as part of the group. Group processes emerge to provide children with experiences of others and opportunities to test their own views and behaviors.
Behavior therapy	Problematic behaviors presented as target symptoms. Prosocial behaviors are trained directly.	Learning new behaviors through direct training, modeling, reinforcement, practice, and role playing. Training in the situations (e.g. at home and in the Community) in which the problematic behaviors occur.
Problem-solving skills training	Cognitive processes and interpersonal cognitive problem-solving skills that underlie social behavior.	Teach problem-solving skills to children by engaging in a step-by-step approach to interpersonal situations. Use of modeling, practice, rehearsal, and role playing to develop problem-solving skills. Development of an internal dialogue or private speech that uses the processes of identifying prosocial solutions to problems.
Pharmacotherapy	Designed to affect the biological substrates of behavior, especially in the light of laboratory-based findings on neurohumors, biological cycles, and other physiological correlates of aggressive and emotional behavior.	Administration of psychotropic agents to control antisocial behavior. Lithium carbonate and haloperidol have been used because of their antiaggressive effects.
Residential treatments	Means of administering other techniques in day treatment of residential setting. Foci of other techniques apply.	Processes of other techniques apply. Also, separation of the child from parents or removal from the home situation my help reduce untoward processes or crises that contribute to the clinical problem.
Family focused		
Family therapy	Family as a functioning system rather than the identified patient serves as focus. Interpersonal relationships, organization, and roles and dynamics of family.	Communication, relationships, and structure within the family and processes as autonomy, problem solving, and negotiation.
Parent management training	Interactions in the home, especially those involving coercive exchanges.	Direct training of parents to develop prosocial behavior in their children. Explicit use of social-learning techniques to influence the child.
Community based		
Community-wide interventions	Focus on activities and community programs to foster competence and prosocial peer relations.	Develop prosocial behavior and connections with peers. Activities are seen to promote prosocial behavior and to be incompatible with antisocial behavior.

demonstrable by both computer tomography and magnetic resonance imaging.

The essential feature of bulimia nervosa is the episodic occurrence of hunger attacks in which large amounts of food of high caloric content are consumed and then usually expelled by means of self-induced vomiting. The patients are aware that their eating behavior is abnormal. They are afraid of not being able to stop eating, often suffer from depressive mood disturbances and self-reproach, worry about their

Table 3. ICD-10 criteria for anorexia nervosa and bulimia nervosa

Anorexia nervosa	Bulimia nervosa
1. Body weight at least 15% under norm, and body mass index ≤ 17.5.[a]	1. Constant preoccupation with eating and hunger attacks in which large quantities of food are consumed in a short time.
2. Weight loss is intentional.	2. Attempts to counteract the fattening effect of eating by various means such as self-induced vomiting, laxative abuse, restrictive diet, etc.
3. Disturbed body image and excessively stressed idea of being too fat.	3. Morbid fear of becoming too fat.
4. Endocrine disorder of the hypothalamic–pituitary–gonadal axis.	4. Often, a past history of anorexia nervosa.
5. In cases beginning before puberty, disturbance of pubertal development, including growth, which is often reversible after remission	

[a]Body mass index = body weight (kg)/[height (m)]^2.

weight, and often have abdominal pain and panic-like anxiety attacks, in addition to guilt feelings after eating. Their eating behavior also often leads to fluctuations of weight, although the average weight is higher than that of patients with anorexia nervosa. Many patients are, in fact, overweight.

Both of these varieties of eating disorder are currently thought to be of multifactorial origin. Twin studies and family studies reveal the importance of genetic factors, although it remains unclear whether the two disorders have a common genetic basis. The psychological and psychosocial influences favoring the occurrence of these disorders include: the esthetic ideal, current in Western countries today, of thinness, which comes into conflict with the increase in body fat that occurs during puberty and is more marked in girls than in boys; the often self-imposed pressure to succeed that is common among girls in our society; premorbid eating disorders in childhood; and family influences, which are now considered to be less important than previously thought. The importance of sociocultural factors is demonstrated by the higher prevalence of anorexia nervosa and bulimia nervosa among immigrants to Western countries and by the increased prevalence of the disorder in groups at special risk because of an exaggerated ideal of thinness (e.g. ballet dancers, gymnasts, actresses and models).

Sexual abuse is often considered to be a factor contributing to the occurrence of anorexia nervosa and may indeed be so in some cases, but it is no more common in girls suffering from anorexia nervosa than in girls with other psychiatric disorders. Thus, sexual abuse cannot be regarded as a specific etiological factor.

Excessive vulnerability of the serotonergic system has been postulated as a common underlying mechanism for both disorders. This assumption is based on the observation that patients with anorexia nervosa who have regained normal weight have elevated central serotonin activity, while patients with bulimia nervosa have diminished central serotonin activity. The hypothesis of faulty serotonergic regulation is also compatible with observations on comorbidity in the eating disorders, which are associated with depressive mood disturbances and with anxiety disorders.

There are both similarities and differences between the modes of treatment of the two disorders. The major initial objective of treatment is weight normalization in the case of anorexia nervosa, normalization of eating behavior in the case of bulimia nervosa. Treatment with antidepressants is not effective in anorexia nervosa (despite the frequent comorbidity with depressive disorders), but it has been found useful in bulimia nervosa, for which selective serotonin reuptake inhibitors (e.g. fluoxetine, fluvoxamine) have been used with success. The most important aspect of the treatment of both disorders is the integration of all treatment measures in an overall plan, which usually consists of several phases. The severity of the disorder and the patient's willingness to cooperate, and that of her family, determine whether treatment is best provided in an ambulatory or inpatient setting.

A multiphase treatment plan has proved useful in the treatment of anorexia nervosa (Remschmidt and Herpertz-Dahlmann 1988; Remschmidt 1992). The first four phases of clinical treatment generally take place over four to six months:

1st phase: Elevation of body weight.
2nd phase: Introduction of an eating plan agreed to by the patient, with monitoring of compliance by treating personnel.
3rd phase: "Self-determination of food intake": in this phase, the patient learns to control her own food intake and eating behavior.
4th phase: In this phase, the emphasis is placed on joint family effort for younger patients, and on the development of independence in all areas of life for older patients. At the same time, preparations are made for discharge.
5th phase: Oupatient follow-up and continuation of treatment measures, as needed, for a further 2 years.

The treatment of bulimia nervosa consists mainly of psychotherapy with cognitive behavior-therapeutic methods. A meticulous problem analysis should be

performed in every case, and the individual problems should be addressed with suitable specific measures. As in anorexia nervosa, a multiphase treatment plan is recommended (Fichter 1989):

1st phase: At first, the major emphasis is on the cessation of pathological eating behavior. This goal is served by the development of a stable relationship with the therapist, the introduction of a structured eating plan with regular meals, thorough education of the patient about the illness and the damage it may cause, the involvement of significant others, and the establishment of alternative modes of behavior, particularly for the times at which the bulimic attacks occur.

2nd phase: This phase is devoted to the discovery of the factors that enabled the disturbed eating behavior to become chronic. A careful causal analysis is performed of all maintaining factors, e.g. inadequate conflict resolution strategies, distortions of body image, and depressive mood disturbances.

3rd phase: Finally, the achieved therapeutic progress is maintained, and future problem situations are anticipated and dealt with constructively through the use of appropriate strategies.

The many longitudinal studies of anorexia nervosa currently available reveal that two thirds of patients no longer suffer from the eating disorder after six to eight years of follow-up, while chronification or, at least, persistence occurs in one third; in the latter patients, the disorder still causes impairment, even though it may no longer fulfill the diagnostic criteria for anorexia nervosa. Remission of the disorder is, in many cases, not synonymous with a return of mental health: even six to eight years after diagnosis, 20% to 30% of patients still suffer from anxiety disorders, drug or alcohol abuse, and personality disorders.

In bulimia nervosa, rates of cure of up to 50% have been reported after follow-up intervals of 2 to 3 years. In the remaining half of all patients, the disorder becomes chronified, or additional mental disorders occur, such as depressive syndromes, alcohol and drug abuse, and personality disorders.

5.5 Mutism

In ICD-10, mutistic behavior is classified under the overall heading of "disorders of social functioning with onset specific to childhood and adolescence" (F94).

The term "mutism" refers to the absence of speech despite the ability to speak. Mutistic behavior occurs mainly in children and adolescents but may also be observed in adults. Absence of speech has a wide variety of causes: diminished drive to speak, fear, reaction to conflict, and defiant refusal to communicate, or, on the other hand, organic disorders and psychoses. *Elective mutism* is distinguished from *total mutism:* in the former, the affected person speaks only with selected persons, in the latter with no one.

Elective mutism is a rare disorder that most commonly arises in childhood between the ages of 5 and 9 years, and rarely in adolescence. The syndrome, narrowly defined, has a prevalence of 0.8 per 1000 children in this age group.

The diagnosis is made on the basis of the clinical history and direct observation. Loss of speech because of organic brain damage, intrinsic mutism, schizophrenic psychoses, and partial or total deafness must be excluded in the differential diagnosis.

Recent discoveries belie the previous assumption that mutism is an exclusively psychogenic disorder. The findings of empirical studies suggest that developmental delay, premorbid abnormalities of personality, and family pathology establish the preconditions for mutistic behavior, which may then be triggered by additional external stress.

There is currently a consensus that the treatment of mutism must always be multidimensional, i.e. it must consist of diverse elements, including treatment of the patient as an individual combined with measures involving his or her environment. In accordance with these general principles, the following treatment elements and methods have been applied successfully:

- *Behavior therapy:* The principle of this form of treatment is the operant reinforcement of speaking with adults and therapists, which may subsequently become generalized. Treatment begins with the promotion of nonverbal communication (eye contact, writing and drawing, facial expression and gestures). In the next step, verbal behavior is systematically promoted and rewarded. In recent years, cognitive behavior therapy has been increasingly used.
- *Family therapy:* The involvement of the family in treatment is very important, because family influences play a major role in the causation and release of mutistic behavior. In the family-therapeutic approach, an effort is made to discern the factors and conditions that have brought about the disorder and continue to maintain it. Admittedly, family therapy cannot be the major form of treatment for mutism, because the patient usually refuses to speak in this setting.

- *Psychosocial interventions:* This heading covers all measures involving the child's living environment, including not only the family but also the kindergarten or school environments, extracurricular activities, and peer groups. In the clinical setting, the patient is involved in creative individual and group activities, physical activity and sport. The establishment of normal communicative behavior is thus promoted on all age-appropriate social levels.
- *Psychotropic drugs:* Many clinicians view mutistic behavior as the result of a causal triad of anxiety (social phobia), obsession, and depression. Psychotropic drugs have been used to treat these putative etiologic factors; the currently available data on such treatment are derived mainly from case reports. The medications given include the MAO inhibitor phenelzine (which is not available in Germany) and the selective serotonin reuptake inhibitor fluoxetine, for which positive results have been reported in both case reports and controlled studies. Fluoxetine may thus be regarded as an effective component in a program of treatment for mutistic behavior, though not as the single or most important treatment measure.

The available data on the prognosis of mutism are inconsistent. While a favorable course has been reported for most patients, in others the manifestations may be extremely tenacious and refractory to treatment.

6 Developmentally Dependent Interaction Disorders

The developmentally dependent interaction disorders include induced sleep disorders and feeding disorders (usually at younger ages), sibling rivalry, separation anxiety disorders, Münchhausen's syndrome, and induced delusional disorder. Some authors also include conduct disorders limited to the family context within this group of disorders.

All mental disorders in children and adolescents have a reactive component, i.e. they are related in some way to the social environment, which plays a major role in their causation, their further course, or both. In the interaction disorders, the contribution of the environment is of another type. It consists of transaction mechanisms by which the disorder, once set on its course, is continually maintained by the behavior of an adult in the child's circle, because the signals emitted by the adult are perceived, interpreted, and internalized by the child in such a way that they go on to exert their effects independently. The primary releaser is relatively unimportant; the decisive factor is that another person is required for the maintenance of the disorder. That which becomes autonomous is thus not, as in other disorders, a particular pattern of behavior of the child alone, but rather a pattern of behavior of the child activated by key stimuli coming from the adult, or else a pattern of behavior of the adult activated by key stimuli coming from the child.

The relationship of the child to the adult is properly emphasized by schemes of classification for mental disorders in infancy that provide a second axis to describe the quality of the relationship. A mother who fears that something might happen to her child during sleep, or that the child might have trouble falling asleep or staying asleep, can induce a sleep disorder in the child (in ICD-10, sleep disorders are classified under the category F51). Frequently asking the child whether it has fallen asleep, frequent checking in the night to see whether the child is still asleep, or nighttime feeding because of the idea that the child has been awakened by hunger systematically maintain the sleep disorder. It is characteristic of induced sleep disorders that the child awakens at specific intervals, namely those at which the mother checks after or feeds the child. These disorders thus maintain the periodicity of early infancy, in which the mother–child or child–mother relationship was still symbiotic.

The treatment of this disorder is directed toward the mother, not the child. She must become more secure in the knowledge that nothing will happen to her child at night, that it is not hungry, that it can fall asleep even without her checking, and so forth. Indeed, many prolonged crying episodes of 2- and 3-month-old infants are induced by mothers who believe they must continually carry, move, and be near the child, and thus prevent the children form learning to be alone (Skuse 1984).

The feeding disorders – encoded in ICD-10 as F98.2 – are a second example of induced disorders which generally result from a conflict between mother and child. These disorders are not induced by the child; they are, at most, supported in their initial phase by the child's behavior. The most important causative factor is that the mother wishes to force her idea of proper eating on the child, i.e. that she feeds the child without consideration of its needs. If the child is fed too often, it becomes accustomed to the frequent interruptions of its daily rhythm and nighttime sleep by feedings, which it comes to expect. If the child is fed in excessive quantities, it develops resistance, and eating turns into a negatively charged experience.

Treatment must accordingly begin with the mother's coming to understand her feelings toward the child and the reasons why she feeds it in this manner. Either she has incorrect notions of the child's needs, or she cannot allow the child to regulate its own feeding behavior. It is no accident that the classic examples of age-specific interaction disorders come from the early childhood years, in which the symbiotic relationship is

still very close. We also see cases in which this symbiotic relationship is maintained beyond infancy and early childhood (Skuse 1994).

An example of the latter is furnished by separation anxiety, classified under category F93.0 in ICD-10. The mother induces the child's anxiety when it becomes separated from her and is alone. As a reaction to separation, the child develops the notion that something might befall the mother, that she might go away or never come back. The child's fears, thus aroused, are overvalued by the mother, possibly to such an extent that separation anxiety disorder is produced in its classical form as school phobia, for which the groundwork may be laid at a very early age. This pattern may be carried on as the child grows older and is generally associated with a lack of independence. The child is happy only when its mother is near.

In milder forms, the mechanism of separation anxiety appears when children who have been raised in part by their grandmothers, or fatherless children, become highly dependent on their mothers and thereby fail to develop independence and have difficulty with separation.

A number of authors draw a parallel between separation anxiety disorder and conduct disorders restricted to the family context. They view the latter not as a preliminary phase of a conduct disorder that goes on to spread outside the family context, but rather as a particular type of parent–child interaction in which the parents persist in applying certain notions of child-rearing at the price of becoming victims of their own children, are hit by them, and so forth, as is well known in the "battered parent syndrome." This pattern seems to be more common with adopted children.

Sibling rivalry (ICD-10 classification: F93.3) is a further example relating not to parents, but to siblings. The presence of the sibling, or the mother's paying attention to the sibling, arouses or maintains abnormal behavior in the child, either regressive or aggressive. The child affected by sibling rivalry cannot give in, or get over its unhappiness, until it is made to realize that its predicament is not really so bad.

Münchhausen's syndrome (ICD-10 classification: F68.1) is a factitious disorder in which the individual intentionally produces or feigns manifestations of illness or disabilities, either physical or psychological. In contrast, Münchhausen-by-proxy syndrome is a rare form of abuse in which a caregiver, usually a parent, produces manifestations of illness in the child or another dependent person. The deception is usually repeated on many occasions, with resulting hospitalizations, morbidity, and sometimes death. The victims include not only unknowing infants, but also older children, who are surprisingly tolerant of their parents' actions, presumably because of fears of loss. A false accusation of sexual abuse directed at a divorced father may constitute a variant of Münchhausen-by-proxy syndrome if the child is led to adopt this idea as its own.

The causative chain in induced delusional disorder (ICD-10 classification: F24) is easily understood. The illness of the mother (or other caregiver) frightens the child and makes it susceptible to her pathological desires for interaction.

The treatment of all of these disorders, diverse as they are, requires an intensive diagnostic assessment of the patterns of interaction within the family. In very young children, a disturbed mother-child relationship is especially likely to produce somatic manifestations, for which organic causes must first be ruled out.

Once the mother-child interaction has been carefully analyzed, measures can be chosen to alter it. Behavior-therapeutic principles have been found particularly useful both in parent counseling and in the treatment of the affected children. Family-therapeutic techniques may also be used (joint family conversation, "behavior tasks," etc.).

7
Early-Onset Adult-Type Disorders

The disorders listed under this heading are illnesses typically occurring in adulthood that may arise in late childhood and early adolescence. They are significantly more common in adolescence.

7.1
Anxiety Disorders

Anxiety disorders are characterized by two features: the development of unusually strong and situationally inappropriate anxiety, and equally marked avoidance behavior.

In ICD-10, phobic disorders are classified separately from other anxiety disorders, and two variants of agoraphobia (with and without panic disorder; F40.0) are distinguished. In DSM-IV, panic disorder occupies a dominant position, and is divided into forms with and without agoraphobia. Both classification schemes have a category for anxiety not related to specific objects or situations, i.e. anxiety neurosis or generalized anxiety disorder (F41.1).

In childhood and adolescence, three kinds of anxiety may be distinguished:

- phobic anxieties related to specific objects and situations,
- fluctuating anxieties (anxiety attacks) that are not related to specific objects or situations, and

- generalized anxieties that do not occur episodically, but rather are persistently present and are not related to specific objects or situations.

The latter form of anxiety is also referred to as free-floating anxiety; it is rare in childhood and becomes significantly more common in adolescence.

Among children and adolescents suffering these disorders, there are significantly more girls than boys (the sex ratio is approximately 3:2). As for the age of onset of illness, monosymptomatic (specific) phobias generally arise in childhood (especially animal phobias); social phobias generally arise around the time of puberty and in early adolescence, in which the themes of childhood anxieties undergo a marked, developmentally typical change toward social situations. In what follows, two groups of anxiety syndromes will be discussed: (1) panic attacks and agoraphobia and (2) generalized anxiety disorder (previously known as anxiety neurosis). Reviews of the anxiety disorders can be found in Klein (1984) and Bernstein et al. (1996) (for a discussion of phobic anxiety disorders, see Sect. 4).

7.1.1 Panic Attacks and Agoraphobia

Panic attacks (anxiety attacks) are severe, recurrent episodes of anxiety that begin suddenly, are not associated with specific situations or circumstances, and therefore cannot be predicted. Their manifestations vary widely from case to case but always include a number of physical symptoms, some of which appear dangerous, such as shortness of breath or a feeling of suffocation, dazedness, worry, a feeling of being about to faint, tachycardia, diaphoresis, trembling, etc. The attacks last only a few minutes and are highly variable in their frequency (a few per month to several per day). Although it is not characteristic for the attacks to be associated with specific situations or circumstances, there are a few patients in which they are. These situations and circumstances are then anxiously avoided (e.g. riding the bus, being in a crowd).

The term "agoraphobia" was originally applied only to fear of open spaces. It is currently a collective designation for fears of the public and of gatherings of people in a wide variety of locations. Thus, the alternative term "multiple situational phobia" has also been suggested for the disorder.

Agoraphobia is among the more common anxiety syndromes. Its longitudinal prevalence lies between 3.4% and 9%. The lifetime prevalence of panic attacks in the general population has been determined to be between 1.4% and 2.4% (Angst and Dobler-Mikola, 1985a,b; Wittchen 1986).

In contrast to the phobic syndromes, there is no clear evidence for the presence of a strong premorbid disposition to anxiety among patients with panic attacks.

It is also questionable whether there is a specific genetic risk factor for panic attacks, which are etiologically more closely related to generalized anxiety disorder (earlier called anxiety neurosis) than to the phobias. Psychophysiological explanatory approaches are based on the observation that panic attacks are often described by patients as beginning with physical symptoms. This has led some to conclude that the physical symptoms are primary and the emotional state of anxiety secondary.

It has been known for some time that agoraphobia, unlike panic attacks, can be induced by stressful life events (Marks 1987), although the patients often cannot remember the acute situation that induced it.

The diagnosis of both types of disorder (panic attacks and agoraphobia) is made on the basis of the clinical history and the symptoms described by the patients. The differential diagnosis must exclude other anxiety syndromes and organic illnesses, particularly those involving the cardiovascular system.

In accordance with current conceptions of the etiology and pathogenesis of panic attacks, in which the auto-observation of physical symptoms precedes the development of anxiety, current methods of treatment are very strongly oriented toward confrontation with physical stimuli. The following approaches have proved useful:

1. confrontational treatment employing strategies for coping with anxiety;
2. the provision of coping strategies;
3. cognitive strategies;
4. drug therapy with tricyclic antidepressants or monoamine oxidase inhibitors. (Benzodiazepines have also been found helpful, but should be used as little as possible because of their potential for abuse.)

Agoraphobia is treated according to the same principles as other phobic syndromes. When it is coupled with panic attacks, the treatments listed above for panic attacks are specifically modified in order to address the agoraphobic component. Confrontation has proved to be the most effective method for treating agoraphobia, as well as panic attacks. It should be carried out in a natural setting whenever possible. For agoraphobia, too, drug therapy with tricyclic antidepressants and benzodiazepines has proved useful, but it should only be applied in combination with behavior therapy (Remschmidt 1997a).

The highest age-specific prevalence of panic attacks and agoraphobia lies between the ages of 20 and 30. Approximately 10% of cases appear before the 16th

year. Both disorders have a strong tendency to become chronified if they are not treated or if they have persisted for a long period of time (e.g. more than one year).

7.1.2 Generalized Anxiety Disorder

The most prominent manifestation of this disorder is generalized, free-floating anxiety that is not associated with any particular situation and does not arise suddenly and episodically, as in panic attacks, but rather persists as a kind of permanent background state and is accompanied by many somatic manifestations such as muscle tension, diaphoresis, trembling, continuous nervousness, palpitations, vertigo, and sometimes upper abdominal symptoms. Fears and worries about the future are often expressed, for example a fear that the patient or a near relative may become ill or have an accident.

Three groups of manifestations of generalized anxiety disorder (ICD-10 classification: F41.1) are listed in the diagnostic guidelines of the ICD-10:

1. worries such as fear of future misfortune, a feeling of being at one's limit, difficulty concentrating;
2. motor tension in the form of physical agitation, tension headache, trembling, and inability to relax;
3. autonomic overexcitability, expressed as diaphoresis, tachycardia, tachypnea, vertigo, and dry mouth, or as dazedness and upper abdominal symptoms.

The DSM-IV diagnostic guidelines emphasize the unrealistic or exaggerated nature of the anxiety and worry about two or more areas of life, which last for at least six months in most patients.

A marked premorbid disposition to anxiety, which occurs on a familial basis, is thought to underlie generalized anxiety disorder. There is often also an accompanying depression, or at least a tendency toward depression. The illness usually begins between the ages of 20 and 30, frequently subsequent to a depressive mood disturbance. It is slightly more common in women.

Because the disorder has no specific situational association, there is no corresponding situational treatment (e.g. by the method of exposure and confrontation). Thus, the aim of treatment is to reduce anxiety in general and to promote the development of coping strategies (Remschmidt 1997a). The following treatment measures have proved useful:

1. relaxation training (e.g. autogenic training or progressive muscle relaxation, after Jacobson);
2. direction of treatment toward physical symptoms (e.g. through biofeedback);
3. drug treatment (with antidepressants, as the disorder often begins after a depressive mood disturbance).

The course of generalized anxiety disorder is variable. In many cases, periods of relative freedom from anxiety alternate with periods in which anxiety is often present. Unlike many phobias, the disorder usually begins in late adolescence or in adulthood, and tends to become chronified. The prognosis with respect to cure is increasingly unfavorable the longer the disorder has been present.

7.2 Obsessive–Compulsive Disorders

Three types of obsessive–compulsive disorders are distinguished in ICD-10: those with predominantly obsessional thoughts or ruminations (F42.0), those with predominantly compulsive acts or obsessional rituals (F42.1), and those with mixed obsessional thoughts and acts (F42.2). The lifetime prevalence of obsessive–compulsive disorders is approximately 2% to 3% in adults. The point prevalence among adolescents is approximately 2%. Obsessive–compulsive disorder usually begins in late childhood and early adolescence. In later adolescence and adulthood, the sex ratio is fairly even, while there is a mild excess of boys in childhood. The clinical picture of these disorders is characterized by a broad spectrum of severity, situational dependence of manifestations (most pronounced in unfamiliar environments), marked comorbidity with other disorders (most commonly anxiety disorders and depressive disorders), abnormal premorbid personality and similar traits in relatives, and, in many cases, a specific provocative factor. Reviews may be found in March and Leonard (1996) and Rapoport et al. (1994).

Family studies have shown that approximately 20% to 25% of children with obsessive–compulsive illnesses have at least one parent who suffers from an obsessive–compulsive illness. First-degree relatives of patients with tic disorders also have an elevated frequency of obsessive–compulsive disorders. These findings imply that genetic factors play an important role. Infections and immunological factors have also been discussed recently among possible etiologic factors for obsessive–compulsive disorders (Swedo and Rapoport 1990). On the neurotransmitter level, these disorders have been attributed to dysfunction of the serotonergic and dopaminergic systems. This hypothesis is supported by the therapeutic effectiveness of selective serotonin reuptake inhibitors and by the high degree of comorbidity of these disorders with tic disorders, for which there is also thought to be an underlying dysfunction of the dopaminergic system.

The treatment is multidimensional in practice; behavior-therapeutic methods are used in combination with drug therapy. The former include desensitization

techniques, particularly in cases with a strong component of anxiety. Moreover, cognitive strategies are combined with action-oriented measures (e.g. thought stopping and saturation methods) (Remschmidt and Niebergall 1997). Among drug therapies, serotonin-specific antidepressants such as clomipramine and selective serotonin reuptake inhibitors such as fluoxetine, fluvoxamine, and paroxetine have been found effective. The therapeutic benefit usually sets in after a delay of three to four weeks from the start of treatment.

The prognosis of obsessive–compulsive disorders in children and adolescents, as in adults, is rather unfavorable. Follow-up studies have revealed that chronification takes place in 30% to 40% of cases of obsessive–compulsive disorders with onset in childhood and adolescence. The further course and prognosis also depend on comorbidity; common comorbid disorders include anxiety disorders, depressive disorders, and personality disorders (Neudörfl and Herpertz-Dahlmann 1996).

7.3 Dissociative Disorders (Conversion Disorders)

In ICD-10, the terms "dissociative disorders" and "conversion disorders" are used synonymously (F44); in DSM-IV, a distinction is drawn between the two. The term "dissociation" designates an interruption of the normal integration of mental functions such as consciousness, perception, memory, identity, personality, and (sometimes) motor control (Blanz et al. 1987; Michelson and Ray 1996).

The discussion in this section is based on a tripartite division of these syndromes:

- conversion disorders (with predominantly physical manifestations),
- dissociative disorder (with predominantly mental manifestations), and
- histrionic personality.

In view of the lack of representative studies on the frequency of these disorders in field samples, one must rely on data from clinical and non-representative reports. Among the patient population of clinics for child and adolescent psychiatry, the prevalence of conversion syndromes and dissociative disorders has been reported as lying between 1.5% and 5%. In most studies, girls predominate, with a sex ratio of 3:1 to 4:1.

Conversion Disorders (Predominantly Physical Manifestations)

Approximately 40% of all conversion syndromes in childhood and adolescence take the form of seizures or movement disorders, while approximately 13% take the form of psychogenic paralysis or sensory deficits.

Psychogenic seizures are the most common type of conversion syndrome in childhood and adolescence, followed by psychogenic gait disturbances and twilight states. Nevertheless, one can often observe patients who suffer both from genuine epilepsy and from psychogenic seizures. The most important features of these disorders are:

1. *Psychogenic seizures* are characterized by their sudden occurrence, their occurrence in the presence of others and tendency to go on longer than epileptic seizures, and the fact that they can be provoked by an experience or stress reaction. Further typical features include bizarre and uncoordinated movements that do not resemble those of known patterns of epileptic seizures, the lack of EEG abnormalities, and the rarity of self-injury.
2. *Psychogenic paralyses and movement disorders* usually involve complex, goal-directed motor processes such as walking, standing, and other voluntary movements. The severity and localization of the movement disorder does not conform to known anatomic patterns of innervation, but rather to lay notions of how the body is constructed. The following psychopathological observations can be made:
   - The manifestations appear to be purposeful and have an ordered content that may be related to the provoking situation.
   - They often have a demonstrative character.
   - Despite the severity of the manifestations, the patient seems remarkably unconcerned about them ("la belle indifférence").
   - The relationship of the manifestations to the provoking situation can often be recognized by outsiders, but not by the patient.
3. *Psychogenic sensory deficits* are found in the form of psychogenic visual disturbances, including abnormal visual perceptions such as micropsia or macropsia, and partial or total psychogenic deafness.

Dissociative Disorders (Predominantly Mental Manifestations)

Dissociation most often affects memory (amnesia), personal integrity (depersonalization, multiple personality) and consciousness (trance states). These disorders are rare in childhood but more common in adolescence. They are characterized by sudden onset (psychogenic amnesia, psychogenic fugue), disturbances of identity (depersonalization), and a disturbed connection to reality, in the absence of psychosis. The individual disorders may be described as follows:

- *Psychogenic amnesia* is characterized by a sudden inability to remember important current and personal events. The loss of memory is so all-encom-

passing that it cannot be explained as ordinary forgetfulness. Exclusion criteria for the diagnosis include the presence of organic brain damage, intoxication, or multiple personality disorder.

- *Depersonalization and derealization* are characterized by a partial loss of reality without an overall disturbance of the patient's connection to reality (as in the schizophrenic psychoses, for example). Depersonalization is a feeling of foreignness in relation to one's own self, often accompanied by the feeling that certain parts of the body no longer belong to one's own body. Derealization is an experience of the environment (persons, objects) as foreign and unreal. When depersonalization occurs in the framework of dissociative disorders, neither hallucinations nor delusions are found. These patients are aware of their experiences of alienation and highly disturbed by them, but they are able to remain rooted in reality.
- *Multiple personality disorder* (dissociative identity disorder) is characterized by the coexistence of two or more "persons" in a single individual. Affected patients perceive the various personalities as discrete units with specific modes of behavior, memories and relationships, and, often, individual names. They often say that one or more of the personalities belong to the opposite sex or are younger than they themselves are. The disorder is extraordinarily rare in childhood and more common in adolescence and young adulthood.

Histrionic Personality Disorder

Histrionic personality disorder can also be mentioned here because it shares certain general features with conversion and dissociation disorders, although it manifests itself not in the form of specific physical or mental symptoms, but rather in abnormal attitudes of the personality as a whole (see Sect. 7.6).

According to Berblinger (1960), hypotheses concerning the etiology and pathogenesis of the conversion syndromes and dissociative disorders can be classified according to three types of postulated underlying mechanism:

1. inactivation of organs and organ systems (e.g. in paralyses, movement disorders, or sensory deficits);
2. increased functional autonomy of the mental apparatus (functional augmentation; e.g. in psychogenic seizures and exaggerated movements); and
3. decreased functional autonomy of the mental apparatus (e.g. in twilight states and amnesia).

A number of etiologic factors have been discussed but have not yet been integrated into a definitive theory. These include familial factors (whether genetic or in the sense of a family tradition), predisposing aspects of personality structure, exposure to role models, conflict-laden or overly demanding situations, and previous medical illnesses, which may influence the "choice of symptoms." The patient's gain from illness has also been emphasized, at least as a factor that may help maintain the disorder, as has the increased frequency of conversion disorders in brain-damaged patients.

The diagnosis is based on meticulous history-taking, negative findings of a thorough neurological examination including ancillary testing, and an extensive psychiatric and psychological examination. The latter must take the patient's cognitive functional level, individual emotional features, and personality structure into account, and must concentrate especially on the possible relationship of the manifestations to typical provoking situations and conflicts.

The differential diagnosis of conversion syndromes and dissociative disorders includes the following illnesses:

1. psychosomatic illnesses (as already described, in all essential features, by Alexander in 1943);
2. paralysis, epileptic seizures, and other movement disorders (video-EEG is useful here);
3. schizophrenic psychoses, which not uncommonly manifest themselves in childhood and adolescents as dissociative disturbances.

If the underlying conflict or the often typical, excessively demanding situation can be identified, then the goal of treatment should be its elimination or neutralization. Thus, treatment always involves a change in the patient's general living conditions. Functional therapy is often required on the level of symptoms as well, e.g. by gradual practice of walking in patients with psychogenic paralysis. One may explain to the patient that he or she has forgotten how to walk because of the illness and must now slowly relearn it.

Treatment is usually provided on an individual basis, because group therapy fulfills the pathological need of many patients for domination and is therefore not indicated, at least in the initial stage of treatment. If additional manifestations are present, such as seizures, treatment is provided according to the same principles, but the role of medications in the overall treatment plan is particularly important. The most significant factors for the success of psychotherapy are the following: the therapist must adapt the therapeutic relationship as precisely as possible to the patient's abilities, and he or she should be cautious about making possibly premature symbolic interpretations. The patient must be provided with alternative coping strategies so that the illness manifestations are no longer necessary.

Conversion syndromes arising for the first time are usually easily treatable and rapidly eliminated.

Treatment is more difficult when the disorder has been present for a longer period and the patient has repeatedly drawn considerable gain from it (Remschmidt 1997b).

7.4 Affective Disorders

The term "affective disorders" encompasses several disorders that differ from one another in etiology, manifestations and course but are similar in the following respects:

1. They are characterized by marked abnormalities of mood and drive, with or without accompanying anxiety. The patient's mood may be altered in the direction of either depression or mania.
2. Furthermore, they are regularly accompanied by cognitive and somatic manifestations.
3. They tend to recur and often become chronified.

The classification of these disorders is still unsatisfactory at present and is based on pragmatic clinical considerations. ICD-10 distinguishes the following types of disorder: depressive episodes (F32), recurrent depressive disorder (F33), manic episodes (F30), bipolar affective disorder (F31), and two types of persistent mood [affective] disorders (F34), namely cyclothymia (F34.0) and dysthymia (F34.1). The DSM-IV essentially follows the same basic scheme; under the designation "depressive disorder," it includes two subtypes – major depressive disorder as a single episode (296.2) and recurrent major depressive disorder (296.3). In both systems, depressive reactions may also be classified as adjustment disorders.

Affective disorders may begin as early as the childhood years but become significantly more frequent in adolescence and then undergo a transition to the adult illness pattern; they are thus discussed in this chapter among other adult-type disorders of early onset. Current reviews are found in Goodyer (1995), Harrington (1993) and Rutter et al. (1986).

The manifestations of depressive conditions can be quite varied. Emotional manifestations other than the sad background mood include marked inhibition of drive, loss of interest, increased anxiety and irritability, sometimes mood fluctuations over the course of the day, and, often, feelings of guilt. Cognitive manifestations include inhibited thinking, brooding, disturbances of concentration, a feeling of helplessness and powerlessness, negative expectations for the future, and suicidal ideation. Finally, there may be somatic manifestations such as sleep disturbances, lack of appetite, weight loss, fatigue, psychomotor slowing or agitation, loss of libido, hypochondriacal symptoms, and other somatic-vegetative complaints such as headache, stomachache, and digestive disturbances.

An understanding of depressive disorders in adolescence requires knowledge of the following special aspects of developmental psychopathology (Rutter et al. 1986):

- Both depressive mood disturbances and well-defined depressive illnesses can occur in childhood but become much more common in adolescence, more in relation to the process of puberty than to chronological age.
- Boys predominate among sufferers from affective disorders in childhood, but the sex ratio shifts in favor of girls over the course of puberty.
- Both suicidal ideation and suicide attempts (within and outside the context of depressive disorders) become considerably more frequent during adolescence, and are more common among girls. Completed suicide is extremely rare before puberty and rises in frequency from adolescence until adulthood.

The single depressive episode is regarded as constituting a disorder in itself, because it is often unclear whether it is a component of a larger disorder. Its characteristic features include the cardinal manifestations of depressive disorders, such as a sad and oppressed basic mood, loss of drive, loss of interest, and sleep disturbances, as well as increased fatigability. Further common manifestations, according to the ICD-10, include difficulties of concentration, diminished self-esteem and self-confidence, guilt feelings, pessimistic perspectives on the future, suicidal ideation or behavior, sleep disturbances, lack of appetite, weight loss, diffuse anxieties, and motor unrest.

The manifestations of recurrent depressive disorders are the same as those of the single depressive episode. The episodes occur repeatedly, and the duration of each is highly variable (weeks or months). The first episode may occur at any age between childhood and old age. In the recurrent depressive disorders, too, various degrees of severity are distinguished (mild, moderate, and severe recurrent episodes; the last may be with or without psychotic manifestations).

The severe forms of recurrent depressive disorders have a great deal in common with the depressive episodes that occur in bipolar illness; thus, the latter must be excluded in the differential diagnosis. The occurrence of hypomanic or manic episodes is regarded as a criterion of exclusion for the diagnosis of a recurrent depressive disorder.

As for manic episodes, these are characterized by excessive drive, lack of distance, aimless activity and hyperactivity, elevated self-esteem and exaggerated conception of oneself, and further by ideas of grandeur or thoroughly unrealistic plans for the future. In these

phases, patients require almost no sleep and are constantly in motion. Mania in children can "flip over" into depression within a few days. In adolescence, mania often displays manifestations that are otherwise typical of schizophrenia, including delusions and hallucinations, and may thus easily be misdiagnosed as a schizophrenic or schizoaffective psychosis. A reliable differential diagnosis is usually feasible only through observation of the patient's further course.

Bipolar disorder is characterized by an alternation of depressive and manic episodes, interspersed with phases of normal mental health (remissions) lasting for variable periods of time. Disorders with exclusively manic or hypomanic manifestations are extremely rare in childhood and adolescence and are also classified among the bipolar disorders. In both classification systems (ICD-10 and DSM-IV), the type and severity of the current episode are classified in individual categories—for example, "bipolar affective disorders, current episode manic without psychotic symptoms" (F31.1). Common precursors of bipolar disorders include conduct disorders and hyperkinetic disorders; this fact makes their diagnosis in childhood more difficult (Remschmidt 1998). Bipolar disorders are extraordinarily rare in childhood and have a prevalence of less than 0.1% in adolescence as well. Their lifetime prevalence is 0.4% to 1%.

The ICD-10 designation "persistent mood [affective] disorders" refers to "persistent and usually fluctuating disorders of mood in which individual episodes are rarely if ever sufficiently severe to warrant being described as hypomanic or even mild depressive episodes. Because they last for years at a time, and sometimes for the greater part of the individual's adult life, they involve considerable subjective distress and disability. In some instances, however, recurrent or single episodes of manic disorder, or mild or severe depressive disorder, may become superimposed on a persistent affective disorder. The persistent affective disorders are classified here rather than with the personality disorders because of evidence from family studies that they are genetically related to the mood disorders, and because they are sometimes amenable to the same treatments as mood disorders" (WHO 1992, p. 128). Cyclothymia and dysthymia are included among these disorders.

Cyclothymia is a persistent instability of mood characterized both by mild depressive episodes and by episodes of elevated mood. The definition of cyclothymia states that it develops in early adulthood, but it may also arise in childhood and adolescence, although the diagnosis is extraordinarily difficult to make in this phase of life. The family history is useful, as cyclothymia is more common in relatives of patients with bipolar affective disorders. Moreover, persons with cyclothymia are more likely to develop bipolar affective disorder. Thus, a preliminary phase of a bipolar affective disorder may be recognized in childhood and adolescence in the form of cyclothymic mood fluctuations.

Dysthymia is a chronic depressive mood disturbance that usually lasts for months but is repeatedly interrupted by periods of good mental health. It is characterized by marked depressive manifestations, generally in the aftermath of a traumatic experience. The traumatic experience may also be a persistent conflict with which the child or adolescent has been preoccupied continuously and for a long period of time, often without being aware of it. Furthermore, marked manifestations of anxiety are often present, or a mixed state of anxiety and depression. Delusions, hallucinations, and daily fluctuations are absent. The general state of chronic depression alternates with periods of good mental health lasting days or weeks. The diagnostic guidelines of the ICD-10 emphasize the chronification of the depressive manifestations, which nevertheless are not severe enough to reach the level of a mild or moderately severe recurrent depressive disorder.

Depressive conditions are diagnosed on the basis of the history and clinical manifestations. Supplementary techniques are available in the form of interviews, checklists, and scales. The differential diagnosis involves both the differentiation of depressive disorders from other psychiatric disorders that may have depressive manifestations (e.g. schizophrenia or organic illnesses) and their differentiation from one another. This is done mainly in reference to the diagnostic criteria of the two classification systems; from the point of view of child and adolescent psychiatry, it should be observed that these systems are not always directly applicable, in unmodified form, to our patient population.

A number of etiologic factors for the affective disorders have been postulated. Family studies have shown that the highest rate of recurrence is among relatives of bipolar index cases (ca. 18%), while that among relatives of unipolar index cases is lower (ca. 7%; Propping 1991). Twin studies have revealed concordance rates among monozygotic twins of 73% for bipolar disorders and 42% for unipolar disorders, and concordance rates among dizygotic twins of 14% for both disorders. Personality traits such as introversion, tendency to anxiety, and neuroticism are positively associated with depressive disorders.

Many models of depression postulate important roles for negative early experiences and current psychosocial stress. The best known is Seligman's (1975) model of learned helplessness, according to which a person unable to control his or her own subjective experiences eventually learns helpless behavior. Such a person feels that he or she is powerless

with respect to self-control and therefore expects to be helplessly exposed to other situations in the future.

According to the current, generally accepted model, all etiologic and pathogenetic factors for depression lead to the end result of an impairment of the biochemical neurotransmitter systems affecting mood, and to neuroendocrine alterations, which finally become manifest as the emotional, cognitive, and somatic features of depression.

Children and adolescents with severe affective disorders should be hospitalized because of the risk of suicide, which is often present in the form of a presuicidal syndrome even before the depressive disorder appears. The presuicidal syndrome of childhood and adolescence consists of marked withdrawal, neglect of interests, and suicidal thoughts, sometimes accompanied by an announcement of suicidal actions.

The use of tricyclic antidepressants to treat depressive children and adolescents remains controversial because the currently available double-blind, placebo-controlled studies have failed to show any therapeutic benefit of antidepressants in this population. They are nevertheless commonly used in clinical practice, and many clinicians are convinced that they are quite useful in individual cases. Their side effects have been very thoroughly studied; the most severe ones are tachycardia and slowing of intracardiac conduction. Thus, an EKG is mandatory both before and during the administration of these drugs, as are follow-up EEG studies, because of their possible influence on cerebral activity. The value of selective serotonin reuptake inhibitors in the treatment of depressive children and adolescents is currently being investigated; no conclusive statement can yet be made about it.

Lithium preparations are used to treat bipolar disorders. They have a prophylactic effect with respect to manic and depressive phases of endogenous, phasic illness. Thus, they can be used to treat acute mania and for prophylaxis in uni- or bipolar affective disorders. Carbamazepine, too, has been found useful in the treatment of certain kinds of affective disorders (prophylaxis against the recurrence of bipolar and unipolar psychotic phases) and for the prevention of recurrence in the schizoaffective psychoses (often in combination with neuroleptics or antidepressants). Finally, there has also been favorable experience with a combination of lithium and carbamazepine for the treatment of affective disorders with psychotic manifestations in adolescence. When such a combination is used, the dose of lithium can be lower than when lithium is given alone. The side effects of lithium, and the feared complication of lithium toxicity, can be avoided by meticulous checking of the serum lithium level, which should lie between 0.8 and 1.2 mmol/l, preferably toward the lower end of this range. Reviews of drug therapy are provided by Kutcher (1997), Nissen et al. (1998), and Remschmidt and Schulz (1995).

The modern forms of behavior therapy, which incorporate cognitive aspects as well, have gained acceptance in recent years for the treatment of depressive disorders in childhood and adolescence and have been found to have a high chance of success. Therapy of this type involves the following elements: establishment of a sturdy therapeutic relationship, reduction of stressful cognitions (thoughts and ideas), establishment and promotion of positive ties and social contacts. For bipolar disorders, however, the cognitive behavioral therapy of depression is indicated only in special circumstances, and only as an auxiliary measure (Harrington et al. 1998).

Additional measures to be considered include all those that take the patient's life circumstances and living environment into account. Two aspects are particularly important in childhood and adolescence: involvement of the parents and involvement of the environment at school or in the workplace.

Severe depressive disorders (i.e. major depression) in children and adolescents are associated with a mean duration of episodes of ca. 7–9 months and are thus considerably more protracted than previously supposed. The currently available studies of disease course all document relapse rates of 40% within 2 years of disease onset and 70% within 5 years. Children and adolescents exposed to a conflict-laden family environment have even higher relapse rates. Furthermore, 20–40% of adolescents who suffer a severe depressive episode develop a bipolar disorder within 5 years (Harrington et al. 1990, 1991).

7.5 Schizophrenia

In the two current classification systems (ICD-10 and DSM-IV), the schizophrenic, schizotypal and delusional disorders are classified in essentially the same way. Schizophrenia is divided into several subtypes. Both systems further provide separate classifications for the schizoaffective disorders (F25) and for schizotypal disorder (F21), which DSM-IV calls schizophreniform disorder. A major difference between the two systems is that DSM-IV includes a temporal requirement (duration of illness at least 6 months). The result of this is that many patients who would be diagnosed as schizophrenic according to the ICD-10 criteria are said to suffer from schizophreniform psychoses according to the DSM-IV criteria. The diagnosis of schizophrenia in DSM-IV is thus more restrictive.

Approximately 4% of all cases of schizophrenia arise before age 15, and 1% before age 10. Among the patients of clinics for child and adolescent psychiatry,

schizophrenic illnesses are present in 1%–2% of children and 2%–5% of adolescents. Boys predominate by a small margin. Studies of clinical samples have revealed that approximately 10% of schizophrenic psychoses appear between the ages of 14 and 20, and 42% between 21 and 30. Approximately three quarters of all schizophrenic patients are between 20 and 40 years old.

In recent years, the findings of empirical studies have cast doubt on the classical subdivision of schizophrenia into hebephrenia, catatonic schizophrenia, paranoid schizophrenia and schizophrenia simplex. Karl Leonhard formulated such doubts years ago and explained his reasons for them in detail in the last edition of his book, *Die Aufteilung der endogenen Psychosen* ("The Classification of the Endogenous Psychoses"). According to Leonhard, the retention of this scheme is partly responsible for the stagnation of research on schizophrenia.

While Leonhard proposes a highly detailed subclassification of the schizophrenic psychoses according to their psychopathological features and course, the simpler concept of positive (Type I) and negative (Type II) schizophrenia – a dichotomous classification of schizophrenia by psychopathological features – has received support from the findings of experimental and clinical studies. The most important features of both types of schizophrenia are listed in Table 4. They differ not only with regard to their clinical manifestations and premorbid personalities, but also with regard to other, experimentally testable functions.

This classification scheme is not entirely free of overlap; manifestations of both types may be present in a single patient.

Despite their relatively vague boundaries, the classical subtypes of schizophrenia mentioned above can be diagnosed in adolescence. In childhood, however, they are difficult to differentiate from one another and rarely appear as "pure cultures." With advancing age, their manifestations become increasingly similar to those seen in adulthood.

The general manifestations of schizophrenic disorders affect the areas of cognition and perception, emotion, language, motor function, and drive (see also the ICD-10 and DSM-IV diagnostic guidelines):

- *Cognition and perception* are frequently disturbed by formal thought disorders, delusions and hallucinations. Delusions become systematized during adolescence, extremely rarely before age 10. Adolescents relatively frequently have somatic hypochondriacal experiences and ideas of interpretation as well as delusions of persecution, of being influenced, and of being poisoned. The hallucinations are most commonly auditory, while visual hallucinations are more common among children. The latter must raise the suspicion of a possible organic etiology.
- *The emotional area, social contact, and conduct* are often disturbed, in adolescent patients, by marked manifestations of withdrawal and isolation. Not uncommonly, the patient's relationship to the outside world is qualitatively restructured, and substitute relationships are established, sometimes with personification of inanimate objects. Affect disturbances, particularly a suspicious and anxious mood, affect lability, negativism, and regression to infantile behavior, are common.
- *Language* disturbances may include changes in the manner of speaking, in logorrhea, a tendency toward perseveration, verbal stereotypies, echolalia, or phonographism. The abnormal speech of childhood schizophrenia may be difficult to tell apart from autistic speech. Neologisms are produced, and the meanings of words may be altered.
- *Spontaneous motor function:* there is often a general lack of coordination (stiffness, jerkiness) and

Table 4. Major features characterizing Type I and Type II schizophrenia

	Type I schizophrenia (positive symptoms, productive manifestations, acute schizophrenia)	Type II schizophrenia (negative symptoms, withdrawal manifestations)
Clinical features	Hallucinations, delusions, positive thought disorders, increased drive, aggression, excitement, bizarre behavior, logorrhea, neologisms	Flat affect, lack of drive, social and emotional withdrawal, apathy, paucity of speech, diminished drive to speak, anhedonia, negative thought disorders (thought inhibition, disconnected thoughts, entangled thoughts)
Attention and sensorium	Increased distractibility	Diminished information-processing capacity
	Broadened attention	Narrowed attention
Hemispheric function	Left hemispheric hypofunction	Frontal dysfunction, bilateral functional disturbance
Premorbid personality	No major cognitive or motor impairment	Cognitive and motor impairment

a reduction of spontaneous movement. Catatonic and cataleptic phenomena are occasionally seen. Stereotyped movements are relatively common (e.g. stereotypic postures or bizarre finger games). Obsessive-compulsive phenomena, which commonly occur as prodromal manifestations of schizophrenia, are often first manifested in the area of motor function.

- *Disturbances of drive* in affected patients include a loss of their earlier spontaneity and initiative. They sit listlessly in their rooms for hours and have no interest at all in pursuing a conversation, reading, or otherwise occupying themselves. The lack of initiative may be so extreme that they sit motionlessly without speaking or eating and no longer even voluntarily urinate or defecate.

The various subtypes of schizophrenia are characterized as follows:

- *Paranoid schizophrenia*, the commonest type, can arise as early as adolescence, but it is most common in adulthood. Delusions and auditory hallucinations predominate. There are also disturbances of cognition and affect. This form of the illness usually does not lead to a change of personality or affect the patient's intelligence. It is the prototype of "positive schizophrenia."
- *Hebephrenia* usually begins after puberty with lack of drive, chaotic thinking, flatness of affect, and a cheerful, silly mood. As the illness progresses, it generally leads to a decline of cognitive and emotional functioning. These patients often already manifest premorbid personality traits (being a "loner," shyness, timidity). The diagnosis of hebephrenia should be made only after several months of observation. The rapid development of negative manifestations places this disorder in the class of the "negative schizophrenias."
- In *catatonic schizophrenia*, the most prominent manifestations are motor phenomena, acute excitement and blocked states (stupor), and mutistic behavior. Command automatism may alternate with negativism. Many other manifestations may also be present, most commonly delusions and hallucinations, which may also be associated with a dreamlike (oneiroid) state.
- *Schizophrenia simplex* leads slowly and inisidiously, without particularly noticeable manifestations, to a condition of permanent mental impairment. The illness usually begins in adolescence; the patients lack drive and have a dull affect, they have no energy or initiative, they are depressed or ill-humored, and they fail in school or at work. They often abandon their usual activities or change jobs, drift along and let themselves go. Schizophrenic illnesses of this type can be very difficult to recognize because the usual diagnostic categories often do not apply to their extremely bland clinical manifestations.
- *Schizoaffective psychoses* are characterized by the simultaneous presence of manic or depressive manifestations and schizophrenic manifestations. These disorders cannot be diagnosed either as schizophrenia or as an affective psychosis. The most important factor for the diagnosis is the *simultaneity* of the schizophrenic and affective manifestations. If these always appear in separate episodes, a schizoaffective psychosis cannot be diagnosed. ICD-10 provides a further level of detail in the classification of these disorders, according to the type of affective manifestations present – thus, schizomanic, schizodepressive, and mixed schizoaffective disorders are distinguished. DSM-IV provides a comparable subdivision (bipolar vs. depressive type). Recurrences are frequent in this form of psychosis, but there is no residual state or lasting mental impairment.

Even though research on schizophrenia is being carried out intensively all over the world, we are still far away from a comprehensive understanding of its etiology. In view of this situation, we can do no more than list the causative factors whose importance has been confirmed by empirical studies, even if we cannot yet integrate them into a unified theory: genetic factors, brain damage, disturbances of cerebral function, and psychogenic factors. For reasons of space, these factors cannot be discussed here in detail. A mere look at this brief list leads one to conclude that the ultimate explanation of the etiology of schizophrenia will be multifactorial.

The currently available empirical data appear to permit the following statements. The functional substrate for the schizophrenic illnesses seems to be a regulatory system for information processing, which is susceptible to disturbances produced by factors of different kinds. These include organic factors (infection, hypoxia), intrapsychic factors (personality, ego structure), genetic factors, and familial and psychosocial factors (deviant communication, "expressed emotions," social stratum, "life events"). One may hypothesize that this system for the selection and processing of information is anatomically located in the limbic system. Evidence in favor of this hypothesis comes from the fact that the limbic system is particularly susceptible to hypoxic damage (including in the pre- and perinatal period). It is also the case that certain viruses, such as cytomegalovirus, have a special affinity for the limbic system, as has been pointed out by the proponents of the "infectious theory of schizo-

phrenia." When the limbic system has been weakened through a combination of influences, the affected persons may develop a schizophrenic psychosis, particularly when they have a genetic predisposition to schizophrenia and have also been exposed to stressful events in their family lives.

The diagnosis of schizophrenia is made on the basis of the clinical manifestations according to the guidelines of ICD-10 and/or DSM-IV. Ancillary scales and interviews are often applied, e.g. Andreasen's scales for the assessment of positive and negative manifestations (Andreasen 1984a,b), or other checklists. Entities to be excluded in the differential diagnosis include psychoses of somatic origin, drug-induced psychoses, and schizoaffective psychoses. Once schizophrenia has been diagnosed, an attempt is made to distinguish between forms with predominantly positive manifestations and those with predominantly negative manifestations, because this distinction is relevant to treatment.

Neuroleptic treatment is necessary in the acute phase of schizophrenia and has been repeatedly shown to be effective. Neuroleptics have the following therapeutically beneficial properties: psychomotor calming, damping of excitement and aggression, and an antipsychotic-antischizophrenic effect (review by Remschmidt et al. 2000).

In the treatment of acute psychotic states with predominantly productive manifestations, the butyrophenone derivatives (particularly haloperidol and benperidol) have proved very useful, as have the phenothiazines perazine, fluphenazine, perphenazine, and chlorproxithene. If the acute psychotic state is accompanied by pronounced agitation, levopromazine is recommended as a means of damping it. The foregoing "classical" concept of the medical treatment of schizophrenia is now beginning to change, however, as so-called atypical neuroleptics are being developed. These substances have a broader receptor binding profile than the typical neuroleptics (most of which are dopamine D_2-receptor antagonists), hardly any extrapyramidal side effects, and a better therapeutic effect on negative manifestations. The most important substances of this type are clozapine, olanzapine, and risperidone.

The greatest amount of experience to date with drug therapy in schizophrenic children and adolescence has been gained with clozapine, which should, however, be used with special caution because of its known adverse effect on the hematopoietic system (risk of agranulocytosis). Thus, complete blood counts must be obtained weekly during the first 18 weeks of treatment, and monthly thereafter. The treatment guidelines further require that clozapine be given only after two other neuroleptics have been tried without success. Under these conditions, and after thorough education of the patients and their parents, the use of clozapine is permissible, and usually very effective. The two other atypical neuroleptics have also been found to be effective. They have no adverse effects on the hematopoietic system and are therefore not subject to the special precautions and conditions for administration that apply to clozapine, but they do not have an equivalent antipsychotic effect. (Table 5 provides an overview of the most important typical and atypical neuroleptics, and corresponding dosages.)

Sulpiride is recommended for the treatment of psychotic states with non-productive manifestations, in which lack of drive, negativism, autistic behavior, inhibition and withdrawal are prominent. Success has also been reported with haloperidol and fluphenazine. Here, too, there is a current movement toward treatment with atypical neuroleptics, not least because of their total or near-total absence of extrapyramidal side effects.

Chronic schizophrenic psychoses are an indication for the use of depot neuroleptics. These have the same effectiveness profile as short-acting neuroleptics and, like them, influence not only the productive manifestations (hallucinations, delusions, thought disorders) but also autistic behavior, withdrawal, and psychomotor inhibition. Haloperidol decanoate, fluphenazine decanoate, fluspirilene, and penfluridol are used as depot neuroleptics. They are given intramuscularly, and their effect lasts for 1 to 4 weeks, depending on the substance. Depot neuroleptics are given in relatively low doses in comparison to short-acting neuroleptics. They have the same side effects. Recently, the atypical neuroleptics have been increasingly used instead of depot neuroleptics.

The common undesired effects of the typical neuroleptics include vegetative effects (hypersalivation, disturbances of accommodation, urinary disturbances, hyperhidrosis), cardiovascular effects, hematological effects, hepatic effects, endocrine effects (e.g. galactorrhea, gynecomastia, amenorrhea), and extrapyramidal effects. Epileptic seizures may occur as a consequence of a lowered seizure threshold, and non-specific mental effects may also occur.

The common undesired effects of clozapine include hypersalivation, weight gain, seizure tendency, and possible hematopoietic effects, which can be avoided by regularly checking the complete blood count. Olanzapine also causes hypersalivation and weight gain but has no effect on the hematopoietic system. Risperidone, unlike clozapine and olanzapine, causes extrapyramidal side effects in a dose-dependent manner and thus occupies an intermediate position between the typical and atypical neuroleptics. Epileptic seizures have also been observed in association with risperidone use.

Supportive psychotherapy is a valuable complement to drug therapy and consists of the following measures:

Table 5. Effectiveness profile of selected neuroleptics

	Sedation	Positive symptoms	Negative symptoms	Neuroleptic potency	Usual oral dose or usual i.m. depot dose
Typical neuroleptics of high potency					
Benperidol	+	++	++	100	1–6 mg/d
Flupenthixol (decanoate)	+	++	++	50	2–10 mg/d (20–100 mg/ 2–4 wk)
Fluphenazine (decanoate)	+(+)	+++	++	30	5–20 mg/d (12.5–100 mg/ 2–4 wk)
Fluspirilene	+	+++	++	300	(2–10 mg/wk)[a]
Haloperidol (decanoate)	+	+++	++	60	2–20 mg/d (50–300 mg/ 2 wk)
Perphenazine (enanthate)	++	+++	++	8	12–64 mg/d (50–200 mg/ 2 wk)
Pimozide	+	+++	++	50	4–20 mg/d
Typical neuroleptics of intermediate to low potency					
Chlorpromazine	+++	+++	++	1	150–600 mg/d
Chlorprothixene	+++	++	++	0.8	150–600 mg/d
Levomepromazine	+++	++	++	0.8	75–600 mg/d
Perazine	++	++	++	0.5	75–600 mg/d
Pipamperone	++			0.2	120–360 mg/d
Promethazine	+++				50–400 mg/d
Sulpiride	+	++	+++	0.5	100–800 mg/d
Thioridazine	+++	++	++	0.7	200–700 mg/d
Tiapride	+				300–600 mg/d
Atypical neuroleptics					
Clozapine	+++	+++	+++	(0.5–2)	25–600 mg/d
Olanzapine	++	+++	+++	(8–20)	10–20 mg/d
Quetiapine	+	+++	+++		150–750 mg/d
Risperidone	+	+++	+++	(50)	1–12 mg/d
Zotepine	++	+++	+++	(2)	75–300 mg/d

+, little or no effect; ++, intermediate effect; +++, strong effect.
[a]Only available in depot form.
Neuroleptic potency is given in relation to chlorpromazine (=1). The neuroleptic potency of the atypical neuroleptics was determined from the mean clinically effective dose rather than the threshold for extrapyramidal symptoms.

psychological guidance of the patient, encouragement, addressing everyday problems and worries, improvement of self-esteem and the ability to make contact and communicate. It has been found to be important not to let the patients regress excessively during supportive psychotherapy, and not to give in to their tendency to withdraw. Psychotherapy with the aim of uncovering putative underlying conflicts is contraindicated in the acute phase of schizophrenic illnesses and, even in the chronic phase, carries the danger of promoting a relapse of psychosis. The reason is that an excessively intense probing of emotional and instinctual impulses may overwhelm the patient's processing abilities.

A further approach in the psychotherapy of schizophrenic patients is directed toward the so-called basic cognitive disturbances, which the patient must first be made to understand and which can then be dealt with by means of a special training program. Cognitive disturbances of thinking, perception, and behavior manifest themselves as blockages, oppressive and racing thoughts, absence of thought, and compulsive ideation, as diffuse, scattered, and easily distracted thinking, or as highly associative thinking lacking direction. The patient's concentration is impaired, as are memory and immediate retention. Perceptions may be distorted both quantitatively and qualitatively. There may be hypersensitivity to light and sound, sensory hypervigilance, blurred, distorted or deformed vision, or similar disturbances. Motor activities may be impeded or even blocked. The ability to carry out automatic or semiautomatic processes is lost; finally, there are coenesthesiae and vegetative disturbances.

Such disorders must be dealt with actively by the therapist, because the patients often fail to complain of them. Making the patient understand the disorder means, in the present context, that the child or adolescent should understand that these disturbances, which he or she can describe quite competently in a

subjective way, are due to the underlying illness and that they can be improved or eliminated by functional therapy and by a treatment of the underlying illness. The integrated psychological therapy program for schizophrenic patients (IPT) of Roder et al. (1992) may serve as an example of a structured therapy programs. Such programs, of which several are available, provide a graduated sequence of training in cognitive differentiation, social perception, and communication, and exercises through which problem-solving techniques may be practiced.

The goal of occupational therapy is to help these frequently introverted and aloof patients become mentally more relaxed and enable them to live in the community again. Through artistic activities (painting, drawing, working with clay), many patients can become less reserved and can, indeed, develop a more optimistic self-image through the often surprisingly good results of their creative efforts. Work therapy is usually used in adolescent patients to adapt them better to their environment through the imposition of progressively more difficult physical demands. It is thus a component of rehabilitation, which is an important goal for all psychiatric patients.

The younger the patient, the more important it is to involve the family in the treatment process. The emphasis is no longer placed on highly ambitious family therapy in which intrafamilial roles are reconstructed, but rather on educating the family about the illness, family counseling, practicing ways of dealing with critical situations, and promoting family strategies for coping with stresses within and outside the family. Only a few such programs have been systematically studied with respect to their use among adolescent schizophrenics.

For adult schizophrenics, treatment programs combining adequately dosed depot medication with structured supportive family programs have been found to be effective. The combination of these two measures has two major effects:

- The structured therapy program with the family may result in the patient's being exposed to fewer inappropriate and hostile emotions on the part of other family members.
- Neuroleptic medication, by its prophylactic effect, helps prevent the patient from being impaired by the emotions that are present.
- 40% of schizophrenic adolescents cannot resume their scholastic or occupational activities, or even return to their home environment, immediately after their discharge from inpatient psychiatric care, either because their illness has become chronified, or because of severe disturbances in their families. This group of patients requires a rehabilitation program designed to enable them, after one to two years of treatment, to become reintegrated into their familiar environment, or to develop a new perspective on advancement in school and at work.

A program of this type was developed at the Marburg Clinic in cooperation with the Leppermühle home for special educational needs. It was shown that this rehabilitation program meets the patients' needs and enables their gradual reintegration into school, the workplace, and the family. The four phases of this program of treatment and rehabilitation are shown in Table 6. The first two phases take place in the psychiatric clinic, the last two in the rehabilitation center.

The course of schizophrenic psychoses beginning in adolescence has been less well studied that that of schizophrenic psychoses beginning in childhood or adulthood. This may be because the psychoses of adolescence have already been grouped together with those of adulthood, as there is not expected to be any great difference between them. Yet this assumption is false, as the few studies published to date have shown: schizophrenic psychoses beginning in the prepubertal period and during adolescence have a less favorable course than those beginning in adulthood. Among cases beginning in adulthood, complete remission was observed in 25%, partial remission in 50%, and a chronic course in 25%; but, among cases beginning in adolescents, complete remission was observed in 23%, partial remission in 25%, and a chronic course in 52%.

7.6 Personality Disorders

The concept of personality disorders has always been problematical in child and adolescent psychiatry (cf. Remschmidt 1978). A historical look at the development of this concept reveals that arguments for and against this type of diagnosis (then termed psychopathy) were advanced decades ago, and the problem is just as real today. In ICD-10, the corresponding chapter is entitled "Disorders of Adult Personality and Behaviour." This title reflects that view that certain particularly marked behavioral abnormalities and disorders of reactive origin are related to the personality disorders. DSM-IV proceeds along similar lines. The most important personality disorders, according to ICD-10, are listed in Table 7.

The frequency of personality disorders in the general population has been estimated as lying between 3% and 8%. Among patients in psychiatric clinics, the rate is about 40%; the disorders most commonly diagnosed are borderline personality disorder, emotionally unstable personality disorder, and anxious personality disorders.

Both biological and psychological factors contribute to the etiology and pathogenesis of the personality

Table 6. The four phases of a treatment and rehabilitation program for adolescents with schizophrenic psychoses, and the objectives of each phase (After Martin and Remschmidt 1983)

Acute phase (clinical treatment)	Remission phase (clinical treatment)	Rehabilitation phase I (care in a home)	Rehabilitation phase II (supervised living group)
Inpatient admission	Further inpatient care	Depot medication, group therapy	Depot medication, becoming independent in the group
Neuroleptic medication	Neuroleptic medication	Individual therapy	
Rapid mobilization	Integration in the group	Practice of daily activities	Self-sufficiency
Individual therapy and care	Participation in a working group	School attendance	Vocational training or apprenticeship
Occupational therapy	School attendance or individual instruction	Finding a vocation	
Maintaining contact with the family	Reality training	Reality training	
Group activities (where possible)	Concentration training Activities in town, becoming independent, leaves of absence, family discussions	Creative support, family contact	
Objective: Improvement of acute manifestations, prevention of withdrawal and chronification	*Objective:* Reintegration in the clinical setting	*Objective:* Reintegration in the larger community, adjustment to reality, finding and preparing for a vocation	*Objective:* Self-sufficiency, vocational development

disorders. As for biological factors, twin and adoption studies have shown that genetic influences are very important in determining personality. Because many personality disorders are characterized by impulsive behavior, lack of self-control, and aggressive impulses, there has been a search for biological markers that might explain these behavioral abnormalities. It is now regarded as scientifically established that they are related to a diminished activity of the serotonergic system. Moreover, neuropsychological deficits corresponding to diffuse or localized disturbances of brain function have been found to be more common in persons with personality disorders than in normal individuals. Finally, persons with personality disorders also have been found to suffer more frequently from attention disorders, and to have a diminished capacity for habituation.

Psychosocial factors such as stressful conditions in early childhood, being raised outside the family, experiences of separation and loss, and physical and sexual abuse – thus, generally speaking, highly unstable developmental conditions – are thought to promote the development of personality disorders in individuals with an underlying (genetic) predisposition. From the point of view of developmental psychology, personality disorders are due to a lack of development of processes of structuring and differentiation.

In the path-breaking conception of August Aichhorn (1925), psychopathy and demoralization were considered to result from ego dysfunction and ego weakness. The therapist's task is accordingly to make up for these developmental deficits. This can be a lengthy process. The concept of pedagogical guidance is most commonly applied in the treatment of personality disorders. Patients, particularly adolescents, with personality disorders can be brought to a relatively good level of adjustment through a suitable configuration of the treatment milieu and environmental influences. Short-term treatment programs are of no use in the treatment of such severe disorders.

There are major differences among personality disorders with respect to their course. The personality disorders that are associated with impulsiveness and antisocial and delinquent behavior have the least favorable prognosis, as was revealed by follow-up studies as early as the 1960's (Robins 1966, Sundby and Kreyberg 1968). On the other hand, many pathological features become milder as patients grow older, as found by Tölle (1966) on long-term follow-up of "psychopathic adults." Approximately two thirds of his sample group ($n = 115$) had achieved a satisfactory mode of living, though sometimes with restrictions. It must be mentioned that the risk of suicide in persons with personality disorders is approximately three times higher than that of the general population; the highest suicide rates are found among patients with borderline or narcissistic personality disorder.

Table 7. Overview of the different types of personality disorders, according to the ICD-10 (WHO 1992)

Specific personality disorder	ICD-10 category	Principal features
Paranoid	F60.0	Suspiciousness and a pervasive tendency to distort experience by misconstruing the neutral or friendly actions of others as hostile or contemptuous
Schizoid	F60.1	Limited capacity to experience and express emotion, anhedonia, emotional coldness, aloofness, social indifference and isolation, shyness and reserve in interpersonal contacts
Dissocial	F60.2	Habitual tendency to deviancy and delinquency, which begins in childhood in many cases; characteristic features such as lack of empathy, callousness, egocentrism, and underdeveloped conscience; also impulsiveness, persistent irritability, unreliability, weakness of attachments, and disregard or violation of social norms
Emotionally unstable	F60.3	*Impulsive type:* lack of impulse control and affect guidance, easy excitability, and, particularly in frustrating situations, a tendency toward outbursts of violent and threatening behavior
		Borderline type: instability of mood and of interpersonal relationships, identity crises, aggressive outbursts under emotional stress, autoaggressive tendencies ranging to drastic self-destructive and parasuicidal acts
Histrionic	F60.4	Dependence on attention from others, confirmation and recognition; labile and superficial affect, tendency toward dramatization, artificiality, and flirting; overconcern with physical attractiveness and inappropriately seductive behavior; suggestibility; lack of consistently maintained goals and values, leading to instability in interpersonal relationships and partnerships
Anankastic	F60.5	Conscientiousness, perfectionism, orderliness, conformity to rules, rigidity, indecisiveness, insecurity; difficult interpersonal relationships
Anxious (avoidant)	F60.6	Insecurity, feelings of inferiority, timidity; shyness, tension, clumsiness in interpersonal contacts; extreme fear of disapproval and rejection
Dependent	F60.7	Inability to lead one's own life independently; inability to assume responsibility for oneself; in close relationships, constant fear of loss and abandonment, accompanied by efforts to adapt to others and comply with their wishes

7.7 Sexual Disorders

The adult-type disorders of early onset include disorders of sexual identity, which may be seen even in childhood and becomes more common in preadolescence (atypical sex-role behavior, transvestitism, transsexualism), and disorders of sexual preference (exhibitionism and fetishism) (Green 1994; Martin and Remschmidt 1997).

In disorders of sexual identity in childhood and preadolescence, there is a discrepancy between the child's anatomical sex and its role behavior. The terms "atypical sex-role behavior" and "atypical psychosexual development" apply to these disorders, and are synonymous. In clothing, habits, and relationships with the opposite sex, affected children adopt modes of behavior that are inconsistent with their own sex. The ICD-10 speaks of a "gender identity disorder of childhood" with the major characteristic that the child wishes to belong to the opposite sex, without desiring to be anatomically transformed, as in transsexualism. The DSM-IV also provides an independent category for this disorder (gender identity disorder), which is very rare. No epidemiologic data are available.

Transvestitism is a deviant sexual behavior in which wearing clothes belonging to the opposite sex generates sexual pleasure, but there is no desire to assume another sexual identity. This disorder is designated by the same term in ICD-10 and DSM-IV. Transvestitism occurs in both male and female children and adolescents. Its prevalence in adults is one per 100,000 males and one per 400,000 females.

Transsexualism is a disorder of sexual identity characterized by an unshakeable desire to belong to the opposite sex, and to achieve this goal by means of a sex-change operation. This desire is a continuation of the mental constellation of the affected individuals as children and adolescents; they view themselves as unlucky individuals who think and feel as women, while being trapped in the body of a man (or vice versa).

The ICD-10 diagnostic guidelines stress that, in addition to the desire to be anatomically transformed,

the transsexual identity must persist for at least 2 years and must not be a manifestation of another disorder, such as schizophrenia or genetic or sex-chromosome abnormalities. DSM-IV provides no independent category for transsexualism and instead subsumes it under the category of gender identity disorder.

Transsexualism occurs in both sexes and is more common in women than in men (sex ratio, 3:2). It is a rare disorder and is thus only rarely seen in child and adolescent psychiatry. When such young people come for an office visit or are brought to the clinic by their parents, the problem is usually very serious, and is often accompanied by further mental disturbances (e.g. suicidal tendencies, depression, aggressive behavior).

The indication for treatment depends on whether the affected young person wishes to develop a sexual identity in accordance with his or her anatomical sex; only then is the patient motivated to undergo treatment. In the absence of such motivation, treatment, in the proper sense of the word, is not possible; the intervention must be limited to educating the young person about the nature of the disorder. At the same time, the parents are a second focus of treatment. They, too, must be educated about the nature of the disorder, and they must learn to accept their child's lack of motivation to undergo treatment. The therapist can best achieve these goals through joint conversations with the parents and the affected young person, in which all open questions can be discussed from both sides.

Some of these cases will pass over into genuine transsexualism, in which there is a desire for a surgical sex change. The prognosis is considerably more favorable when the patient is motivated to acquire a sexual identity in accordance with his or her anatomical sex. In such cases, the frequency of secondary mental disorders is also lower than in cases that progress to genuine transsexualism.

7.8 Substance Abuse and Drug Dependency

Mental and behavioral disorders due to psychoactive substance use are classified in ICD-10 under the category F1. By international agreement, these disorders are classified according to the substance used. They are discussed in detail in Vol. 3, Part 2, and we will therefore not describe them more extensively in this chapter on child and adolescent psychiatry. We will restrict the discussion here to the question of comorbidity with other types of developmental psychopathology (Schulz and Remschmidt 1999). Knowledge of the relationship between substance abuse and dependency and comorbid psychiatric disorders in childhood and adolescence opens new possibilities for the understanding and early recognition of risk factors for addiction. A complex interaction takes place between personality traits, increased vulnerability for mental disorders, and familial and psychosocial stress.

The problem of comorbidity of drug abuse in adolescence may be described as follows:

- Many psychopathological manifestations may be consequences of drug abuse (e.g. amotivational syndrome in cannabis abuse, drug-induced psychoses).
- A number of psychiatric illnesses in childhood and adolescence make drug abuse more likely and influence its course (e.g. hyperkinetic syndrome, conduct disorder, dissocial personality disorder).
- Drug abuse, in turn, influences the course of various psychiatric disorders (e.g. depressive syndromes, bulimia nervosa).
- Familial and psychosocial risk factors favor the development and maintenance of a drug problem (e.g. a broken home, addiction and dissocial personality disorders in other family members, deprivation syndromes, etc.).

Currently available longitudinal studies of the period from kindergarten to adolescence show that children at risk of developing drug addiction can be identified at an early age. As early as kindergarten, these children manifest aggressive-expansive behavior, lack of self-control, and a high degree of impulsiveness. They are easily excited and blind to danger; their behavior is characterized by hasty action, an intense search for immediate reinforcement, and an elevated sensitivity to external stimuli, and they often behave inconsiderately toward others. These findings should be used to develop preventive measures that can then be put into practice.

8 References

Agras S, Sylvester D, Oliveau D (1969) The epidemiology of common fears and phobias. Compr Psychiatry 10: 151–156

Aicardi J (ed) (1992) Diseases of the nervous system in childhood. McKeith, London (Clinics in developmental medicine 115/118)

Aichhorn A (1925) Verwahrloste Jugend. Huber, Bern

Aichhorn A (1971) Verwahrloste Jugend, 7th edn. Huber, Bern

*Ainsworth MD (1978) Patterns of attachment : a psychological study of the strange situation. Erlbaum, Hillsdale/NJ

Alexander F (1943) Fundamental concepts of psychosomatic research. Psychosom Med 5: 205–210

Andreasen NC (1984a) Scale for the assessment of negative symptoms (SANS). University of Iowa, Iowa City

Andreasen NC (1984b) Scale for the assessment of positive symptoms (SAPS). University of Iowa, Iowa City

Angst J, Dobler-Mikola A (1985a) The Zurich Study. V. Anxiety and phobia in young adults. Eur Arch Psychiatry Neurol Sci 235: 171–178

Angst J, Dobler-Mikola A (1985b) The Zurich Study. VI. A continuum from depression to anxiety disorders ? Eur Arch Psychiatry Neurol Sci 235: 179–186

*Beidel DC, Turner SM (1998) Shy children, phobic adults. Nature and treatment of social phobia. American Psychological Association, Washington

Berblinger K (1960) Hysterical crisis and the question of hysterical character. Psychosomatics 1: 270–279

**Bernstein G, Borchardt CM, Perwien AR (1996) Anxiety disorders in children and adolescents: a review of the past 10 years. J Am Acad Child Adolesc Psychiatry 35: 1110–1119

Biederman J, Faraone SV, Milberger S, Jetton JG, Chen L, Mick E, Greene RW, Russell RL (1996) Is childhood oppositional defiant disorder a precursor to adolescent conduct disorder? Findings from a four-year follow-up study of children with ADHD. J Am Acad Child Adolesc Psychiatry 35: 1193–1204

Bird H, Canino G, Rubio-Stipec M, Gould MS, Ribera J, Sesman M, Woodbury M, Huertas-Goldman S, Pagan A, Sanchez-Lacay A et al (1988) Estimates of the prevalence of childhood maladjustment in a community survey in Puerto Rico. Arch Gen Psychiatry 45: 1120–1126

Bishop DVM (1985) Age of onset and outcome in 'acquired aphasia with convulsive disorder' (Landau-Kleffner syndrome). Dev Med Child Neurol 27: 705–712

Bishop DVM (1994) Developmental disorders of speech and language. In: Rutter M, Taylor E, Hersov L (eds) (1994) Child and adolescent psychiatry. Blackwell, London, pp 546–568

Blanz B, Lehmkuhl G, Lehmkuhl U, Braun-Scharm H (1987) Hysterische Neurosen im Kindes- und Jugendalter. Z Kinder Jugendpsychiatr 15: 97–111

Bouchard TJ Jr, McGue M (1990) Genetic and rearing environmental influences on adult personality: an analysis of adopted twins reared apart. J Pers 58: 263–292

*Bundesarbeitsgemeinschaft für Rehabilitation (ed) (1994) Rehabilitation Behinderter. Wegweiser für Ärzte und weitere Fachkräfte der Rehabilitation, 2nd edn. Deutscher Ärzteverlag, Cologne

Christophersen ER, Rapoff MA (1983) Toileting problems of children. In: Walker CE, Roberts MC (eds) Handbook of clinical child psychology. Wiley, New York, pp 593–615

Clore ER, Hibel J (1993) The parasomnias of childhood. J Pediatr Health Care 7: 12–16

**Cohen DJ, Volkmar FR (eds) (1997) Handbook of autism and pervasive developmental disorders, 2nd edn. Wiley, New York

Deonna T, Ziegler A, Malin-Ingvar M, Ansermet F, Roulet E (1995) Reversible behavioural autistic-like regression: a manifestation of a special (new) epileptic syndrome in a 28-month-old child: a 2-year longitudinal study. Neurocase 1: 1–9

Eiberg H, Berendt I, Mohr J (1995) Assignment of dominant inherited nocturnal enuresis (ENUR 1) to chromosome 13q. Nat Genet 10: 354–356

*Esser G (1991) Was wird aus Kindern mit Teilleistungsschwächen? Der langfristige Verlauf umschriebener Entwicklungsstörungen. Enke, Stuttgart

Evans D, Noam G, Wertlieb D, Paget K, Wolf M (1994) Self-perception and adolescent psychopathology: a clinical-developmental perspective. Am J Orthopsychiatry 64: 293–300

Fichter MM (1989) Bulimia nervosa und bulimisches Verhalten. In: Fichter MM (ed) Bulimia nervosa. Grundlagen und Behandlung. Enke, Stuttgart, pp 1–10

**Garner DM, Garfinkel PE (eds) (1997) Handbook of treatment for eating disorders, 2nd edn. Guilford, New York

Garrison CZ, Weinrich MW, Hardin SB, Weinrich S, Wang L (1993) Post-traumatic stress disorder in adolescents after a hurricane. Am J Epidemiol 138: 522–530

Giaconia RM, Reinherz HZ, Silverman AB, Pakiz B, Frost AK, Cohen E (1995) Traumas and posttraumatic stress disorder in a community population of older adolescents. J Am Acad Child Adolesc Psychiatry 34: 1369–1380

Gillberg C (1995) Rett-Syndrome. In: Gillberg C (ed) Clinical child neuropsychiatry. Cambridge University Press, Cambridge, pp 235–243

Gilles de la Tourette G (1885) Étude sur une affection nerveuse caracterisée par l'incoordination motrice accompagnée d'écholalie et de coprolalie (Jumping, Latah, Myriachit). Arch Neurol (Paris) 9: 19–42, 158–200

*Goodyer IM (ed) (1995) The depressed child and adolescent. Developmental and clinical perspectives. Cambridge University Press, Cambridge

Green R (1994) Atypical sexual development. In: Rutter M, Taylor E, Hersov L (eds) Child and adolescent psychiatry. Blackwell, London, pp 749–758

*Häfner H (1983) Allgemeine und spezielle Krankheitsbegriffe in der Psychiatrie. Nervenarzt 54: 231–238

Hansson S (1992) Urinary incontinence in children and associated problems. Scand J Urol Nephrol Suppl 141: 47–57

*Harrington R (1993) Depressive disorder in childhood and adolescence. Wiley, Chichester

Harrington R, Fudge H, Rutter M, Pickles A, Hill F (1990) Adult outcomes of childhood and adolescent depression: I. Psychiatric status. Arch Gen Psychiatry 47: 465–473

Harrington R, Fudge H, Rutter M, Pickles A, Hill F (1991) Adult outcomes of childhood and adolescent depression. II. Links with antisocial disorders. J Am Acad Child Adolesc Psychiatry 30: 434–439

Harrington R, Wood A, Ferduyn C (1998) Clinically depressed adolescents. In: Graham P (ed) Cognitive behaviour therapy for children and families. Cambridge University Press, Cambridge, pp 156–193

Heller T (1908) Über Dementia infantilis (Verblödungsprozeß im Kindesalter). Z Erforsch Behandl Jugendl Schwachsinn Wissenschaftl Grundl 2: 17–28

Herpertz-Dahlmann B, Remschmidt H (1988) Somatische Störungen bei Anorexia nervosa. Monatsschr Kinderheilkd 136: 732–738

Hebebrand J, Barth N, Coners H, Hinney A, Himmelmann GW, Remschmidt H (1997a) Strategien zur Aufklärung der genetischen Prädisposition zur Anorexia nervosa. In: Willenberg H, Hoffmann SO (eds) Handeln: Ausdrucksform psychosomatischer Krankheit und Faktor der Therapie. VAS, Frankfurt (Main), pp 97–101

Hebebrand J, Klug B, Fimmers R, Seuchter SA, Wettke-Schäfer R, Deget F, Camps A, Lisch S, Hebebrand K, von Gontard A, Lehmkuhl G, Poustka F, Schmidt M, Baur MP, Remschmidt H (1997b) Rates for tic disorders and obsessive-compulsive symptomatology in families of children and adolescents with Gilles de la Tourette syndrome. J Psychiatr Res 31: 519–530

Herpertz-Dahlmann B, Remschmidt H (1989) Anorexia nervosa und Depression: Zur Gewichtsabhängigkeit der depressiven Symptomatik. Nervenarzt 60: 490–495

Hinman F (1986) Non-neurogenic neurogenic bladder (the Hinman syndrome) – 15 years later. J Urol 136: 769–777

Hjälmås K (1992) Urinary incontinence in children: suggestions for definitions and terminology. Scand J Urol Nephrol Suppl 141: 1–6

*Howlin P (1997) Prognosis in autism: do specific treatments affect long-term outcome. Eur Child Adolesc Psychiatry 6: 55–72

Jenkins RL (1968) The varieties of children's behavioral problems and family dynamics. Am J Psychiatry 124: 1440–1445

Kanner L (1943) Autistic disturbances of affective contact. Nerv Child 2: 217–250

Kazdin AE (1987) Treatment of antisocial behavior in children: current status and future directions. Psychol Bull 102: 187–203

Kazdin AE (1995) Conduct disorders in childhood and adolescence, 2nd edn. Sage, Thousand Oaks

*Kazdin AE (1997) Psychosocial treatments for conduct disorder in children. J Child Psychol Psychiatry 38: 161–178

King BH, State MW, Shah B, Davanzo P, Dykens E (1997) Mental retardation: a review of the past 10 years, part I. J Am Acad Child Adolesc Psychiatry 36: 1656–1663

Klackenberg G (1987) Incidence of parasomnias in children in a general population. In: Guilleminault C (ed) Sleep and its disorders in children. Raven, New York, pp 99–113

Klein RG (1994) Anxiety disorders. In: Rutter M, Taylor E, Hersov L (eds) Child and adolescent psychiatry. Blackwell, London, pp 351–374

Kohlberg L (1984) The psychology of moral development: nature and validity of moral stages. Harper and Row, San Francisco

Korhonen T (1984) A follow-up study of Finnish children with specific learning disabilities. Acta Paedopsychiatr 50: 255–263

*Kutcher SP (1997) Child and adolescent psychopharmacology. Saunders, Philadelphia

Lader MH, Wing W (1966) Physiological measures, sedative drugs and morbid anxiety. Oxford University Press, London

Lazarus RS (1966) Psychological stress and the coping process. McGraw-Hill, New York

Lazarus RS (1981) Streß und Streßbewältigung – ein Paradigma. In: Filipp SH (ed) Kritische Lebensereignisse. Urban and Schwarzenberg, Munich, pp 198–232

Leckman JF, Cohen DJ (eds) (1999) Tourette's syndrome. Wiley & Sons, New York Chichester Weinheim Brisbane Singapore Toronto

*Leonhard K (1995) Aufteilung der endogenen Psychosen und ihre differenzierte Ätiologie, 7th edn. Thieme, Stuttgart

Lerman P, Lerman-Sagie T, Kivity S (1991) Effect of early corticosteroid therapy for Landau-Kleffner Syndrome. Dev Med Child Neurol 33: 257–260

Levine MD (1975) Children with encopresis: a descriptive analysis. Pediatr 56: 412–416

Lundberg U (1986) Stress and type A behavior in children. J Am Acad Child Psychiatry 25: 771–778

Magnusson D, Stattin H, Allen UL (1986) Differential maturation among girls and its relation to social adjustment: a longitudinal perspective. In: Baltes PB, Featherman DL, Lerner RM (eds) Life-span development and behavior, vol 7. Academic, New York

March JS, Leonard HL (1996) Obsessive-compulsive disorder in children and adolescents: a review of the past 10 years. J Am Acad Child Adolesc Psychiatry 35: 1265–1273

Marks IM (1987) Fears, phobias, and rituals: panic, anxiety, and their disorders. Oxford University Press, New York

Martin M, Remschmidt H (1983) Ein Nachsorge- und Rehabilitationsprojekt für jugendliche Schizophrene. Z Kinder Jugendpsychiatr 11: 234–242

Martin M, Remschmidt H (1997) Störungen der Sexualentwicklung und des Sexualverhaltens. In: Remschmidt H (ed) Psychotherapie im Kindes- und Jugendalter. Thieme, Stuttgart, pp 250–258

Matt E (1995) Episode und "Doppel-Leben": Zur Delinquenz Jugendlicher. Monatsschrift Kriminol Strafrechtsreform 3: 153–164

Maughan B, Yule W (1994) Reading and other learning disabilities. In: Rutter M, Taylor E, Hersov L (eds) Child and adolescent psychiatry. Blackwell, London, pp 647–665

Michelson LK, Ray WJ (eds) (1996) Handbook of dissociation. Plenum, New York

Moffitt TE (1993) Adolescence-limited and life-course-persistent antisocial behavior: a developmental taxonomy. Psychol Rev 100: 674–701

Neudörfl A, Herpertz-Dahlmann B (1996) Der Verlauf von Zwangserkrankungen im Kindes- und Jugendalter – Eine Literaturübersicht. Z Kinder Jugendpsychiatr 24: 105–116

*Neuhäuser G, Steinhausen HC (eds) (1998) Geistige Behinderung. Grundlagen, klinische Symptome, Behandlung und Rehabilitation, 2nd edn. Kohlhammer, Stuttgart

Nissen G, Fritze J, Trott GE (1998) Psychopharmaka im Kindes- und Jugendalter. Fischer, Stuttgart

Olbing H (ed) (1993) Enuresis und Harninkontinenz bei Kindern. Marseille, Munich

Parry-Jones B, Parry-Jones WL (1992) Pica: symptom or eating disorder? A historical assessment. Br J Psychiatry 160: 341–354

Patterson GR, De Baryshe BD, Ramsey E (1989) A developmental perspective on antisocial behavior. Am Psychol 44: 329–335

Perry J, Cooper S (1989) An empirical study of defense mechanisms. Arch Gen Psychiatry 46: 444–452

*Propping P (1991) Psychiatrische Genetik. Springer, Berlin Heidelberg New York

Pynoos RS, Eth S (1986) Witness to violence: the child interview. J Am Acad Child Psychiatry 25: 306–319

Rapoport JL, Swedo S, Leonard H (1994) Obsessive-compulsive disorder. In: Rutter M, Taylor E, Hersov L (eds) Child and adolescent psychiatry. Blackwell, London, pp 441–454

Reinecker H (1990) Soziale und spezifische Phobien. In: Reinecker H (ed) Lehrbuch der klinischen Psychologie. Modelle psychischer Störungen. Hogrefe, Göttingen, pp 49–72

Rekers G (1982) Shaping your child's sexual identity. Baker, Grand Rapids/MI

Remschmidt H (1978) Die "Psychopathie" in der Kinder- und Jugendpsychiatrie. Z Kinder Jugendpsychiatr 6: 280–301

Remschmidt H (1988) Der Krankheitsbegriff in der Kinder- und Jugendpsychiatrie. In: Remschmidt H, Schmidt MH (eds) Kinder- und Jugendpsychiatrie in Klinik und Praxis, vol I, Thieme, Stuttgart, pp 143–152

*Remschmidt H (1992) Psychiatrie der Adoleszenz. Thieme, Stuttgart

Remschmidt H (1997a) Angstsyndrome. In: Remschmidt H (ed) Psychotherapie im Kindes- und Jugendalter. Thieme, Stuttgart, pp 198–219

Remschmidt H (1997b) Konversionssyndrome. In: Remschmidt H (ed) Psychotherapie im Kindes- und Jugendalter. Thieme, Stuttgart, pp 243–249

Remschmidt H (1998) Bipolar disorders in children and adolescents. Curr Opin Psychiatry 11: 379–383

Remschmidt H (ed) (2000) Schizophrenia in children and adolescents. Cambridge University Press, Cambridge

Remschmidt H (in press) Specific developmental disorders in childhood and adolescence. In: Gelder M (ed) Oxford textbook of psychiatry. Oxford University Press, Oxford

Remschmidt H, Hebebrand J (1993) Das Tourette-Syndrom: Eine zu selten diagnostizierte Tic-Störung? Deutsches Ärzteblatt 24: 1805–1810 (edition A1) (18. Juni) (edition A2: 1854–1860; B: 1287–1291; C: 1175–1179)

Remschmidt H, Herpertz-Dahlmann B (1988) Anorexia nervosa im Jugend-alter. Monatsschr Kinderheilkd 136: 718–723

Remschmidt H, Niebergall G (1997) Zwangssyndrome. In: Remschmidt H (ed) Psychotherapie im Kindes- und Jugendalter. Thieme, Stuttgart, pp 220–231

Remschmidt H, Remschmidt U (1974) Symptomatologie, Verlauf und Prognose von Tic-Erkrankungen im Kindes- und Jugendalter. Klin Pädiatr 186: 185–199

*Remschmidt H, Schmidt MH (eds) (1994) Multiaxiales Klassifikationsschema für psychische Störungen des Kindes- und Jugendalters nach ICD-10 der WHO, 3rd edn. Huber, Bern

Remschmidt H, Schulz E (1995) Psychopharmacology of depressive states in childhood and adolescence. In: Goodyer IM (ed) The depressed child and adolescent. Developmental and clinical perspectives. Cambridge University Press, Cambridge, pp 253–279

Remschmidt H, Walter R (1990a) Psychische Auffälligkeiten bei Schulkindern. Eine epidemiologische Untersuchung. Z Kinder Jugendpsychiatr 18: 121–132

*Remschmidt H, Walter R (1990b) Psychische Auffälligkeiten bei Schulkindern. Hogrefe, Göttingen

Remschmidt H, Martin M, Hennighausen K, Schulz E (2000) Treatment and rehabilitation. In: Remschmidt H (ed) Schizophrenia in children and adolescents. Cambridge University Press, Cambridge, pp 192–267

Rett A (1966) Über ein eigenartiges hirnatrophisches Syndrom bei Hyperammonämie im Kindesalter. Wien Med Wochenschr 116: 723–726

**Robins LN (1966) Deviant children grown up. Williams and Wilkins, Baltimore

Robson P (1988) Self-esteem: a psychiatric view. Br J Psychiatry 153: 6–15

Roder V, Brenner HD, Kienzle N et al. (1997) Integriertes Psychologisches Therapieprogramm für schizophrene Patienten, 4th edn. Psychologie Verlags Union, Weinheim

Rutter M, Tizard J, Whitmore K (eds) (1970) Education, health and behaviour. Longman, Harlow

*Rutter M, Yule P, Whitmore K (1976) Research report: Isle of Wight studies 1964–1974. Psychol Med 6: 313–332

**Rutter M, Izard CE, Read PB (eds) (1986) Depression in young people: developmental and clinical perspectives. Guilford, New York

Scarr S, McCartney K (1985) How people make their own environments: a theory of genotype-environment effects. Child Dev 54: 424–435

Schilling A (1963) Sprech- und Sprachstörungen. In: Berendes J, Link R, Zöllner F (eds) Hals-Nasen-Ohrenheilkunde, vol 2, part 2. Thieme, Stuttgart

Schulz E, Remschmidt H (1999) Substanzmißbrauch und Drogenabhängigkeit im Kindes- und Jugendalter. Dtsch Ärztebl 96: A414–418

Seligman MEP (1975) Helplessness. On depression, development and death. Freeman, San Francisco

Shapiro T, Huebner HF (1985) Motorische Störungen. In: Remschmidt H, Schmidt MH (eds) Kinder- und Jugendpsychiatrie in Klinik und Praxis, vol III. Thieme, Stuttgart, pp 70–76

Silberg J, Rutter M, Meyer J, Maes H, Hewitt J, Simonoff E, Pickles A, Loeber R, Eaves L (1996) Genetic and environmental influences on the covariation between hyperactivity and conduct disturbance in juvenile twins. J Child Psychol Psychiatry 37: 803–816

Skuse D (1994) Feeding and sleeping disorders. In Rutter M, Taylor E, Hersov L (eds) Child and adolescent psychiatry. Blackwell, London, pp 467–489

State M, King BH, Dykens E (1997) Mental retardation: a review of the past 10 years, part II. J Am Acad Child Adolesc Psychiatry 36: 1664–1671

Sundby HS, Kreyberg PC (1968) Prognosis in child psychiatry. Williams and Wilkins, Baltimore

Swedo SE, Rapoport JL (1990) Neurochemical and neuroendocrine considerations of obsessive-compulsive disorders in childhood. In: Deusch SI, Weizman A, Weizman R (eds) Application of basic neurosciences to child psychiatry. Plenum, New York, pp 275–284

Tinbergen N, Tinbergen EH (1984) Autismus bei Kindern. Parey, Berlin

Tölle R (1966) Katamnestische Untersuchungen zur Biographie abnormer Persönlichkeiten. Springer, Berlin Heidelberg New York

Torgersen S (1979) The nature and origin of common phobic fears. Br J Psychiatry 134: 343–351

Tourette Syndrome Association International Consortium for Genetics (1999) A complete genome screen in sib pairs affected by Gilles de la Tourette syndrome. Am J Hum Genet 65: 1428–1436

Tramer M (1942) Lehrbuch der allgemeinen Kinderpsychiatrie: einschließlich der allgemeinen Psychiatrie der Pubertät und Adoleszenz. Schwabe, Basel

Uldall B, Hansen FJ, Tonnby B (1993) Lamotrigine in Rett-Syndrome. Neuropediatrics 24: 339–340

*Vandereycken W, Beumont P (eds) (1998) Treating eating disorders: ethical, legal and personal issues. New York University Press, New York

van Gool JD, de Jonge GA (1989) Urge syndrome and urge incontinence. Arch Dis Child 64: 1629–1634

*Volkmar FR (1992) Childhood disintegrative disorder: issues for DSM-IV. J Autism Dev Disord 22: 625–642

von Gontard A (1998) Genetik der Enuresis nocturna. Med Gen 3: 415–416

von Gontard A, Eiberg H, Hollmann E, Rittig S, Lehmkuhl G (1998) Molecular genetics of nocturnal enuresis: clinical and genetic heterogenity. Acta Paediatr 87: 571–578

Warnke A (1987) Behandlung der Legasthenie im Kindesalter. Monatsschr Kinderheilkd 135: 302–306

Warnke A (in press) Rechenschwäche (Dyskalkulie). In: Lauth GW, Brack UB, Lindenkamp F (eds) Praxishandbuch Verhaltenstherapie bei Kindern und Jugendlichen. Beltz, Weinheim

Warnke A, Niebergall G (1997) Legasthenie und Rechenstörungen. In: Remschmidt H (ed) Psychotherapie im Kindes- und Jugendalter. Thieme, Stuttgart, pp 322–334

Warnke A, Roth E (2000) Umschriebene Lese-Rechtschreibstörung. In: Petermann F (ed) Lehrbuch der Klinischen Kinderpsychologie und -psychotherapie, 4th edn. Hogrefe, Göttingen, pp 453–476

Watson BU, Watson CAS, Fredd R (1982) Follow-up studies of specific reading disability. J Am Acad Child Psychiatry 21: 376–382

Weber D (1970) Der frühkindliche Autismus unter dem Aspekt der Entwicklung. Huber, Bern

Weber D (1985) Autistische Syndrome. In: Remschmidt H, Schmidt MH (eds) Kinder- und Jugendpsychiatrie in Klinik und Praxis, vol II. Thieme, Stuttgart, pp 269–298

Weber D, Remschmidt H (1997) Autismus. In: Remschmidt (ed) (1997) Psychotherapie bei Kindern und Jugendlichen. Thieme, Stuttgart, pp 356–374

Weinschenk C (1965) Die erbliche Lese-Rechtschreibschwäche und ihre sozialpsychiatrischen Auswirkungen, 2nd edn. Huber, Bern

Werry JS, Carlielle J, Fitzpatrick J (1983) Rhythmic motor activities (stereotypies) in children under five: etiology and prevalence. J Am Acad Child Psychiatry 22: 329–336

**WHO (1992) The ICD-10 classification of mental and behavioural disorders, clinical descriptions and diagnostic guidelines. World Health Organization, Geneva

WHO (1993) The ICD-10 classification of mental and behavioural disorders, diagnostic criteria for research. World Health Organization, Geneva

WHO (1996) Multiaxial classification of child and adolescent psychiatric disorders. Cambridge University Press, Cambridge

Wilsher CR (1986) The nootropic concept in dyslexia. Ann Dyslex 36: 118–137

Wittchen HU (1986) Epidemiology of panic attacks and panic disorder. In: Hand I, Wittchen HU (eds) Panic and phobias. Springer, Berlin Heidelberg New York

Woolston JL (1991) Eating and growth disorders in infants and children. Sage, Newbury Park (Developmental clinical psychology and psychiatry, vol 24)

Yule W (1992) Post-traumatic stress disorder in child survivors of shipping disasters: the sinking of the 'Jupiter'. Psychother Psychosom 57: 200–205

Zappella M, Genazzani A, Facchinetti F, Hayek G (1990) Bromocriptine in the Rett Syndrome. Brain Dev 12: 221–225

Zigler E (1967) Familial mental retardation: a continuing dilemma. Science 155: 292–298

Zitrin CM, Klein DF, Woerner MG, Ross DC (1983) Treatment of phobias. I. Comparison of imipramine hydrochloride and placebo. Arch Gen Psychiatry 40: 125–138

CHAPTER

8

H. Helmchen, H. Lauter

Diagnostic Problems in Geriatric Psychiatry

Translator: E. Taub

1 Introduction

There are three main areas of difficulty in the diagnosis of mental illness in elderly patients: (1) the differentiation of mental disorders from the mental changes of nonpathological aging, (2) the differentiation of mental disorders from the signs and symptoms of physical illness and impairment, and (3) the differentiation of mental illnesses from one another.

2 Differentiation of Mental Disorders from the Mental Changes of Nonpathological Aging

There are a number of methodological and factual problems complicating this area of differential diagnosis. The first methodological problem is that there is no accepted definition of "normal aging." If we were to define the norm as a statistical average, i.e. the mean values of quantitative indices of mental function across the entire population of individuals in a certain age category, then we would find that an impairment of cognitive ability satisfying the definition of dementia is "normal" for individuals over 100 years old, for example. To avoid such paradoxes, we may prefer to define the norm as the mean value of age-associated changes in healthy elderly individuals only. The distinction between health and illness is often blurred in old age, however, because nearly all individuals over 70 years old carry several medical diagnoses (Borchelt et al. 1996; Linden et al. 1996a). It is difficult to tell whether relatively mild and slowly progressive mental and intellectual changes in old age are products of the normal aging process or of disease, because many diseases of old age begin insidiously. Categorical diagnoses are therefore often based on threshold values of severity along one or more dimensions. The setting of these threshold values and the choice of the dimensions themselves are conventions that are admittedly influenced by contemporary trends and by other uncontrolled factors, such as subjective assessment of the state of health (Borchelt et al. 1996) or the recognition of social roles in the assessment of everyday activities.

Thus the recent trend toward increased life expectancy and better health in old age have led to the splitting of the "third life phase," i.e. age 70 and beyond, into two phases: a new "third life phase" from age 70 to approximately age 85, typified by an active life style (e.g. travel, sport, clothing), and then a "fourth life phase" of increasing physical frailty (P.B. Baltes 1997). In the multicultural context of a society that includes immigrants from many different parts of the world, specific cultural expectations for the behavior of older people may affect the way normal and diseased aging are distinguished from each other. The threshold for a diagnosis of dementia may be affected, for example, by the (still) high degree of respect for the elderly in Asian cultures or by culturally determined degrees of social support for the elderly, which may be sufficient to compensate for impaired performance of everyday activities (one of the dimensions of the diagnostic scheme).

In any case, psychiatric norms for old age are generally based on the second type of model discussed above, i.e. on the notion of "average mental health," however defined – even in the absence of adequate empirical evidence for many such parameters. Such evidence is vitally important, because parameters in different areas change in different ways over the course of aging (see Chap. 1, Vol. 2, Part 1), or some parameters (e.g. many laboratory values) may indeed remain quite stable from young adulthood to extreme old age (Kage et al. 1996).

The first concept of the norm seems to be the dominant one in the popular understanding, in which dementia is seen as a "normal," inevitable component of aging, and depression as a "normal" reaction to the common experiences of loss in old age; neither is considered a sign of disease. The many consequences of this assessment of aging, which does not effectively distinguish health from disease and thus amounts to a negative stereotype, are discussed at length in the Anglo-American literature under the heading of "ageism" (Butler 1990; Illhardt 1993). Even scientifically trained physicians are influenced by such prejudices, when they fail to recognize the mental illnesses of old age as illnesses (Helmchen et al. 1999), or when they wrongly consider present impairments to be irreversible and fail to treat them (Linden et al. 1996b). In contrast to this negative stereotype of aging, which is still widespread in the medical community, a very different positive stereotype of "the golden autumn of life" has come about in the context of psychological research on aging (Thomae and Maddox 1982; Lehr and Thomae 1987). This equally erroneous stereotype probably arose because these studies were performed largely on people who were healthy enough to respond to the advertisements for study volunteers and mobile enough to visit the research sites without external help.

Thus stereotypes of aging might affect the investigator's diagnostic threshold values, i.e. the diagnosis of mental illness in old age is susceptible not only to variations in diagnostic criteria, resulting from lack of scientific knowledge, but also to subjectively colored, method-dependent variations in observation.

A further methodological problem is posed by the cohort-specific appropriateness of instruments (Israel et al. 1984). Satzger et al. (1996) showed that the reference values of the Hamburg-Wechsler intelligence test (HAWIE), which are set in accordance with a sampling of the population at large, have risen by a full standard deviation in the last 50 years; in other words, the population has become "more intelligent" in terms of the HAWIE, and the reference values have therefore been corrected upward. The use of the new, corrected instrument in very old individuals will therefore yield an underestimate of their intelligence, because the altered reference values were derived from a different age cohort. On the other hand, Knäuper and Wittchen (1994) showed that depression in individuals over 80 years old is inadequately detected by the Diagnostic Interview Schedule (DIS), because the linguistic usage of the test instrument no longer corresponds to that of the elderly test-takers. Moreover, the considerable amount of psychiatric morbidity that falls below the threshold values set by the current systems for operational diagnosis (ICD-10, DSM-IV) – so-called subthreshold morbidity – leads us to ask whether these systems give an accurate picture of the extent of psychiatric morbidity in the elderly. According to the Berlin Old Age Study (*Berliner Altersstudie*, BASE; Helmchen et al. 1999), 17% of a random age- and sex-stratified sample of the 70- to 100-year-old population of West Berlin had subthreshold psychiatric morbidity. Even if this prevalence figure is not significantly higher than that of younger age-groups (Maier et al. 1994), it is still remarkably high. Above all, further studies are needed to better define the border between subthreshold psychiatric morbidity and lesser degrees of impaired well-being that are not the product of illness (Häfner 1997; Helmchen and Linden 2000), as well as to confirm or refute the hypothesis that there are age-specific difficulties in diagnosis, not only because of the increased prevalence of somatic illness and co- and multimorbidity in old age (see below), but also because of variable changes in affect, motivation, and cognitive performance.

Jacob (1986) vividly described the experiential change of the elderly, which takes place on the background of the generalized multimorbidity of old age. He stressed the need to complement the usual clinical and psychological tests of disease manifestations and impaired performance with the exploration and analysis of internal perceptions and experiences. Certain features of inner experience are common to normal and pathological aging, according to Jacob; the structure of inner experience in the elderly is typified by "the progressive infiltration of the past into the present, leading to the increasing availability of long-remembered information, and the decreasing relative availability of information that is currently being acquired." He drew particular attention to the "bland indifference, unshakeable passivity, and aloof attitude" of patients with certain diseases of the brain, and to their "selective inattention or semi-attention, non-perception, and unawareness of central nervous impairments."

Subthreshold morbidity in old age most often falls into the category of the affective disorders. Subthreshold affective abnormalities differ from the operationally defined affective disorders not only quantitatively, in the lesser severity of disease manifestations (such as feelings of guilt, feelings of worthlessness, and suicidal ideation) and in their shorter duration (more fluctuating course of depression), but probably also quantitatively, in the psychopathological pattern of worry, lack of well-being, and suffering (Geiselmann and Bauer 2000; Linden et al. 1998). Such abnormalities are difficult to label as pathological. Depressed affect in old age, often referred to as "body-centered complaints," "mild resignation," or "unhappiness," may be considered an appropriate reaction to the patient's changing life circumstances and finite life prospects (Palmore 1988), while elevated affect, perceived as "friendly remoteness," or a perhaps remarkable cheerfulness, despite physical or socioeconomic hardship, may be interpreted as a "wise" adaptive mechanism, as an expression of the protective effects of self-perception, self-esteem, and individual personality traits in the elderly, or as a "protective illusion" (Staudinger et al. 1996). An apparently unreasonable cheerfulness may, in fact, be pathological euphoria accompanying an organic brain disease (Alsen 1960); in the elderly, it is not always easy to distinguish cheerful imperturbability from uncritical indifference (see Chaps. 11 and 12, Vol. 2, Part 2). Above all, dejection, or loss of the will to live, should not simply be taken as an appropriate reaction to the patient's life situation, for depression in the elderly is known to be too rarely recognized and treated (Helmchen et al. 1999). Linden and Barnow (1998), authors of the BASE, concluded that expressions of being "tired of life," or of thoughts about death, are relatively uncommon even in the elderly and should always be followed up by an exploration of possible death wishes or suicidality, which may be symptoms of depression requiring treatment.

The relation between age-specific and dementia-induced changes in cognitive performance has also been inadequately studied to date, particularly in the extremely aged (Helmchen and Reischies 1998). Because both old age and dementia give rise to cognitive deficits, a problem of differential diagnosis arises. In the diagnostic process, it is important to bear in mind the concept of cognitive aging and the distinction between the clinical dementia syndrome on one hand and cerebral dementing disease on the other, as well as empirical findings regarding (a) the patient's current

neuropsychological state, (b) his or her premorbid intelligence, and (c) the rate of progression of the deficit.

Aging alone considerably reduces the speed of cognitive processes in healthy individuals (Lindenberger et al. 1993; Salthouse 1985). Salthouse found that the maximum speed of a 70-year-old was only about 40% of that of a 20-year-old (Salthouse 1982). A 60% reduction of speed, however, by no means implies a 60% reduction of other aspects of cognitive performance; these were essentially unchanged (Reischies and Lindenberger 1996). It would not be considered a significant cognitive impairment if a 60-year-old could converse freely just as well as 20-year-old, while taking twice as much time (though perhaps the 20-year-old might think so). On the other hand, qualitative intellectual impairment, e.g. reduction of the cognitive reserve capacity, as revealed by the "testing the limits" method (M.M. Baltes et al. 1992), and impairments of such functions as orientation, the use of tools, the recognition of known faces, the use of language, and word-finding, are signs of dementia.

Baseline adult intelligence is obviously important. A very old individual of low baseline intelligence might be assigned the diagnosis of a dementia syndrome in the absence of any causative dementing illness. A low level of education or vocational training is thus considered a risk factor for dementia (Fratiglioni et al. 1991; Kay et al. 1964; O'Connor et al. 1991; Parsons 1965), although there is some controversy on this point (Moritz and Petitti 1993). It remains unclear whether education simply alters the threshold for clinical manifestations, because highly educated individuals can (perhaps) compensate for an impairment of cognitive functioning longer than others, or whether there is also an effect of education on the disease process itself: a high educational level might be associated with a larger amount of neuronal stimulation, or with a more healthful lifestyle, and thus perhaps lead to a lesser risk of the chronic illnesses or deficits that give rise to dementia (Berkman et al. 1986; Mortimer 1990; Helmchen et al. 1999).

As for the processes that lead to impairment of cognitive performance, and ultimately to the diagnosis of the dementia syndrome, an important new piece of evidence was provided by the recent study by Davidson et al. (1995). Dementia is significantly more common in elderly schizophrenics than in the elderly population in general. Davidson and colleagues found, and others subsequently confirmed, that the neuropathological findings in demented elderly schizophrenics have no features in common with those of patients with neurodegenerative dementing illnesses. It was concluded that the dementia of elderly schizophrenics is the product of a different pathophysiological mechanism (Arnold et al. 1998; Purohit et al. 1998; Rajkowska et al. 1998). The important consequence of this fact for nosologic diagnosis is that the differential diagnosis of dementia must take factors other than neurodegenerative illness into account.

The normal aging process leads to a deterioration of cognitive performance over the eighth, ninth, and tenth decades of life by approximately one to two standard deviations below the normal values for retirement age (Reischies and Lindenberger 1996). It therefore follows that the cognitive performance of a less gifted person, or one suffering from a persistent, mild cognitive disturbance because of schizophrenia or prior brain injury (Harvey et al. 1995), may deteriorate to a very low level over the course of normal aging, with the result that the affected person may be unable, or barely able, to adapt to the changing demands of everyday life (see Chap. 10, Vol. 2, Part 2). Such individuals will become increasingly dependent on the help of others, and some diagnosticians might consider them to have a mild dementia syndrome, in accordance with the accepted diagnostic criteria for dementia, even though no dementing illness per se is present. Conversely, the development of a dementia syndrome in a person of high baseline adult intelligence may be undiagnosable by standard diagnostic criteria until the dementing illness has reached a relatively advanced stage. In both of the cases just discussed, an assessment of the patient's present cognitive state alone, without taking its temporal course into account, would lead to an erroneous diagnosis.

It thus becomes necessary to consider the rate of deterioration of cognitive performance. Longitudinal studies have shown that the test performance of many elderly individuals is essentially constant, i.e. does not decline to any psychometrically detectable degree after an interval of 1 year (Brayne et al. 1995; Reischies and Geiselmann 1997; Schmand et al. 1995). This state of affairs is clearly distinguishable from the rapid deterioration of patients with Alzheimer's dementia, whose score on the Mini-Mental Status Examination (MMSE) declines by 3 to 4 points per year (Berg et al. 1990; Reischies and Schaub 1997).

This rate-of-change criterion for the early diagnosis of dementia would be more useful if tests were available that were more sensitive to small changes, and if empirically determined threshold values for the rate of deterioration could be applied. The MMSE cannot adequately assess the cognitive performance of highly capable individuals, who have to suffer considerable deterioration before any change in the MMSE score is noticeable; it also remains to be demonstrated that a rate of change in MMSE scores of 1 point or less per year, or of 3 or more points per year (say), is indeed a valid criterion for the diagnosis of normal or pathological cognitive aging, respectively.

Despite the methodological difficulties of using a rate-of-change criterion, including its high variability and its susceptibility to learning effects, and despite its lesser applicability to patients with a fluctuating course, or to those, such as stroke patients, who remain stable at a low level of function, it is nevertheless clear that consideration of the rate of progression may be very useful in the diagnosis of dementia, particularly when the cognitive deficit is still relatively mild.

3 Differentiation of Mental Disorders from the Symptoms and Signs of Physical Illness and Impairment

A further problem complicating the psychiatric assessment of the elderly arises from the fact that older individuals often suffer from one or more somatic illnesses, whose effects are not always easily distinguished from the somatic and psychic consequences of mental illness, and the resulting performance impairment and disability. Participants in the BASE underwent a comprehensive physical examination by medical specialists in several disciplines and received an average of eight medical diagnoses, of which an average of two were considered severe (Linden et al. 1995); 30% of patients 70 years and older suffered from five or more moderately severe or severe illnesses (Steinhagen-Thiessen and Borchelt 1996). It had already been found in a previous study that depressed patients have more somatic illnesses than mentally healthy individuals (Stevens et al. 1995), and this finding was confirmed in the BASE (Helmchen et al. 1999). Along with this somatic comorbidity, elderly depressed patients also have a higher mortality rate (Ernst and Angst 1995), which is attributable to the poorer somatic health of this group and, more specifically, to an elevated cardiovascular morbidity (see Chap. 14, Vol. 2, Part 2). In the BASE, demented patients had worse results than individuals without a psychiatric diagnosis on renal, pulmonary, and thyroid function tests, as well as on the complete blood count and parameters of functional impairment, but the mean number of chronic illnesses of demented patients was not significantly higher than that of control subjects (Helmchen et al. 1999).

The somatic comorbidity of the elderly is accompanied by the further problem of polypharmacy: 87% of BASE participants aged 70 and older took at least one medication daily, and 24% took five or more medically prescribed drugs (Steinhagen-Thiessen and Borchelt 1996). These included drugs with relatively frequent adverse effects, including depressive manifestations (methyldopa, beta blockers, corticosteroids, levodopa) or cognitive impairment (benzodiazepines, haloperidol, phenothiazine, barbiturates, meprobamate, propranolol, methyldopa, reserpine) (Borchelt and Geiselmann 1995). Participants diagnosed as having depression took significantly more medications daily than those without a psychiatric diagnosis; in these depressed patients, too, the risk of depressive manifestations or cognitive impairment as an adverse effect of medication was significantly higher than in the controls (Helmchen et al. 1999).

In view of the frequency of somatic comorbidity and polypharmacy among the elderly, it may be asked whether the individual items used to establish the presence of a depressive syndrome in scales such as the Hamilton Depression Scale, or to aid in the differential diagnosis of depressive illness according to the ICD-10 or DSM-IV criteria, are indeed valid diagnostic indicators of depression. Certain somatic and vegetative disturbances among these items – such as sleep disturbance, weight loss, loss of libido, somatic complaints, and loss of motivation – may well be the direct results of simultaneously present somatic illnesses or else adverse effects of the medications used to treat them.

The multidisciplinary approach of the BASE offered a good opportunity for closer study of the problem of somatic comorbidity in elderly depressed patients. For all patients who had both depressive manifestations and a somatic illness, an internist was asked to judge whether the positive items of the Hamilton Depression Scale (out of a total of 21 scale items) were "possibly," "probably," or "probably not" the result of a somatic illness or of adverse effects of medications (Linden et al. 1995).

As expected, the results differed for different items of the Hamilton Depression Scale. Nine of the 21 items on the scale were judged by the internist to be a "possible" or "probable" direct effect of somatic illness or medication in more than 50% of patients. This was true to a marked degree for the item "general physical symptoms," which was judged as a purely psychopathological phenomenon in a mere 7% of patients. The same was also true, albeit to a lesser extent, of typical features of depression such as agitation or loss of interest. Even the two cardinal manifestations of depression, "depressive mood" and "[impairment of] work and other activities," were relatively frequently thought to be possible direct results of somatic illness or medication. Items least frequently thought to be due to somatic causes included "guilt feelings," "suicidal ideation," "daily fluctuations," and "hypochondria." The mean overall score on the Hamilton Depression Scale dropped from 5.3 to 4.4 when all items that were "probably" the result of somatic illness or medication

were removed from the calculation, and to 2.4 when "possible" items were also removed.

The question remains unanswered, of course, whether the participants' disease manifestations were actually caused by somatic illnesses or by medications, rather than by depression itself. As for the medications, the possibility cannot be ruled out that some of them were prescribed to treat depressive symptoms that were not recognized as such, but were instead attributed to somatic illness, as is often the case when depression and somatic illness coexist (Lobo and Campos 1997). In such patients, the prescription of medication would not be the cause, but rather the effect of depression. Furthermore, this study reveals that the diagnosis of depression in old age is susceptible not only to the "false-negative" misattribution of subthreshold and otherwise unclassifiable depressive manifestations to other causes, but also, in many patients with somatic comorbidity, to the "false-positive" misattribution of symptoms of other origins to depression. Such errors, and the misguided therapeutic efforts that they produce, can only be avoided by basing the diagnosis of depression in patients with somatic comorbidity exclusively on those disease manifestations that have been found to be specific, reliable criteria for depression and which are, at most, very rarely caused by somatic illness or by medication use. These include guilt feelings, daily fluctuations, hopelessness, anhedonia, and lack of emotional responsiveness.

4 Differentiation of Mental Disorders from One Another

The earliest attempts at a systematic differential diagnosis of mental illness in the elderly were made in the 1950s (Madden et al. 1952; Roth 1955). Since then, considerable progress has been made, initially in the clinical sphere and, in the last decade, in the area of standardized neuropsychological testing.

The most important diagnostic distinction between mental illnesses in old age is that between depression and dementia, not only because they are by far the most common mental illnesses, but also because these two diagnoses carry substantially different prognostic and therapeutic implications. The differential diagnosis of these two entities is complicated mainly by the frequent simultaneous occurrence of depressive manifestations and cognitive disturbances in the elderly. Many manifestations, especially those affecting behavior, such as lack of motivation, social withdrawal, and loss of communication and motivation, are nonspecific (Helmchen and Linden 1993; Lishman 1987) or may be signs of depression occurring in the context of subcortical dementia (Folstein and McHugh 1978). Hierarchical stratification rules ("dementia before depression") and negative stereotypes ("advanced age equals dementia") confer a high risk of overlooking the diagnosis of a treatable dementia or of wrongly considering it untreatable.

If the frequency and intensity of depressive and cognitive symptoms are such as to fulfill the diagnostic criteria for both depression and dementia, then one speaks, in the current terminology, of comorbidity. This phenomenological designation, however, leaves open the question of whether two independent diseases are truly present. This can only be concluded with a degree of confidence when there is a long-standing prior history of depressive episodes without cognitive disturbance, or of the development of cognitive impairment independent of depression. When only the patient's present state is considered, there is generally no way to determine whether the problem is dementia with depressive manifestations, or depression with dementia-like manifestations. Reifler et al. (1982) suggested naming these two possibilities "type 1" and "type 2." Emery and Oxman (1992) introduced the concept of a dementia spectrum of depression, consisting of interlocking continua of depression, cognitive impairment, and degenerative dementia, on the basis of which they defined five prototypes:

1. Major depression without depressive dementia
2. Depressive dementia
3. Degenerative dementia without depression
4. Depression in the presence of degenerative dementia
5. Independent, simultaneous occurrence of dementia and depression

In this scheme, the second category, "depressive dementia," corresponds to the particularly interesting syndrome considered by other authors under the name "dementia syndrome of depression" (Folstein and McHugh 1978) or "depressive pseudodementia."

The concept of "depressive pseudodementia" (Bulbena and Berrios 1986; Kiloh 1961; Post 1962; Wells 1979) arose from the observation that cognitive disturbances sometimes improve after the remission of depressive manifestations. The term has, however, largely been abandoned in recent years (Zimmer and Lauter 1984), because its definition was too broad, too vague, and too variable to be of use, because the cognitive disturbances mostly do persist after remission of the depressive manifestations (see Chap. 10, Vol. 2, Part 2; Reifler et al. 1989), and because depressive disturbances commonly occur both prodromally and at the onset of dementing illnesses (Devanand et al. 1996; Helmchen et al. 1999; Henderson 1990; Kral

1983; Kral and Emery 1989; Liston 1977; Shraberg 1978). A quantitative meta-analysis of 16 studies of "reversible dementia" published between 1972 and 1994 revealed that the disease states so designated are predominantly the result of depression or of drug intoxications, and more rarely of other mental illnesses such as dissociative disorders, and that they are quite rare – when strict criteria are applied, "reversible dementia" accounts for less than 1% of the mental illness seen in outpatients (Weitingh et al. 1995). Depressive disturbances with significant cognitive impairment are, accordingly, now usually designated as dementia with depressive mood disturbance (DSM-IV: 290.13; ICD-10: F00.13).

There have been several reports of complete remission of cognitive disturbances after the disappearance of depression, but approximately half of such patients nonetheless turn out to have persistent cognitive disturbances after a follow-up of 2–3 years (Reischies et al. 1997), and as many as 90% are demented after a longer follow-up averaging 8 years (Rabins and Pearlson 1994). In these cases of "depression-induced cognitive impairment (DICI)" (Rabins et al. 1994), it remains unclear whether depression leads to functional or structural changes in the brain, or whether depression is itself the result of other, as yet unknown processes. A further possibility is a combination of these two mechanisms: perhaps depression lowers the level of cognitive functioning below a critical threshold in individuals with an already low cognitive reserve capacity (Lishman 1987).

Tests of cognitive performance can be used to distinguish between dementia and depression when they are not narrowly restricted to memory, but also include assessment of language, attention, motivation, speed, and psychomotor variables (Christensen et al. 1997). According to a recent meta-analysis (Christensen et al. 1997), the number of errors made by depressed patients on such tests is more than half a standard deviation higher than that of mentally normal control subjects, while patients with Alzheimer's disease make approximately twice as many. Tests such as the Inglis Paired Associate Learning Task, the Anomalous Sentences Repetition Test, the Boston Naming Test, and the Wechsler Adult Intelligence Scale (WAIS) Block Design Test can be used to distinguish depressed patients from demented patients, but not from normal subjects, while the Halstead Category Test and Halstead-Reitan Finger Tapping, Wechsler Memory Scale (WMS), Animal Naming, and Number-Symbol Tests can be used to distinguish depressed patients from normal subjects, but not from demented patients. The assessment of memory deficits with the aid of verbal or visual memory and recognition tasks does not, in general, distinguish depressed from demented patients, but some differentiation is nonetheless possible by means of tasks that are relatively undemanding (requiring little motivation or attention), relatively unpleasant, and temporally unconstrained. The differentiation of depression from dementia was more difficult in very old and very depressed patients and in those who had undergone electroconvulsive therapy.

In conclusion, after decades of effort, there is still no neuropsychological testing procedure that is practical for clinical use and yields a sharp differentiation between dementia and depression. The clinician must bear in mind that depression can both lead to cognitive disturbances and be masked by them, particularly when complaints of memory loss, intellectual slowing, and lack of spontaneity dominate the clinical picture. In addition, depression is commonly overlooked in elderly patients because it may be mild and of long standing, and somatic complaints may be most prominent. The clinical criteria shown in Table 1 for the recognition of depression in the presence of cognitive impairment therefore retain their usefulness (Lauter and Dame 1991).

In summary, this discussion of "depressive pseudodementia" has drawn attention to the fact that cognitive disturbances in old age need not be symptoms of a dementing illness, but may rather be an expression of any of several reversible mental disorders, and must therefore be differentiated in terms of their different etiologies and pathophysiologies. In the clinical field, the important fact to remember is that depressive manifestations in patients with cognitive disturbances may be treatable, whether they are primary or secondary (Reifler et al. 1989). In addition, all other treatable causes of reversible dementia must be excluded (see Chap. 2, Vol. 2, Part 2).

Diagnostic difficulties also occasionally arise from the different manifestations and course of certain mental illnesses in patients of advanced age (see Chap. 9, Vol. 2, Part 1), either because the "gestalt" of a long-present chronic illness may change as the patient ages, or because the first signs of a newly appearing disease in an elderly patient may be quite different from those more commonly seen in younger patients.

Thus 25%–50% of patients with Down's syndrome over the age of 50 develop dementia (Johannsen et al. 1996), while individuals with mental retardation of other causes develop dementia approximately four times as frequently as the normal population (Cooper 1997a). Retarded patients also have a larger number of additional somatic illnesses and a lesser capacity for adaptive behavior (Cooper 1997a; Moss and Patel 1997). Behavioral disorders, eating disorders, and the appearance or intensification of aggression, as well as a high rate (27.6%) of psychotic complications (delusions of being robbed or persecuted, visual hallucinations of strangers in the house), are quite common among mentally retarded individuals who have become

Table 1. Differentiation of the cognitive disturbances accompanying depression and dementia

	Depression	Dementia
Complaints Patient's description	Rather extensive and detailed; "I don't know"-type answers	Rather trivializing, circumstantial
Content	Difficulty thinking, impairment of memory	"Perhaps a bit more forgetful"
Findings behavior	Plaintive but also lacking initiative, diminished language output, social withdrawal, not unkempt, discrepancy between complaints of cognitive dysfunction and objective neuropsychological test findings	Loss of initiative, language disturbances (paraphasias etc.), unkempt, neglect of personal hygiene, marked cognitive disturbance but relatively mild disturbance of affective communication (maintained facade)
Psychological testing	Cognitive dysfunction barely noticeable in undemanding tasks without time pressure	Cognitive dysfunction even in undemanding tasks without time pressure
Onset	Depressive mood alteration *before* cognitive disturbances	Cognitive disturbances *before* depressive mood alteration
Course history	Earlier affective disturbances	No preceding affective illnesses, usually fairly long prior history of progressive cognitive decline
Follow-up	Improvement, at least partial remission in the affective area	Progression, though fluctuations may also occur
Nosology	Dissociative disorder, e.g. Ganser syndrome with depressive mood alteration, depressive delusion of being demented Depression-induced cognitive disturbance Dementia syndrome of depression	Dementia with depressive mood alteration, subcortical dementia with apathetic features

demented (Cooper 1997b; Moss and Patel 1997; Prasher and Chung 1996). Nevertheless, caretakers are usually unaware of the connection between these behavioral changes and dementia (Duggan et al. 1996).

Schizophrenics over the age of 65 also develop dementia much more frequently than the general population, though much more slowly than patients with Alzheimer's disease (Ciompi and Müller 1976; Harvey et al. 1995). When delusional paranoid manifestations first appear in old age, their chronicity and limited responsiveness to treatment, in association with only mild impairment of personality structure and everyday activities, are evidence against (late) schizophrenia, while the lack of psychopathological manifestations and of somatic findings of cerebral disease are evidence against a productive psychotic development of dementia. On the other hand, the impression of incoherence in cases of mania, with flight of ideas, that are of the dysphoric-irritative rather than the hyperactive type may occasionally lead to an erroneous diagnosis of late schizophrenia, or indeed of dementia.

Primary personality disorders are, as a rule, easy to distinguish from mental illness on the basis of the history alone. In contrast, chronically compensated personality disorders that become decompensated only in old age, or the age-dependent exacerbation of certain personality traits, cannot always be distinguished from an organically caused personality change merely by a single examination of the patient's present mental status (see Chap. 12, Vol. 2, Part 2).

Finally, there are clearly forms of mental suffering in old age that have yet to be recognized and described, to say nothing of their classification and nosologic definition (see Chap. 9, Vol. 2, Part 1).

5 Conclusion

The differential diagnosis of mental illness, even in old age and extreme old age, is a task that cannot afford to be neglected today, because different illnesses require different methods of treatment and chronic care and also have different prognoses. The process of accurate differential diagnosis and treatment is carried out in several discrete steps. The physician must first suspect that a mental illness may be present and then evaluate the patient's problem differentially in terms of the particular disease manifestations and their intensity, arrive at a nosologic or etiological-pathogenetic diagnosis, and finally design and carry out a plan of treatment appropriate to the situation of the individual elderly patient.

6 References

Alsen V (1960) Euphorie. Hüthig, Heidelberg

Arnold SE, Trojanowski JQ, Gur RE, Blackwell P, Han LY, Choi C (1998) Absence of neurodegeneration and neural injury in the cerebral cortex in a sample of elderly patients with schizophrenia. Arch Gen Psychiatry 55: 225–232

Baltes MM, Kühl KP, Sowarka D (1992) Testing for limits of cognitive reserve capacity: a promising strategy for early diagnosis of dementia? J Geront Psychol Sci 47: 165–167

Baltes PB (1997) Die unvollendete Architektur der menschlichen Ontogenese: Implikationen für die Zukunft des vierten Lebensalters. Psychol Rundschau 48: 191–210

Berg L, Coben LA, Smith DS, Morris JC, Miller JP, Rubin EH, Storandt M (1990) Mild senile dementia of the Alzheimer type. 3. Longitudinal and cross-sectional assessment. Ann Neurol 28: 648–652

Berkman LF, Berkman CS, Kasl S, Freeman DH, Leo L, Ostfeld AM, Cornony-Huntley J, Brody JA (1986) Depressive symptoms in relation to physical health and function in the elderly. Am J Epidemiol 124: 372–388

Borchelt M, Geiselmann B (1995) Are there specific health characteristics in depression versus dementia in old age? In: Bergener M, Brocklehurst JC, Finkel S (eds) Aging, health and healing. Springer, Berlin Heidelberg New York, pp 427–440

Borchelt M, Gilberg R, Horgass AL, Geiselmann B (1996) Zur Bedeutung von Krankheit und Behinderung im Alter. In: Mayer KU, Baltes PB (eds) Die Berliner Altersstudie. Akademieverlag, Berlin, pp 449–474

Brayne C, Gill C, Paykel ES, Huppert F, O'Connor DW (1995) Cognitive decline in an elderly population – a two wave study of change. Psychol Med 25: 673–683

*Bulbena A, Berrios GE (1986) Pseudodementia: facts and figures. Brit J Psychiatry 148: 87–94

Butler RN (1990) A disease called ageism. J Am Geriatr Soc 38: 178–190

*Christensen H, Griffiths K, Mackinnon A, Jacomb P (1997) A quantitative review of cognitive deficits in depression and Alzheimer-type dementia. J Int Neuropsychol Soc 3: 631–651

Ciompi L, Müller C (1976) Lebensweg und Alter der Schizophrenen. Springer, Berlin Heidelberg New York

Cooper SA (1997a) High prevalence of dementia among people with learning disabilities not attributable to Down's syndrome. Psychol Med 27: 609–616

Cooper SA (1997b) Psychiatric symptoms of dementia among elderly people with learning disabilities. Int J Geriatr Psychiatry 12: 662–666

Davidson M, Harvey PD, Powchik P, Parella M, White I, Knobler HY, Losonczy MF, Keefe RSE, Katz S, Frecska E (1995) Severity of symptoms in chronically institutionalized geriatric schizophrenic patients. Am J Psychiatry 152: 197–207

*Devanand DP, Sano M, Tang MX, Taylor S, Gurland, BJ, Wilder D, Stern Y, Mayeux R (1996) Depressed mood and the incidence of Alzheimer's disease in the elderly living in the community. Arch Gen Psychiatry 53: 175–182

Duggan L, Lewis M, Morgan J (1996) Behavioural changes in people with learning disability and dementia: a descriptive study. L Intellect Disabil Res 40: 311–321

*Emery VO, Oxman TE (1992) Update on the dementia spectrum of depression. Am J Psychiatry 149: 305–317

Ernst C, Angst J (1995) Depression in old age. Is there a real decrease in prevalence? Eur Arch Psychiatr Clin Neurosci 245: 272–287

Folstein M, McHugh P (1978) Dementia syndrome of depression. In: Katzman R, Terry RD, Bick KL (eds) Alzheimer's disease: senile dementia and related disorders; aging. Raven, New York

Fratiglioni L, Grut M, Forsell Y, Grafström M, Holmén K, Eriksson K, Viitanen M, Bäckman L, Ahlbom A, Winblad B (1991) Prevalence of Alzheimer's disease and other dementias in an elderly urban population: relationship with age, sex and education. Neurology 41: 1886–1892

Geiselmann B, Bauer M (2000) Subthreshold depression in the elderly: qualitative or quantiative distinction? Compr Psychiatry 41 [Suppl 1]: 32–38

Häfner H (1997) Was tun mit Krankheiten, die keine sind? Subdiagnostische Störungen und unversorgte Morbidität. Münch Med Wochenschr 139: 158–160

Harvey PD, White L, Parella M, Putnam KM, Kincaid MM, Powchik P, Mohs RC, Davidson M (1995) The longitudinal stability of cognitive impairment in schizophrenia. Mini-mental state scores at one- and two-year follow-ups in geriatric in-patients. Br J Psychiatry 166: 630–633

Helmchen H, Linden M (1993) The differentiation between depression and dementia in the very old. Ageing Soc 13: 589–617

Helmchen H, Linden M (2000) Subthreshold disorders in psychiatry: clinical reality, methodological artifact, and the double-threshold problem. Compr Psychiatry 41[Supp 1]: 1–7

Helmchen H, Reischies FM (1998) Normales und pathologisches kognitives Altern. Nervenarzt 69: 369–378

*Helmchen H, Baltes MM, Geiselmann B, Kanowski S, Linden M, Reischies FM, Wagner M, Wernicke T, Wilms HU (1999) Psychiatric illnesses in old age. In: Baltes PB, Mayer KU (eds) The Berlin Aging Study. Aging from 70 to 100. Cambridge University Press, New York, pp 167–196

*Henderson SA (1990) Co-occurrence of affective and cognitive symptoms: the epidemiological evidence. Dementia 1: 119–123

Illhardt FJ (1993) "Ageism": Vorurteile gegen das Alter. Z Gerontol 26: 335–338

Israel L, Kozarevic D, Sartorius N (1984) Source book of geriatric assessment, vols I and II. Karger, Basel

Jacob H (1986) Krankheitserleben und Altersstörungen. Akt Neurol 13: 28–34

Johannsen P, Christensen JE, Mai J (1996) The prevalence of dementia in Down syndrome. Dementia 7: 221–225

Kage A, Nitschke I, Fimmel S, Köttgen E (1996) Referenzwerte im Alter: Beeinflussung durch Alter, Medikation und Morbidität. In: Mayer KU, Baltes PB (eds) Die Berliner Altersstudie. Akademieverlag, Berlin, pp 405–428

Kay DWK, Beamish P, Roth M (1964) Old age mental disorders in Newcastle-upon-Tyne. II. A study of possible social and medical causes. Br J Psychiatry 110: 668–682

Kiloh LG (1961) Pseudodementia. Acta Psychiatr Scand 37: 336–351

Knäuper B, Wittchen HU (1994) Diagnosing major depression in the elderly: evidence for response bias in standardized diagnostic interviews? J Psychiatr Res 28: 147–164

Kral VA (1983) The relationship between senile dementia (Alzheimer type) and depression. Can J Psychiatry 28: 304–306

*Kral VA, Emery OB (1989) Long-term follow-up of depressive pseudodementia of the aged. Can J Psychiatry 34: 445–446

Lauter H, Dame S (1991) Depressive disorders and dementia: the clinical view. Acta Psychiatr Scand Suppl 366: 40–46

Lehr U, Thomae H (eds) (1987) Formen seelischen Alterns: Ergebnisse der Bonner Gerontologischen Längsschnittstudie (BOLSA). Enke, Stuttgart

Linden M, Barnow S (1998) The wish to die in very old persons near the end of life: a psychiatric problem? Results from the Berlin Aging Study. Int Psychogeriatr 9: 291–307

Linden M, Borchelt M, Barnow S, Geiselmann B (1995) The impact of somatic morbidity on the Hamilton Depression Rating Scale in the very old. Acta Psychiatr Scand 92: 150–154

Linden M, Gilberg R, Horgas AL, Steinhagen-Thiessen (1996a) Die Inanspruchnahme medizinischer und pflegerischer Hilfe im hohen Alter. In: Mayer KU, Baltes PB (eds) Die Berliner Altersstudie. Akademieverlag, Berlin, pp 475–496

Linden M, Maier W, Achberger M, Herr R, Helmchen H, Benkert O (1996b) Psychische Erkrankungen und ihre Behandlung in Allgemeinarztpraxen in Deutschland. Ergebnisse aus einer Studie der Weltgesundheitsorganisation (WHO). Nervenarzt 67: 205–215

*Linden M, Kurtz G, Baltes MM, Geiselmann B, Lang F, Reischies FM, Helmchen H (1998) Depression bei Hochbetagten. Ergebnisse der Berliner Altersstudie. Nervenarzt 69: 27–37

Lindenberger U, Mayr U, Kliegl R (1993) Speed and intelligence in old age. Psychol Aging 8: 207–220

Lishman WA (1987) Organic psychiatry. The psychological consequences of cerebral disorders, 2nd edn. Blackwell, Oxford

Liston EH (1977) Occult presenile dementia. J Nerv Ment Dis 164: 263–267

Lobo A, Campos R (1997) Managing the psychiatry/primary care interface. In: Robertson MM, Katona CLE (eds) Depression and physical illness. Wiley, Chichester, pp 39–66

Madden JJ, Lohan JA, Kaplan LA, Manfredi HM (1952) Nondementing psychoses in older persons. JAMA 150: 1567–1570

Maier W, Herr R, Lichtermann D, Gansicke M, Benkert O, Faust G (1994) Brief depression among patients in general practice. Prevalence and variation by recurrence and severity. Eur Arch Psychiatry Clin Neurosci 244: 190–195

Moritz DJ, Petitti DB (1993) Association of education with reported age of onset and severity of Alzheimer's disease at presentation: implications of the use of clinical samples. Am J Epidemiol 137: 456–462

Mortimer JA (1990) Epidemiology of dementia: cross-cultural comparisons. Adv Neurol 51: 27–33

Moss S, Patel P (1997) Dementia in older people with intellectual disability: symptoms of physical and mental illness, and levels of adaptive behaviour. J Intellect Disabil Res 41: 60–69

O'Connor DW, Pollit PA, Treasure FP (1991) The influence of education and social class on the diagnosis of dementia in a community population. Psychol Med 21: 219–224

Palmore EB (1988) The facts on aging quiz: a handbook of uses and results. Springer, Berlin Heidelberg New York

Parsons PL (1965) Mental health of Swansea's old folk. Br J Prev Soc Med 19: 43–47

Post F (1962) The significance of affective symptoms in old age. Oxford University Press, London

Prasher VP, Chung MC (1996) Causes of age-related decline in adaptive behavior of adults with Down syndrome: differential diagnosis of dementia. Am J Ment Retard 101: 175–183

Purohit DP, Perl DP, Haroutunian V, Powchik P, Davidson M, Davis KL (1998) Alzheimer disease and related neurodegenerative diseases in elderly patients with schizophrenia: a postmortem neuropathologic study of 100 cases. Arch Gen Psychiatry 55: 205–211

*Rabins PV, Pearlson GD (1994) Depression induced cognitive impairment. In: Burns A, Levy R (eds) Dementia. Chapman and Hall, London, pp 667–679

Rajkowska G, Selemon LD, Goldman-Rakic PS (1998) Neuronal and glial somal size in the prefrontal cortex: a postmortem morphometric study of schizophrenia and Huntington disease. Arch Gen Psychiatry 55: 215–224

Reifler BV, Larson E, Hanley R (1982) Coexistence of cognitive impairment and depression in geriatric outpatients. Am J Psychiatry 139: 623–626

Reifler BV, Ter L, Raskind M, Veith R, Barnes R, White E, McLean P (1989) Double-blind trial of imipramine in Alzheimer's disease patients with and without depression. Am J Psychiatry 146: 45–49

Reischies FM (1998) Age related cognitive decline and the dementia threshold. In: Lomranz J (ed) Handbook of aging and mental health. Plenum, New York, pp 435–448 (in press)

Reischies FM, Geiselmann B (1997) Age-related cognitive decline and vision impairment affecting the detection of dementia syndrome in old age. Br J Psychiat 171: 449–451

Reischies FM, Lindenberger U (1996) Grenzen und Potentiale kognitiver Leistungen im hohen Alter. In: Mayer KU, Baltes PB (eds) Die Berliner Altersstudie. Akademieverlag, Berlin, pp 351–377

Reischies FM, Schaub T (1997) Epidemiologische Verlaufsuntersuchungen der Demenz. In: Rösler M, Retz W, Thome J (eds) Proceedings des Aloys Alzheimer Symposiums Würzburg. Deutscher Studienverlag, Weinheim, pp 58–66

Reischies FM, Geiselmann B, Geßner R, Kanowski S, Wagner M, Wernicke TM, Helmchen H (1997) Demenz bei Hochbetagten. Ergebnisse der Berliner Altersstudie. Nervenarzt 68: 719–729

*Roth M (1955) The natural history of mental disorder in old age. J Ment Sci 101: 281–301

Salthouse TA (1982) Adult cognition, an experimental psychology of human aging. Springer, Berlin Heidelberg New York

Salthouse TA (1985) A theory of cognitive aging. North Holland, Amsterdam

Satzger W, Dragon E, Engel RR (1996) Zur Normenäquivalenz von HAWIE-R und HAWIE. Diagnostika 43: 119–138

Schmand B, Lindeboom J, Launer L, Dinkgreve M, Hooijer C, Jonker C (1995) What is a significant score change on the mini-mental-state examination? Int J Geriatr Psychiatry 10: 411–414

Shraberg D (1978) The myth of pseudodementia: depression and the aging brain. Am J Psychiatry 135: 601–603

Staudinger U, Freund AM, Linden M, Maas I (1996) Selbst, Persönlichkeit und Lebensgestaltung im Alter: Psychologische Widerstandsfähigkeit und Vulnerabilität. In: Mayer KU, Baltes PB (eds) Die Berliner Altersstudie. Akademieverlag, Berlin, pp 321–350

Steinhagen-Thiessen E, Borchelt M (1996) Morbidität, Medikation und Funktionalität im Alter. In: Mayer KU, Baltes PB (eds) Die Berliner Altersstudie. Akademieverlag, Berlin, pp 151–183

Stevens DE, Merikangas KR, Merikangas JR (1995) Comorbidity of depression and other medical conditions. In: Beckham EE, Leber WR (eds) Handbook of depression. Guilford, New York, pp 147–199

Thomae H, Maddox GL (eds) (1982) New perspectives on old age: a message to decision makers. Springer, Berlin Heidelberg New York

Weitingh MD, Bossuyt PMM, van Crewel H (1995) Reversible dementia: more than 10% or less than 1%? A quantitative review. J Neurol 242: 466–471

*Wells CE (1979) Pseudodementia. Am J Psychiatry 136: 895–900

Zimmer R, Lauter H (1984) Zum Problem der depressiven Pseudodemenz. Z Gerontol 17: 109–112

CHAPTER 9

R. Ferszt, S. Kanowski

Aging of People with Mental Illness

Translator: E. Taub

1 Introduction

The term "aging of people with mental illness" might be taken to refer to either the influence of time on the course of psychiatric diseases ("natural history") or, alternatively, the importance of involutional changes after an arbitrarily chosen age, e.g. age 60. Moreover, in the present era of increasing life span, it may be of interest whether diseases first appearing in the presenium have a specific course in old age, because patients developing mental illness at advanced ages today may have a subsequent course of illness lasting as long as 30 years. In this review, we aim to describe the influence of involution on preexisting diseases, but we shall also attempt to present the available data on the latter question.

2 Fundamental Problems of Longitudinal Studies

Despite the large number of publications on dementing processes that have appeared in recent years, there have been relatively few studies on the closely related problem of aging in the mentally ill. There are probably many external reasons for this discrepancy, but we feel that one important point is the state of current psychiatric nosology, which is largely descriptive. Because many psychiatric diagnoses are defined and validated on the basis of clinical symptoms in younger patient groups, the assessment of changes related to aging, i.e. the consideration of these diseases as *processes*, may be difficult. Some patients, for example, despite persistence of their subjective impairments, may migrate from one diagnostic category to another as they grow older or may drop out of the nosological system entirely, not because they have recovered, but because the combination of signs and symptoms they show in old age cannot be adequately classified.

If there is no morphological or pathophysiological correlate whose course can be followed, longitudinal studies are performed only on those comparatively few patients who are available for repeated clinical examination despite their physical multimorbidity.

The downsizing of psychiatric hospitals and the closing of wards for the aged and chronically ill lessen the availability of these groups of patients in systematic follow-up studies, particularly if they are being nursed at home.

Field studies may suffer from cohort problems if the study is to be performed over a time span of 50 years or more. Only a small number of patients survive to be studied in high old age, and this group cannot be considered representative of the entire original patient group. Furthermore, age-dependent changes of memory and informational behavior are to be expected. There will not only be memory impairments of organic origin; psychopathological episodes will be recounted differently from the viewpoint of old age than at the time they are experienced, and different age cohorts, because of their differing education, will give information about certain psychopathological disturbances in diverse ways. Thus a tangle of substantive and methodological problems complicates the study of the course of psychiatric diseases in old age. (For more on this topic, see also Chap. 8, Vol. 2, Part 1.)

3 Schizophrenic Psychoses

The relatively good state of our knowledge of this topic is mainly due to three studies (Bleuler et al. 1976; Ciompi 1980a,b, 1984, 1985; Harding et al. 1988), which document a wide spectrum of disease courses and contradict the traditional notion that schizophrenic illnesses have a uniformly poor prognosis.

Ciompi distinguished eight types of disease course, four of which tended to have an unfavorable prognosis and accounted for 48.6% of his patients; the corresponding figures for poor outcome were somewhat lower in the Burghölzli study (Bleuler et al. 1976) (40%) and the Vermont study (Harding et al. 1988) (38%). Mortality was elevated in schizophrenic patients of all types (1.5- to 1.7-fold that of the general population); catatonic types had a significantly higher mortality rate than paranoid types (2.5- and 1.4-fold that of the general population, respectively); and female patients and patients who became ill after the age of 40 also had a somewhat higher mortality rate. As might have been expected from clinical experience, the frequency of suicides among schizophrenics was higher than that in the general population, as was the frequency of death of unknown cause. More than half of the originally documented clinical symptoms improved in patients surviving until advanced age; stupor, anxiety, and agitation were the symptoms most likely to regress with age, and abulia, disturbances of affect, phonemes, mannerisms, and delusions were the least likely to do so (Ciompi 1980a). With advancing age, nonproductive residual states overshadowed the originally most prominent symptoms which had originally led to the classification as a particular type of schizophrenia; these were attenuat-

ed, and the clinical syndromes seemed to level off until they were unrecognizable. In fact, about 20% of the patients had recovered to such an extent by the time they reached old age that specific schizophrenic symptoms could no longer be demonstrated; a further 43% had clearly improved, 30% were unchanged, and only 6% had worse schizophrenic symptoms in old age than previously.

Davidson et al. (1995) retrospectively studied 393 chronic schizophrenics with histories extending back more than 60 years in some cases. The unfavorable courses in this group appeared to be characterized by an increase in negative symptoms and a rather modest decrease of positive symptoms, which clearly failed to "burn out," as might have been intuitively expected.

The prevalence of dementia syndromes in the general population over 65 years of age is exceeded by that among schizophrenics: approximately 25% of patients in this age-group in the Lausanne follow-up study suffered from moderate to severe organic cerebral psychosyndromes (Ciompi et al. 1980a,b, 1984, 1985); 17% had significant thought and memory disorders, including partial disorientation, and a further 8% had severe amnestic psychosyndromes with almost total spatial and temporal disorientation (Ciompi 1980b). Dementia syndromes were even more frequent in types of schizophrenia with a primarily unfavorable course. In their chronically hospitalized group, Davidson et al. (1995) found a steady, remarkably slow decline of cognitive function, quantitatively estimated as 1.2–4.6 points on the Mini-Mental Status Examination (MMSE) scale per decade of life. The average MMSE score of younger patients was 27.6, while the corresponding figure for patients over 65 was 16 points and for patients over 85, 9.6 points. If we consider that normal individuals aged 80 or older generally have MMSE scores of at least 24, then the degree of impairment in these elderly hospitalized patients was sufficient to support a diagnosis of dementia. The very slow decline of cognitive performance would be quite uncharacteristic for the course of a neurodegenerative disease; the annual rate of decline mentioned above is approximately one tenth of that expected in Alzheimer's disease, for example.

Uncertainty remains, therefore, about the pathogenesis of this not at all uncommon disturbance. Purohit and coworkers performed an autopsy study of 100 chronic schizophrenics with an average age of 76.5 years (range, 52–101 years). Over the course of their illness, 72 patients had developed dementia of at least moderate severity (score 2 on the Clinical Dementia Rating Scale; Hughes et al. 1982). Nonetheless, in the great majority of these patients, no morphological correlate of dementia could be found. An Alzheimer-type encephalopathy was diagnosed in 9% of the patients, and other relevant neuropathological abnormalities were found in 4%, but in all other patients (87%), histological investigation yielded no clues as to the pathogenesis of dementia (Purohit et al. 1998). Other authors have also found that neurodegenerative lesions are no more common in elderly schizophrenics than in the elderly in general (Arnold et al. 1994). This is in line with Ciompi's observation (Ciompi 1980b) that cognitively impaired schizophrenic patients generally do not give the impression of having an "organic" type of cognitive impairment; the cognitive deficit of elderly schizophrenics does not fit into any known category of organic brain disease. There has been speculation about many possible pathogenetic factors, including chronically insufficient cognitive challenge, chronic neuroleptic use, and the effects of electroconvulsive therapy or psychosurgical procedures, but this problem has not yet been satisfactorily solved. Now that morphological and pathophysiological abnormalities in schizophrenia are being described increasingly often, the hypothesis of a developmentally dependent vulnerability of schizophrenics to some kind of "functional" brain disturbance, remaining compensated until old age, has also been advanced (Arnold et al. 1994; Akbarian et al. 1996). Davidson et al. (1995) found that symptoms of dementia appeared mainly in patients who had already had negative symptoms earlier in life and proposed that negative symptoms and dementia might have a common organic origin. On the other hand, it must also be asked whether the clinical methods and psychological tests generally used in the assessment of dementia are at all adequate to delineate the disturbances found in aging schizophrenics, and it is tempting to raise the concept of "schizophrenic pseudodementia" for discussion once again.

Special considerations apply to patients with schizophrenia of late onset, i.e. onset after the age of 45. These account for approximately one fifth of all schizophrenics; indeed, 3%–5% become ill only after the sixth decade of life (Jeste et al. 1995). Late-onset schizophrenia preferentially affects women, not infrequently those with a relatively high premorbid performance level (Yassa and Suranyi-Cadotte 1993). Systematizing paranoid syndromes, combined with bizarre delusional ideas and commentating phonemes, are typical in late-onset patients, as are a less common occurrence of loose association, inadequate affect, and negative symptoms; formal thought disorders are very rarely described (Almeida et al. 1995). The prognosis is hence more favorable, at least by societal criteria. These late-onset patients are relevant here in that their further course in old age, especially with respect to cognitive impairment, seems to be benign.

Nonetheless, Jeste et al. (1995) were unable to support the concept of late-onset schizophrenia as an independent nosological entity on the basis of family history, descriptions of premorbid development, and neurophysiological studies. Almeida et al. (1995), on the other hand, are proponents of the independence of this syndrome from the core group of schizophrenias.

4 Non-schizophrenic Paranoid Psychoses

The percentage of individuals with chronic non-schizophrenic paranoid psychoses who undergo psychiatric treatment is unknown. Only a small fraction of those treated remain under medical care in the long term. Thus assessment of the age-dependent course of these syndromes is fraught with uncertainty. Opjordsmoen (1989) described the course, over 30 years, of 301 patients who had been predominantly delusional at the time they entered the study, at an average age of approximately 35 years; a paranoid disorder of non-schizophrenic type had been diagnosed in 53 of these patients (average age, 40.6 years). Twenty-seven years later, 80% of this subgroup had married and had an average of two children, and 52% were employed. Among the 33 patients still available to follow-up after 30 years, 55% were not undergoing psychiatric treatment, and 27% were being treated by their family physicians. Many of them had adapted their lifestyles to the essentially unchanged paranoid symptoms. It is instructive to compare these figures with those for schizophrenics on 30-year follow-up: only 43% of the latter ever married, they remained childless significantly more frequently (average, 0.7 children), and only 20% were still employed. Out of 75 schizophrenics in the study, 45 were available to follow-up at 30 years; of these, 45% were under inpatient psychiatric treatment, and only 24% were not under psychiatric treatment. There were no cases of suicide among the 19 non-schizophrenics who had died, but among the 18 schizophrenics who had died, there had been three definite and two possible suicides. According to the Norwegian authors' findings, strongly reactive momentum of a non-schizophrenic delusional disorder at the time of its onset predicts a favorable outcome, while the continuous presence of symptoms for more than 6 months at the onset of the disease predicts an unfavorable long-term outcome (Opjordsmoen and Retterstol 1991). The prognosis of delusional, non-schizophrenic disorders is considerably more favorable than that of schizophrenia, although it seems to have remained essentially the same despite the introduction of neuroleptics (Opjoerdsmoen and Retterstol 1993).

5 Affective Disorders

The contradictory nature of the literature has already been pointed out by Müller (1989). Some of the problems lie in the unsatisfactory nosological classification of depressive syndromes. While the group of clearly phasic, bipolar illnesses possess a certain degree of homogeneity, highly disparate varieties of disease may lie concealed behind the diagnoses of unipolar or chronic depression (Goodwin and Jamison 1990). Moreover, we share Murphy's opinion (Murphy 1989) that there is still no satisfactory classification scheme for depressive disorders in all age categories.

In the *phasic* affective disorders, according to several authors, the frequency and amplitude of phases increases with age (Angst 1966, 1987), while the phases themselves become longer (Foster and Reisberg 1984). The latter effect, however, seems to be reversed after the age of 60, when, according to Matussek and colleagues (1965), depressive phases become shorter again. Müller found a reduction of phase frequency after age 65 in almost two thirds of his patients, no change in one fifth, and an increase in only one fifth (Müller 1989). Amenson and Levinsohn studied almost 1000 depressive patients with disease courses extending back more than five decades. They found an increasing prevalence of depressive syndromes with advancing age and attributed this largely to the fact that women with a prior history of an affective disorder are especially prone to a relapse in old age. This sex difference seemed to even out only after age 70 (Amenson and Lewinsohn 1981; Gurland 1976).

Partial remissions are more common in old age, largely because of the increased representation of depressions that have become chronic in this age-group, partly reinforced by the high prevalence of physical illness in the elderly (Angst 1966; Roth and Kay 1956). The common association of depressive symptoms with dementing illnesses in old age is probably not due to the group of aging persons with phasic illnesses; in any case, the literature provides no convincing evidence that this group is particularly subject to the development of neurodegenerative disease (Migliorelli et al. 1995; Van Ojen et al. 1995). A notable exception is that of the 22 patients with depressive "pseudodementia," aged between 62 and 78 years, described by Kral (1982), who were followed for an average of 8 years after the end of the acute

phase. Six patients from the entire group went on to have further, reversible depressive phases associated with severe cognitive disturbances; four of these six patients, however, and a further 16 from the entire group, were eventually given a diagnosis of Alzheimer's disease. The prior history of these patients was not reported in detail, but it would be highly relevant to the interpretation of the results. All of them apparently fulfilled the criteria for "endogenous" depression, and monopolar phases were present, at least in the subgroup of patients who suffered relapses. Possible explanations include a transient unmasking of a gradually developing encephalopathy, i.e. a comorbidity of "endogenous" depression and Alzheimer's disease, or a mainly organic cause for the phasic illnesses, possibly aggravated by anticholinergic drugs. Unfortunately, it remains unclear to what extent these patients had already suffered from affective illnesses in their youth (Kral 1982).

As other authors have stressed, within the spectrum of depressive symptoms, it is precisely the cognitive problems that tend to last the longest in a given phase, outlasting others by many months (Marcos et al. 1994). For now, it remains unclear whether these cognitive deficits, occurring in the context of depression in a minority of aging, phasically ill patients, can ultimately become irreversible even if no neurodegenerative illness supervenes, as the findings of a few authors seem to suggest (Marcos et al. 1994; Abas et al. 1990).

A change in the pattern of manic symptoms with advancing age has been described by several authors. States of vivacious, elevated mood become rarer and are replaced by irritable, aggressive, dysphoric states of increased drive. Expansive delusions and flight of ideas occur more frequently (albeit with progressively lesser intensity and velocity), as does incoherence, particularly when cognitive disturbances supervene (Post 1984; Baldwin and Tomenson 1995).

The relatively low prevalence of affective illnesses probably does not reflect the *true extent of depressive dysfunction* in the elderly (Gurland 1976; Kay et al. 1985). In a field study, Girling et al. (1995) studied 1773 individuals aged 77 and older. Of these, 6% stated that they felt depressed most of the time, 22% complained of irritability, and 35% complained of disturbing restlessness and inner tension; 21% at least sometimes felt that they did not want to continue living any longer. The origins of these low-grade chronic depressive states are varied and many have not been adequately studied. Are they long-standing disorders that remain compensated and asymptomatic in the patients' younger years, finally surfacing in old age, or do they actually first arise in the senium, constituting what the conventional classification schemes have labeled "subdiagnostic morbidity"?

6 Dysthymias

Opinions differ as to the development of low-grade depressive syndromes in old age – the dysthymias, or so-called depressive neuroses. Ernst and Ernst (1968), in their extensive review of chronic neurotic processes, state that "old age more frequently has an alleviating effect on neurotic anxiety than an aggravating effect . . . and . . . often places a relieving distance between the patients and their symptoms" (p. 41). This was later confirmed by a field study, in which the prevalence of dysthymic disorders declined from 3% in the general adult population to approximately 1% in those over 65 (Weissman et al. 1988). On the other hand, Stefansson and colleagues (1991) found a much higher prevalence (difference, 6.4%) in individuals over 50, particularly in women. In a group of 224 outpatients of a special clinic for depression, Devanand and colleagues (1994) made a diagnosis of major depression in 51% of patients over 60 and of dysthymia in just 18% of the same group. The reportedly high prevalence of dysthymic disorders and the increased frequency of the diagnosis of dysthymia in elderly patients are probably accounted for not by previously dysthymic patients who grow old, but by previously healthy individuals developing dysthymia in old age: elderly dysthymic patients had generally become dysthymic relatively late (in the sixth decade), while the onset of dysthymia before age 21, with continuation into old age, was rare. Younger dysthymic patients had frequent comorbidities with axis I and axis II disorders (major depression, anxiety, personality disorders), but such comorbidities only occurred in exceptional cases in the elderly patients. The typical finding was of a chronic depressive condition that fulfilled the DSM-III-R criteria for dysthymic disorder, but differed from the form seen in younger patients with respect to possibly important nosological features: the prevalence of the classical form of dysthymia, which begins at an early age, declines over time, but a new and "purer" form of dysthymia, i.e. one with a lesser degree of psychopathological comorbidity, apparently comes to the fore after approximately age 55, and its prevalence does not decline thereafter (Devanand et al 1994).

7 Anxiety Disorders

A description of the broad spectrum of anxiety disorders – generalized anxiety disorder, phobias, panic disorders, obsessive-compulsive disorders,

post-traumatic stress disorder (PTSD) – and of their course as patients grow older – even excluding the dysthymias – would go beyond the scope of this chapter, the more so as the problems of classification discussed in the introductory sections are particularly acute here, and there are still no relevant long-term follow-up studies utilizing the DSM-III criteria. Müller, in his review, discussed the evidence for a basically favorable course of these disorders as the patients grow older (Müller 1989); as Ernst and Ernst (1968) described, aging had a positive influence on the symptoms and signs of aged neurotics ($n = 57$); in particular, phobias and "obsessive–compulsive phenomena" (p. 407) largely resolved. Overall, half of Müller's patients were regarded as cured, and the great majority as improved. The small group of neurotics with unchanged symptoms included those with hypochondriac and, to a lesser extent, those with depressive symptoms and signs. Foster and Reisberg (1984) obtained similar findings with respect to the anxiety disorders.

Long-standing, chronic conditions of these types seem to wane in intensity and to become stabilized, while patients in whom such conditions are of late onset tend to experience a worsening of symptoms over time. In a group of men over 65, anxiety disorders were the second most common psychiatric diagnosis after dementia; in a group of women over 65, they were the most common diagnosis (Myers et al. 1984). Other field studies yielded similar findings (Regier 1988; Lindesay et al. 1991). While more specific phobias, which typically originate in the first third of life, tend to regress, the prevalence of anxiety disorders increases with age because of the occurrence of anxiety disorders of late onset. Among the latter, the most prominent types are fear of public transport and of travel. These agoraphobic syndromes first arise in old age after traumatic experiences (syncopes, accidents, assaults), tend to worsen, and are associated with significant dysfunction. Approximately one third of patients with anxiety disorders of late onset are found to have accompanying depressive symptoms; panic attacks are rare, while accompanying somatic illnesses are common. As in the dysthymias, there seem to be two types of disease course in the anxiety disorders: in the type often described by earlier authors, "neurotic" anxieties and obsessive–compulsive symptoms arising in youth tend to abate over time (Ernst and Ernst 1968), while, in the other type, symptoms arising after the age of 60 have a less favorable course. Much of what the elderly endure in terms of physical morbidity and social unsettlement can produce anxiety, but it remains unclear why one group of patients with long-standing anxiety disorders should have a stable course in the face of these challenges, while another, previously healthy group becomes symptomatic with age. A basic diagnostic difficulty in this age-group lies in the interaction between understandable anxieties about cardiovascular disease, for example, and anxiety disorders that secondarily assume an independent existence (Lindesay 1991).

Because the category of PTSD came into widespread use only after the publication of the DSM-III, relatively few longitudinal studies are available with follow-up into old age. There is, however, an extensive literature on war veterans and victims of Nazi persecution. These studies unequivocally show that, unlike other psychogenic disorders, PTSD tends to become chronic and, not uncommonly, to intensify over periods of up to 50 years and does not share the pattern of regression in old age that is typical of many psychogenic disorders. In 1997, Engdahl and colleagues (1997) studied 262 elderly veterans of the Second World War and the Korean War who had been exposed to severe trauma while being held prisoner by the enemy. A total of 53% had subsequently had PTSD at some time in their live, and 29% still had PTSD at the time of study, approximately 50 years after the traumatizing events took place. The incidence of PTSD was even higher in individuals who had endured extreme trauma in Japanese prisoner-of-war camps: among these, 84% had had PTSD at some time, and 59% had had the syndrome continuously for over 50 years and still had it at the time of the study. The psychiatric comorbidity was predictably high: 45% had other axis I disorders in old age, and only 34% had no history of axis I disorders. The frequency of chronically persistent PTSD and mortality were similarly correlated with the intensity of wartime trauma in a 50-year follow-up study of 107 veterans (Lee et al. 1995). In a cohort of 124 survivors of the Holocaust who had been persecuted on supposedly "racial" grounds and who had no history of bipolar illnesses, obsessive–compulsive disorders, or organic syndromes, examination 45 years after the primary traumatization revealed PTSD in 46%, of which the most common symptoms were sleep disturbances, nightmares, and uncontrollable memories. The prevalence of PTSD in old age was three times higher in those who had been held in concentration camps than in those who had not (Kuch and Cox 1992). A retrospective study of the course of post-traumatic symptoms in old age, including 73 individuals older than 55 (of whom 37 were older than 66), showed that the affected individuals were relatively well adapted socially, but their psychopathological problems, which were mainly attributable to PTSD, changed very little over time; furthermore, 67 subjects reported depressive mood disturbances, 29 reported suicidal thoughts, and 50 reported memory disturbances. PTSD typically appeared in individuals who had been persecuted in adolescence, and nearly all subjects showed remarkable constancy of symptoms as

they grew older, with a gradual increase in their intensity in old age (Freudenberg 1991). PTSD patients, compared to the general population, have an elevated psychiatric and somatic comorbidity lasting for many years (Boscarino 1997; Alarcon et al. 1997) and incline toward unhealthy habits such as smoking, medication, and alcohol abuse (Beckham et al. 1997). They also tend to have endocrinologic abnormalities, particularly with respect to cortisol regulation, even many years after trauma, and, at least according to the report by Bremner and colleagues, have a higher frequency of structural abnormalities in the hippocampus (Bremner et al. 1995).

8 Alcoholism and Other Addictions

While the problem of alcohol and substance abuse in old age is increasingly being recognized (Ruppert 1996; Ades and Lejoyeux 1994; Bristow and Clare 1992), most longitudinal studies provide a follow-up of no more than 10 years' duration (Finney and Moos 1992; De Soto et al. 1989). As Müller (1989) pointed out, hardly any studies have been performed with follow-up into old age, so that we are obliged to extrapolate from the trends in the course of these diseases as they affect younger individuals. Müller attributed the lack of suitable studies to the greatly elevated mortality rate in alcoholics, due to frequent somatic comorbidity (Feuerlein et al. 1994) and a suicide rate ten times that of the general population (Rossow and Amundsen 1995). Among those who lived to age 65 in the Lausanne follow-up study, Müller noted a favorable course in old age, with relatively little somatic and psychoorganic damage, even if total abstinence had not been achieved. The same selection process was observed in substance-dependent individuals, whose mortality rate may be as high as 29 times that of the general population: no further deterioration was observed in abusers of narcotics or sleeping pills who survived until old age (Müller 1989). A study in which 286 college students and 456 inhabitants of large cities were regularly followed until age 60 revealed that many surviving individuals were still abusing alcohol to a clinically relevant extent even in the seventh decade of life. A total of 205 had become alcoholics by age 47, and, by age 60, 18% of the students and 28% of the large-city inhabitants in this subgroup had died; among those who survived to age 60, 70% and 39%, respectively, were still regularly drinking alcohol, and 59% and 28%, respectively, were alcohol abusers. The course of the latter two subgroups remained constant for decades (Vaillant 1996). In another study, 9% of 650 patients older than 65 were alcohol abusers, and most of these had been so since their youth; only 10% reported having decreased their alcohol consumption in old age (Bristow and Clare 1992). The high prevalence of substance abuse, including sleeping-pill abuse and alcohol dependency, is becoming even higher according to recent studies; in several reports, the figures lie above 10% (see above for references). This fact is likely due to the contribution of surviving long-term sufferers from these conditions to the total. It was found in a French study, however, that one third of subjects had become alcohol abusers only after age 60, typically after devastating losses, such as the death of a spouse. These late-onset alcoholics had a bland and rather stable course and little psychiatric comorbidity (Ades und Lejoyeux 1994). Alcoholics surviving until old age and late-onset alcoholics were similar in this respect. In summary, the few available studies support the notion of a stable, comparatively benign course in the small group of alcoholics who survive to old age, while showing a surprisingly high percentage of late-onset alcoholics.

9 Personality Disorders

In the early and intermediate stages of life, personality disorders are important vulnerability factors for psychiatric disease and important determinants of disease course (Quinton et al. 1995). While the accentuation of personality traits in old age is frequently described, the development of personality disorders is a very complicated issue. If we consider that aging itself is an important stress factor, associated, as it is, with impairment of physical health, independence, and the ability to carry out activities of daily life, then we may expect inflexible personality types to decompensate. It is also conceivable, however, that personality styles held to be deviant at younger ages might confer certain advantages in old age. Paranoid or schizoid personalities would cope better with loneliness, while dependent personalities would gladly cast off their occupational duties and assume a newly legitimated dependency. Empirical studies must take account not only of cohort effects and selective mortality (as in borderline personality disorder), but also of altered behavioral criteria in old age. Impulsive, suicidal, or self-destructive behavior may, in the elderly, take forms that do not fit easily into current systems of psychiatric classification, appearing, for example, as noncompliance or refusal of treatment. Fearful, clinging behavior that would be pathological in a younger person may seem appropriate, or at least understandable, in old age.

Several studies have revealed a low prevalence of most personality disorders after age 55 (Cohen et al. 1994; Gurland 1984); this is particularly true of paranoid, schizoid, schizotypal, antisocial, and borderline disorders, and less so of histrionic and dependent disorders. In a group of 427 patients with "psychopathic personalities," Müller found that the great majority of features associated with personality disorders improved over the course of 30 years, less because of late maturation than because of a decline in the intensity of these features, and a relaxation of affect, as the patients grew older (average age, 71.9 years). Hypochondriac syndromes and hysterical personalities, however, developed in the opposite direction (Müller 1981). These observations from Lausanne were borne out by the findings of Tölle's study in Germany of a group of patients who were approximately 15 years younger (Tölle 1986) and by a long-term follow-up study from the Soviet Union (Semke 1964). Tölle found an unfavorable course in 33.9% of patients and a favorable course in 31.3%, while the remaining 34.8% coped with life by means of compromise strategies, involving a mixture of adaptation to, and narrowing of, their living conditions. McGlashan (1986) followed the life histories of 81 patients with the diagnosis of borderline personality disorder. Their ages at the time of diagnosis ranged from 16 to 55, and follow-up examinations were performed 2–32 years later, so that a number of patients could be observed up to and during old age. After a stable phase in approximately the second decade after diagnosis, these patients' conditions deteriorated again in old age (the age of involution). McGlashan and colleagues saw this process as the expression of a decline in cognitive performance and of the associated degree of social compensation in occupational life, through which the patients became increasingly dependent on their typically fragile and deficient personal relationships. It was thus not the personality style itself that had changed over the course of life, but rather the capacity to deal with, and partly compensate for, the difficulties created by it.

10 Conclusion

The intuitively predictable alleviation and waning dynamics of many disease processes in old age was described by early authors and is confirmed by the more recent studies. The traditional notion that schizophrenia has a uniformly poor prognosis certainly cannot be retained. In a fairly large subgroup of patients, however, cognitive impairments appear that are difficult to define clinically and to classify. Although many elderly, non-schizophrenic psychotics are relatively well adapted, their psychotic symptoms do not respond to pharmacotherapy. The class of patients with affective disorders, heterogeneous enough at any age, is even more so in advanced age, and this situation is reflected in the contradictory literature; the likeliest conclusion from the available data is that phasic processes have a tendency to become chronic, while bland, chronic processes of many years' duration have a tendency to become stabilized. In patients with affective disorders, too, as in schizophrenics, cognitive residual syndromes tend to occur whose classification and possible relation to organic changes in the brain are problematic. Aging generally has a favorable effect on preexisting psychogenic syndromes and personality disorders, with a few notable exceptions, particularly the post-traumatic syndromes. Aged alcoholics and substance addicts, despite the massive physical and mental stresses to which they are subjected, constitute a group of survivors whose further course is relatively unremarkable, even though abstinence is often not achieved. With the increased life expectancy of large segments of the population, the time over which the aging process can interact with mental illness has become considerably longer. Only a few longitudinal studies have as yet been able to address this recent development. Such studies inevitably run up against the inadequacy of current systems of psychiatric classification, which are not always able to accommodate the realities of geriatric psychiatry. The generally favorable course of elderly patients who were mentally ill before they reached old age stands in contrast to the considerable morbidity of patients who become mentally ill only once they are old.

11 References

Abas MA, Sahakian BJ, Levy R (1990) Neuropsychological deficits and ct scan changes in elderly depressives. Psychol Med 20: 507–520

Ades J, Lejoyeux M (1994) Addictive behavior in the elderly. Rev Prat 44(11): 1439–1442

Akbarian S, Kim JJ, Potkin SG, Hetrick WP, Bunney WE Jr, Jones EG (1996) Maldistribution of interstitial neurons in prefrontal white matter of the brain of schizophrenic patients. Arch Gen Psychiatry 53: 425–436

Alarcon RD, Deering CG, Glover SG, Ready DJ, Eddleman HC (1997) Should there be a clinical typology of posttraumatic stress disorder? Aust N Z J Psychiatry 31 (2): 159–167

Almeida OP, Howard RJ, Levy R, Anthony SD (1995) Psychotic states arising in late life (late paraphrenia) psychopathology and nosology. Br J Psychiatry 166: 205–214

Amensohn CS, Lewinsohn PM (1981) An investigation into the observed sex difference in prevalence of unipolar depression. J Abnorm Psychol 90: 1–13

Angst J (1966) Zur Aetiologie und Nosologie endogener depressiver Psychosen. Springer, Berlin Heidelberg New York

*Angst J (1987) Verlauf der affektiven Psychosen. In: Kisker KP, Lauter H, Meyer JE, Müller C, Strömgren E (eds) Psychiatrie der Gegenwart, vol 5. Springer, Berlin Heidelberg New York, pp 115–136

Arnold SE, Franz BR, Trojanowski JQ (1994) Elderly patients with schizophrenia exhibit infrequent neurodegenerative lesions. Aging 15: 299–303

Baldwin RC, Tomenson B (1995) Depression in later life. Br J Psychiatry 167: 649–652

Beckham JC, Kirby AC, Feldman ME, Hertzberg MA, Moore SD, Crawford AL, Davidson JR, Fairbank JA (1997) Prevalence and correlates of heavy smoking in Vietnam veterans with chronic posttraumatic stress disorder. Addict Behav 22(5): 637–647

Bleuler M, Huber G, Gross G, Schüttler R (1976) Der langfristige Verlauf schizophrener Psychosen. Gemeinsame Ergebnisse zweier Untersuchungen. Nervenarzt 47: 477–481

Boscarino JA (1997) Diseases among men 20 years after exposure to severe stress implications for clinical research and medical care. Psychosom Med 59(6): 605–614

Bremner JD, Randall P, Scott TM, Bronen RA, Seibyl JP, Southwick SM, Delaney RC, McCarthy G, Charney DS, Innis RB (1995) MRI-based measurement of hippocampal volume in patients with combat-related posttraumatic stress disorder. Am J Psychiatry 152(7): 973–981

Bremner JD, Licinio J, Darnell A, Krystal JH, Owens MJ, Southwick SM, Nemeroff CB, Charney DS (1997) Elevated CSF corticotropin-releasing factor concentrations in posttraumatic stress disorder. Am J Psychiatry 154(5): 624–629

Bristow MF, Clare AW (1992) Prevalence and characteristics of at-risk drinkers among elderly acute medical in-patients. Br J Addict 87(2): 291–294

**Ciompi L (1980a) The natural history of schizophrenia in the long term. Br J Psychiatry 136: 413–420

**Ciompi L (1980b) Catamnestic long-term study on the course of life and aging of schizophrenics. Schizophr Bull 6(4): 606–618

Ciompi L (1984) Zum Einfluss sozialer Faktoren auf den Langzeitverlauf der Schizophrenie. Schweizer Archiv für Neurologie Neurochirurgie und Psychiatrie 135(1): 101–113

**Ciompi L (1985) Aging and schizophrenic psychosis. Acta Psychiatr Scand Suppl 319: 93–105

Cohen BJ, Nestadt G, Samuels JF, Romanowski AJ, McHugh PR, Rabins PV (1994) Personality disorder in later life. A community study. Br J Psychiatry 165: 493–499

Davidson M, Harvey PD, Powchik P, Parrella M, White L, Knobler H, Losonczy MF, Keefe RSE, Katz S, Frecska E (1995) Severity of symptoms in chronically institutionalized geriatric schizophrenic patients. Am J Psychiatry 152(2): 197–207

*Devanand DP, Nobler MS, Singer T, Kiersky JE, Turret N, Roose SP, Sackheim HA (1994) Is dysthymia a different disorder in the elderly? Am J Psychiatry 151: 1592–1599

De Soto CB, O'Donnell WE, De Soto JL (1989) Long-term recovery in alcoholics. Alcohol Clin Exp Res 13(5): 693–697

Engdahl B, Dikel TN, Eberly R, Blank A Jr (1997) Posttraumatic stress disorder in a community group of former prisoners of war; a normative response to severe trauma. Am J Psychiatry 154(11): 1576–1581

Ernst K, Ernst E (1968) Eine vergleichende Literaturübersicht. In: Ernst K, Kind H, Rotach-Fuchs M (eds) Ergebnisse der Verlaufsforschung bei Neurosen. Springer, Berlin Heidelberg New York, pp 1–84

Feuerlein W, Kufner H, Flohrschutz T (1994) Mortality in alcoholic patients given in-patient treatment. Addiction 89(7): 841–849

Finney JW, Moos RH (1992) The long-term course of treated alcoholism. II. Predictors and correlates of 10-year functioning and mortality. J Stud Alcohol 53(2): 142–153

Freudenberg N (1991) Alterswandel psychischer Vefolgungsschäden. In: Stoffels (ed) Schicksale der Verfolgten. Springer, Berlin Heidelberg New York, pp 44–62

Foster JR, Reisberg B (1984) Effects of aging on psychiatric disorders beginning earlier in life. In: Kay Burrows (ed) Handbook of studies on psychiatry and old age. Elsevier, Amsterdam, pp 265–276

Goodwin FK, Jamison KR (1990) Manic depressive illness. Oxford, New York

Girling DM, Barkley C, Paykel ES, Gelhaar E, Brayne C, Gill C, Mathewson D, Huppert FA (1995) The prevalence of depression in a cohort of the very elderly. J Affect Dis 34: 319–325

Gurland BJ (1976) The comparative frequency of depression in various adult age groups. J Gerontol 31: 283–292

Gurland BJ (1984) Personality disorders in old age. In: Burrows K (ed) Handbook of studies on psychiatry and old age. Elsevier, Amsterdam

Harding CM (1988) Course types in schizophrenia: an analysis of European and American studies. Schizophr Bull 14(4): 633–643

Hughes CP, Berg L, Danziger WL, Coben LA, Martin RL (1982) A new clinical scale for staging dementia. Br J Psychiatry 140: 566–572

Jeste DV, Harris JH, Krull A, Kuck J, McAdams LA, Heaton R (1995) Clinical and neuropsychological characteristics of patients with late-onset schizophrenia. Am J Psychiatry 152: 722–730

Kay DW, Henderson AS,Scott R, Wilson J, Rickwood D, Grayson DA (1985) Dementia and depression among the elderly living in the Hobart community: the effect of diagnostic criteria on the prevalence rates. Psychol Med 15(4): 771–788

*Kral VA (1982) Depressive Pseudodemenz und senile Demenz vom Alzheimer-Typ. Nervenarzt 53: 284–286

Kuch K, Cox BJ (1992) Symptoms of PTSD in 124 survivors of the Holocaust. Am J Psychiatry 149(3): 337–340

Lee KA, Vaillant GE, Torrey WC, Elder GH (1995) A 50 year prospective study of the psychological sequelae of world war II combat. Am J Psychiatry 152(4): 516–522

*Lindesay J (1991) Phobic disorders in the elderly. Br J Psychiatry 159: 531–541

Marcos T, Salamero M, Gutierrez F, Catalan R, Gasto C, Lazaro L (1994) Cognitive dysfunctions in recovered melancholic patients. J Affect Dis 32: 133–137

Matussek P, Halbrach A, Troeger V (1965) Endogene Depression. Eine statistische Untersuchung unbehandelter Fälle. Urban and Schwarzenberg, Munich

**McGlashan TH (1986) The chestnut lodge follow up study III. Long-term outcome of borderline personalities. Arch Gen Psychiatry 41(6): 586–601

Migliorelli A, Teson A, Sabe L, Petracchi M, Leiguarda R, Starkstein SE (1995) Prevalence and correlates of dysthymia and major depression among patients with Alzheimer's disease. Am J Psychiatry 152: 37–44

Müller C (1981) Psychische Erkrankungen und ihr Verlauf sowie ihre Beeinflussung durch das Alter. Huber, Bern

**Müller C (1989) Altersveränderungen vorausgeganger psychischer Erkrankungen. In: Kisker KP, Lauter H, Meyer JE, Müller C, Strömgren E (eds) Alterspsychiatrie Psychiatrie der Gegenwart, vol 8. Springer, Berlin Heidelberg New York, pp 397–411

Murphy E (1989) Depressionen im Alter. In: Kisker KP, Lauter H, Meyer JE, Müller C, Strömgren E (eds) Alterspsychiatrie Psychiatrie der Gegenwart, vol 8. Springer, Berlin Heidelberg New York, pp 225–253

Myers JK, Weissmann MM, Tischler GL, Holzer III CE, Leaf PJ, Oorvaschel H, Anthony JC, Boyd GH (1984) Six month prevalence of psychiatric disorder in three communities: 1980–1982. Archives Gen Psychiatry 41: 959–967

*Opjordsmoen S (1989) Delusional disorders. I. Comparative longterm outcome. Acta Psychiatr Scand 80: 603–612

Opjordsmoen S, Retterstol N (1991) Delusional disorder: the predictive validity of the concept. Acta Psychiatr Scand 84(3): 250–254

*Opjordsmoen S, Retterstol N (1993) Outcome in delusional disorder in different periods of time. Possible implications for treatment with neuroleptics. Psychopathology 26(2): 90–94

Post F (1984) Affective psychoses. In: Burrows K (ed) Handbook of studies on psychiatry and old age. Elsevier, Amsterdam, pp 278–289

Purohit DP, Perl D, Haroutounian V, Powchik P, Didson M, Davis KL (1998) Alzheimer disease and related neurodegenerative diseases in elderly patients with schizophrenia. Arch Gen Psychiatry 55: 205–221

Quinton D, Gulliver L, Rutter M (1995) A 15–20 year follow up of adult psychiatric patients. Br J Psychiatry 167: 313–315

Regier DA, Boyd JH, Burke JD (1988) One month prevalence of mental disorders in the United States: based on five epidemiologic catchment area sites. Arch Gen Psychiatry 5: 997–986

Rossow I, Amundsen A (1995) Alcohol abuse and suicide: a 40-year prospective study of Norwegian conscripts. Addiction 90(5): 685–691

Roth M, Kay DWK (1956) Affective disorder arising in the senium. II. Physical disability as an aetiological factor. J Ment Sci 102: 141–150

Ruppert SD (1996) Alcohol abuse in older persons: implications for critical care. Crit Care Nurs Q 19(2): 62–70

Semke VY (1964) The course of psychopathy diseases in old age. Zh Nevropatol Psykhiat 64: 1688–1696

Stefansson JG, Lindal E, Bjornson JK, Guomundsdottir A (1991) Lifetime prevalences of specific mental disorders among people born in Iceland in 1931. Acta Psychiat Scand 84: 142–149

Tölle R (1981) Persönlichkeitsstörung und Neurose. In: Mester H, Tölle R (eds) Neurosen in der Psychiatrie. Springer, Berlin Heidelberg New York

Vaillant GE (1996) A long-term follow-up of male alcohol abuse. Arch Gen Psychiatry 53(3): 243–249

Van Ojen R, Hooijer C, Bezemer D, Jonker C, Lindeboom J, Van Tilburg W (1995) Late life depressive disorder in the community. Br J Psychiatry 166: 316–319

Weissman MM, Leaf PJ, Bruce ML, Florio L (1988) The epidemiology of dysthymia in five communities: rates, risks, comorbidity, and treatment. Am J Psychiatry 145: 815–819

Yassa R, Suranyi-Cadotte B (1993) Clinical characteristics of late onset schizophrenia: comparison with delusional disorder with and without hallucinations. Schizophr Bull 19: 701–711

CHAPTER

10

M. Blanchard, N. Graham

Old-Age Depression

1 Introduction

Depression in older people is probably no more but no less frequent than in younger people. The detrimental effect it has on life remains the same. It is a powerful determinant of quality of existence and is strongly associated with suicide. Although its aetiology may be somewhat different, based on the disabilities and losses which are increasingly frequent with older age, it remains a condition that is eminently treatable by multidisciplinary professionals. The major problem appears to be the suffering that occurs in silence – older people do not realise that their distress can be helped, and some professionals are likewise frozen with therapeutic nihilism. This chapter outlines areas where depression in old age is different from the condition encountered earlier in life, but the overriding emphasis must be the similarity across the age span in the importance of recognising those who are depressed and offering appropriate help.

2 Epidemiology

The important National Institute of Mental Health/Epidemiologic Catchment Area (NIMH/ECA) community studies suggested that major depression, as they identified it, declined in older age-groups, where it was therefore less of a clinical problem. This finding did no service to the many clinicians who encounter depression so frequently among their older patients. But who is to be believed? Apparent failure to identify morbidity in epidemiological studies can be founded upon the idiosyncrasies of case definition, sampling frame and means of assessment. These problems are perhaps greater across the age span, where, because of clinically identified differences in the presentation of depression and the confusion that physical illness brings in the aetiology of symptoms, separate epidemiological measures have been used to determine the prevalence of "caseness" for those above the arbitrary age of 65 years.

With the pathway to care and the "filters" through which people need to pass in order to become patients and receive care, there are three levels at which surveys into depression can measure morbidity: the community, primary care and hospitals (outpatients and inpatients). The results of surveys at each level are not comparable because of the distribution within the referral system; large numbers of cases remain hidden in the community, many accumulate in primary care and an excess of more severe and resistant cases tend to collect in clinics. We will therefore examine prevalence at each of these levels.

2.1 Community

Depressive symptoms do appear to become more frequent with increasing age (Gaitz and Scott 1972) and its concomitant disability. However, as already mentioned, results from the NIMH/ECA studies indicated that depressive syndromes, as they identified and defined them, were less prevalent in adults aged 65 years or older than in younger age-groups (Weissman et al. 1985). The Diagnostic Interview Schedule (DIS) was used to assess depression, and somatic symptoms, which may have been due to accompanying physical illness, were excluded from the diagnostic algorithm. The prevalence of major depression in old age was reported as 1.1% (excluding bereavement), which is two or three times less than younger age-groups. There have been many criticisms of the findings of the NIMH/ECA studies. Although large samples were used with good response rates (approximately 74%), a quarter of the interviews were incomplete and/or completed by informants. Moreover, the findings from older people in institutions were not included. The DIS excluded from diagnosis minor episodes of depression and those complaints, as indicated, which could conceivably be associated with physical disorders.

Blazer (1989) used results from the NIMH/ECA with those of his own studies and stated that the majority of older adults do not fit Diagnostic and Statistical Manual (DSM) criteria for depression, but rather have depressive symptoms associated with physical illness and/or adjustment to life stress. Even if this proves to be true – and it is disputed – any health policy based only on the prevalence of defined major depressive disorders will be ineffective if significant losses of function are associated with depressive symptoms that are substantial and widespread but not congruent with the diagnosis of a major disorder (Kennedy et al. 1989).

Because of the difficulties in identifying clinically relevant depression in older populations, schedules specifically designed for older people have been developed. The Comprehensive Assessment and Referral Evaluation (CARE) (Gurland et al. 1983) was developed to identify cases of depression in older people severe enough to warrant some form of intervention (probable pervasive depression). It was used to demonstrate very similar levels of depression in New York (13%) and London (12.9%). Copeland et al. (1987) utilised the "psychiatric" concept of caseness in the development of the Geriatric Mental State (GMS), which is linked to a diagnostic computer

programme (AGECAT) in order to analyse in a repeatable fashion the responses to this semi-structured interview. Rates of 3.0% more severe depression and 8.3% milder depression were identified in the older community in Liverpool (UK). From the results of a 3-year follow-up, an estimated incidence of GMS depression caseness of 23.7 per 1000 per year was made (Copeland et al. 1992). Blanchard et al. (1994) used a different methodology but calculated a similar incidence rate (18.2 per 1000 per year) within inner London.

Cross-cultural comparisons in Europe, the EURO-DEP studies (Copeland et al. 1999), have examined the variation in the prevalence of depression between countries in people aged 65 years and over. They have compared the clinical features and the mode of presentation of depression in each centre and, where possible, have studied the social support net works, adverse life events, daily life stresses and risk factors, and reasons for failure to receive treatment in primary care. The GMS interview, with its AGECAT determination of caseness, was used to make comparisons in the original studies. To allow further meaningful comparisons to be made with centres which did not use the GMS, a method which identified common ground between the different measures used was developed, the Euro-D.

Differences in the prevalence of depression were found across Europe: Amsterdam, 12%; Berlin, 17.6%; Dublin, 11.9%; Iceland, 8.8%; Liverpool, 10%; London, 17.3%; Munich, 23.6%; Verona, 18.3%; and Zaragoza, 10.7%. However, the cause of this variation was not obvious. The prevalence in women predominated over men, but there was no constant association between prevalence and age. A meta-analysis ($n = 13,808$) gave an overall prevalence of 12.3%, 14.1% for women and 8.6% for men.

With the closure of National Health Service (NHS) psychiatric long-stay hospital beds, more mentally ill and frail older people are being cared for in residential and nursing homes within the community in the United Kingdom. Ames (1990) examined the residents in 12 part III residential homes in North London using the Brief Assessment Schedule (a derivative of the CARE). A total of 24% of the residents had evidence of depression. The screened cases were then further assessed for specific diagnosis: half were found to have a primary affective disorder, while one third had depressive symptoms in the presence of an organic mental disorder. At 1 year follow-up, a quarter of the participants had died, while only 28% of those assessed for depression showed evidence of recovery. This study highlights the fact that community residential homes contain a large number of disabled residents with severe and enduring depression who require specialist support and management.

2.2 Primary Care

There are few surveys of depression in older people in primary care. Subjects in this setting have decided to consult their doctor, and consultation may well be for a physical condition. Recognition of any depression will depend upon general practitioner diagnosis; those who are recognised as depressed become part of the conspicuous morbidity in primary care, while those not recognised constitute part of the hidden morbidity. General practice surveys are therefore of self-selected groups who are likely to have an increased physical morbidity. Iliffe et al. (1991) examined a random sample of patients aged 75 years or older registered with London general practitioners and used the short version of CARE (Gurland et al. 1984). They found evidence of depression in 22% of subjects (total morbidity). The majority of these cases of depression were mild in nature. A study from the United States (Borson et al. 1986) used the Zung Self-Reporting Depression Scale (ZSRDS; Zung et al. 1983) to measure depression in people attending veterans' primary care facilities. They discovered 24.4% caseness in these men aged over 60 years, with an estimated 10% prevalence of DSM major depression. Only 1% of the depressed reported the use of mental health services.

That a large proportion of depression does remain "hidden" from mainstream medical services has been indicated by several studies (Koenig et al. 1988; Blanchard et al. 1994). Blanchard and co-workers studied the "declaration" of depressive symptoms by older people to general practitioners and discovered that only one third of depressed older people believed that their doctor was aware of their current depression. Iliffe et al. (1991) examined general practitioners' notes and found that only 6% of their "screened" depressed cases were recorded as such. There is clearly a need to improve the identification of depressed older people in primary care, as without this subsequent management or referral cannot occur. The means to do this could involve the use of screening questionnaires such as the self-CARE (Bird et al. 1987) and the development of training programmes for primary care physicians and other primary care personnel (Gask 1992; Waterreus et al. 1997).

2.3 Psychiatric Inpatients

The population of older patients admitted to hospital for care is heavily determined by the health system operating within individual countries. Most depressed

older people in Britain are dependent upon primary care referral in order to obtain specialist care. The majority of psychiatric research, which inevitably forms our knowledge base of depression, has been carried out on this rather select group of patients. Depression in older people is associated with an increased use of primary care services, but there is no indication that depression itself leads to preferential referral to specialist services. There are probably many factors that determine the referral or otherwise of depressed older people to specialist facilities, but these have not been investigated in any systematic way. Probable factors would include some measures of severity of depression, e.g. psychotic symptoms, lack of social support in the light of observed self-neglect, mention of suicidal ideas, attempts at self-harm, request for psychiatric help by the patient or carers, repeated annoyance of the general practitioner by patient or carers and a concomitant psychiatric condition such as alcoholism or phobia.

Depression is the commonest single condition dealt with by most specialist old-age mental health services. The locus of care for an individual will depend upon the current organisation of local services. The trend is towards expanding community-based nursing support with a concomitant reduction in inpatient facilities. The established efficacy of electroconvulsive therapy (ECT) for certain depressive subtypes, and its use in life-threatening states and where pharmacotherapy is contraindicated or ineffective, is also likely to indicate the need for inpatient status in this new climate.

2.4
Physically Ill

There has also been considerable research on depression in older physically ill people in hospital settings. Patients with untreated depression may have impaired physical recovery and reduced compliance to physical treatments. Koenig et al. (1988) examined older patients admitted to a medical ward and identified major depression in 11.5% and other depressive syndromes in 23%. Those more likely to be depressed were older, had less formal education, had cognitive dysfunction, suffered from more severe medical illness (particularly recent myocardial infarction), had a history of psychiatric illness, were unmarried and had more severe disability. Within the outpatient setting, Kukull et al. (1986) used the ZSRDS and resurveyed a cohort of older general medical clinic outpatients at 33 months. They found point prevalence of 20% depression (score, >60) at both times with half of the patients depressed at both points, a marked level of chronicity. The number of new physical diagnoses was the best predictor of depression at retest with positive associations between chronicity of depression and the number of co-existing physical diagnoses.

3
Comparative Phenomenology

3.1
Young Versus Old

The possible existence of a depressive syndrome particular to older people – differing from younger adults in terms of aetiology and phenomenology – has been debated. Depression in old age has at times been thought to be more "melancholic", psychotic or hypochondriacal. However, there are difficulties in determining differences in the phenomenology between older and younger depressed patients because of the need to examine equivalent populations. Kramer-Ginsberg et al. (1989) studied 70 consecutive patients aged 60 years or older who were admitted to an acute psychogeriatric unit diagnosed as suffering from DSM major depression. A total of 60% of these patients had hypochondriacal symptoms as measured on the Hamilton Depression Rating Scale (HDRS), and 12% of these were of delusional intensity. In marked contrast to this, Blanchard et al. (1994) found, within their cases of depression identified by GMS in a community population of older people, only three out of 59 (5%) reached subcase level, and another three out of 59 (5%) reached case level on the hypochondriasis scale. These conflicting results demonstrate clearly the importance of population selection upon the phenomenological profile of depression. Two studies by Oxman et al. (1990) and Musetti et al. (1989) compared samples of old and young patients with depression and found similar symptom profiles; a stereotype of old-age depression as one of somatisation and hypochondriasis was not confirmed.

In a consensus document produced by the Royal Colleges of General Practitioners and Psychiatrists (Katona et al. 1995), the difficulties in diagnosing depression in some older people and the traps that physicians may fall into are highlighted. In particular, it should be noted that somatic symptoms and cognitive deficits might be the presenting features for some depressed older people and confusingly depressed mood as such may be absent.

3.2
Early Versus Late Onset

Baldwin and Tomenson (1995) compared early-onset (59 years or younger) and late-onset major depression

in 57 depressed inpatients. They found that symptoms between the two groups differed little; however, heritability was greater in the early-onset group, and there was a striking association of vascular disease and/or risk with late-onset patients. Thus, although phenomenology may be similar, there is the possibility that depression in later life has more of a 'biological' basis in late-onset illness, and more research is needed in this area.

Koenig et al. (1993) profiled the depressive symptoms of younger and older medical inpatients with major depression. There were great similarities in their symptom profiles, but in older people it was found that somatic symptoms might be more important in this group in diagnosing major depression, than previously realised.

3.3 Depression Versus Dementia

Depression and dementia are the two most common clinical problems in the psychiatric care of older people, and difficulties in diagnosis can occur as not only can the symptoms of the syndromes overlap, but up to about a quarter of patients with cognitive impairment warrant a diagnosis of depression, and up to a fifth of depressed patients appear to have cognitive problems. Yesavage (1993) states that more than a quarter of patients with dementia may be initially misdiagnosed as having an affective disorder, and inaccuracy in the diagnosis of dementia may be as high as a third. Symptoms such as flattened affect, paucity of speech, slowness, poor concentration, persecutory delusions, hallucinations, aggressivity or irritability, diurnal rhythm changes and anxieties can all occur in both conditions. There are, however, symptoms primarily associated with depressive disorder, namely pervasive unhappiness, excessive guilt and suicidal ideation, which along with factors in the clinical presentation can help with the differential diagnosis (Yesavage 1993). There is currently no biological test or marker which can help in diagnosis, but O'Brien et al. (1994) suggest that temporal lobe magnetic resonance imaging may be useful.

The term "pseudodementia", which has been used widely in the debate about the depression/dementia interface, and which literally means a false dementia, is based on the ideas of non-progression, non-organicity and reversibility. It may have a variety of causes, one of which is depression. It implies a falseness of deficit, almost an absence of pathology, and the authors suggest that the term no longer be used. It is more useful to discuss the genuine depressive dementia that can occur in older people. Emery and Oxman (1992) discuss the idea of three intersecting continua of depression, cognitive impairment and degenerative dementia. They suggest that five conditions may occur:

1. Major depression with minimal cognitive deficits (noting that depression in individuals aged over 40 years nearly always involves some cognitive disadvantage)
2. Depressive dementia possibly associated with the pathology of mood disorders and controversially with the physical changes associated with late-onset mood disorders
3. Degenerative dementia without depression
4. The depression of degenerative dementia either directly caused or secondary to demoralisation
5. Independent co-occurrence

The need to make an exact differential diagnosis is eased to some extent by taking this continuum perspective and recognising depressive dementia, dual diagnosis and the idea of excess disability within dementia. The fact that depression may precede dementia is another intriguing factor in this complicated topic (Jorm et al. 1991).

4 Aetiology

4.1 Role of Physical Illness

There is a strong association between depression and physical ill health in older people. Certain illnesses may have a direct aetiologic role, either through physiological or cognitive mechanisms, while others may act as perpetuating factors. It is therefore very important to assess and manage physical illness as an integral part of the management of depressed older patients. Specific illnesses which have particular relevance to older people and depressive illness are stroke (Starkstein and Robinson 1989), Parkinson's disease (Dooneief et al. 1992), hypothyroidism (Tappy et al. 1987), dementia (Wragg and Jeste 1989) and hip fracture (Shamash et al. 1992).

Kennedy et al. (1989) used the Centre for Epidemiological Studies Depression Scale (CES-D) to study older community residents and found a hierarchy of characteristics associated with the substantial levels of depressive symptoms: illness, disability, isolation, bereavement and poverty. The prevalence and relative risk of depressive symptoms was related to the number of medical conditions, the number of problems in activities of daily living (ADL), disability, perceived health and the number of visits to a physician. The fact that people felt that they had little or no control over their health also appeared to be important. There is

evidence that it is disability and handicap that are the important correlates of depression rather than physical illness per se. Prince et al. (1997) discovered from their household enumerated population of older people that impairment, disability and handicap were all strongly associated with depression and that this association was strongest with handicap, with an odds ratio of 24.2 in the most handicapped quartile.

Physical illness as an important associate with depression also has an impact on prognosis. Kennedy et al. (1990) studied characteristics that may predict persistence or remission of depressive symptoms in their community population. They re-interviewed 1577 people at 24 months and compared the characteristics of 97 people whose depressive symptoms persisted over the 2 years with 114 people whose symptoms remitted. Increasing disability and declining health preceded the emergence of depressive symptoms and accounted for 70% of the variance in their discriminant analyses. Advanced age and worsening health were associated with persistent symptoms, and improved health with remission.

4.2 Role of Bereavement

Old age is a time of bereavements and losses which could naturally be expected to precipitate depressive states. It has been argued, however, that this type of life event is to be expected in old age and therefore may be less damaging in its psychological impact than when it occurs earlier in life. The vast majority of bereaved people do pass through the natural process of grief without recourse to doctors, least of all psychiatrists. However, given the fact that the majority of depression in older people is not recognised or treated, it is important to examine the relationship between bereavement and depression in a representative community sample. Bruce et al. (1990) used NIMH/ECA data and compared depressive episodes and dysphoria between newly bereaved ($n = 39$) and married ($n = 1047$) people aged 45 years and older. They found that bereavement greatly increased the risk of depression and dysphoria, and that the "bereaved depression" patients reported significantly fewer symptoms of guilt when compared to the non-bereaved depressed. None of the bereaved depressed patients had experienced a prior depressive episode. Blanchard et al. (1994) reported that 27 out of 96 depressed community patients became depressed in the 6 months following a bereavement, and of these, 23 had their depressive symptoms for more than 2 years. In an attempt to determine the relationship between recent bereavement and suicide, Bunch (1972) studied 75 consecutive cases of suicide and compared them with 150 matched controls living in the same area. She found that significantly (up to five times) more suicides than controls had been widowed within the previous 3 years and that more widows than widowers killed themselves. It is probable that bereavement does play a greater aetiologic role in depression and suicide in older people than perhaps is generally appreciated and that this has important implications for its management.

4.3 Role of Social Factors

There are few studies in older people using detailed methodology such as that employed by Brown and Harris (1978); community studies tend to use methodologically simpler techniques for the measurement of social factors and to use measures of depressive symptoms rather than clinical caseness criteria. Despite this, there is evidence that social factors can influence onset and course of depression in older people. Murphy (1982) compared a group of 100 older depressed people referred to local psychiatric services with a group of age- and sex-matched non-depressed people from the general population. Taking 1 year before interview or 1 year before onset of depression as the time periods of interest, 48% of patients compared with 23% of non-depressed subjects had a severe life event in the year before onset; 42% of patients and 19% of the non-depressed had a major non-health difficulty; 39% of patients and 26% of the non-depressed had major health difficulties. Emmerson et al. (1989) studied 101 patients with DSM-III major depression and compared them with 85 non-depressed community residents who scored 4 or less on the General Health Questionnaire. They found that significantly more of the depressed patients (24% vs. 7%) reported at least one severe life event in the 3 months prior to onset of illness. The fact that life difficulties were rare in their sample was attributed to the fact that they were studying a relatively affluent population. Pahkala (1990) studied 1529 people born in Ahtari, Finland in 1923 or before. The ZSRDS was used as a postal screen and then anyone scoring 40 or more, plus a random sample of those scoring less, were interviewed. This cross-sectional study demonstrated that depression was associated with retirement through sickness, lack of intimate friendships and the occurrence of long-standing and current social stressors.

Social factors are also important in the prognosis of depression. Kivela and Pahkala (1989) performed a prospective follow-up with a mean duration of 14.9 months on 264 community depressed older people. Outcome was good in 41% of patients. Poor

outcome was associated with low social participation, low frequency of visiting contacts and low perceived health. Oxman et al. (1992) attempted to examine the effect of the characteristics of social networks and support on depressive symptoms. They utilised results on surveys carried out on 1962 community residents aged 65 years and older first seen in 1982 and followed up in 1985. Baseline depression, functional disability in 1982 and change in disability by 1985 contributed mainly to variance and required adjustment along with sociodemographic variables for their multiple regression analyses. Loss of a spouse, inadequate emotional support and its change between 1982 and 1985 made the largest social contributions to depression and its outcome. Other significant factors were "tangible" support adequacy and its change, loss of a confidant, number of children making weekly visits and any change in this, and the absence of a confidant in 1982. This study highlights the need to consider specific dimensions of social support. Ong et al. (1987) investigated the role of a support group for depressed older people and found that a simple weekly support group resulted in significant reductions of relapse and readmission over 1 year. The effect of re-establishing social interaction has been demonstrated by Schonfield et al. (1985), who examined two groups ($n = 42$, $n = 47$) aged 55–91 years. They followed one group through a mental health treatment programme, which emphasised the strengthening of social networks, and the other through a nutrition programme. Initially, the mental health programme group scored significantly higher on the Beck Depression Inventory and had fewer friends as measured on the Social Support Network Inventory. "Graduates" of the mental health programme improved significantly in their Beck scores with a concomitant increase in friends. Their conclusion was that continued socialisation in later years might serve to allay depression.

4.4
Role of Genetic Factors

When the presence of a family history of depression in older depressed patients is assessed, a difference appears between the unipolar depression that first occurs late in life and the depression that continues from earlier in the life cycle. These findings were concluded for hospital patients, and it must be remembered that the determination of a previous history of depression in older relatives may be difficult. Baron et al. (1981) examined age at onset data in 1468 first-degree relatives of 252 probands with unipolar and bipolar affective disorders. Early-onset (< 40 years) probands had more early-onset relatives and a greater risk for affective disorder in their relatives than late-onset probands. This suggested a familial factor in early-onset illness tending to breed true, but of course an entire follow-up of relatives would be needed to complete the picture with respect to late-onset depression. Maier et al. (1991) also investigated the notion that late-onset depression (>60 years) was believed to be associated with less risk of depression in first-degree relatives. They compared older inpatients with unipolar major depression ($n = 92$) with aged-matched controls ($n = 33$). They also discovered that relatives of probands with late-onset depression were at less risk of depression than those with early-onset depression. Similar differences, with late-onset depression being more "event precipitated", having less evidence of family history and earlier "personality inadequacies", are reported by Musetti et al. (1989) and Brodaty et al. (1991).

4.5
Brain Biochemistry, Physiology and Structure

Developments in imaging and neurobiochemical research techniques have stimulated investigations into the depression of older people. It does appear that there are specific subgroups of older people with depression who differ from the depressed population as a whole in terms of structural brain changes and neurotransmitter status. These results indicate the possibility of a "biological propensity" towards onset and/or poor prognosis of depression in some older people.

Jacoby et al. (1983) examined an index of brain tissue density from the computed tomography (CT) scans of 37 elderly depressed patients calculated using Hounsfield units (HU) in 12 predefined areas of brain. They compared their study group with 36 healthy controls and 23 subjects with dementia. The controls showed the highest HU values and dementia sufferers the lowest. There was no correlation between HU values and age. In the depressed patients, ventricular dilatation was associated with lower HU levels, but this was not the case in controls. The authors suggested that these results might indicate a specific pathological process underlying depression in a certain subgroup of depressed older people, especially as the changes did not appear to be related to future dementia on follow-up. Significantly, 2-year mortality was increased in the depressed people with ventricular enlargement as shown on CT scan (Jacoby and Bird 1981). Greenwald et al. (1996) studied patients with depression ($n = 48$) and discovered that they were more likely to manifest magnetic resonance imaging (MRI) abnormalities than non-depressed older people with similar demographic features ($n = 39$). After allowing for age and gender,

depressed patients had a greater degree of hyperintensities in the subcortical grey matter than controls. It appears that depression in late life can be associated with a remarkable amount of overt pathologic change in the brain, and this has been reviewed by O'Brien et al. (1996).

Older and younger people appear to differ in their aminergic systems, differences likely to lead to greater levels of depression in older people when considered within the context of the current amine paradigm. Robinson et al. (1971) examined the monoamine oxidase (MAO) activity in blood plasma and platelets of 113 healthy subjects aged between 21 and 84 years. Activity was found to correlate highly with increasing age when measured in plasma and platelets. Women had higher activity than men only in the platelet measures. In addition, Nemeroff et al. (1988) discovered a marked reduction in the platelet-tritiated imipramine-binding sites in older depressed patients. In terms of possible neuroendocrine abnormalities in depression in older people, Heuser et al. (1996) compared the effects of 6-week treatment with amitriptyline on hypothalamic-pituitary-adrenocortical (HPA) regulation in older depressed patients ($n = 39$) and age-matched controls ($n = 14$). The depressed patients had profoundly abnormal HPA responses, which began to disappear after 1 week of treatment with amitriptyline. In contrast, amitriptyline did not affect neuroendocrine regulation in the comparison subjects at any time during the test period. This raises the possibility that the normalisation of the HPA feedback control is related to the antidepressive effect of amitriptyline and is therefore significant in aetiology.

Another physical measure used to differentiate "biological" depression from depressive symptoms is that of sleep patterns. A recent study has cast doubt upon the nosology of depression and bereavement reactions found in various research and clinical diagnostic criteria. This is an important issue in older people where so much depression is discounted as a "normal reaction" to circumstances. Reynolds et al. (1992) studied 31 older volunteers with recent spousal bereavement who were stratified by the presence ($n = 15$) or absence ($n = 16$) of major depression; they had no personal history of psychiatric disorder. Those with major depression had lower sleep efficiency, more early morning wakening, shorter rapid eye movement (REM) latency and a greater percentage of REM. These depressed patients resembled older patients with recurrent unipolar major depression. Sleep in bereavement without depressive symptoms was similar to 15 healthy controls. The findings suggest that the current concept of uncomplicated bereavement is unconfirmed, because many of the depressed bereaved would have fallen into this category and yet they had identical sleep patterns to those found in major depressive episodes.

5 Prognosis

5.1 Relative Prognosis

Studies on the prognosis of depression in older people can again be separated into those based upon hospital populations and those examining depression within a community setting. Initial studies appeared to show a poor overall prognosis in the more severe cases of depression in older people, but further research has indicated that reality may not be quite so gloomy. Alexopoulos et al. (1996) compared older and younger depressed inpatients and found them to have an equal recovery rate of 60% at 6 months. Cole and Bellavance (1997) conducted a comprehensive review of the literature on prognosis from 1955 to 1993. Their meta-analysis indicated that about 60% of patients treated in the hospital setting either remained well or had treatable relapses. Only one in five patients developed chronic symptoms.

5.2 Secondary Care Prognosis

Murphy (1983) performed a 1-year prospective study of 124 older patients referred to psychiatric services for the first time; only one third of her group had a good outcome. Poor outcome was associated with the severity of the initial illness and the presence of psychotic symptoms and with physical health problems and severe life events in the follow-up year. Baldwin and Jolley (1986) performed a retrospective analysis of between 42 and 104 months on 100 elderly patients admitted to hospital with severe, non-neurotic depressive states, none with previous hypomania. They found that 60% remained well or had one or more further episodes followed by full recovery, while 7% suffered continuous depressive symptoms. This more optimistic result contrasts with that found by Murphy (1983) and may in part be due to their less rigorous methodology (retrospective) and differences in treatments (more ECT used by these authors). Male sex and poor physical health at presentation or developing subsequently were the only predictors of poor outcome. Cole (1983) examined age, age at onset and the course of primary depression following hospital admission among older people. He found that the effect of age was not significantly related to the

prognosis of depression in the absence of persistent organic signs and severe physical illness. The severity of illness and continuing physical ill health were the two most important factors determining outcome. People with late onset depression were more likely to remain completely well during follow-up than those with early life onset (55% vs. 25%). Cole (1985) also studied the course of 55 older depressed outpatients whom he followed for 24–63 months. A total of 38 (69%) remained well for more than 60% of the follow-up period, and 17 patients (31%) remained chronically ill. Treatment compliance, absence of physical disability, long-term follow-up and maintenance antidepressant therapy were associated with favourable outcome.

5.3 Community Prognosis

Copeland et al. (1992) examined the prognosis of depression in the community by following their cohort over 3 years. They found an increased mortality and a marked chronicity (30% remained cases) among their depression patients. This apparently poor prognosis probably reflects the lower overall severity of depression found in the community compared to inpatient studies and also the lack of recognition and paucity of treatment known to occur (Livingston et al. 1990; Blanchard et al. 1994).

5.4 Suicide as an Outcome

Although suicide is increasingly more frequent among young men, suicide rates are highest among older people throughout the world (Moscicki 1995). In England and Wales, the suicide rate is estimated to be about 10–20 per 100,000 per year. In a survey of older people who committed suicide in the United States, 64% used violent means, whereas in the rest of the Western world, where firearms are less accessible, 85% committed suicide by self-poisoning, jumping, hanging or car exhaust inhalation. There does appear to be a clearer relationship between depression and suicide in older people than is the case with younger people (Barraclough et al. 1974; Conwell and Caine 1991). This theoretically could make strategies for reducing the rates of suicide in older people less complex. Other important factors include sex (male), separation or divorce, previous suicide attempts, a family history of psychiatric illness or suicide and recent severe life events. Recent major political changes in Eastern Europe appear to have increased the rate of suicide in this area; however, among those aged 75 years or over, the rates have decreased (Sartorius 1995).

6 Treatment

6.1 General Principles

Many of the guiding principles for the treatment of depression in younger people are true for older people as well. However, differences do include the fact that the majority of first assessments and follow-up appointments of depressed older people are carried out in their own homes and that informant interviews and an assessment of family support may be more pertinent. Once a diagnosis of depression has been reached, then a comprehensive assessment of the patient's physical, psychological and social history and current state needs to be made before it is possible to produce a management plan. It is essential that the depression is treated within its context; no two depressions are the same in this respect.

6.2 Pharmacotherapy

McCusker et al. (1998), in a review article, examined acute-phase pharmacological and psychological treatments in older people in outpatient, community or nursing home settings. Forty out of 233 articles met their selection criteria, 26 on pharmacological treatment and 14 on psychological treatment. Antidepressants were found to be more effective than placebo, and rational therapies more effective than a no-treatment control.

Older patients may be more prone to the side-effects of antidepressants and, certainly with the older tricyclics, dosage may require careful titration. This is less of a problem with the newer selective serotonin re-uptake inhibitors (SSRI). Older patients may take longer to recover (Schneider et al. 1994), so that a reasonable course of antidepressants may be longer than in younger adults, lasting up to 6–8 weeks.

Anstey and Brodaty (1995) stated that there are no high-quality randomised, placebo-controlled trials of drug or psychological therapy in the acute phase of old-age depressive illness of whatever severity, and none could now be done for ethical reasons. Jacoby (1998), in a commentary to the McCusker article, makes some important points about older depressed people: that they are a heterogeneous group, that many studies are far too short, lasting only a few weeks, and

that the Hamilton Depression Scale is commonly used as an outcome measure, a scale that has a heavy weighting on somatic items, which makes it less suitable as a measure, as non-depressed older people can score positively on these items. He goes on to consider that, certainly among those with mild to moderate depression, time spent talking might in some ways be as effective as any drug.

Callahan et al. (1994) studied the effect that training in the recognition and management of depression in older people had on the actions of primary care physicians. They provided patient-specific treatment recommendations and guidelines for prescribing to their intervention group and compared them with a usual care group. Intervention patients were more likely to receive a written diagnosis of depression and be prescribed antidepressants (nortryptyline or fluoxetine); however, over a 9-month follow-up, there were no differences in outcome between the two groups. This may have been due to problems of compliance or, as the authors suggest, there may be a need to provide more integrated interventions, including attempts to target psychosocial stressors, in the management of depression in older people; this is probably outside the purview of primary care as services are currently organised.

6.3 Electroconvulsive Therapy

ECT remains an important treatment for the few older people with severe depression in whom a rapid response is necessary or who are resistant to other treatments. It is a relatively safe procedure with clinical improvement in approximately 80% of those who receive it (Katona 1994). Cognitive side-effects can be a particular problem in older people, especially in those with organic mood disorders, but in most cases any amnesia resolves almost completely (Benbow 1989). There are no absolute contraindications to ECT, but great care needs to be taken in patients with uncontrolled cardiac disease, recent stroke or raised intracranial pressure from another cause. The relative risks need to be weighed up against alternative treatments or no treatment, and, with the advances in psychopharmacology such as the SSRI, it is the authors' clinical impression that the need for ECT is diminishing.

6.4 Psychotherapies

There is no reason to assume that research findings on the efficacy of behaviour and cognitive therapy in younger people should not be applied to older people. However, earlier this century, Freud made an observation which unfortunately, because of his influential status, had a powerful inhibitory effect on the development of psychotherapy for older adults for decades. "Near and above the fifties, the elasticity of the mental processes on which the treatment depends is, as a rule, lacking - old people are no longer educable, and, on the other hand, the mass of material to be dealt with would prolong the duration of the treatment indefinitely" (Freud 1924). His conclusion was that mental health problems were not as amenable to psychological interventions in old age as in youth. In contradiction to these beliefs, Gallagher-Thompson et al. (1990) reported on a 2-year follow-up of 91 older adult volunteers with major depression who had been treated with 4 months of brief cognitive, behavioural or psychodynamic psychotherapy. Seventy per cent of them remained non-cases, and there was no difference in outcome according to the type of treatment received. However, this was a relatively young (mean age, 67 years) and physically healthy group of people.

In his extensive review of psychological therapies in older people, Woods (1992) highlights the importance of coping with physical illness and disability in old age and the relative neglect of this as a major psychological issue requiring more attention from therapists and researchers. Mossey et al. (1996) examined the response to psychological intervention of subdysthymic depression, which occurs in 20%–50% of older people hospitalised for physical illness. Such depression is associated with physical and social disability, delayed recovery and excess health service use, and yet little work has been done either to describe its nature or to examine its response to treatment. In a randomised clinical trial, the efficacy of interpersonal counselling (IPC), a short-term psychotherapy, delivered following hospital discharge by psychiatric clinical nurse specialists, was assessed in patients with a Geriatric Depression Scale (GDS) score of more than 10 but not meeting DSM-III-R criteria for major depression or dysthymia. At 6 months, a statistically significant difference in the rate of improvement in GDS, indicated by scores of 10 or less, was observed for IPC compared to control (usual care) members (60.6% vs. 35.1%).

In terms of treatment for depression in older people living in residential care, Ames (1990) was unable to positively influence the depression he encountered. However, there is some evidence from Rosen et al. (1997) that changes involving the milieu of the home (a recreational therapist working 1 h a day) can benefit depressed residents. Unfortunately, the benefit disappeared with the ending of the intervention.

Steuer (1982) listed the modifications which psychotherapists currently believe are necessary for the treatment of older people:

- The therapist needs to take an active role in directing any therapy, at the same time allowing the patient to actively seek solutions to their own problems.
- Current problems should be the focus of treatment.
- Education of patients about normal ageing and intervention by helping to plan daily activities and acting as a social advocate is necessary.
- The goals of treatment may be more limited than in younger patients, e.g. symptom reduction, acceptance of greater dependency, learning new coping skills and increasing self-esteem all may be useful to improve the patient's quality of life.

Assertion training, problem-orientated treatment and cognitive-behavioural therapies may therefore all have a place in management. Certainly, acute intervention in the face of physical illness and multiple losses has been strongly advocated for older people, as has the supportive value of therapeutic groups (Ong et al. 1987).

6.5 Multidisciplinary "Packages" of Care

Most work in secondary care old-age psychiatry occurs within the community and involves multiple interventions rather than simple prescription. Blanchard et al. (1995) examined the effect on depression identified by screening within the community of brief, individually tailored packages of care developed by a secondary care multidisciplinary team. The team included doctors, a social worker, psychiatric nurses and a psychologist, with occupational therapy and physiotherapy advice available. Initial assessments and treatment were carried out in the older people's own homes, the former by a research psychiatrist and the latter by a community nurse in liaison with the primary health care teams. At 3 months after intervention, they discovered a significant benefit in depression when compared with normal primary care management, and this benefit was particularly marked in older people with long-standing depression. The prescription of antidepressants as part of the treatment package (which was suggested in just under half of the depressed people) was met with resistance from both older people and their general practitioners. However, psychological interventions were accepted by everyone offered them. In a similar study to examine the effect of multidisciplinary management on depression in older people who were in receipt of home care in the community, Banerjee et al. (1996) discovered a marked treatment benefit when compared to a "normal treatment" control. Again, antidepressant medication did not appear to be a significant factor in the treatment effect. These two pieces of work counter the therapeutic nihilism frequently met with when working with depressed older community residents. They demonstrate a possible way forward to improve the prognosis of this group, with cooperation between secondary and primary care teams developing multidisciplinary care packages which are delivered in the home and coordinated by a "key-worker".

7 Public Health Issues

Western societies are oriented towards youth and the future and are increasingly rewarding productivity and accomplishment – the establishment of "meritocracies". This, along with the negative therapeutic attitudes as discussed above, engenders negative stereotypical attitudes towards older people. Among older people there is a tendency to attribute depressive symptoms to growing old or physical illnesses. Depressed older people either avoid seeking help or search for medical rather than psychological treatments; they carry with them society's false belief that discomfort and unhappiness are to be expected in later years. Consequently, older people tend not to be referred to mental health services, but if they are, it may only be once their problems have become chronic. When health care intervention occurs, it is invariably shared by several medical specialists, social workers and other health care professionals. It is sometimes unclear who is assuming responsibility for coordinating the efforts of all concerned. Without such coordination, the care delivery is fragmented and occasionally actions are counter-productive.

It is important to advance a mental health policy to counter this negative stance taken by our society. There should be primary preventative measures such as education about ageing during school years, preretirement counselling, education of medical students (particularly those who are to become primary care physicians) and education of the general public through television and newspapers. Secondary prevention measures should include the necessary screening questionnaires, training in various interviewing techniques and easy access to, or the development of, specialist multidisciplinary mental health teams with attached care managers, whose role would be to carefully coordinate management (WHO 1996). All mental health professionals need training in the recognition, assessment and treatment of depression in older people. Novel ways of trying to manage older people with depression or those at risk of depression in their own homes need to be investigated. One example could be the use of volunteer "befrienders" to decrease the isolation which is commonly associated with

depression, and another would be more socially oriented local "drop-in" centres, which could start to foster a sense of community among the old. Tertiary prevention will require the establishment of psychosocial rehabilitative programmes within residential and nursing homes to prevent the suffering that currently occurs from this treatable mental illness.

8 References

Alexopoulos GS, Meyers BS, Young RC, Kakuma T, Feder M, Einhorn A, Rosendahl E (1996) Recovery in geriatric depression. Arch Gen Psychiatry 53(4): 305–312

Ames D (1990) Depression among elderly residents of local-authority residential homes; its nature and the efficacy of intervention. Br J Psychiatry 156: 667–675

Anstey K, Brodaty H (1995) Antidepressants and the elderly: double blind trials 1987–92. Int J Geriatr Psychiatry 10: 265–279

Baldwin RC, Jolley DG (1986) The prognosis of depression in old age. Br J Psychiatry 149: 571–583

Baldwin RC, Tomenson B (1995) Depression in later life. A comparison of symptoms and risk factors in early and late onset cases. Br J Psychiatry 167(5): 649–652

*Banerjee S, Shamash K, Macdonald AJ, Mann AH (1996) Randomised controlled trial of effect of intervention by psychogeriatric team on depression in frail elderly people at home. BMJ 313(7064): 1058–1061

Baron M, Mendelwicz J, Klotz J (1981) Age of onset and genetic transmission in affective disorders. Acta Psychiatr Scand 64: 373–380

*Barraclough M, Bunch J, Nelson B et al (1974) One hundred cases of suicide-clinical aspects. Br J Psychiatry 125: 355–373

Benbow SM (1989) The role of electroconvulsive therapy in the treatment of depressive illness in old age. Br J Psychiatry 155: 147–152

Bird A, Macdonald A, Mann A, Philpott M (1987) Preliminary experience with the SELFCARE (D): a self-rating depression questionnaire for use in elderly, non-institutionalised subjects. Int J Geriatr Psychiatry 2: 131–138

Blanchard M, Waterreus A, Mann AH (1994) The nature of depression amongst older people in inner London, and the contact with primary care. Br J Psychiatry 164: 396–402

*Blanchard M, Waterreus A, Mann AH (1995) The effect of primary care nurse intervention upon older people screened as depressed. Int J Geriatr Psychiatry 10: 289–298

Blazer D (1989) Depression in the elderly. N Engl J Med 320: 164–166

Borson S, Barnes RA, Kukull WA, Okimoto JT, Veith RC, Thomas IS, Carter W, Raskind MA (1986) Symptomatic depression in elderly medical outpatients. I. Prevalence, demography and health service utilization J Am Geriatr Soc 34: 341–347

Brodaty H, Peters K, Boyce P, Hickie I, Parker G, Mitchell P, Wilhelm K (1991) Age and depression. J Affect Disord 23: 137–149

Brown GW, Harris TO (1978) Social origins of depression. Tavistock, London

Bruce ML, Kim K, Leaf PJ, Jacobs S (1990) Depressive episodes and dysphoria resulting from conjugal bereavement in a prospective community sample. Am J Psychiatry 147: 608–611

Bunch J (1972) Recent bereavement in relation to suicide. J Psychosom Res 16: 361–366

*Callahan CM, Hendrie HC, Dittus RS, Brater DC, Hui SL, Tierney WM (1994) Improving treatment of late life depression in primary care: a randomised clinical trial. J Am Geriatr Soc 42(8): 839–846

Cole M (1983) Age, age of onset and course of primary depressive illness in the elderly. Can J Psychiatry 28: 102–104

*Cole M (1985) The course of elderly depressed outpatients. Can J Psychiatry 30: 217–220

**Cole M, Bellavance F (1997) The prognosis of depression in old age. Am J Geriatr Psychiatry 5: 4–14

Conwell Y, Caine E (1991) Rational suicide and the right to die: reality and myth. N Engl J Med 325: 1100–1103

**Copeland J, Dewey M, Wood N, Searle R, Davidson I, McWilliam C (1987) Range of mental illness amongst the elderly in the community. Prevalence in Liverpool using the GMS-AGECAT package. Br J Psychiatry 150: 815–823

**Copeland J, Davidson I, Dewey M, Gilmore C, Larkin B, McWilliam C, Saunders P, Scott A, Sharma V, Sullivan C (1992) Alzheimer's disease, other dementias, depression and pseudodementia: prevalence, incidence and three-year outcome in Liverpool. Br J Psychiatry 161: 230–239

Copeland J, Beekman A, Dewey M et al (1999) Depression in Europe. Geographical distribution among older people. Br J Psychiatry 174: 312–321

Dooneief G, Mirabello E, Bell K et al (1992) An estimate of the incidence of depression in idiopathic Parkinson's disease. Arch Neurol 49: 305–307

*Emery V, Oxman T (1992) Update on dementia spectrum of depression. Am J Psychiatry 149(3): 305–317

Emmerson JP, Burvill P, Finlay-Jones R, Hall W (1989) Life events, life difficulties and confiding relationships in the depressed elderly. Br J Psychiatry 155: 787–792

Freud S (1924) On psychotherapy. Hogarth, London

Gaitz C, Scott J (1972) Age and the measurement of mental health. J Health Soc Behav 13: 55–67

*Gallagher-Thompson D, Hanley-Peterson P, Thompson LW (1990) Maintenance of gains versus relapse following brief psychotherapy for depression. J Consult Clin Psychol 58(3): 371–374

Gask L (1992) Training general practitioners to detect and manage emotional disorders. Int Rev Psychiatry 4: 293–300

Greenwald BS, Kramer-Ginsberg E, Krishnan RR, Ashtari M, Aupperle PM, Patel M (1996) MRI signal hyperintensities in geriatric depression. Am J Psychiatry 153(9): 1212–1215

*Gurland B, Copeland J, Kuriansky J et al (1983) The mind and mood of aging. Croom Helm, London

Gurland B, Golden R, Teresi J, Challop J (1984) The SHORT-CARE: an efficient instrument for the assessment of depression, dementia and disability. J Gerontol 39: 166–169

Heuser IJ, Schweiger U, Gotthardt U, Schmider J, Lammers CH, Dettling M, Yassouridis A, Holsboer F (1996) Pituitary-adrenal-system regulation and psycho-pathology during amitriptyline treatment in elderly depressed patients and normal comparison subjects. Am J Psychiat 153(1): 93–99

Iliffe S, Haines A, Gallivan S, Booroff A, Goldenberg E, Morgan P (1991) Assessment of elderly people in general practice. 1. Social circumstances and mental state. Br J Gen Pract 41: 9–12

Jacoby R (1998) Commentary on a review: heterocyclic antidepressants and rational psychological therapies are effective in older patients with mild to moderate depression. Evidence-Based Mental Health 1(3): 77

Jacoby R, Bird JM (1981) Computer tomography and the outcome of affective disorder: a follow-up study of elderly patients. Br J Psychiatry 143: 124–127

*Jacoby R, Dolan R, Levy R, Baldy R (1983) Quantitative computed tomography in elderly depressed patients. Br J Psychiatry 143: 124–127

Jorm A, Van Duijn C, Chandra V et al (1991) Psychiatric history and related exposures as risk factors for Alzheimer's disease: a collaborative re-analysis of case-control studies. Int J Epidemiol 20 [Suppl 2]: 43–47

*Katona C (1994) Depression in old age. Wiley, Chichester

Katona C, Freeling P, Hinchcliffe K, Blanchard M, Wright A (1995) Recognition and management of depression in late life in general practice: consensus statement. Prim Care Psychiatry 1: 107–113

Kennedy G, Kelman H, Thomas C et al. (1989) Hierarchy of characteristics associated with depressive symptoms in an urban elderly sample. Am J Psychiatry 146: 220–225

Kennedy G, Kelman H, Thomas C (1990) Persistence and remission of depressive symptoms in late life. Am J Psychiatry 148: 174–178

Kivela S-L, Pahkala K (1989) The prognosis of depression in old age. Int Psychogeriatr 1: 119–133

*Koenig HG, Meador KG, Cohen HJ, Blazer D (1988) Depression in elderly hospitalised patients with medical illness. Arch Intern Med 148: 1929–1936

Koenig HG, Cohen HJ, Blazer DG, Krishnan KR, Sibert TE (1993) Profile of depressive symptoms in younger and older medical inpatients with major depression. J Am Geriatr Soc 41(11): 1169–1176

Kramer-Ginsberg E, Greenwald BS, Aisen PS, Brod-Miller C (1989) Hypochondriasis in the elderly depressed. J Am Geriatr Soc 37: 507–510

Kukull WA, Koepsell TD, Invit S et al (1986) Depressive and physical illness among elderly general medical clinic patients. J Affect Dis 10: 153–162

*Livingston G, Thomas A, Graham N, Blizard R, Mann A (1990) The Gospel Oak Project: the use of health and social services by dependent elderly people in the community. Health Trends 2: 70–73

Maier W, Lichtermann D, Minges J, Heun R, Hallmayer J, Klingler T (1991) Unipolar depression in the aged: determinants of familial aggregation. J Affect Dis 23: 53–61

*McCusker J, Cole M, Keller E, Bellavance F, Berard A (1998) Effectiveness of treatments of depression in older ambulatory patients. Arch Intern Med 13(158): 705–712

Moscicki E (1995) Epidemiology of suicide. Int Psychogeriatr 7(2): 137–148

Mossey JM, Knott KA, Higgins M, Talerico K (1996) Effectiveness of a psychosocial intervention, interpersonal counseling, for subdysthymic depression in medically ill elderly. J Gerontol A Biol Sci Med Sci 51(4): M172–178

**Murphy E (1982) Social origins of depression in old age. Br J Psychiatry 141: 135–142

*Murphy E (1983) The prognosis of depression in old age. Br J Psychiatry 142: 111–119

Musetti L, Perugi G, Soriani A, Rossi V, Cassano G, Akiskal H (1989) Depression before and after age 65: a reexamination. Br J Psychiatry 155: 330–336

Nemeroff CB, Knight DL, Krishnan RR, Slotkin TA, Bissette G, Melville ML et al (1988) Marked reduction in the platelet-tritiated imipramine binding sites in geriatric depression. Arch Gen Psychiatr 45: 919–923

O'Brien J, Desmond P, Ames D et al (1994) The differentiation of depression from dementia by temporal lobe magnetic resonance imaging. Psychol Med 24: 633–640

*O'Brien J, Ames D, Schwietzer I (1996) White matter changes in depression and Alzheimer's disease: a review of magnetic resonance imaging studies. Int J Geriatr Psychiatry 11: 681–694

Ong YL, Martineau F, Lloyd C, Robbins I (1987) Support group for the depressed elderly. Int J Geriatr Psychiatry 2: 119–123

Oxman TE, Barrett JE, Barrett J, Gerber P (1990) Symptomatology of late-life minor depression among primary care patients. Psychosomatics 31: 174–180

*Oxman TE, Berkman LF, Kasl S, Freeman DH Jr, Barrett J (1992) Social support and depressive symptoms in the elderly. Am J Epidemiol 135: 356–368

Pahkala K (1990) Social and environmental factors and depression in old age. Int J Geriatr Psychiatry 5: 99–113

*Prince MJ, Harwood RH, Blizard RA, Thomas A, Mann AH (1997) Impairment, disability and handicap as risk factors for depression in old age. The Gospel Oak Project V. Psychol Med 27(2): 311–321

Reynolds CF, Hoch CC, Buysse DJ, Houck PR, Schlernitzauer M, Frank E, Mazumdar S, Kupfer DJ (1992) Electroencephalographic sleep in spousal bereavement-related depression of late life. Biol Psychiatr 31: 69–82

Robinson DS, Davis JM, Nies AA et al (1971) Relation of sex and aging to monoamine oxidase activity of human brain, plasma and platelets. Arch Gen Psychiatry 24: 536–539

*Rosen J, Rogers JC, Marin RS, Mulsant BH, Shahar A, Reynolds CF 3rd (1997) Control-relevant intervention in the treatment of minor and major depression in a long-term care facility. Am J Geriatr Psychol 5(3): 247–257

Sartorius N. (1995) Recent changes in suicide rates in selected eastern European and other European countries. Int Psychogeriatr 7(2): 301–308

Schneider L, Reynolds C, Lebowitz B, Friedhoff A (1994) Diagnosis and treatment of depression in late life. American Psychiatric Press, Washington DC

Schonfield L, Garcia J, Streuber P (1985) Factors contributing to mental health treatment of the elderly. J Appl Gerontol 4: 30–39

Shamash K, O'Connell K, Lowy M, Katona C (1992) Psychiatric morbidity and outcome in elderly patients undergoing emergency hip surgery. Int J Geriatr Psychiatry 7: 505–509

*Starkstein S, Robinson R (1989) Affective disorders and cerebrovascular disease. Br J Psychiatry 154: 170–182

Steuer J (1982) Psychotherapy with the elderly. Psychiatr Clin North Am 5: 199–213

Tappy L, Randin J, Schwed P et al (1987) Prevalence of thyroid disorders in psychogeriatric inpatients: a possible relationship of hypothyroidism with neurotic depression but not with dementia. J Am Geriatr Soc 35: 526–531

Waterreus A, Mann A, Blanchard M, Aquilina C (1997) Out of the darkness: a training manual in the recognition and management of old age depression for primary care workers. Institute of Psychiatry, London

*Weissman MM, Myers JK, Tischler GL et al (1985) Psychiatric disorders (DSM-III) and cognitive impairment among the

elderly in a US urban community. Acta Psychiat Scand 71: 366–379

*Woods RT (1992) Psychological therapies and their efficacy. Rev Clin Gerontol 2: 171–183

*World Health Organization (1996) Psychiatry of the elderly: a consensus statement. World Health Organization, Geneva

Wragg R, Jeste D (1989) Overview of depression and psychosis in Alzheimer's disease. Am J Psychiatry 146: 577–586

Yesavage J (1993) Differential diagnosis between depression and dementia. Am J Med 94 [Suppl 5A]: 23–28

Zung WWK, Magill M, Moore JT et al (1983) Recognition and treatment of depression in a family medicine practice. J Clin Psychiatry 44: 3–9

CHAPTER

11

J. Wertheimer

Psychiatric Treatment and Rehabilitation of the Elderly Mentally Ill

1 Therapeutic and Rehabilitation Strategies

Psychiatric disorders associated with old age encompass both pathologies which began in earlier periods of life and those arising after 65 years of age. One of their principal characteristics is that they frequently develop within complex contexts where physical and psychiatric co-morbidity is frequently encountered and where interpersonal, social, environmental and economic factors are much in evidence. While a clinical approach reveals signs and symptoms allowing us to arrive at a diagnosis, an in-depth analysis reveals the extent to which a particular patient's situation is unique, defined as it is by a multiplicity of parameters. A therapeutic approach must take this data into account.

An investigation into the components of an elderly person's psychopathological situation and the drawing up of a suitable treatment follow two phases – analysis and synthesis. The actual process is rather more complicated than that. In practice, through the relationship established with the patient, a doctor makes a number of observations which he or she then combines in such a way as to progressively create a coherent overall image of the patient's situation, allowing a therapeutic plan of action to be established. It is not enough therefore to simply submit a patient to a whole battery of scales, collecting information on symptoms, signs, functional disorders, degree of dependence, level of family responsibility etc. The juxtaposition of this sort of information merely serves to conceal the qualitative subtlety inherent in their interdependence as well as the fact that they emanate from an individual human being. In short, synthesis and analysis must be contemporaneous.

An analysis based on case history, an examination, subsequent observation and further examinations will result in a classification of psychiatric and somatic entities and a description of deficits and their consequences on everyday life in terms of disabilities, handicaps and functional disorders. An analysis will further elucidate the relational dimension, notably the patient's capabilities in areas of human contact. This implies both instruments of perception (sight, hearing) and cognition (e.g. attention, memory, language) as well as personality traits and the human network comprising family, friends and medical contacts. Finally, an analysis will supply information on the practicalities of the patient's existence – the location and layout of his or her home, the degree of isolation, economic resources and use of care and support organizations and of the social services.

In view of such complexity, one speaks of a therapeutic strategy rather than a treatment. This presupposes a structuring of the various procedures employed in achieving a common end result. The term "strategy" also implies a time element. This time element is intimately integrated into the provision of medical care, a term which takes on broader dimensions allying clinical and therapeutic observation, analysis and synthesis, and fact and intuition.

Therapeutic strategy in geriatric psychiatry provides a good illustration of medical complexity where it is not only the illness that is in question, but the patient him- or herself. The older the patient, the more numerous the parameters implicated in defining his or her pathological situation. Rare are the cases where the simple treatment of a disease such as depression or delirium will result in the return of good health. In the majority of cases, doctors finds themselves in a system where causality is not linear but circular. For instance, delirium caused by an infection will increase if the fever-induced dehydration is not corrected. Consequences are transformed into additional causes.

This complexity is also related to the diversity of types of treatment employed. Simultaneous use is frequently made of biological means, relational techniques and measures aimed at rehabilitation, the latter producing an associated psychological impact. Finally, a therapeutic strategy needs to be adapted to each individual patient depending on the multiplicity of criteria which constitute his or her psychiatric and somatic illnesses and the behavioural pattern resulting from the disorders or influenced by character and motivation.

Any such therapeutic strategy must, therefore, encompass a logical coordination of various types of treatment as a means to an end which is unique in each individual case. This strategy could be that of rehabilitation defined as a number of medical, psychiatric, psychological and social measures coordinated into a joint programme whose aim is the restitution or conservation of optimum autonomy. This definition includes the aspect of complexity, the necessity for flexibility between therapies and the need for a realistic outcome, the aim of which is not a return to normality (which would be largely utopian in this age category), but to a reasonably adapted degree of autonomy. Finally, this strategy must be sufficiently supple to allow adaptation to individual problems and to the degree of evolution existing at the time of rehabilitation.

Elderly and very elderly individuals possess certain characteristics which indicate a degree of adaptation of normal rehabilitative measures. The tendency towards psychomotor slowing dictates the use of slower rhythms of activity than those used with young adults. The goal is rarely that of return to some professional activity. Hence, a rehabilitation treatment aims at

better adaptation to normal day-to-day activities and better social integration, including an assertion of the individual role. The problem of patients' motivation in cooperating in what is being done for them is particularly crucial. An apparent absence of such motivation is, however, no reason for giving up. It can be simply a manifestation of a disorder such as depression. The strategy then needs to be adapted to include a progressive resurgence of motivation. In such cases, rehabilitation will tend to channel its activities towards creating adaptation within the framework of daily life. Consequently, rehabilitation measures would be modulated so as to correspond to the nature of the disorder, its intensity and the patient's potential ability to cooperate.

The techniques employed in any therapeutic strategy may be divided into four categories (Table 1). The first of these comprises types of biological treatment which aim at pathological processes and are based on knowledge in the domains of the neurosciences, particularly neuropathology, neurobiology and neuropharmacology. The second concerns what patients, their friends and family and their carergivers are living through. It requires structured psychotherapy, psychological support or counselling. The third category brings together a group of techniques which tend to reproduce situations aimed at re-establishing skills, e.g. training in daily tasks of living, occupational therapy or social skills training. The final category of techniques aims at adapting the environment to disabilities and handicaps. It is of particular help to patients in teaching them to cope with their habitat and its space, and household and leisure equipment, while at the same time increasing safety. This type of passive adjustment encourages a greater margin of autonomy.

Table 1. Treatments in geriatric psychiatry

Biological treatments
General medication
Psychotropic medication
Medication for dementia
Electroconvulsive therapy
Light therapy
Psychotherapeutic treatments
Psychotherapeutic attitude
Individual psychotherapy
Group psychotherapy
Psychotherapeutic approach to dementia
Psychological support to the family
Psychological support to the carers
Re-adaptive treatments
Corporal approach
Occupational therapy
Sociotherapy
Music therapy
Memory re-adaptation
Adaptation to the environment
Environment acceptance
Telethesis
Gerontechnology

2 Biological Treatments

2.1 Medication

In general, the prescription of medicines to the elderly should be limited to substances that are absolutely essential. The high frequency of polymorbidity results in patients being exposed to multiple prescriptions with the concomitant risks of drug interaction and poor compliance.

Pharmacokinetics change with age. There is a slowing down of absorption, transfer, metabolism, distribution and excretion of medicines. Such changes can have an important effect on the plasma levels of psychotropic substances such as tricyclic anti-depressants. The resorption of these substances can be slowed down by their own effect on intestinal peristalsis, with their transfer by albumin being reduced by 15%–20%. This results in an increase in the free fraction of the substances, with their half-life being increased – due to their accumulation in the fatty tissues of the organism – from 10% at 20 years of age to 50% at 60. To that can be added a possible reduction in enzyme breakdown by the liver and a slowing down in the rate of elimination due to a reduction in the glomerular filtration rate, which falls by 50% between 20 and 70 years of age.

Compliance is influenced by numerous factors, particularly in elderly patients suffering from mental disorders. These factors are of three main orders. The first of these concerns the patients themselves – their cognitive ability, particularly memory, the ability to understand the objective pursued by the treatment and the ability to organize taking medication. The necessary motivation to take care of themselves is also implied here. This can be severely compromised where patients' awareness of their illness is low or non-existent, as can happen in cases of dementia and psychotic disorders or where a severe depressive disorder results in notions of incurability. The second factor influencing compliance is the doctor, who needs to reinforce his or her prescription by clear explanations concerning the reason for the choice of medication, the times when the medicines should be taken, their mode of action, the risk of side-effects and what attitude to adopt should they appear. Ideally, an elderly

patient should not be required to take more than three medicines at one time. The third factor comprises the medicines themselves: their size, taste and any undesirable effects. The doctor must take care not to complicate an already complex treatment by multiplying the galenic forms of the prescribed medication.

2.1.1 General Medication

Co-morbidity results in important reciprocal influences between psychic and somatic disorders. Some physical illnesses can lead to mental disorders, particularly in the elderly. Delirium is an example of a disorder which can result from an infectious disease (e.g. bronchopneumonia), a metabolic disturbance (e.g. dehydration, hyper- or hypoglycaemia) or medication (tricyclic anti-depressants, antiparkinsonian drugs). Another example, hypothyroidism, can be the cause of a depressive disorder. Conversely, mental disorders can have somatic consequences, particularly of a general nature, due to alterations in alimentary behaviour – negligence in the variety of food ingested by a demented person or lack of appetite in a depressive. On the other hand, the unexpected arrival of a physical illness concomitant with a mental disorder may aggravate a state of dependence and accentuate the psychiatric symptoms in the form of a regression, for instance. The diagnosis and treatment of physical illnesses are, therefore, indispensable in the context of geriatric psychiatry.

2.1.2 Psychotropic Medication

The whole range of psychotropic medication – neuroleptics, anti-depressants and tranquilizers – are accessible to the elderly. However, some precautions need to be respected because of the risk of side-effects. Most neuroleptics can trigger extrapyramidal syndromes (EPS) or increase the intensity of pre-existing EPS. Tricyclic and tetracyclic anti-depressants (TCA) have atropinic effects and are contra-indicated in cases of closed-angle glaucoma, cardiac conduction disorders and prostatism. Benzodiazepine tranquilizers (BZD), because of their dual sedative and muscle-relaxant action, can lead to patients falling down. These substances, which are among the most commonly taken by the elderly, can also lead to memory impairments. This assertion must, however, be treated with reserve, since the anxiolytic effect inherent in these drugs can also lead to beneficial consequences on memory functions by removing the anxiety which inhibits them. Of the three groups of psychotropic medication mentioned so far, only BZD are likely to bring about dependence.

A corollary of the high side-effect potential of psychotropic drugs is the on-going research into the development of less toxic substances. Unfortunately, powerful but harmless medication is not yet available to us. However, considerable progress in this direction has been made over the last 10 years. Among neuroleptics, clozapine has established itself as an anti-psychotic without EPS but with a risk of agranulocytosis. To the range of anti-depressants drugs have been added that are totally lacking in atropinic effects. These include a selective monoamine oxidase inhibitor (MAOI-A) called moclobemide, serotonin reuptake inhibitors (SSRI), such as fluoxetine, fluvoxamine, citalopram and sertraline, and serotonin and noradrenaline reuptake inhibitors (SNRI) such as venlafaxine. However, SSRI and SNRI produce side-effects in the digestive system.

It must be said that clinical pharmacological investigations involving groups of elderly persons are few and far between. Those that exist usually concern subjects of less than 70 years of age and conform to strict exclusion criteria. However, the clinical reality is such that the majority of candidates for psychotropic medication suffer from polymorbidity and a high percentage of them are very elderly. Their introduction to psychotropic substances must therefore be largely empirical. However, a certain number of guiding principles should be followed. In general, the initial dose should be a third of the normal adult dose. This will be adapted progressively as a function of tolerance. Where intercurrent somatic diseases occur, the dosage must again be modulated, particularly where neuroleptics have been administered to the mentally ill to calm their agitation. Preference should be given to substances whose side-effects are compatible with the patient's physical state. The administration of neuroleptics and anti-depressants can be monitored by regular measurements of serum level.

Psychotic Disorder

Stabilization of chronic psychotic illnesses among the elderly is usually achieved using strong doses of neuroleptics – identical to those of a young adult – either orally or by depot injection. Tardive delusions, the themes of which are usually theft and intrusion by third parties, only necessitate medication where great behavioural changes occur and where anxiety is present. In such cases, suitably adapted doses of sedative neuroleptics (chlorprothixene, thioridazine, clozapine) would be used where anxiety predominates, and incisive neuroleptics (e.g. haloperidol, moperone, pipotiazine) where it is absent.

Depressive Disorder

The administration of anti-depressants is imperative in the presence of depressive disorders. First-intention medication is likely to consist of SSRI, SNRI, selective MAOI or mianserine. SSRI share both the same method of action and the same therapeutic impact so

that they are frequently difficult to distinguish from one another. This is not the case for tricyclics and tetracyclics, of which some (imipramine, nortryptiline) tend to have a stimulating effect, whereas others (amitryptiline, maprotiline) have a tranquilizing effect. Despite their atropinic effects, these drugs should not be prescribed. They may be employed as a second resort or as a first resort where they have already proved their effectiveness in previous attacks, assuming there to be no contra-indication. Mianserine is indicated for depression disorder with anxiety. Where severe depression is accompanied by delusions, an association of an anti-depressant and an incisive neuroleptic is indicated.

In cases of resistant depression (absence of therapeutic response after 3 months and at least two changes in medication), an association with lithium (0.5–0.7 mmol/l) can be tried and, if that has produced no result after 2 weeks, electroconvulsive therapy (ECT) should be envisaged. Thymoregulators (lithium, valproic acid, carbamazepine) are still indicated in cases of bipolar disorders. A maintenance dose of anti-depressant should be continued for at least 2 years following remission of recurring depression.

Depressive disorder can be associated with mild, moderate or even severe cases of dementia. The prescription of SSRI is particularly indicated in cases of dementia associated with Alzheimer's disease where a serotonin deficit exists. Moclobemide may also be indicated here, since this type of dementia is accompanied by an increase in MAO.

Hypomania, regardless of whether it is episodic or has a durable influence on dementia, reacts well to haloperidol.

Delirium

As far as possible, treatment of delirium should only be directed at the cause of the illness: infectious disease, metabolic disorder or stopping the involved medication. Treatment can also be helped enormously by placing patients in reassuring surroundings with particular emphasis on helping them to orientate themselves correctly with respect to time, place and situation. However, a high intensity of agitation may necessitate treatment using medication. In such a case, oral administration of low-toxicity clomethiazole may be recommended.

Anxiety Disorder

Anxiety disorder is often associated with depressive disorder. Panic disorder is less common and may be treated, as for young adults, with SSRI. Medication used to treat anxiety is above all the benzodiazepine group (BZDs), which may be used, where necessary, in association with an anti-depressant. Depending on the clinical situation, BZD can be prescribed with medium-to-long-term action (bromazepam, lorazepam, alprazolam, oxazepam) or very long term action (prozepam, ketazolam, clobazam). Molecules with medium-to-long term action are used when symptoms require it, whereas those with very long term action cover states of permanent anxiety. It is important to limit the period of administration as much as possible and reduce the dosage progressively. Alternatives to BZDs are the sedative neuroleptics (thioridazine, chlorprothixene, levomepromazine) administered in small doses. These have the advantage of possessing few side-effects at such doses and do not induce dependence. Another possibility is buspirone, the anxiolytic effect of which is not optimal until after 7–10 days.

Sleep Disorder

Insomnia in the elderly is very often either the symptom of a physical or psychiatric illness or the manifestation of a psychological problem. Any prescription of sleep-inducing medication should be preceded by an investigation into the causes of the problem, causes on which it should be possible to act, e.g. cardiac insufficiency treatment, pain-killing. On the other hand, it is important to bear in mind the physiological changes as regards sleep that take place with age, notably a lengthening of the latency period for sleep to come after 70 years of age and an increase in the number of wakenings during the night.

Where a sleep disorder is associated with a depressive disorder, the administration of a tricyclic anti-depressant (amitryptiline, trimipramine) just prior to the patient's bedtime may suffice. Where recourse to a hypnotic is judged necessary, preference should be given to substances with which the risk of weaning and insomnia rebound symptoms is small (chloral hydrate, zolpidem, zoplicone). Occasionally, it is impossible to avoid administering BZD. In such cases, a choice has to be made between BZD with short- to very short term action (midazolam, triazolam) where the problem is that of experiencing difficulties in getting to sleep, or with medium-term (lormetazepam) or very long term action (flunitrazepam) where the problem is one of disturbed sleep or premature wakening. It is also possible to associate zolpidem, zopiclone or chloral hydrate with a retard form of neuroleptic (30 mg thioridazine retard). The administration of hypnotics should be limited in time. However, it may be necessary to allow some patients who have grown accustomed to taking a sleeping pill at night to continue to do so indefinitely.

2.2 Electroconvulsive Therapy

ECT is a treatment which comes up against prejudices that are of a strictly emotional nature. Yet it remains a

highly effective and often last-ditch means of treating serious psychiatric situations in the elderly. Indications concern severe depressive disorders with agitation and/or delirious ideas of ruin, blame or incurability. The sudden appearance of such a clinical picture can indicate an ECT as a first-line treatment. This treatment is used more frequently for resistant depressions and where the severity of the depression is such that the patient's life may be in danger. Less frequent indications are manias resistant to neuroleptics with accompanying lithium intolerance and depression associated with Parkinson's disease.

There are few contra-indications other than space-occupying lesions, although cardiac conduction disorder and recent myocardial infarction require caution. It should not be forgotten that a spontaneous epileptic seizure rarely has any serious effect on general health. Where ECT sessions are conducted at a rhythm of two per week for 3 or 4 weeks, the only complication is likely to be narcosis. Reversible confusional states and amnesia limited to the circumstances of a particular treatment episode are occasionally observed. Where dementia is associated with depressive disorder, unilateral ECT may be used. This implies less risk of confusion and memory disorder (Wilkinson 1994).

2.3 Light Therapy

Light therapy is indicated in cases of seasonal depression, even beyond 65 years of age. Such depressions tend to occur towards wintertime when there is less sunshine. They are frequently accompanied by an over-consumption of sweet things.

3 Psychotherapeutic Treatments

The patient–doctor relationship has in itself psychotherapeutic potential. During the first contact with a doctor, the patient often talks of physical complaints which may well be an initial movement towards tackling the themes of psychical and existential suffering. The patient's interlocutor must thus adopt an attitude of listening and observation, hence manifesting availability. At the same time, this attitude demonstrates the doctor's recognition of the patient as an individual.

Like all other health care personnel, doctors must be aware of their potential role as a mirror in which patients who have lost their self-esteem may rebuild it again. Hence psychotherapeutic input is not the sole business of specialists. It is implicit in all exchanges between carers and patients. This dimension must be taught at the same time as the characteristics inherent in each of the three periods of life, including old age, within a developmental perspective of the individual from birth to death. Indeed, the issue is not only to recognize and accept the losses accompanying old age, but also to restore a coherent, integral vision of the self, without a break and without abandoning what has become integrated into a whole life throughout its life span.

In principle, all types of psychotherapy applicable to adults can be used with the elderly, whose population is in fact very heterogeneous. It must be realized that, due to their cognitive and affective resources, some elderly individuals are able to undergo classical formalized psychotherapies implying a profound restructuring within themselves. This is often true of patients who have experienced such approaches during their active life and would like to continue with or reactivate them. However, many elderly patients requiring psychotherapeutic help are impaired in their perception (sight, hearing), in their capacity to move around or in their cognitive faculties. Such limitations must in no way be used as arguments for renouncing psychological help and abandoning patients in disarray.

Several problems lie behind the more or less conscious reticence of the psychotherapist to treat elderly patients. One of them is historical in nature. Despite the fact that he continued his self-analysis up to his death at the age of 83, Freud maintained that beyond the age of 40 psychic rigidity prevented any reappraisal of an individual. This prejudice – which is no longer accepted – can nevertheless, within the framework of an orthodox position, serve as an alibi either to abandon any attempt at treatment or remain on a superficial level even where exploration at greater depth is eminently possible. Another obstacle is the difference in age and generation between the younger therapist and his or her patient. The ambiguity here lies in the fact that the patient's life experience is likely to be much broader than the therapist's and that the patient is in a phase of existence way outside the latter's experience. In addition, the transferential relationship is itself the object of a further ambiguity since the therapist may well be identified successively or even simultaneously with a paternal or maternal image and a filial one. Another problem is that of time. Is it worthwhile undertaking steps the hypothetical benefit of which is limited by the approach of death? A corollary to this question is, for the therapist, the inevitable confrontation with the certitude of the fatal issue of existence. One final point to add is that the elderly patient can be experienced as an object of conflictual identification to the extent that, for the therapist, he or she may personify parents with which

relationships are difficult. This ensemble of questions contributes to the situation where psychotherapeutic help to the elderly is an area in which too many psychiatrists and psychotherapeutic psychologists refuse to become involved – a curious paradox in the context of our ageing society.

The needs are indeed enormous, given that the elderly are particularly exposed to situations of loss during a period in their existence where their capacity for adaptation is reduced. Some contribution of a psychotherapeutic nature ought to be available to everyone. Such a proposition is only possible from a pragmatic viewpoint where basic notions have been communicated to all those concerned so that, through indirectly relating various therapeutic techniques, contributions of a psychotherapeutic nature can be transmitted. The trained psychotherapist would intervene in particular situations and act as referent for the medical staff.

3.1 The Psychotherapeutic Attitude

The basic attitude to adopt conforms with general principles. The first of these is to consider elderly individuals, whether or not they are ill, as unique individuals with their own history and their own experience consisting of successes, failures and suffering. The second is to be available, i.e. attentive to the patient's expressed or implicit needs, and to be open to dialogue. The third is to maintain a suitable distance so as to avoid any harmful identification with the patient and to allow the necessary free reflection in order to arrive at an appropriate course of action. The fourth is to establish and maintain contact, whatever the sensory, cognitive, emotional or behavioural obstacles. The fifth is to appear consistent in one's way of behaving and communicating. The aim of these few rules is to encourage patients to develop a positive attitude towards themselves and to minimize the likelihood of the onset of depressive self-depreciation.

3.2 Types of Psychotherapy

The complex nature of the problems encountered necessitates flexibility in the classical psychotherapeutic concepts which advocate neutrality. Once again, even if the strict application of this principle should not be excluded, with some elderly patients these concepts are more likely to be laid aside in favour of joint therapies consisting of a mixture of psychodynamic interpretation, advice, concrete medico-social interventions and the prescription of medication. This flexibility also needs to be extended to the setting. A consultation should be face-to-face. It can take place on the therapist's premises or at the home of the patient, particularly if he or she lives in an institution. Consultations last between 20 and 45 min. The frequency of sessions depends on the severity of the symptoms and can vary between once or twice a week to once a month or less. When regular meetings are no longer necessary, the therapist remains available as a potential stand-by. This provides reassurance and a structuring point of reference. Consultations at regular intervals, arranged on the patient's initiative, also act as reference points and possess undoubted preventative value. In this way, a psychotherapy normally does not come to an end, but remains an on-going available resource.

3.2.1 Individual Psychotherapy

There are three main trends in individual psychotherapy for the elderly. These are psychotherapy based on support, on psychoanalytical methods and of a cognitive–behavioural nature.

Support-Based Psychotherapy

Support-based psychotherapy is probably the most frequently employed method of individual psychotherapy. It is particularly indicated where insight capacity is poor or when the patient is handicapped from the cognitive and sensory points of view. The mere fact of being listened to, implying recognition, tends to reassure patients. It helps and guides them in their decision-making, in approving whatever they undertake and in helping to provide whatever assistance they require. It is in short an approach aimed at recovery and tries to encourage elderly individuals to have recourse to their own resources.

Analysis-Based Psychotherapy

Analysis-based psychotherapy brings out a patient's faculties for introspection. The value of this model lies not only in the treatment it gives rise to but, also in its role as a reference point in the understanding of a given situation. The emphasis is laid on unconscious defence mechanisms, particularly those dealing with loss of faculties and fears accompanying old age. This period of very great vulnerability results in anxiety which may well be linked with similar previous experiences, particularly during childhood or adolescence, both phases of life characterized by fragility. Whereas support-based psychotherapy focuses on the present time, analysis-based psychotherapy aims at establishing links between the present experience and significant facts of past life. This method of approach is particularly indicated in depressive disorders.

Cognitive-Behavioural Psychotherapy

The same is true of cognitive-behavioural psychotherapy, which focuses on what the depressed patient is experiencing in his or her thoughts. This obviously assumes a capacity for introspection on the part of the patient. These thoughts are invaded by cognitive distortions in the form of exaggerations, of disproportionate reactions, of false interpretations - all impregnated with pessimism and feelings of guilt, indignation and incompetence. This realistic, pragmatic method of approach comprises four components: learning to identify negative thoughts, establishing a link between these negative thoughts and depressive feelings, learning to identify and modify the types of deformed thoughts which govern the negative feelings and incorporating this procedure in the development of a point of view which is less deformed and more adapted to oneself and to one's place in the world of today and the future (Meador and Davis 1996).

3.2.2 Group Therapy

Individual psychological problems often feature a social dimension in the broadest sense possible. This may have a negative influence on the individual's integration into society or into his or her own group structures or it may result in a loss of equilibrium within the immediate entourage, namely family and medical staff. Group therapy therefore has several potential objectives, including the reinforcement of self-esteem through interaction with others, better self-knowledge through common introspective procedures, giving value to oneself, one's decisions, one's actions and abilities through confrontation with others, mutual psychological support, exchange of experiences, exercises in sociability, adjustment in the relationships within families and couples, particularly in crisis situations, and the psychological support of medical teams.

This multiplicity of objectives lies at the origin of the large number of techniques employed. Hence one can distinguish groups centred on current psychological problems, on memory recall, on facts of day-to-day life and on interpersonal relationships - both those with a psychoanalytical orientation and those using group analysis or cognitive-behavioural therapy. To these must be added self-help groups for patients having the same pathology (e.g. Parkinson's disease) or for families of chronically ill (dementia) patients. Family intervention, which usually uses systematic concepts, is employed in critical situations when, for instance, a psychiatric pathology suddenly affects an elderly patient. Its objective is to help the family group to mobilize and harmonize its resources not only to solve the current problem but also to reinforce the group so as to be able to deal with possible similar situations in the future in an autonomous way.

4
Re-adaptive Treatments

Whereas medicine-based treatments are targeted on symptoms and psychotherapeutic approaches are essentially subjective, re-adaptive treatments attack functional disorders resulting from the illness. However, while this always remains the main objective of such techniques, they may also have an impact in the same areas as the two other categories of treatment.

4.1
The Corporal Approach

The corporal approach provides a good illustration of the potential diversity of impacts mentioned above. With the onset of old age, relationships with the body may change in so far as the capacity to perform is lowered and motor and sensory handicaps appear. As old age progresses, a retrogression in bodily functions may appear with a lowering of motor activity. Some organic cerebral problems - notably parietal lesions - are accompanied by disorders in the perception of the corporal image. Finally, the subjective perception of the body is intimately intermingled with psychiatric states such as depression, psychosis, somatoform disorders and hypochondria. The number of techniques adapted to these various situations is legion. One common element which should be more or less exploited is the relationship which invariably develops in this type of approach.

4.1.1 Physiotherapy

Physiotherapy used in geriatric psychiatry often assumes the double role of the re-adaptation of motor functions and as a means for psychotherapy. Without being verbalized, several subjective parameters are implicated, e.g. the degree of intrusion by third parties that can be tolerated or how to manage distance. A psychotherapeutic treatment also frequently helps in reinforcing self-awareness through the body and reactivating motor potential, all of which implies an enhancement of a patient's self-image. This lever is particularly useful in depressive states where verbal expression is inhibited.

4.1.2 Psychomotor Therapy

Psychomotor therapy aims at "improved corporal awareness (feelings, emotions, speech and thought)" (Bovier and Ramseier 1995). Observation of the motor function constitutes the first phase. This provides an image of the patient's "body style" and leads to the elaboration of a course of treatment acting on the main bodily functions – space/time motoricity, sensory impressions, feelings, movement coordination and visceral functions. This type of therapy applied to elderly subjects aims at encouraging them to adapt to their ageing body by accepting its limits and reorganizing their own capabilities. In addition, it reveals the patient's unique physical semiology and its link with psychiatric disorders and allows a remedial approach to be developed.

4.1.3 Relaxation Techniques

Relaxation techniques (Schultz's autogenic training or the psychotherapeutic relaxation of de Ajuriaguerra) or similar (Alexander's eutonia) may also be used with elderly patients. Their common factor is the use of bodily knowledge to increase self-knowledge.

4.1.4 Gymnastics

Gymnastics helps both to maintain physical fitness and to discover and practice the possibilities inherent in physical movement. In addition, being part of a group is valuable from a socialization point of view. Taking part in ball games in a sitting position introduces stimulation through emulation and competitiveness. This is particularly useful with patients whose psychiatric illness isolates them from contact with others.

4.1.5 Occupational Therapy

Occupational therapy employs a technique of confrontation between the patient and the outside world with a view to encouraging his or her adaptation to day-to-day existence. It can be used equally well at home or in a clinic, hospital or long-term institution. Depending on the particular objective in view, three types of occupational therapy may be defined – functional, psychiatric and dementia specific. The first of these is destined more specifically for re-adaptation of the upper limbs following traumatic, neurological or rheumatological damage. It employs supervised activities aimed at practising certain movements, e.g. weaving. In addition, it concentrates on an adaptation to a normal living environment by encouraging day-to-day activities and modifying this environment so as to accommodate handicapped persons.

Psychiatric occupational therapy is mainly concerned with relationships and creativity. In the course of various activities, patients establish contact with the occupational therapist and with other patients. In addition, through their own creative efforts, they learn to control the external world once more and re-establish a sense of values through the objects they have created.

Dementia-specific occupational therapy adapts activities to patients' abilities. It attempts to mobilize not only the abilities they use spontaneously but also potential, underemployed ones and those not employed at all. A potential pitfall is overstimulation, which may result in catastrophic reactions. These techniques may be applied both in individual and group situations. An important aspect of the therapy is observation with a view to recognizing areas where remedial care is necessary. A simple example is to arrange clothes in the order they should be put on for a patient suffering from dressing apraxia. Subsequent treatment would depend on the gravity of memory and attention-span disorders. The principle is to encourage patients to act rather than acting for them.

In the field of occupational therapy too, it is often difficult to separate what is functional and what is psychiatric re-adaptation in geriatric psychiatry. In reality, occupational therapy can usually be said to be of an integrated type, with the driving force being relationships and where the objectives of physical mobilization and discovery or re-development of an awareness of personal capabilities have an implicit psychotherapeutic impact.

4.2 Sociotherapy

Sociotherapy forms part of the broader concept of animation, i.e. the attempt to encourage some form of activity. In concrete terms, this is manifested by what is undertaken, be it at home or in a group environment, to maintain the natural rhythms of life alternating between activity and rest, effort and relaxation, action and passivity, constraint and freedom, wakefulness and sleep. Thus, in the particular case of an institutional situation, life must be modulated so as to integrate a multitude of rhythms, including influences due to habits, to what is planned and what is unforeseen and to daily, weekly or more distant events based on monthly, seasonal or annual rhythms.

The implicit aim is to encourage an awareness of time by creating points of reference in the present, past and future and developing memory and anticipation.

Sociotherapy is a technique which employs group dynamics. It is organized in a particular place and structures time by employing various activities such as stimulating interest in what is going on in the world, games and activities in areas such as physical exercise (e.g. gymnastics, walking), creativity and housekeeping (cooking). A patient has the possibility of influencing decisions. While the general aim remains the animation of the patient, sociotherapy also encourages self-awareness and self-esteem by virtue of the roles he or she assumes, the mobilization of his or her abilities and experience and the confrontation with others both as individuals and as a group.

4.3 Art Therapies

Art therapies are methods of approaching rehabilitation that employ the therapeutic potential of artistic activities. They operate simultaneously on cognitive faculties, affectivity and bodily perception.

Music is capable of mobilizing perception as much as expression and communication. It implies, through rhythm, an awareness of time and, through melody, the emotional sensitivity that is unique to each individual. Finally it stimulates memory from the points of view of recall, of affective recollection of the past and of procedural retention.

Music therapy calls on all of these aspects and endeavours to exploit them to the full in a carefully planned, rigorous and methodical fashion (Verdeau-Pailles 1995). These methods are based on listening or on improvisation and creativity. The chosen approach depends on the pathology and personality of the patient. A depressed patient, for instance, is likely to be more open to an adagio than an allegro. Music therapy is especially indicated where patients' ability to give verbal expression to their experiences is poor, notably in cases of psychosomatic disorders. Hence its field of action is very wide indeed. It includes dementia, in which music has proved to be a surprisingly effective tool capable of stimulating seriously ill patients to remember melodies and words of songs. Their participation in improvisation groups, using instruments of percussion or simple stringed instruments such as a dulcimer, provides them with a non-verbal means of expression and reawakens in them an awareness of rhythm, and hence time. Finally, music is used in dance, which can remobilize dormant rhythmic abilities often intact in demented patients.

Art therapy is fundamentally creative. Mainly employing drawing, painting and modelling, it provides a glimpse of the non-verbal area of a patient's inner life. As such, it reveals conflict as well as positive development. The created work or work in progress may also lend itself as a useful means to initiate a dialogue with considerable therapeutic potential. Even if the production goes no further than a few simple lines, as in the case of severe dementia, it has value as a manifestation of attention and intention. Art therapy may be practised as an end in itself or as an integrated part of occupational therapy or sociotherapy.

4.4 Memory Re-adaptation

Memory is a complex function comprising a number of subsystems said to be of the sensory, primary (short-term, work-associated), secondary (long-term), procedural (implicit) and prospective types. Normal elderly individuals employ memory-training techniques aimed at improving performance through the use of strategies which help recall things to mind. Use can be made of external means (diaries) or internal procedures, which may be verbal or employ visual imagery. Memory training may be carried out in groups ("memory workshops"), where emulation is obtained through the use of exercises based on different sensory registers. The advantage of this technique is that it adds a certain sociotherapeutic value. Memory complaints form part of an ensemble in which affectivity plays an important part.

4.5 Auxiliary Means and Adapting the Environment

Sensory disabilities produce negative consequences on an individual's control of his or her surroundings which may affect emotional susceptibility. Correcting or improving visual and auditive disorders using spectacles, crystalline lens implants or hearing aids is of primary importance. The employment of auxiliary means to help in walking (canes, crutches, walkers), instruction in the use of a wheelchair or the wearing of braces are indispensable methods to improve mobility and combat the psychological effects of a reduction in autonomy.

The environment in which patients live must also be adapted to their disabilities. This implies the modification of any obstacles which may occupy their living space and principally concerns the occupational therapist. To these modifications based on individual requirements need to be added more general adaptations in the architectural and ergonomic arrangements

of the patient's environment. From the architectural point of view, hospitals and similar establishments must be designed taking the characteristic needs of elderly patients into account. Ideally, the surroundings should provide the best quality of life possible. Particular attention should be paid to access, volume of living space, lighting, safety and signposting to facilitate orientation. Ergonomic factors are the business of furniture and hospital equipment designers. Ergonomics is at the origin of the discipline called gerontechnology, which introduces technical progress into the gerontological domain and adapts it accordingly. Examples are teletheses ("electronic prostheses"), which allow some operations to be carried out at distance (e.g. door opening). Gerontechnology also looks into the adaptation of equipment used by elderly handicapped persons in their everyday life (e.g. household appliances, TV, radio) and tries to find technical solutions to problems of rehabilitation. In short, all these measures of adaptation aim at remedying the fact that normally constructed environments, means of communication and consumables are not designed to take the disabilities of handicapped persons into account. Such measures must be integrated into the basic principles of rehabilitation in geriatric psychiatry.

4.6 Rehabilitation Organization

Rehabilitation in geriatric psychiatry employs the skills of a wide number of professionals, notably doctors with a variety of specializations (e.g. psychiatrists, psychogeriatricians, geriatricians, neurologists, internists), psychologists, social workers, nursing staff, occupational therapists, sociotherapists, physiotherapists, psychomotoricians, speech therapists and so on. To this vast array of disciplines has to be added the complex nature of the network of care and support in which the patient exists. The obvious risk that may be run in such a situation is that of therapeutic anarchy if communication and information exchange is not well organized. The danger may be further increased if the same philosophy towards patient care is not shared by all concerned. Hence the organization of rehabilitation must be coherent, must create a framework for discussion and must transcend individual organizations, i.e. must encompass the whole network.

The path followed by a particular patient may pass through several stages such as out-patient services, day care or short- and long-term hospitalization. A consistent approach in both objectives and attitudes is essential throughout these various stages. In addition, the multi-disciplinary team must coordinate the interventions of its members as a function of a common, rehabilitative objective which may be, for instance, the patient's remaining at home. The achievement of this objective will depend on the achievement of the specific objectives fixed by each of the professions as a function of its technical competence. Hence the patient is the object of a multidisciplinary project, the methodology of which (Rufini and Gaillard 1996) could well be that presented in Fig. 1.

A philosophy of patient care with the objective of rehabilitation requires the multi-disciplinary team to establish an interpretation of the patient's situation

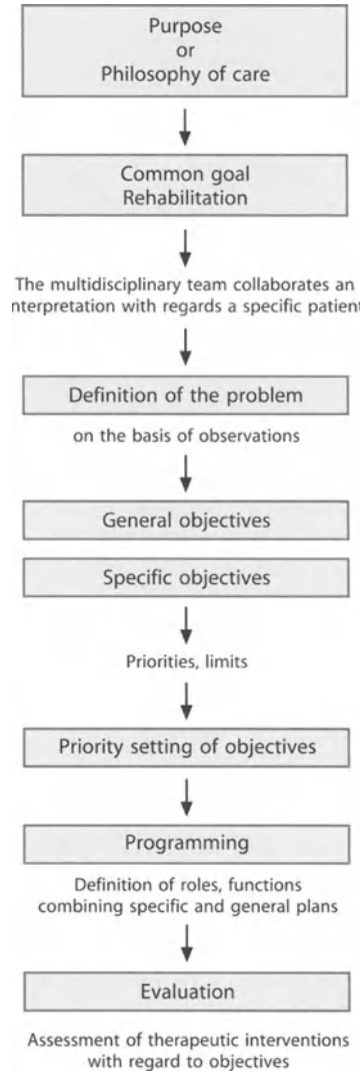


Fig 1. Possible methodology of multi-disciplinary treatment. (From Rufini and Gaillard 1996)

based on observation. It then proceeds to define the problems to be solved and the general and specific objectives to be pursued, taking priorities and limitations into account. This leads to a hierarchization of the objectives, to the drawing up of a programme and to the definition of roles. This whole process is subject to a series of evaluations which may modify the definition of problems to be solved.

This process implies on the one hand a shared involvement on the part of the different disciplines and on the other a clear definition for each of them. A mutual understanding of their individual competence allows the territory of each discipline to be well defined, hence avoiding the encroachment of one on the territory of another and ensuring overall treatment of the highest possible order.

5 References

Bovier P, Ramseier E (1995) Champs d'application de la thérapie psychomotrice. In: Richard J, Rubio L (eds) La thérapie psychomotrice. Masson, Paris, pp 151–166

Meador KG, Davis CD (1996) Psychotherapy. In: Busse EW, Blazer DG (eds) Textbook of geriatric psychiatry. American Psychiatric Press, Washington, pp 395–412

Rufini J, Gaillard M (eds) (1996) Pratique psychogériatrique. La genèse d'une équipe multidisciplinaire. L'Harmattan, Paris

Verdeau-Pailles J (1995) Musicothérapie. In: Richard J, Rubio L (eds) La thérapie psychomotrice. Masson, Paris, pp 117–125

Wilkinson DG (1994) Electroconvulsive therapy (ECT). In: Copeland JRM, Abou-Saleh MT, Blazer DG (eds) Principles and practice of geriatric psychiatry. Wiley, New York, pp 569–574

CHAPTER

12

T. Fuchs, H. Lauter

Psychiatric Aspects of the End of Life

Translator: E. Taub

1 Introduction

End-of-life psychiatry remains an inadequately developed branch of research and clinical practice in the field. This is despite the frequency not only of psychological problems, but also of mental disorders requiring treatment in the last stages of life. Terminally ill and dying patients rarely receive psychiatric treatment, and the majority of psychiatrists have less experience with the problems of treatment in this field than many palliative care physicians.

Psychiatrists are most often consulted with questions of the pharmacological or psychotherapeutic treatment of dying patients. Usually, however, this occurs only in individual cases; an American study, for example, revealed that psychiatric consultation was requested for only 1.9% of cancer patients in internal medical or surgical wards (Levine et al. 1978). Still more surprisingly, psychiatric advice is rarely requested even in cases involving medical decisions about assisted suicide and euthanasia on demand; a poll among Dutch psychiatrists revealed that they had been consulted in only 3% of such cases (Groenewoud et al. 1997).

A number of trends imply, however, that end-of-life psychiatry will become increasingly important in the future. The development of thanatopsychology, of palliative medicine, and of the hospice movement reveals an increasing awareness of the significance of the last phase of life. More and more questions and problems are being posed that require the interdisciplinary cooperation of oncologists, palliative care physicians, psychiatrists, and psychotherapists. Further, the increasing prevalence of psychiatric illnesses of organic origin, especially dementia, makes it the psychiatrist's responsibility to be able to answer the relevant therapeutic and ethical questions competently, including in the advanced stages of illness.

Lastly, problems of euthanasia have recently become the focus of international debate and are of special relevance to psychiatry. In several countries, psychiatric expertise has been utilized in the drafting of recommendations for regulations on active euthanasia or assisted suicide or in the formulation of regulations that are currently already in effect (Fuchs and Lauter 1997; Fuchs 1997; see also Vol. 1, Part 2, Chap. 18). In view of these developments, a consideration of the psychiatric aspects of the end of life seems more urgent now than ever. The present chapter deals first with the way patients come to terms with impending death and terminal illness and then with the psychiatrist's responsibilities in this situation; finally, an overview is provided of human experiences in the immediate proximity of death.

2 Significance of Death in Psychiatry

Death is hardly as remote a concern to psychiatry as it may seem. Psychiatrists may, indeed, have little to do with the medical, biological side of death, but they are intimately concerned with its existential dimension. The horizon of death, however distant, lends every moment of life a certain qualification and orientation; life contains the possibility of dying within itself. Mental suffering results when human beings, in the midst of their lives, are suddenly and disturbingly confronted with the alternatives of hope or despair, success or failure, meaningfulness or futility of existence. When talking with patients, a psychiatrist who has been sensitized to this dimension will readily perceive that they are not merely interested in the treatment of their illnesses, but also in achieving a positive final outcome for their lives as the end draws near. In suicide, the central form in which death manifests itself in psychiatry, the patient seeking death confirms its role as a last judgment on the value and success of life.

The theme of the finiteness of existence is also a visible or latent determinant of the central neurotic conflicts. It was no coincidence that psychoanalysis began with Breuer and Freud's *Studies on Hysteria* in 1895 (Breuer and Freud 1955), a study of patients (among them Anna O.) whose symptoms had developed immediately after an experience of death or loss. Freud himself certainly did not view the fear of death as the original source of anxiety and instead attempted to derive it from "castration anxiety." It was mainly Otto Rank and Melanie Klein who countered this theory with the concept of a fundamental fear of disintegration, loss of self, and annihilation, already present in earliest childhood (Rank 1931; Klein 1948). Thereafter, however, the psychoanalytic tradition largely ignored the problem of the awareness of death.

It was only recently that J.E. Meyer (1979) showed the great extent to which the fear of death contributes to the generation of many neurotic illnesses. The course of such "thanatophobic neuroses" as cardiac phobia, panic disorder, hypochondria, or obsessive-compulsive disorder is characterized by the gradual concealment of the originally manifest, elementary fear of death behind more tangible anxieties or psychosomatic complaints. Roth (1959) found that half of all patients with anxiety and depersonalization disorder had experiences of death or severe illness in the period immediately preceding onset. Skoog (1965) made the same finding in most patients with severe obsessive-compulsive disorder; Straus (1960) and von Gebsattel (1954) viewed the obsessive-compulsive patient's fear

of decay, illness, rottenness, and dirt as a latent fear of death, which the patient seeks to banish through rituals and obsessive thoughts. Individuals exposed to an extremely stressful situation and thereafter suffering from post-traumatic stress disorder often carry an awareness of the nearness of death with them throughout their lives ("death imprint"; Lifton 1979). Finally, many Holocaust victims nearing the end of their lives suffer a collapse of previously adequate coping and defense mechanisms, associated with a worsening of the mental suffering that is the lasting result of their earlier persecution (Lauter et al. 1999).

The importance of the awareness of death for psychopathology is particularly emphasized in existential psychotherapy, as Yalom recently wrote (1989, p. 179): "The attempt to escape the fear of death lies at the core of the neurotic conflict." Yalom views two forms of defense against the fear of death as central to the development of neurosis (Yalom 1989, pp. 143ff.): the belief in the "final rescuer," a person or mythical figure whose power and presence can save the individual from death, as long as this individual can successfully fuse with him or her; and the narcissistic belief in one's own specialness and invulnerability, which finds expression in the quest for independence and autonomy. The collapse of these forms of defense during life crises or in the immediate proximity of death expresses itself in many different psychopathological symptoms, but also gives the patient the chance to take a stand against the loneliness of death with appropriate therapeutic help and, in so doing, to gain in maturity (Yalom 1989, pp. 243ff.).

3
When Death Is Near: Stages of Dealing with Terminal Illness

The diagnosis of a disease that is expected to be fatal transforms the patient's general awareness of mortality to a sense of the immediate proximity of his or her own death. This is a radical change in the temporal perspective on life. The past, which, until now, could always be "left behind," suddenly assumes the nature of immutable fact: the course of life can no longer be fundamentally changed; it is now time to make up the accounts. The future, on the other hand, splits into two segments: the short time left to live, in which pressure is felt to make the most of life, and the anticipated later time "without me," and with it the questions, "What will remain of my life? What trace will I leave behind? In what way or ways will I live on?"

People today have much less effective defenses against this existential anxiety and distress than they once had. The prevailing doubt with regard to traditional notions of the afterlife has shrunk the period of time on which we can project our lives – even though the average life expectancy has increased. Life now seems to be the "last chance" (Gronemeyer 1993), and what we achieve or fail to achieve within its confines now assumes the character of the final and irrevocable.

The enormous abundance of thanatopsychological, popular scientific, and autobiographical literature on this subject in the last two decades bears witness to the intensive efforts that have been made to achieve mastery over the last phase of life, or at least to remove its terror. Not least, the works of E. Kübler-Ross (1974), reports of personal experiences such as those of Peter Noll (1984), Maxie Wander (1980), and Simone de Beauvoir (1965), and S.B. Nuland's demythologizing depiction of the biological and clinical process of dying (Nuland 1994) have contributed to the current, more realistic approach to the subject. The "phase model" proposed by Kübler-Ross has now given way to a more differentiated view of dying as a process that differs strongly among individuals depending on their personalities, their life histories, and the individuals around them. The following section includes a discussion of a number of typical attitudes and stages that combine, in highly varying ways, to shape the end-of-life process of each individual.

3.1
Initial Reactions

The diagnosis of a fatal illness is usually experienced as a "judgment" or death sentence; the immediate reaction is characterized, as a rule, by profound emotional distress that may rise to the level of a feeling of being annihilated and by massive anxiety and worry. This first reaction is often followed by defense processes, including an intensive effort to find a way out by trivializing or denying the illness. A few weeks usually pass before the reestablishment of an unstable equilibrium. The patient may avoid talking or thinking about the illness, while remaining aware of it; he or she may partly deny the illness by ignoring its fatal prognosis, or largely or totally repress it from consciousness – indeed, repression of symptoms by the patient may already have contributed to a delay in making the diagnosis. Switching back and forth between avoidance, partial or total denial, and acceptance may occur at any phase of disease and thereby create a problem for those around the patient.

In the past, the diagnosis and prognosis of fatal illnesses were often kept secret or conveyed only to the patient's family. Today, however, frank, caring, and

repeated conversations with the patient have become usual, so that there should be no unnecessary worry and anxiety caused by contradictory medical information or the "knowing silence" of others. Keeping the diagnosis a secret deprives patients of the chance to make the best use of the time remaining, to say goodbye to family members and to life itself in whatever way they choose, to settle unresolved conflicts, and, whenever possible, to acquire greater depth and maturity in the process. It would be just as wrong, however, to deliver an unfavorable diagnosis in a perfunctory and abrupt manner. In all phases of an incurable illness, there remains a certain amount of hope that must never be completely taken away from the patient, no matter how ill-founded it may seem. It must also be remembered that men over 69 have an elevated risk for suicide immediately after the delivery of the diagnosis (Fox et al. 1982).

3.2
Rebellion – Anxiety – Depression

Emotional reactions in the further course of the process include feelings of being wronged and offended by the injustice of an "undeserved" death sentence, disappointment over one's fate, and anger and rebellion against God or the world. Resentment, bitterness, and envy of the healthy, who will be allowed to live on, may become manifest as ill humor, irritability, or even latent enmity toward others. It is important for family members, and just as important for therapists, to meet such apparently unjustified emotional expressions with understanding and acceptance, whenever possible. They may thereby prevent the patient from adopting a destructive, spiteful attitude that will ultimately make the dying process very much more difficult.

Anxiety accompanies nearly all phases of dying, but it may take very different forms, all of which must be properly recognized and identified as a precondition to successful treatment:

- Fear of disease manifestations: Most severely ill patients fear pain, suffering, shortness of breath, and deformity caused by operative procedures. Fear is thus directed at the threatened loss of bodily integrity associated with the further progression of disease; when amplified by a lack of information and avoidance of medical explanation, this fear may often take on catastrophic proportions.
- Fear of isolation: The anonymity and solitude of dying in a hospital ward lead to the (often justified) fear of social exclusion and loneliness; this may be compounded by the fear, born of distrust, of being lied to or deprived of legal capacity by their family or physicians. Such fears are usually an expression of long-standing disturbances of interpersonal relationships and of trust.
- Fear of loss of autonomy: Severe illness means reliance on others, dependence, and often helplessness. The loss of control and autonomy is perceived by the patient as a threat and a humiliation; the idea that one's own dignity may also be confirmed by the recognition and caring assistance of others is often difficult to accept. Emotions of embarrassment and shame may then rise to the feeling of being no more than a burden to others, so that a "quick end" may be secretly desired. An amplified form of the fear of losing one's autonomy is the fear of ceasing to exist mentally, of fading away into unconsciousness or dementia. This fear has become very common as life expectancies have risen and dementing illnesses have become more prevalent.
- Fear of death: In its vital or "blind" form, the fear of death is probably present in all highly developed living beings. In human beings, however, it acquires anticipatory components: fear of the loss of, and final separation from, loved ones; fear of not having any further experiences, of no longer participating in what goes on in the world; and, finally, the elementary fear of simply passing into nothingness, of extinction, of "ceasing to be" – which has now largely replaced the earlier, religiously motivated fear of divine retribution in Hell.

In addition to these fears, there is also the retrospective anxiety aroused by the "irreparable," compounded by feelings of guilt, regret, and despair with respect to lost opportunities, unfulfilled wishes, and missed areas of personal experience. Thus Hinton (1975) asked 60 patients with advanced cancer about their feelings of satisfaction and fulfillment in life. It was found that the less satisfied they were with their lives, the greater their fear of death was.

Depression, like anxiety, may occur at any phase of the end of life. Psychic causes of depression include the stresses already mentioned: isolation and loneliness, feelings of guilt and worthlessness, the feeling that one's life has not been worthwhile, disappointed hopes of recovery or cure, persistent pain and other forms of bodily suffering perceived as uncontrollable often result in depression. Moreover, cancer patients often experience the tumor not only as a threatening and destructive enemy, but also as an evil and malevolent object within themselves; they may transfer these feelings onto themselves and perceive their illness as a punishment for their own imagined offenses or failures. Depressive episodes are most often provoked by recurrences of disease or by mutilating operative procedures (e.g. mastectomy, colostomy, amputation); such situations may prove

to be beyond the capacity of whatever coping mechanisms have been in place until that time.

3.3 Coping and Acceptance

As soon as the diagnosis is communicated, various coping strategies go into effect with which patients – insofar as they do not deny the existence of the disease – attempt to adapt to the new situation. At first, they may do all they can to postpone the moment of death and to influence the course of their illness with new information and with scientific medical, or paramedical, treatment processes. More effective in the long run, however, are internal processes of reinterpretation and reevaluation that allow patients to see the situation from a different perspective. These include emphasizing the present, using the awareness that one has a "last chance" to intensify personal relationships and strengthen one's sense of life, searching for opportunities that life still affords, resetting priorities, and making corresponding changes in life style. Patients must often come to terms with their own life histories, i.e. take stock of their lives and thereby find a meaning for life and death ("life review"; Butler 1963). Many patients draw deep satisfaction from the settling of "unfinished business," last conversations with relatives and friends, and the fulfillment of long-held wishes. Saying goodbye and grieving in anticipation of the loss of all that was dear and valued are also ways of coping with the last phase of life.

Nonetheless, the phase of acceptance, which Kübler-Ross (1974) described as being practically normative, is rarely completely achieved and should not be presented to patients as a challenge or a requirement. More often, death is viewed with resignation as the inevitable end. A Dutch study of 191 terminally ill patients revealed that only about one third were able to accept death, while one fourth abandoned life resignedly and a further fourth remained militant against death, or in denial, until the end. Women were found to be significantly more able to accept death than men (Bruning and Hesselink 1986, cited in Munnichs 1989).

3.4 Terminal Phase

The last phase of the disease is often heralded by the collapse of defense mechanisms, by generalized organic functional disturbances, by the failure of therapy, and by the transition to palliative measures; it is often signaled to patients by transfer to a private room, to a hospice, or to their own homes. The steady weakening of resistance and of the will to live leads to a state of emotional exhaustion, resignation, and quiet sorrow. Patients whose perception and articulation are impaired often have no further opportunity to come to terms with death by cognitive or communicative means. Their last words are often limited to brief exchanges or to trivial matters of daily routine. Nonetheless, contact through touch and voice is still possible, particularly when a relationship of trust already exists between the patient and the caregivers. In the agonal phase, consciousness is intermittently clouded, and reality gradually pales. In the Dutch study mentioned above, 45% of patients died while drowsy, sleeping, or comatose, 27% with quiet acceptance, but 21% while agitated, fearful, or anxious (Bruning and Hesselink 1986, cited in Munnichs 1989).

4 Objectives and Problems of Psychiatric Treatment

Studies reveal a high prevalence of mental disorders in the severely ill. In a multicenter study of 215 cancer patients, Derogatis et al. (1983) found psychiatric diagnoses according to the DSM-III in 47% of patients: 32% of the patients had anxious or depressive reactions, 6% had severe depressive disorders, 4% had organic brain syndromes, 3% had personality disorders, and 2% had severe anxiety syndromes. Several other studies have shown that the incidence of severe depressive disturbances among cancer patients is on the order of 25% and increases, in later stages of disease, to as high as 75% (Plumb and Holland 1977; Bukberg et al. 1984; Massie and Holland 1990). The prevalence of delirium syndromes among cancer patients in the terminal phase varies from 40% to 85% (Massie et al. 1983). Still more frequent are psychiatric complications in the course of acquired immunodeficiency syndrome (AIDS) (Tross and Hirsch 1988).

4.1 Anxiety

A number of causes of anxiety syndromes in the severely ill have already been mentioned. Anxieties of mental origin may manifest themselves as diffuse inner restlessness, irritability, aggression, or somatic complaints, but even a seemingly "composed" and well-adapted attitude may actually be the product of an intensive effort to conceal latent despair under the appearance of normality. Anxieties of mental origin rarely produce erethic restlessness, agitation, and

panic; such manifestations are more characteristic of anxiety syndromes of organic origin, particularly incipient delirium, arterial hypoxia, dyspnea due to pulmonary metastases or embolism, or an incipient reaction to the withdrawal of benzodiazepines or opiates. The differential diagnosis of such syndromes may require a thorough medical and neurological evaluation.

Pharmacological therapy is indicated for organically induced anxiety, but may also be used as an adjunct to psychotherapeutic intervention. Short-acting benzodiazepines such as lorazepam, alprazolam, and diazepam are recommended for use in the severely ill because of the slowing of metabolism and the consequent risk of a cumulative dosage effect. They should be given multiple times a day in order to ensure an uninterrupted therapeutic effect. Clonazepam is a longer-acting agent that has been found useful for the treatment of patients who additionally have seizures or neuropathic pain (Breitbart and Passik 1993). If benzodiazepines are ineffective, or if their use raises the concern of respiratory depression, neuroleptics of low potency may be used. For panic disorders, tricyclic antidepressants such as imipramine or the newer serotonin antagonists are effective.

Psychotherapy for the treatment of anxiety and depression in severely ill patients is of a supportive nature and is directed toward enhancement of self-esteem and short-term crisis intervention. The therapist faces a major challenge with regard to his or her own life experiences and anxieties. Therapy can encourage the patient to verbalize anxiety and other threatening affects, give vent to grief over imminent separations, and overcome the barriers to others that are posed by shame or distrust. The therapist's own experience with the dying, and even literary descriptions such as Tolstoy's *The Death of Ivan Ilyich*, may help the therapist guide the patient through the processes of reinterpretation mentioned above.

One important approach, "life review therapy," makes use of the patient's own life history; the patient's achievements and skills are emphasized, and attempts are made to correct the patient's negative views of his or her life (Fuchs 1992a). If certain subjects are too heavily charged with anxiety to be verbally accessible, nonverbal processes such as painting and art therapy may be valuable means of expression (Dreifuss and Meerwein 1984). Finally, direct methods of reducing anxiety include relaxation techniques, directed imagery, and hypnosis.

All of these psychotherapeutic interventions not only can improve the mental state of the patient in the last phase of life, but also, on occasion, prolong life. Thus Spiegel et al. (1989) showed that patients with breast cancer who participated in supportive group therapy lived significantly longer than those who did not.

4.2 Depression

Depressive disorders often arise in the last phase of life, but are not easy to diagnose. Severe depression often manifests itself as an apathetic exhaustion syndrome that is difficult to distinguish from the manifestations of the physical illness, as are the typical vegetative symptoms (loss of appetite or weight, sleep disturbance). On the other hand, phases of sadness, dejection, and anxiety are normal accompanying phenomena of dying, as discussed above.

The diagnosis of depression masked by a bodily illness is supported primarily by psychic symptoms such as continuous hopelessness, negativism, excessive guilt feelings with ideas of punishment, feelings of worthlessness, and, above all, suicidal ideation; these symptoms are nearly always associated with treatable depression (see below). Finally, the differential diagnosis must include side effects of chemotherapeutic agents (e.g. vincristine, steroids, interferon) and other medications, endocrine disturbances, and paraneoplastic syndromes (see also Chaps. 17, 18, Vol. 2, Part 2).

Even though depression cannot always be distinguished from the manifestations of physical illness or from the undesired side effects of somatic treatments, all cases of moderate or severe depression should be treated with medication; the notion that depression is a normal reaction to cancer is not a valid argument against such treatment, as the effectiveness of antidepressants in cancer patients is well documented (Popkin et al. 1985; Massie and Holland 1990). Lower doses (25–125 mg) have generally been found to be sufficiently effective in such patients; higher doses may lead to toxic effects because of the lowered metabolic rate. Tricyclic agents and serotonin antagonists (beware of nausea and agitation!) are also used; trazodone and trimipramine have proved their worth as agents with both analgesic and sedating effects.

4.3 Painful Conditions

Persistent and insufficiently controlled painful conditions are a major source of suffering for severely ill patients. The frequency of such conditions in cancer patients is approximately 30% and rises to 60%–90% in the terminal phase. Cancer patients with persistent pain have twice as many psychiatric complications as other patients, primarily depressive syndromes (Derogatis 1983). Sedating antidepressives and anxiolytics are often indicated in such situations and may help prevent a premature elevation of opiate dosage (Kocher 1984).

Psychotherapeutic interventions, relaxation techniques, imagery, and hypnosis may influence the emotional component of pain perception and thereby help guard against an excessive fixation on medications.

4.4 Organic Brain Syndromes

A total of 15%–29% of hospitalized cancer patients have mental disturbances of organic origin (Levine et al. 1978); more frequent types include delirium, dementia, affective disorders of organic origin, hallucinoses, and personality changes. Metabolic disturbances such as hypercalcemia or hypothyroidism, toxic effects of medications (steroids, opiates, chemotherapeutic agents), radiation toxicity, or brain metastases in the final phase of melanoma or other cancers are often associated with organic brain syndromes, which present as cognitive impairment, acute anxiety, or hallucinatory and paranoid states.

The AIDS–dementia complex, which occurs in two thirds of all AIDS patients at some point in their course, consists of a combination of memory disturbance, cognitive slowing, motor disturbance, and depression or psychosis. Psychopathological changes in the agonal phase include reduced drive, vigilance, and attention, confusion, anxious restlessness of a vegetative character, and many types of exogenous reaction (Bonhoeffer's symptom).

Neuroleptics are the most suitable agents for the psychiatric treatment of these organic brain syndromes. In marked agitation, parenteral haloperidol (0.5–2 mg) may be given every hour until the symptoms have become less severe, and may then be given orally. In addition, the patient's potential for orientation should be maximized by means of a well-structured, individually arranged environment, continuity of nursing staff, frequent contacts with relatives, and gentle correction of misinterpretations.

4.5 Suicidality and Death Wish

The desire of terminally ill patients to end their lives prematurely is one of the most serious problems confronting the psychiatrist and is currently a matter of public concern in the ongoing debate over euthanasia. Generally speaking, elderly and very elderly persons seldom have death wishes or suicidal ideation, which, when they do occur, are almost without exception signs of a diagnosable psychiatric illness; despite multiple stresses and losses, the great majority of such patients feel no desire to die as the end of life approaches (Linden and Barnow 1998).

There are, however, hardly any studies of representative sample populations regarding the prevalence and reasons for death wishes in the severely ill. Brown et al. (1986) interviewed 44 patients with terminal cancer in a palliative care ward; all ten patients who expressed a wish to die prematurely were suffering from clinical depression. Chochinov et al. (1995) studied 200 patients with advanced malignancies in a palliative care ward; just under half (45%) reported having brief and transient death wishes, and 17 patients (8.5%) had more serious and lasting death wishes. The latter patients also complained significantly more frequently of pain and lack of family support; ten of them suffered from clinical depression. Finally, Emanuel et al. (1996) questioned 155 cancer patients and found that more than one quarter (27%) had already thought seriously of asking their physicians for euthanasia or assistance with suicide; 12% had actually discussed this subject with their physicians. The latter 12% of patients had a significantly higher frequency of depression than the other patients.

Figures on suicide yield an indication of the prevalence of death wishes: cancer patients in advanced stages of disease have a significantly higher suicide rate than the age-matched general population (Farberow et al. 1971; Luohivuori and Hakama 1979). Risk factors include preexisting personality disorders, alcohol abuse, social isolation, depression, and inadequately treated pain (Breitbart 1990). Significantly elevated suicide rates have also been found among AIDS patients (Marzuk et al. 1988; Coté et al. 1992).

Among the 2.7% of annual deaths in the Netherlands that are the result of active euthanasia on request or of medically assisted suicide, three quarters occurred in the setting of advanced cancer. A sample of such patients less often gave pain as the reason for their desire to die than feelings of meaninglessness, dependence, and lack of dignity (van der Maas et al. 1996; Fuchs 1997).

The available studies imply that, among the many motives underlying the death wishes of severely ill and dying patients, the more important ones are depression and anxiety, lack of family support and security, feelings of the indignity and meaninglessness of their own suffering, and inadequately treated pain. The desire to die is often an expression of the need for attention and esteem, a cry for help to others, or a protest against unbearable suffering. Finally, a problem of communication with the physician, the nursing team, or family may motivate the desire for "salvation by injection," through which the patient wrongly thinks he or she will relieve others of an unwanted burden.

The marked fluctuation of death wishes (Chochinov et al. 1995), their many-layered motivation, and their dependence on successful or unsuccessful communi-

cation with others all imply that supposedly simple solutions, such as the championing of "self-determination" and rapid termination of life through active euthanasia, can hardly do justice to the complexity of the situation of dying patients. Not all, but certainly many patients who want to die have a psychiatric disturbance that must be recognized and treated.

Moreover, the experience from palliative care wards and hospices for dying patients reveals that adequate medical care, competent treatment of pain, and good emotional support can strongly influence the patient's attitude toward living and dying. It is doubtless important to discuss the patient's desire to die openly and non-judgmentally; the patient may gain significant relief through such an interaction alone. It may also be reasonable to inform family members gently of the patient's dire situation. The physician can retain credibility and help the patient through times of desperation only after coming to terms with his or her own personal attitudes toward premature termination of life.

4.6 Withholding Life-Prolonging Medical Treatment

Psychiatric problems of the end of life also include the question of whether life-prolonging medical treatments should be withheld when a fatal illness has already severely impaired the patient's life functions, and death can no longer be prevented ("passive euthanasia"). Patients' unwavering desires to terminate, or not to initiate, medical treatment measures must surely be respected; the physician, however, has the duty to ascertain whether such a decision has come about because of the influence of others, external pressure, or depression.

The consultation of a psychiatrist is mandatory when there is doubt as to the patient's decision-making competence and the withholding of treatment would be associated with serious risks or disadvantages. Particularly with patients in advanced stages of dementia, the question arises as to whether life-prolonging measures ought to be continued even if the health and well-being of the patient are impaired by a second illness or if the patient's survival depends on the artificial provision of nutrition and hydration (Karger and Haupt 1997). Such decisions may be made easier if the patient, while still able to make such decisions for him- or herself, has already clearly expressed the desire to forego life-saving treatment in the case of a severe and incurable illness and if this can be credibly attested by family members. Even then, the question remains as to what extent the patient's previously expressed wishes still correspond to present, possibly quite different wishes and needs, and whether they can be applied to the current situation even though it could not have been foreseen in every detail.

If the patient's attitudes and preferences are unknown, the physician, family members, or legal guardian must try to ascertain the patient's interests as well as they can. The ethical principles underlying such determinations are controversial (Beleites 1998). The decision taken in the individual case depends, among other things, on whether the proposed medical intervention – ranging from the administration of antibiotics or infusions to the insertion of a gastrostomy, to dialysis and surgical procedures – offers a sufficient prospect of palliating of the patient's suffering, or whether it essentially serves only to prolong the patient's life and may actually produce further, iatrogenic suffering. Clearly, when such decisions are made, attention should be paid to the body language with which a demented patient may convey his or her distress.

On the other hand, it must not be assumed a priori that the quality of life of a patient in an advanced stage of dementia is necessarily negative, particularly because so little is known about the experiential world of the demented. Indeed, the current emphasis on cutting expenditures for medical and nursing care only increases professional caregivers' responsibility to afford demented patients special protection. It may be ethically justifiable to withhold artificial nutrition or medical interventions in some cases, but this should never become routine practice, as such measures have a very important symbolic value: they bear witness to the respect in which the patient is still held as long as his or her personal identity, though overshadowed by illness, remains perceptible; and they raise a necessary barrier against the human tendency to neglect weak, powerless individuals and exclude them from society.

4.7 Environment of the Dying Person

A further important aspect of psychiatric assistance at the end of life consists of the support of family members and professional caregivers. The family of the dying person is exposed to severe stress by his or her illness. The emotional and nursing support of the patient, the associated restrictions of life style and financial burden, uncertainty over the course of the illness, and anticipatory grief over the impending final separation often lead to a continuous conflict between hope and despair and, finally, to exhaustion of the ability to help. Further stress is created by family members' uncertainty over how openly they should discuss the grave situation with the patient and deliberately say goodbye – over whether the family members themselves might be overcome with grief in

such a situation or potentially undermine the patient's remaining will to live.

Such conflicts, and the need to fend off feelings of hopelessness and depression, often lead family members to withdraw inward and away from the dying person; the patient may note this and desire more strongly to die. According to a study by Krant and Johnson (1977/1978) on near relatives of 75 terminally ill patients, only one fifth of the relatives had spoken with the patients about the possibility of death. A study of couples (Hinton 1981) revealed that two thirds of the partners had spoken only incompletely, or not at all, with the patients about their impending death. Many physicians, too, react to the patient's worsening condition and their own therapeutic powerlessness with an unconscious retreat from the patient.

A study by Schulz and Aderman (1976) revealed that the determination of a poor prognosis significantly changes the behavior of staff and family members (shorter visits, less verbal communication, fewer contacts). Nursing care is made more difficult by certain coping styles, character idiosyncrasies, or personality disorders of severely ill patients, which often lead to an increased need for control in response to the fear of dependence, distrust, or a withdrawal inward.

The psychiatrist can recognize interpersonal conflicts of these types and, by maintaining contact with the individuals caring for the patient, providing accurate medical information, and advising on the best ways to cope with the situation, can contribute to the maintenance and reinforcement of the emotional support system that the patient needs from his or her social environment. The psychiatrist thus plays an important mediating role by promoting dialogue among all those involved in the dying process. Furthermore, the psychiatrist can encourage family members to discuss stressful topics and uncertainties with the patient and motivate them to anticipatory grief. Repression of impending death can make it difficult or impossible to cope with the loss when it finally occurs. Such pathological grief reactions are most commonly found among family members whose bond to the dying patient is ambivalent or not emancipated. It is particularly important to bear this possibility in mind in view of the elevated morbidity and mortality of widowed partners (particularly men) in the first year after death; timely psychotherapeutic help should be made available.

4.8 Hospices

The hospice movement arose in England in the late 1960s as an attempt to counteract the increasing anonymity of death in our society. Its purpose is to create a secure space in which the dying can be accompanied on their last journey with appropriate medical, psychological, and pastoral care, just as the medieval hospices gave shelter to pilgrims and travelers. In 1967, Dame Cicely Saunders founded St. Christopher's Hospice in London; the first hospice in Germany opened in 1985. Since then, approximately 30 inpatient hospices and as many palliative care wards have been established in Germany, along with approximately 270 ambulatory care services and more than 180 volunteer initiatives with some 10,000 volunteers (Saunders 1993; Student 1993; for hospice movements in other countries, see also Aranda 1999; Haber1999).

In hospices, patients with incurable, fatal illness are treated under a different set of objectives requiring a different approach from medical personnel. The relief of physical and mental suffering is the primary concern; to this end, new medical, physiotherapeutic, and psychotherapeutic techniques for the treatment of pain are tested and further developed. The principle of timely and often prophylactic analgesic treatment, in coordination with acupuncture, hypnosis, and other techniques, enables the successful treatment of painful conditions in most cases, without making patients confused or sedated.

The fundamental concept of the hospice movement is that of using brief inpatient stays to optimize medical treatment, especially the treatment of pain, so that patients can be cared for in their home environment again as soon as possible, in coordination with specially trained nurses, lay assistants, and family members. A hospice is thus not a "death house," but an institution that is open to the community and oriented toward outpatient care. Assessment studies have shown that the subjective and objective life quality of hospice patients is positive (Greer et al. 1986; Morris et al. 1986). The potential for good medical care at the end of life that the hospice movement affords is certainly underutilized.

5 Near-Death Experiences

This discussion of psychiatric aspects of the end of life will be brought to a close with a look at so-called "near-death experiences," which are not only of psychiatric interest, but also of great personal concern to many people who are dying.

As early as the end of the nineteenth century, there were reports by people who had been through externally life-threatening situations, such as falls while mountain-climbing, saying that they had experienced states of inner alienation and absence of feeling that

would today generally be referred to as depersonalization and dissociation. Noyes and Kletti (1976a,b) collected more than 200 such reports of accident victims in life-threatening situations and derived the following essential components of the near-death experience:

1. Slowing of the subjective passage of time down to a complete standstill
2. In contrast, a considerable speeding up of the train of thought
3. Sudden inner calm and freedom from pain and fear, associated with a feeling of not being involved and of unreality
4. Alienation from one's own body
5. A rapidly running inner panorama of life memories

The actual end-of-life experiences associated with out-of-body states were first made known to a broader public through their description by Moody (1975) and, later, mainly through the research findings of Ring (1980). According to several studies now available, comparable extraordinary perceptions and experiences are reported by 30%–40% of people who were resuscitated after cardiac arrest, lay in a coma, had a life-threatening accident, or seriously attempted suicide (Ring 1980; Ring and Franklin 1981; Greyson and Stevenson 1980; Sabom 1983; review in Roberts and Owen 1988). In these studies, there was no association of specific experiences with demographic and social characteristics of the subjects, such as age, sex, ethnic origin, education, occupation, religion, or beliefs, or with possible previous knowledge concerning near-death experiences. It is possible that we are dealing here with a common, basic pattern of human experience in the immediate proximity of death that is independent of individual orientation and socialization (Fuchs 1997).

The experience typically begins with a feeling of peace, well-being, and freedom from pain, which may rise to the level of ecstatic joy. Almost without any intervening transition, this is followed by the sensation of hovering outside one's body and of perceiving what is happening all around without any emotional participation. Remarkably, the subjective passage of time becomes profoundly slowed, often so much so that a "feeling of eternity" is experienced. This out-of-body experience then gives way, in most individuals, to a tunnel or darkness experience, finally followed by an encounter with a person-like apparition of light or by a fast-forward-type sequence of remembered autobiographical images ("life panorama"). Very often, affected individuals retrospectively ascribe a special religious or spiritual meaning to this experience. There may be a fundamental resetting of life priorities and lasting changes of personality.

The possible explanations of these phenomena may be divided into neurophysiological, psychodynamic, and transpersonal approaches.

Some of the essential components of the near-death experience also occur in toxically or functionally altered brain states such as temporal lobe epilepsy and states produced by the administration of narcotics, such as ketamine, or hallucinogens, such as mescaline, LSD, and phencyclidine. This fact is the basis for the hypothesis of a disturbance and activation of the temporal limbic system caused by terminal hypoxia or hypercapnia, in association with the endogenous production of endorphins (Rodin 1980; Carr 1982).

Psychodynamic explanatory approaches are based on the model of a dissociation of consciousness into a participating and an observing self, along with a regressive defense against the threat of death by means of hallucinatory wish fulfillment (Ehrenwald 1974; Noyes and Kletti 1976b).

In transpersonal interpretations, near-death experiences are viewed as a kind of "altered state of consciousness" belonging to an archetypal psychic matrix that is shared across cultures and has manifested itself in all historical periods in religious, visionary, and mystical experiences (Grosso 1983).

Even if none of these theories provides an entirely satisfactory and conclusive explanation of near-death experiences at present, the significance of these experiences is undiminished. They represent a profound transformation of human spatial, temporal, and emotional experiences and can have long-lasting effects on the human personality. They provide an impressive demonstration of the central, and ultimately therapeutic, importance of death for human life.

6 References

Aranda S (1999) Global perspectives on palliative care. Cancer Nurs 22: 33–39

Beleites E (1998) Wegweiser für ärztliches Handeln. Grundsätze der Bundesärztekammer zur ärztlichen Sterbebegleitung. Dtsch Ärzteblatt 95: 1851–1853

Breitbart W (1990) Suicide in cancer patients. Oncology 1: 49–53

*Breitbart W, Passik SD (1993) Psychiatric aspects of palliative care. In: Doyle D, McDonald N (eds) Oxford textbook of palliative medicine. Oxford Medical Publications, Oxford, pp 609–626

Breuer J, Freud S (1955) Studies on hysteria. Complete psychological works of Sigmund Freud, vol II. Hogarth, London

Brown JH, Henteleff P, Barakat S, Rowe CJ (1986) Is it normal for terminally ill patients to desire death? Am J Psychiatry 143: 208–211

Bruning H, Hesselink J (1986) Omgaan met sterven, samen leven, samen sterven [Dealing with dying, living together, dying together]. Zon Uitgeverij, Leiden

Bukberg J, Penman D, Holland JC (1984) Depression in hospitalized cancer patients. Psychosom Med 46: 199–212

Butler RN (1963) The life review. An interpretation of reminiscence in the aged. Psychiatry 26: 65–76

Carr D (1982) Pathophysiology of stress-induced limbic lobe dysfunction: a hypothesis for NDEs. Anabiosis 2: 75–90

*Chochinov HM, Wilson KG, Enns M et al (1995) Desire for death in the terminally ill. Am J Psychiatry 152: 1185–1191

Coté TR, Biggar RJ, Dannenberg Al (1992) Risk of suicide among patients with AIDS. JAMA 268: 2060–2068

de Beauvoir S (1965) A very easy death. Pantheon, New York

Derogatis LR, Morrow GR, Fetting J et al (1983) The prevalence of psychiatric disorders among cancer patients. JAMA 249: 751–757

Dreifuss E, Meerwein F (1984) Die Psychotherapie Sterbender – der Beitrag der Psychoanalyse. In: Spiegel-Rösing I, Petzold H (eds) Die Begleitung Sterbender. Junfermann, Paderborn, pp 259–277

Ehrenwald J (1974) Out-of-the-body-experiences and the denial of death. J Nerv Ment Dis 159: 227–233

Emanuel EJ, Fairclough DL, Daniels ER, Clarridge BR (1996) Euthanasia and physician-assisted suicide: attitudes and experiences of oncology patients, oncologists, and the public. Lancet 347: 1805–1810

Farberow NL, Ganzler S, Cuter F, Reynolds D (1971) An eight year survey of hospital suicides. Suicide Life-Threatening Behavior 1: 194–201

Fox BH, Stanek EJ, Boyd SC, Flannery JT (1982) Suicide rates among cancer patients in Connecticut. J Chronic Dis 35: 89–100

Fuchs T (1992a) Erinnerungstherapie im Alter. Psychother Psychosom Med Psychol 42: 295–370

Fuchs T (1992b) Außerkörperliche Erfahrungen bei Reanimation. Fundamenta Psychiat 10: 100–107

Fuchs T (1997) Euthanasie und Suizidbeihilfe. Das Beispiel der Niederlande und die Ethik des Tötens. In: Spaemann R, Fuchs T (eds) Töten oder sterben lassen? Herder, Freiburg, pp 31–107

Fuchs T, Lauter H (1997) Der Fall Chabot: Assistierter Suizid aus psychiatrischer Sicht. Nervenarzt 68: 878–883

Greer DS, Mor V, Morris JN et al (1986) An alternative in terminal care: results of the National Hospice Study. J Chron Dis 39: 9–26

Greyson B, Stevenson I (1980) The phenomenology of near-death experiences. Am J Psychiatry 137: 1193–1196

Groenewoud JH, van der Maas PJ, van der Wal G et al (1997) Physician-assisted death in psychiatric practice in the Netherlands. N Engl J Med 336: 1795–1801

Gronemeyer M (1993) Das Leben als letzte Gelegenheit. Wissenschaftliche Buchgesellschaft, Darmstadt

Grosso M (1983) Jung, parapsychology, and the near-death experience: toward a transpersonal paradigm. Anabiosis 3: 3–38

Haber D (1999) Minority access to hospice. Am J Hosp Palliat Care 16: 386–389

Hinton JM (1975) The influence of previous personality on reactions to having terminal cancer. Omega 6: 95–111

Hinton JM (1981) Sharing or withholding of dying between husband and wife. J Psychosom Med 25: 337–343

Karger A, Haupt M (1997) Sterbehilfe bei Demenz. Ethische Überlegungen zwischen Paternalismus und Autonomie. Nervenarzt 68: 907–913

Klein M (1948) A contribution to the theory of anxiety and guilt. Int J Psychoanal 29: 114–123

Kocher R (1984) The use of psychotropic drugs in the treatment of cancer pain. In: Zimmermann M, Drings P, Wagner C (eds) Pain in the cancer patient. Springer, Berlin Heidelberg New York, pp 118–126

Krant MJ, Johnston L (1977/1978) Family members' perceptions of communications in late stage cancer. Int J Psychiatr Med 8: 203–216

*Kübler-Ross E (1974) Interviews mit Sterbenden. Kreuz, Stuttgart

Levine PM, Silberfarb PM, Lipowski ZJ (1978) Mental disorders in cancer patients. Cancer 42: 1385–1391

Lauter H, Zacher D, Fuchs T (1999) Biographie und Lebensschicksal. In: Folkerts H, Schonauer K, Tölle R (eds) Dimensionen der Psychiatrie. Thieme, Stuttgart, pp 19–24

*Lifton RJ (1979) The broken connections: on death and the continuity of life. Simon and Schuster, New York

Linden M, Barnow S (1998) The wish to die in very old persons near the end of life: a psychiatric problem? Results from the Berlin Aging Study (BASE). Int Psychogeriatr 9: 291–307

Luohivuuori KA, Hakama M (1979) Risk of suicide among cancer patients. Am J Epidemiol 109: 59–65

Massie MJ, Holland J (1990) Depression and the cancer patient. J Clin Psychiatry 51: 12–17

Massie MJ, Holland J, Glass E (1983) Delirium in terminally ill cancer patients. Am J Psychiatry 140: 104–1050

Marzuk PM, Tierney H, Tardiff K et al (1988) Increased risk of suicide in persons with AIDS. JAMA 259: 1333–1337

*Meyer JE (1979) Todesangst und das Todesbewußtsein der Gegenwart. Springer, Berlin Heidelberg New York

Moody RA (1975) Life after life. Mockingbird, Georgia

Morris JN, Suissa S, Sherwood S et al (1986) Last days: a study of the quality of life of terminally ill cancer patients. J Chron Dis 39: 47–62

Munnichs JMA (1989) Sterben und Tod. In: Platt D (ed) Handbuch der Gerontologie, vol 5: Neurologie, Psychiatrie. Fischer, Stuttgart, pp 461–471

*Noll P (1984) Diktate über Sterben und Tod. Pendo, Zürich

Noyes R, Kletti R (1976a) Depersonalization in the face of life-threatening danger: a description. Psychiatry 39: 19–27

Noyes R, Kletti R (1976b) Depersonalization in the face of life-threatening danger: an interpretation. Omega 7: 103–114

Nuland SB (1994) How we die: reflections on life's final chapter. Knopf, New York

Plumb MM, Holland J (1977) Comparative studies of psychological function in patients with advanced cancer. I. Self-reported depressive syndrome. Psychosom Med 39: 264–276

Popkin MK, Callies AL, Mackenzie TB (1985) The outcome of antidepressant use in the medically ill. Arch Gen Psychiatry 42: 1160–1163

Rank O (1931) Technik der Psychoanalyse. III. Die Analyse des Analytikers. Deuticke, Leipzig

Ring K (1980) Life at death: a scientific study of the near death experience. Coward, McCarn and Geoghegan, New York

Ring K, Franklin S (1981) Do suicide survivors report near-death experiences? Omega 12: 191–208

*Roberts G, Owen J (1988) The near-death experience. Br J Psychiatry 153: 606–617

Rodin EA (1980) The reality of near-death experiences. A personal perspective. J Nerv Ment Dis 168: 259–263

Roth M (1959) The phobic anxiety-depersonalization syndrome. J Neuropsychiatry 1: 293–306

Sabom MB (1983) Erinnerung an den Tod: Eine medizinische Untersuchung. Kindler, Munich

Saunders C (1993) Hospiz und Begleitung im Schmerz. Herder, Freiburg

*Schulz R, Aderman S (1976) How the medical staff copes with dying patients: a critical review. Omega 7: 11–21

Skoog G (1965) Onset of anankastic conditions. Acta Psychiatr Scand Suppl 184: 5–82

Spiegel D, Bloom JR, Kramer HC, Gottheil E (1989) Effect of psychosocial treatment on survival of patients with metastatic breast cancer. Lancet ii: 89–891

Straus E (1960) Ein Beitrag zur Pathologie der Zwangserscheinungen. In: Straus E (ed) Psychologie der menschlichen Welt. Springer, Berlin Göttingen Heidelberg, pp 187–223

Student JC (1993) Das Hospiz-Buch, 3rd edn. Lambertus, Freiburg

Tross S, Hirsch DA (1988) Psychological distress and neuropsychological complications of HIV infection and AIDS. Am Psychol 43: 929–934

*van der Maas PJ, van der Wal G, Haverkate I et al (1996) Euthanasia, physician-assisted suicide, and other medical practices involving the end of life in the Netherlands, 1990–1995. New Engl J Med 335: 1699–1705

von Gebsattel V (1954) Die Welt des Zwangskranken. In: von Gebsattel V (ed) Prolegomena einer medizinischen Anthropologie. Springer, Berlin Göttingen Heidelberg, pp 74–127

*Wander M (1980) Leben wär' eine prima Alternative. Luchterhand, Darmstadt

**Yalom ID (1989) Existentielle Psychotherapie. Edition Humanistische Psychologie, Cologne

CHAPTER

13

I.F. Brockington, M. Lanczik

Psychiatric Illnesses in Women

The sections by M. Lanczik were translated by E. Taub.

1
Psychiatry of Menstruation

1.1
Introduction (M. Lanczik)

The reciprocal relationship between gynaecological and psychopathological processes is particularly evident in the setting of menstruation. Alterations or disorders of rhythm in either the psychological or the biological sphere may affect events in the other sphere. This occurs, for example, when a woman suffering from melancholia loses her normal rhythm, so to speak, not only with respect to "temporally rhythmic living conditions" – a term used by Pauleikhoff (1986) in allusion to Tellenbach (1983) – but also in her menstrual cycle, or when psychic inhibition leads to inhibition of menstruation and, in extreme cases, to amenorrhoea. On the other hand, the normally functioning menstrual cycle can give rise to psychopathological phenomena occurring in the same rhythmic pattern, such as the relatively common pre-menstrual dysphoric syndrome and, in extreme cases, the rare menstruation psychoses.

The fact that mental illnesses may bear a temporal relationship to certain phases of the menstrual cycle was one of the earliest discoveries of biological psychiatric thinking. Hippocrates explained mood alterations related to menstruation as the consequence of an obstruction to the flow of menstrual blood. Menstruation-related mental illness became a subject of scientific investigation in the early eighteenth century, with the submission of a dissertation by Georg Ernst Stahl in the year 1702. In the nineteenth century, psychiatrists assumed that 10% of all organic psychosyndromes in women were causally related to menstrual disorders (Splett 1998). Robert C. Frank, in 1931, was the first to postulate a direct aetiological relationship between psychopathology and ovarian hormones. Collins et al. (1985) simultaneously measured all of the relevant cyclically secreted hormones and plasma catecholamines and were able to show that depressive and anxious mood disturbances occur primarily in the luteal phase. Mental well-being, on the other hand, is observed significantly more frequently in the follicular phase of the menstrual cycle.

1.2
Pre-menstrual Dysphoric Syndrome (M. Lanczik)

The most frequently reported psychopathological disorder bearing a temporal relationship to changes in sex hormone concentrations is the so-called pre-menstrual dysphoric syndrome. Research on this mood disorder was long hindered by the inadequacy of its definition. It was first included in a system of psychiatric classification, the third edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-III), in 1987. In the revision of DSM-III, it was named late luteal dysphoric syndrome. In DSM-IV, it was given the clinically accurate designation of pre-menstrual dysphoric syndrome (Spitzer et al. 1989).

1.2.1 Clinical Features

Hamilton et al. (1984) list approximately 150 different mental and physical manifestations that may occur in the pre-menstrual period. There is, however, a fairly characteristic, clinically relevant pattern of manifestations, including sadness (56%), anxiety (36%) and lability of affect (26%) (Freeman et al. 1985). The latter condition expresses itself clinically as suddenly occurring phases of tearfulness and increased irritability (48%), which may frequently lead to interpersonal conflict with the patient's life partner and also interfere with her professional activities. Depressive mood alterations are frequently reported, expressing themselves as self-reproach or other self-demeaning thoughts, and as a feeling of hopelessness that may, in the long term, result in withdrawal from social relationships. Interest in normal activities, such as work and hobbies, is diminished (Moos 1968a,b). Moreover, patients complain of a subjective feeling of difficulty concentrating, easy fatigability and a feeling of tension.

Suicidal behaviour may occur as a very serious complication of this disorder. The incidence of parasuicidal behaviour has been found to be elevated in women suffering from pre-menstrual dysphoric syndrome who are subjected to a non-specific stress (Chaturvedi et al. 1995).

Pre-menstrual changes in appetite express themselves less commonly as loss of appetite (as in endogenous depression) than as a craving for certain types of foods; there is a significant association between the intensity of such cravings and the severity of the mood disorder (Dye et al. 1995). Increased appetite is, however, also observed in connection with the menstrual cycle in women who do not suffer from pre-menstrual mood changes.

Sleep disorders occurring frequently in the pre-menstrual period include both insomnias and hypersomnias.

Finally, there are often a number of somatic symptoms, such as breast pain and headache (23%), swelling and abdominal complaints (Freeman et al. 1985). The latter, however, almost never attain the

intensity of the somatic symptoms accompanying dysmenorrhoea. Some women suffer both from a pre-menstrual syndrome of psychopathological significance and from a gynaecologically relevant dysmenorrhoea; in this context, the simplifying term "peri-menstrual syndrome" is often used (Bancroft et al. 1993).

1.2.2 Diagnostic Evaluation and Classification

The decisive factor for the provision of a diagnosis and for differential therapeutic planning is not the fairly typical clinical picture, but rather the course of the disorder, i.e. the regular appearance of its manifestations – in varying degrees of severity – in the late luteal phase of the menstrual cycle and their remission after the onset of menstruation. Manifestations never occur in the week after the menses. Thus these women are free of mental disturbances during only 1 week of the menstrual cycle.

For most women with pre-menstrual dysphoric disturbances, the psychopathological changes are not severe enough to affect their social lives or their professional performance. The diagnosis of pre-menstrual dysphoric syndrome is, therefore, only given when the manifestations of the disorder significantly limit social or professional performance ability and when they have occurred in the majority of menstrual periods in the year preceding diagnosis. Nevertheless, there are as yet no uniform guidelines for determining the degree of severity of pre-menstrual syndrome or for the validation of inter- and intra-individual studies of the course and treatment of the disorder (Hurt et al. 1992). In DSM-IV, a distinction is newly drawn between pre-menstrual syndrome and pre-menstrual dysphoric syndrome: the former is said to be distinguished from the latter by its lesser severity and lower frequency of exacerbations. In our opinion, this dichotomy is not justified from the psychopathological, the course-typological or (especially) the aetiological point of view and should therefore be discarded.

It is noteworthy that the psychopathological manifestations are the same as those of the so-called post-partum blues; we have accordingly coined the designation "post-partum dysphoric syndrome" for the latter disorder (Lanczik et al. 1992; Heidrich et al. 1994; Lanczik and Brockington 1999).

There is some evidence that pre-menstrual dysphoric syndrome is a heterogeneous disorder. It is not yet clear, for example, whether subtypes of pre-menstrual dysphoric syndrome tending towards depression, dysphoria or anxiety are psychopathologically distinct.

1.2.3 Differential Diagnosis

Pre-menstrual dysphoric syndrome differs from dysmenorrhoea in that the latter is clinically characterised essentially by pain and physical discomfort rather than by a pattern of psychopathological manifestations. Moreover, dysmenorrhoea occurs not pre-, but peri-menstrually.

Pre-menstrual dysphoric syndrome is distinguished from other affective disorders such as dysthymia or endogenous depression primarily by its course. Dysthymia and endogenous depression do, in fact, occur at increased frequency in the late luteal phase, but, unlike pre-menstrual dysphoric syndrome, they do not remit with the onset of the menses (Brockington et al. 1988; Endicott 1993). It should be remembered that the affective manifestations of pre-menstrual syndrome may have the intensity, though not the duration, of a phase of endogenous depression.

In rare cases, it may be difficult to distinguish an affective disorder of "rapidly cycling" type from a combination of pre-menstrual dysphoric syndrome with an oligo-ovulatory cycle. This example suffices to make clear that these patients require attentive gynaecological and psychiatric care.

Disorders of adaptation, like affective illnesses, also tend to be worse pre-menstrually.

1.2.4 Course and Prognosis

The earliest manifestation of pre-menstrual dysphoric syndrome may appear at any time after menarche. Its course in interrupted only by pregnancy, during which these women feel exceptionally well mentally; they tend, however, to develop depression shortly before the first post-partum menstruation (O'Hara 1987).

When pre-menstrual dysphoric syndrome remains untreated, its intensity increases over time, although the severity of manifestations often varies from cycle to cycle. The period of greatest vulnerability for the development of pre-menstrual dysphoric syndrome is between the ages of 25 and 35. Women usually present for treatment between the ages of 30 and 40. The manifestations usually become less severe after age 45 and generally disappear after menopause or hysterectomy (Casson et al. 1990; Freeman et al. 1995). Bäckström et al. (1981), however, reported cases in which manifestations persisted after oophorectomy.

1.2.5 Epidemiology

Precise data on the prevalence of pre-menstrual dysphoric syndrome are not yet available in adequate

quantity. This is partly the result of the use of different definitions of this affective disorder in the various studies published to date. Nonetheless, it has been clearly shown to occur very frequently in comparison to all other mental illnesses.

A total of 40%–73% of all women are said to suffer from pre-menstrual mood alterations, which are, in most cases, not severe enough to affect these women to a major degree in their everyday lives. Such mood alterations are regarded by many women as a quasi-normal pre-menstrual phenomenon. However, 28% of all women report that their work is adversely affected by pre-menstrual mood alterations, 5%–8% consult a physician for this problem and 2.1% are unable to work in the pre-menstrual period (N. Wood et al. 1982; Andersch et al. 1986; Rohde et al. 1992).

1.2.6 Aetiology

Biological Factors

The possible causes of pre-menstrual dysphoric syndrome that have been considered to date are quite diverse. Bancroft and Rennie (1995) found that pre-menstrual mood alterations were positively correlated with the quantity of menstrual bleeding and the occurrence of pre- and peri-menstrual pain, although the cause-and-effect relationship between these pathological phenomena remains unclear, and it has not been determined whether they might all be produced by a common causal factor.

It has been known for some time that oestrogens raise the overall concentration of cerebral neurotransmitters (Distler and Graf 1986) by inhibiting the enzyme monoamine oxidase (MAO) (Luine and Rhodes 1983), and they should, therefore, have an antidepressant effect (Klaiber 1979). Oestrogens also have serotonergic effects (Sherwin 1990). They increase the number of cerebral serotonin receptors as well as serotonin receptor sensitivity (Kow and Pfaff 1985).

The central nervous system also possesses specific receptors for progesterone (Steiner 1987); unlike the oestrogens, progesterone increases MAO activity (Luine et al. 1975). Nonetheless, through its activity as a serotonin re-uptake inhibitor and an independent slowing effect on serotonin metabolism, it simultaneously increases serotonin activity and may, therefore, also have an antidepressant effect.

The results of these studies are theoretically compatible with the conclusion that the higher hormone concentrations in the follicular phase of the menstrual cycle have an antidepressant effect (Biegon et al. 1983; Luine and Rhodes 1983; Kow and Pfaff 1985). In the late luteal phase, the concentrations of these hormones are lower, and this protective factor is lacking.

In general, however, the temporal coincidence with the pre-menstrual drop in hormone concentrations does not explain why some women suffer from an affective syndrome and others do not. No direct relationship has yet been demonstrated between the syndrome and changes in the concentrations of the steroid sex hormones (Rubinow et al. 1988; Schmidt et al. 1991). Different groups of researchers have obtained inconsistent results. Pre-menstrually dysphoric women have been found to have both lower progesterone levels and higher oestrogen levels in the luteal phase than psychopathologically unaffected women. Bäckström et al. (1983) and Rubinow et al. (1988), however, found that these patients had the same peripheral blood levels of gonadal hormones as normal control subjects (reviewed in Dinane and O'Keane 1991). These observations also accord with the findings of Schmidt et al. (1991) that steroid hormone concentrations are not abnormal in pre-menstrual dysphoric syndrome and that an interruption of the late luteal phase by progesterone supplementation does not improve the psychopathological manifestations. Only Eriksson et al. (1992) found that affected women had higher progesterone levels in the late luteal phase than mentally healthy control subjects. In any case, it seems that pre-menstrual dysphoric syndrome cannot be explained as solely the result of elevated levels of one steroid hormone or depressed levels of another. There are even reports that manifestations may persist in the rhythm of the menses after hysterectomy with simultaneous oophorectomy (Bäckström et al. 1981), although, in the overwhelming majority of cases, the manifestations subside with menopause or after the pharmacological or surgical prevention of ovulation (Muse et al. 1984; Bancroft et al. 1987a; Casson et al. 1990).

Some patients with pre-menstrual dysphoric syndrome have corpus luteum insufficiency, which is characterised pathophysiologically by a disturbance of follicular maturation and by inadequate secretion of the corpus luteum products progesterone and oestradiol. This disorder, which is thus associated with labile ovarian function, also occurs at increased frequency post partum and pre-menopausally, i.e. in phases in which mental illnesses are also more prevalent.

There is increasing evidence that chronobiological factors also play a role in the aetiology and pathogenesis of pre-menstrual dysphoric syndrome. Patients are reported to have a low melatonin level and a premature termination of melatonin secretion (Parry et al. 1990).

It remains unclear to what extent thyroid endocrine function is involved in the aetiology of pre-menstrual dysphoric syndrome or of post-partal mental illnesses (Okano 1999). It is currently known only that women

with pre-menstrual dysphoric syndrome have a greater variability of thyroid hormone levels (Roy-Byrne et al. 1987; Schmidt et al. 1993; Girdler et al. 1995), as do women with post-partal mental illnesses (Pop et al. 1991).

Biography, Psychodynamics and Personality

There is no relevant evidence to date suggesting that pre-menstrual dysphoric syndrome can be induced by pschodynamic factors of any kind. Early psychoanalytic explanatory models postulated a suppressed conflict in the perception of the female role, which expressed itself in the manifestations described above on the onset of the menses. Against this idea is the fact that the psychopathologically relevant manifestations actually subside when menstruation begins. Furthermore, this disorder is also observed in hysterectomised women with intact ovarian function. Rohde et al. (1992) found no significant difference between fertile and infertile women regarding the incidence of this disorder. Furthermore, the large majority of studies documented no correlation between any specific personality type or personality disorder and pre-menstrual dysphoric syndrome, although 10% of affected women fulfil the DSM-IV criteria for a personality disorder of some kind (Gise et al. 1990).

It remains controversial whether the finding that women with pre-menstrual dysphoric syndrome generally have a high score for neuroticism, as reported by Keye and Trunnell (1986), Hallman et al. (1987) and Chuong et al. (1988), is relevant to the aetiology of the disorder. Schwarzer and van der Ploeg (1987) maintain that such a finding is not a constitutive element of the disorder and that it lacks causal significance because women have the same pattern of manifestations, and to a comparable degree of severity, regardless of whether their neuroticism scores are normal or elevated.

The sociobiographical traits listed by Appelt (1988) as possible determinants of the experience of menstruation, including age, social stratum, education, occupation and experiences of giving birth, were not confirmed as such in the more recent studies by Bergant et al. (in press). The significance of the experience and expression of anger in the causation of various somatic illnesses, which has received attention in recent years, leads directly to the question of the extent to which these might be operative in the pre-menstrual period, a time of increased psycho-endocrinological vulnerability. Despite frequently expressed opinions to the contrary, however, Bergant et al. (in press) were able to show that there is no significant difference in the ways that women with and without pre-menstrual symptoms deal with anger. The same study provided evidence that objectively and subjectively elevated risk factors, such as double stresses in professional and private life, multiple somatisation tendencies, negative experience of menarche and attitude toward menstruation and a tendency towards depressive mood changes are among the predisposing variables for this disorder.

In summary, the research findings discussed above concerning the pathogenesis of pre-menstrual dysphoric syndrome indicate that it most likely has endocrine-biological causes, though the chain of causation may also include psychological components, and supportive psychotherapeutic care of the patient may be necessary.

Genetics

It has been found that 70% of all daughters of patients with pre-menstrual dysphoric syndrome also have this condition, as opposed to 37% of daughters in a control group. When a twin suffers from pre-menstrual dysphoric syndrome, her co-twin is more likely to be affected than a non-twin sister: the concordance rate is 93% for monozygotic twins, 44% for dizygotic twins and 31% for non-twin sisters. The difference between dizygotic twins and non-twin sisters is, however, not statistically significant (Dalton et al. 1987). These data should not be overinterpreted, as the severity of the syndrome was not assessed as part of the study, but they clearly indicate that genetic factors play a role in aetiology and pathogenesis (Condon 1993).

1.2.7 Co-morbidity of Pre-menstrual Dysphoric Syndrome with Other Mental Illnesses

There is still controversy over the extent of co-morbidity of pre-menstrual dysphoric syndrome with other affective illnesses, e.g. with endogenous depression (de Jong et al. 1985; Warner et al. 1991). Bancroft and Rennie (1995), in a prospective study, did not confirm the presence of a relationship, which had been suggested by earlier retrospective analysis, between the occurrence of pre-menstrual mood disturbances and a history of previous manifestations of endogenous depression. On the other hand, the findings by Graze et al. (1990) and Pearlstein et al. (1990) do support the hypothesis of co-morbidity with endogenous depression. The lifetime prevalence of post-partum depression has been reported to be twice as high in women who suffer from pre-menstrual dysphoric syndrome (Pearlstein et al. 1988; O'Hara 1987). Hallman (1986) goes so far as to consider pre-menstrual dysphoric syndrome a variety of endogenous depression with an atypical course.

All clinicians are aware that endogenous depression is exacerbated pre-menstrually (Malikian et al. 1989) and that hospital referrals and suicide attempts cluster in the pre-menstrual period (Dalton 1959). It is

still unknown, however, whether there is a direct (genetic or other) aetiological or pathogenetic relationship between pre-menstrual dysphoric syndrome and manic-depressive illness. Evidence in favour of such a relationship is provided by the studies carried out by Harrison et al. (1985) and Freeman et al. (1990). Harrison et al. (1985) found an elevated prevalence of endogenous depression in first-degree relatives of women with pre-menstrual dysphoric syndrome, and Freeman et al. (1990) found a positive psychiatric family history in 45% and a family history of alcoholism in 40%. Women with pre-menstrual dysphoric syndrome have a significantly elevated prevalence of other affective illnesses – 56% according to Freeman et al. (1990), 60% according to Halbreich and Endicott (1985) and 30% according to de Jong et al. (1985), who further documented a 45% prevalence of psychiatric disorders of all types in these patients.

1.2.8 Treatment

Almost fifty more or less rational methods of treatment for pre-menstrual dysphoric syndrome have been proposed. They vary from psychotherapeutic to pharmacological, hormonal and even surgical interventions. Only a few controlled therapeutic trials have been performed. Thus little can be said to date about the potential efficacy of treatments other than psychopharmacologic and hormonal.

Psychopharmacotherapy

Antidepressants. Among the tricyclic antidepressants, therapeutic studies are available for clomipramine, which acts mainly through a serotonergic effect, and nortryptiline, a relatively specific inhibitor of noradrenaline re-uptake. Clomipramine was described as effective in doses as low as 25–75 mg per day by the research group of Eriksson et al. (1990) and Sundblad et al. (1992, 1993). The dosage of nortryptiline is given as between 50 and 125 mg per day, but therapeutic successes have been reported at doses as low as 10 mg per day (Harrison et al. 1989). The dosage of these tricyclic antidepressants may be titrated upwards or downwards depending on the severity of disease manifestations; it is clear that lower doses are generally required than for endogenous depression. Desipramine is also effective in pre-menstrual dysphoric syndrome, but it is not as effective as the selective serotonin re-uptake inhibitors. A study by Freeman et al. (1996a) revealed that the constellation of adverse effects of desipramine, which is typical of the tricyclic antidepressants, is less well tolerated by this group of patients than by those with endogenous depression. Other 5-HT_2 agonists have been found to be less effective in pre-menstrual dysphoric syndrome, with the exception of nefazodone (Freeman et al. 1994), which also inhibits serotonin re-uptake and is reported to have a good therapeutic effect, including an anxiolytic effect (Eriksson et al. 1995; Freeman et al. 1996a).

Selective inhibitors of serotonin re-uptake are reported to be effective in 60%–70% of patients with pre-menstrual dysphoric syndrome (Steiner et al. 1995) and are the treatment of first choice (Barnhart et al. 1995). Their effectiveness has been very well demonstrated scientifically; 20 mg daily of fluoxetine has been found to be effective (Rickels et al. 1990; Stone et al. 1991; Menkes et al. 1992; S.H. Wood et al. 1992; Su et al. 1997). A comparable effect can be achieved with sertraline at a dose of 50–150 mg per day (Freeman et al. 1996a; Yonkers et al. 1996, 1997) or paroxetine at 10–40 mg per day (Eriksson et al. 1995; Sundblad et al. 1997). Fluvoxamine, at a dose of 100 mg per day, leads to a significant improvement of pre-menstrual dysphoric syndrome manifestations (Freeman et al. 1996b).

A possible advantage of sertraline over fluoxetine in the treatment of pre-menstrual dysphoric syndrome is its shorter half-life, which facilitates the temporal restriction of treatment to the late luteal phase of the menstrual cycle; unlike tricyclic antidepressants, selective serotonin re-uptake inhibitors have been shown to be therapeutically effective when taken only during the late luteal phase. According to a recent study by Steiner et al. (1997), 75% of patients respond to intermittent treatment with fluoxetine. Undesired effects, which may be caused by active metabolites and may not appear until after menstruation, can thus be kept to a minimum (Halbreich and Smoller 1997; Lenzinger et al. 1997). Intermittent treatment with selective serotonin re-uptake inhibitors, i.e. treatment during the luteal phase only, is superior to continuous treatment throughout the menstrual cycle (Sundblad et al. 1993), particularly with regards to accompanying somatic manifestations (breast swelling, abdominal complaints).

With respect to physical symptoms, noradrenaline and serotonin re-uptake inhibitors appear to work equally well (Sundblad et al. 1992; S.H. Wood et al. 1992; Sundblad et al. 1993). Tricyclic antidepressants also continue to have a place in the pharmacological treatment of pre-menstrual dysphoric syndrome.

With respect to side-effects, it should be noted that patients with endogenous depression suffer more from their condition than patients with pre-menstrual dysphoric syndrome and are thus more likely to tolerate side-effects of medication. For this reason, too, selective serotonin re-uptake inhibitors are the agent of first choice for pre-menstrual dysphoric syndrome. Nevertheless, they are not entirely free of side-effects:

many patients complain of sleep disturbances that first appear when the medication is taken. Disturbances of sexual function that may occur in connection with pre-menstrual dysphoric syndrome, such as loss of libido and anorgasmia, may be positively influenced by the use of selective serotonin re-uptake inhibitors, but may also be brought on or exacerbated by it.

The psychostimulant and serotonin agonist fenfluramine has been reported to alleviate food cravings and to have an antidepressant effect at a dose of 15 mg twice a day. Nevertheless, it seems to be less effective than other selective serotonin re-uptake inhibitors against other manifestations of the disorder, such as lability of affect with increased irritability, tearfulness etc. (Brzeznski et al. 1990). It is not on the market in all countries and has been removed from the market in some.

The listing of antidepressants in this section is not meant to imply that others not mentioned here are ineffective, but there are as yet no studies on the use of other agents in the treatment of pre-menstrual dysphoric syndrome.

There are, surprisingly, no controlled studies on the treatment of pre-menstrual dysphoric syndrome with monoamine oxidase inhibitors. At present, only case reports on the two irreversible MAO inhibitors phenelzine and tranylcypromine are available. Nothing is known concerning the possible effectiveness of the reversible MAO inhibitor moclobemid in pre-menstrual dysphoric syndrome.

Anxiolytics and Psychostimulants. Benzodiazepines have been prescribed for many years as a treatment for pre-menstrual dysphoric syndrome. An inadequate number of therapeutic trials were available until the introduction of the atypical benzodiazepine alprazolam (Harrison et al. 1987), which has both anxiolytic and antidepressant effects. Most, but not all studies (Schmidt et al. 1993) indicate that 4 mg per day is effective when taken starting 6–14 days before menstruation. The dose should be reduced stepwise, by 25%, each day after the onset of the menses, in order to prevent a rebound effect with increased anxiety. Interestingly, alprazolam was found to be less effective in patients who had anxiety disorders and depressive mood changes in the follicular phase, as well as in those who became ill only in the late luteal phase (Berger and Presser 1994). Freeman et al. (1995) found a larger therapeutic effect than with progesterone treatment when alprazolam was given at a dose of only 0.25 mg four times daily. The fact that alprazolam has no clinically active metabolites simplifies the management and temporal restriction of treatment. The currently available studies of intermittent alprazolam use by patients with pre-menstrual dysphoric syndrome suggest that the danger of physical dependence is small (Smith et al. 1987; Harrison et al. 1990).

The partial 5-HT_{1A} receptor antagonist buspirone, which is said to have a good anxiolytic effect, particularly in long-term use, was shown by the studies of Rickels et al. (1989) to be especially effective against the dysphoric components of the syndrome, in the narrow sense of the term. Intermittent treatment is recommended for 12 days before the menses at a dose of 25 mg daily (Rickels et al. 1989). This medication, too, has a short half-life, which facilitates the management of treatment.

The advantage of treatment with alprazolam and buspirone over treatment with most of the antidepressants is that they are very well suited to intermittent use (Freeman et al. 1995), although the selective serotonin re-uptake inhibitors are otherwise more effective (Pearlstein et al. 1997).

Lithium. Lithium prophylaxis against pre-menstrual dysphoric syndrome is ineffective in most cases (Singer et al. 1974). This may be because the blood lithium level varies over the course of the menstrual cycle despite constant dosing and is, in fact, lowest in the pre-menstrual period. Perhaps the subgroup of patients who respond consists of those suffering not from pre-menstrual dysphoric syndrome in the narrowly defined sense, but from a pre-menstrual relapse or exacerbation of endogenous depression or cyclothymia (Sletten and Gershon 1966; Singer et al. 1974; Steiner et al. 1980; de Leon-Jones et al. 1982). It has not yet been scientifically studied whether, as we suspect, the lithium responders are precisely those patients with a positive family history of bipolar affective disorders (Lanczik 1995).

Hormonal Treatment

Case reports and controlled scientific studies of hormonal treatment for pre-menstrual dysphoric syndrome have yielded partly conflicting results. There are many reports of therapeutic success with progesterone. The suppression of ovulation by the continuous administration of gestagen–oestrogen combinations is still considered the most promising method of hormonal treatment for the alleviation, if not total suppression, of disease manifestations (Andersch and Hahn 1981; Graham and Sherwin 1987; Walker and Bancroft 1990). This may be achieved by the use of contraceptives, which suppress ovulation while preserving menstruation, or by the percutaneous application of oestradiol at an anovulatory dose of 0.2 mg daily (Watson and Studd 1990). According to Corney and Stanton (1991), oral contraception improves both the physical symptoms and mental concentration in 31% of patients and causes a deterioration of well-being in 20%. Graham and Sherwin (1992) confirmed this finding and pointed out that

oral contraception is more effective against the somatic than against the psychic manifestations. Maddocks et al. (1986) found no therapeutic effect with progesterone supplementation alone and even, in some cases, an anti-therapeutic effect. Oestrogen supplementation alone, however, was found to improve mental well-being (Herzberg and Coppen 1970; Winston 1973).

Anovulatory cycles may also be obtained by down-regulating pituitary secretion of the gonadotropins with gonadotropin-releasing hormone (GnRH) analogues, which, when given continuously, result in complete hypogonadism. Oestrogen and progesterone must be given in addition. The presence of undesired vegetative side-effects may lead to the usefulness of this method in obtaining the desired mental effect to be questioned. This method is, in our opinion, only of theoretical interest at present; it serves to highlight the importance of the hypothalamic–anterior pituitary–ovarian functional loop for the pathogenesis of pre-menstrual dysphoric syndrome (Muse et al. 1984; Mortola et al. 1991; Hammarbäck and Bäckström 1988).

Two groups recommend hysterectomy with bilateral oophorectomy as a treatment of last resort in rare cases, such as in very severely affected patients with demonstrated non-response to GnRH analogues (Casson et al. 1990; Casper and Hearn 1990).

Psychotherapy

The response to supportive psychotherapeutic measures is said to be correlated with the response to placebo. Yet this disorder, whose origin is clearly pathophysiological rather than psychopathological, subjects many patients to interpersonal stresses, particularly in their relationships with partners and in their sexual lives, which may be accessible to treatment by supportive psychotherapy; a number of authors favour group therapy (Walton and Youngkin 1986). The motto for the treatment of pre-menstrual dysphoric syndrome should thus not be "medical treatment instead of psychotherapy", because, in many cases, medical therapy in combination with supportive psychotherapy is the best option.

Phototherapy

Phototherapy, which has been found to be an effective treatment for the subgroup of seasonally occurring cases of endogenous depression, has also been reported to alleviate the manifestations of pre-menstrual dysphoric syndrome (Parry et al. 1989, 1990, 1993).

Conclusion

In conclusion, treatment with psychoactive medications is the single best therapeutic approach in pre-menstrual dysphoric syndrome. Combinations with hormonal, psychotherapeutic and other adjuvant methods may yield better results than psychopharmacologic treatment alone.

1.3
Menstrual Psychosis (I.F. Brockington)

Menstrual psychosis disorder has the following features:

- Acute onset, against a background of normality
- Brief duration, with full recovery
- Psychotic features, i.e. confusion, delusions, hallucinations, stupor and mutism, or a manic syndrome – not merely pre-menstrual tension or depression
- A circa-mensual periodicity, in rhythm with the menstrual cycle

The first observations appeared in the eighteenth century, and as early as 1827 menstrual mood disorder was used as a defence in filicide (Hitzig 1827). In 1902, von Krafft-Ebing, in his monograph *Psychosis Menstrualis*, proposed a temporal classification. The present classification is a modification of this, organised by onset within the menstrual cycle and stage of reproductive life.

1.3.1 Classification by Timing Within the Menstrual Cycle

Pre-menstrual Psychosis. These psychoses start during the second half of the cycle and often end suddenly at menstrual onset.

Catamenial Psychosis. These begin with menstrual flow.

Para-menstrual Psychosis. These are psychoses with variable timing, always in harmony with the menstrual cycle. An excellent example, with 35 timed episodes, was published by Ewald (1922). There are instances of a systematic change from one epoch to another, e.g. from pre-menstrual to mid-cycle.

Mid-cycle Psychosis. These are comparatively uncommon (Wollenberg 1891).

Epochal Menstrual Psychosis. This is a term introduced by von Krafft-Ebing to denote bipolar psychoses lasting throughout the cycle, with switches linked to menstruation.

1.3.2 Classification by Stage of Reproductive Life

Pre-pubertal Cases. These girls develop circa-mensual episodes before the menarche (Friedmann 1894; Schönthal 1892).

Onset After Childbirth. This has been termed, in the literature on puerperal psychosis, "the menstrual relapse phenomenon". A menstrual psychosis can also begin in the puerperium, without a preceding puerperal episode.

Circa-mensual Psychosis During Amenorrhoea. There are examples of a menstrual psychosis continuing during months in which menstruation failed to appear, and others which occurred only during amenorrhoea.

Menopause. There are possible cases which began after the menopause or oophorectomy.

1.3.3 Features of the Illness

Menstrual psychosis is rare, but perhaps not excessively rare. It is not a specific disease entity. The most typical examples have manifested non-menstrual bipolar disorder at another stage of life. Clinically, it is similar to puerperal psychosis, with manic features, stupor, catatonia, schizo-affective depression and cycloid episodes. The close relationship between these two psychoses, linked to female reproduction, is also demonstrated by women who develop puerperal and menstrual mental illness at different times in their lives (at least 20 in the literature).

A major endocrinological investigation was conducted in Japan (Kitayama et al. 1984). This made the important observation that most closely studied patients had anovulatory cycles.

As for treatment, there are claims of success with progesterone or oral contraceptives, but they are not always effective. Arresting menstruation with danazol or gonadorelin agonists should be tried. In patients with anovulatory cycles, clomiphene, which promotes normal menstruation, has been used. This disorder offers an opportunity for unconventional treatments. They should be prescribed in the context of a long-term study of individual patients, with daily ratings or exact timing of events.

2
Psychiatry of Childlessness and the End of Child-Bearing (I.F. Brockington)

2.1
Infertility

Motivation for motherhood is among the strongest and most universal. Many women unable to bear children suffer greatly from this misfortune: infertility is the most upsetting experience of their lives, and the yearning for children dominates everything. Infertile couples often suffer from self-reproach over sexual indiscretions, abortions, contraception or venereal disease or simply from a tremendous sense of guilt over the fact of infertility itself. It is difficult to deal with the envy felt towards fertile couples.

Infertility differs from other stresses in its duration. The psychological reaction unfolds in stages over years, with initial denial and disbelief, followed by embarrassment, bitterness and recrimination. When treatment begins, there is a cycle of optimism and hope. Sexual functioning comes under strain during the investigation. The discovery of azoospermia is especially stressful. Eventually there is acceptance or resignation.

2.1.1 Artificial Insemination

Artificial insemination using the husband's semen has been available from the late eighteenth century, and insemination by donor (AID) since 1884. The psychological effects of the latter, on the husband and the marriage, seem minimal. Husbands rarely react with jealousy to the birth of the baby, any more than to an adopted child. The proof that the experience is entirely acceptable to most couples is that it is often repeated. One of the principles of treatment is privacy, ensuring that donor and couple never meet and remain ignorant of each other's identity.

2.1.2 In Vitro Fertilisation

In vitro fertilisation (IVF) was first performed in 1978. The success rate is low. Some women, however, report satisfaction at having tried everything possible; it is easier to cope with infertility, and they are able to focus their minds on adoption.

2.1.3 Surrogate Motherhood

Surrogate motherhood is another innovation, with psychological complications. It has two meanings:

- A woman contracts with a couple to be inseminated (artificially or naturally) with the husband's semen and to surrender the child to the genetic father and adoptive mother. The surrogate provides both oocyte and womb and is a substitute spouse.
- The wife is the genetic mother, donating a fertilised oocyte to the surrogate gestational mother. This method, involving IVF and embryo transfer, is the only method of enabling a wife with working

ovaries, but without a uterus, to have a child which is genetically her own. The first pregnancy by this method was reported in 1989.

Surrogate pregnancy has stirred up an intense ethical debate. Apart from religious objections, there is concern about the physical and psychological consequences for the gestational mother, who endures the discomforts and complications of pregnancy and may bond to the child during gestation. Even if she is a friend or a relative, she may be under psychological coercion. There are endless opportunities for custody disputes, and there may be complications for the laws of parental responsibility and property.

2.2 Pseudo-cyesis

When a woman believes herself to be pregnant and develops symptoms and signs of pregnancy, this is called pseudo-cyesis. Bivin and Klinger wrote their classic monograph in 1937.

The differential diagnosis includes the following:

- Delusions of pregnancy (Vié and Bobé 1932), in which there are no somatic signs of pregnancy. This is a common delusion, which can occur in a variety of psychoses. There is often a history of hallucinatory intercourse. Delusions of pregnancy also occur in men, with and without the somatic phenomena of pseudo-cyesis.
- Simulated pregnancy – for social, mercenary or legal purposes, e.g. to escape the capital penalty. A famous example is Olympe de Gouges, who wrote one of the first feminist tracts, *The Rights of Woman*, in 1791; in spite of her claim to be pregnant, she was guillotined.

The clinical features include the following:

- A firm belief in the pregnancy, usually lasting until a false labour at 9 months, after which the disorder usually resolves.
- Amenorrhoea
- Morning sickness and pica
- Enlargement of the breasts and nipples, even a discharge of colostrum
- Abdominal enlargement, caused by muscular contraction, tympanites, fat or pathological lesions, e.g. ascites or fibroids, but without effacement of the navel
- An illusion of foetal movements
- Enlargement of the uterus to the size of a 6-week pregnancy

Modern diagnostic tests, such as ultrasound, have greatly reduced its frequency.

The psychological basis is usually an intense desire for children, especially in older childless women. In some cases, however, a guilty fear of pregnancy has been the background; this has occasionally led to dangerous attempts at abortion in women who are not pregnant (von Neugebauer 1912; Percheval 1911).

Pseudo-cyesis is a demonstration of the influence of psyche over soma, mediated by hormonal secretion. It occurs in dogs, cattle and rodents, due to persistence of the corpus luteum. This explains the breast changes, moderate uterine enlargement and secretory endometrium. In humans, this has also been demonstrated by surgical examination (Courrier et al. 1927; Reeb 1933). Persistence of the corpus luteum, however, is not the only explanation. Hormonal measurements have been made in at least 30 patients and have shown chronic anovulatory states, hyper-prolactinaemia and hyper-androgenism as alternatives.

These women require psychotherapy. Simply revealing the diagnosis is unsatisfactory because the patient may go to another doctor with the same symptoms or develop a recurrence. The underlying conflicts must be explored, helping the patient to face the fact that she is not pregnant. Some require antidepressant medication.

2.3 Sterilisation

Women can be prevented from bearing children by various operations on the uterus and fallopian tubes. None of these methods is completely effective, with failure rates of about 0.5%, with an increased risk of ectopic pregnancy. Sterilisation can be reversed by repairing the fallopian tubes, and, with modern microsurgery, this is often successful. As sterilisation becomes more common, it is carried out in younger women. As the divorce rate climbs, its reversibility is becoming an increasingly important issue.

The indications are contraceptive, medical, eugenic and psychiatric. Contraception is by far the commonest, and sterilisation is the most effective and widespread contraceptive method. The medical grounds include diabetes, rhesus incompatibility, hypertension and renal and heart disease. Eugenic sterilisation for hereditary medical or psychiatric disorders is rare. Sterilisation of women with severe learning disability has been legal in certain countries at certain times. This remains a controversial issue, which is becoming more important. With the closure of institutions and the greater tolerance of sexual activity outside marriage, there is an increased risk of pregnancy in women with severe learning difficulties, with the spectre of inherited disorders and problems in moth-

ering. Yet these women greatly desire children, not having the same resources to compensate for lack of them.

Some early studies reported a high frequency of depression, but modern prospective studies have failed to confirm this. Cooper et al. (1982) in Oxford interviewed 201 women 4 weeks before non-puerperal tubal sterilisation, done for contraceptive reasons; 190 were re-interviewed 6 months later and 193 again 18 months after the operation. The number with psychiatric illness fell from 21 before the operation to nine 6 months later and then rose again to 18 at 18 months. Not surprisingly, the presence of psychiatric disorder before the operation was a predictor of its continued presence. The World Health Organisation has conducted a collaborative project to assess the effects of sterilisation in different cultures. In the report from the Nottingham field centre (Bledin et al. 1984), there was little evidence of psychiatric disorder either before or after sterilisation: nine out of 138 sterilised women were "cases" before the operation, and there were only three new cases 6 weeks afterwards and four more at the 6-month assessment, less than in the control group.

Regret is more common in the following groups of women:

- Those with fewer children (Ekblad).
- Younger women.
- Those in whom sterilisation was the condition for a termination. This barbaric and punitive practice used to be the rule in some countries. It is doubtful whether sterilisation should ever be carried out simultaneously with termination, because there is little time for discussion, and a woman may take a rash decision in the extremity of her anxiety about an unwanted pregnancy.
- Those sterilised at parturition. It is often difficult to make a balanced judgement when a new child is about to be born.
- Those under external pressure.
- Those with learning difficulties.
- Those sterilised for medical reasons. Particularly poignant are the feelings of women who have undergone tubal ligation because of an inherited disorder, which has since become diagnosable by amniocentesis.
- Those with psychiatric illness. Psychiatric illness often impairs judgement and, after recovery, decisions may be regretted.
- Those who seek sterilisation in a context of marital disharmony. In almost all cases undertaken to save the marriage, this was unsuccessful.
- Those with religious scruples.

Several have commented on a special group of women with an intense fear of pregnancy (tocophobia). Most of these women were greatly relieved by sterilisation (Binder 1937; Hoppeler 1955).

2.4 Hysterectomy

Hysterectomy is one of the commonest operations, performed in about 10% of women. Among the usual indications are menorrhagia due to fibroids and various types of abdominal, pelvic or back pain. It would not be surprising if the loss of the womb had psychological effects on feminine identity. In younger women, the loss of fertility can be a source of discontent. Hysterectomy is sometimes performed expressly to achieve sterilisation, having the advantages of greater certainty and the elimination of a non-functioning organ which could later become diseased.

Twenty years ago, there was a widespread belief that hysterectomy caused depression. Since then this idea has been thoroughly and systematically refuted. Several prospective investigations have shown that mental health improves after hysterectomy (Gath et al. 1995; Martin et al. 1980). The ranks of women with "post-hysterectomy" depression are swollen by those who seek a surgical remedy for psychosomatic complaints.

Apart from depression, hysterectomy may have an effect on libido, but on the whole this is also a myth. Authoritative prospective studies in St. Louis and Oxford showed an increase in the frequency of intercourse and of enjoyment.

3 Psychiatry of Pregnancy (I.F. Brockington)

3.1 Pregnancy Adjustment

The psychopathology of pregnancy needs to be understood in terms of the adjustment which all women have to make when they conceive. Pregnancy is not only a biological event, but also an adaptive process (Cohen 1988). A pregnant woman must not only carry the baby through safely, but square up to the sacrifices that motherhood demands. She must ensure the acceptance of the child by the family, develop an attachment to the baby within and prepare for the birth. She must adjust to the alteration in her physical appearance and develop a somewhat different relationship with the child's father.

Many pregnancies are unplanned and not initially welcomed. Many react to conception with grief and

anger. A recent random sample of English mothers showed that 44% of pregnancies were unintentional, including 17% which ended by legal abortion. The range was from 80%–84% intended pregnancies in married women aged 25–29 with one child born less than 4 years previously to 26% in the unmarried (Cartwright 1988).

The planning of pregnancy and its acceptance are two different things. The fact of planning does not guarantee acceptance; 6%–12% of those who planned their pregnancies switch to rejection. Usually, the change is in the other direction, and there is often an immediate acceptance of an unplanned pregnancy. Even if the initial response is negative, it is usually replaced by gradual acceptance. In a small proportion, however, rejection continues to the end. Although there is a shift to acceptance, this may be superficial, involving an ambivalence which continues after the child is born.

Pregnancy has a profound effect on the relationship with the child's father. At every stage, this relationship is of the highest importance. A pregnant woman needs increased attention and care and is sensitive to perceived rejection. For many women, the baby is viewed as a gift to the husband. Pregnancy alters other relationships. The need for attention and care extends to family members and friends. Many women become closer to their families of origin and their in-laws. The change in shape is sometimes distressing, but others take pride and pleasure in their pregnant appearance.

Pregnancy is accompanied by medical disorders, and in all there is an interaction between physical and psychological factors. Pica is common: 20% have cravings for inedible substances. Geophagia (eating earth or clay) can lead to iron deficiency anaemia and bowel obstruction. Hyper-emesis can lead to Wernicke's encephalopathy (Dupouy and Courtois 1930; Selitzky 1925). Pre-eclamptic toxaemia is more common in first-time mothers; its dreaded complication was eclampsia, which was sometimes followed by confusional or manic psychoses (see below).

3.1.1 Denial of Pregnancy

Failure to recognise pregnancy is common in the early stages, especially in those pregnant accidentally or for the first time; in a few it continues until delivery (Brezinka et al. 1994). One must distinguish between the following:

- Those who are obese or near the menopause – who simply do not notice the pregnancy
- Those who conceal it
- Those who, against all the evidence, remain obstinately unaware of it

Among the psycho-social factors which predispose, isolation is important because there is no-one to notice and comment on the change in appearance. Another is a moral climate antagonistic to extra-marital sexual relations. Low intelligence may contribute. The psychological mechanisms which enable pregnancy to remain unacknowledged include denial and dissociation. There may also be the belief that the baby had died.

The late discovery of an unwelcome pregnancy carries a small risk of suicide. The mother is also at risk of all those complications of delivery which, with modern obstetrics, have become rare. For the child there are increased hazards at the time of the birth, including neonaticide.

3.1.2 Prenatal Attachment

Recently it has been realised that the mother "bonds" or "affiliates" to the unborn child in a way analogous to the formation of the mother–infant relationship after birth (Arbeit 1975; Cranley 1981; Leifer 1977). She begins to have fantasies about the baby and often talks affectionately to him or her. She may engage the husband and other children in "playing" with the child. She begins to imagine what it will be like to assume responsibility for baby care.

There is a pathology of the affiliative stage. In some mothers, there is minimal attachment even at term. The foetus is viewed as an intrusion, whose movements annoy or distract the mother, disturb her sleep or make her feel ill. When the mother's attitude to the pregnancy is obstinately rejecting, therapists can direct attention to the relationship with the child.

3.1.3 Foetal Abuse

When a mother deeply resents her pregnancy, she may try to harm the foetus. This occurs, with determined intent, in self-induced abortion. It may be a manifestation of rage against the infant within. Condon (1987), who described this phenomenon, stated that it aroused the same incredulity as did child abuse in the 1950s.

It is not only the mother who may "batter" the foetus. Domestic violence is common. Abuse may increase during pregnancy, when kicks and blows are directed at the abdomen, rather than the face. The main factors are sexual frustration, unreadiness for

fatherhood, jealousy, the mother's irritability and substance abuse.

3.2 Mental Illness During Pregnancy

3.2.1 Anxiety

For many mothers, pregnancy is a time of considerable anxiety. The first trimester may involve an anguished decision whether to continue or terminate. Women who have previously miscarried or suffered infant loss reach a peak of apprehension at the time the earlier pregnancy ended. In the third trimester, there is an escalation of anxiety, centred on three main themes – the fear of parturition (tocophobia), of foetal abnormality and of not coping with motherhood.

These anxieties will usually be managed by ventilation and support, but anxiolytic medication can be used cautiously. Benzodiazepines are contraindicated in the last stage, because they can cause prolonged foetal intoxication ("the floppy infant syndrome"), especially in premature infants. Propranolol is best avoided, because of reports of intra-uterine growth retardation and neonatal cardiac and respiratory symptoms.

3.2.2 Depression

Although pre-partum depression has not aroused the same interest as post-partum depression, it is no less common (Cooper et al. 1988). Depression is common in all women in the reproductive age-group, and its frequency during pregnancy may be no more than this general tendency. It can be recurrent: there are at least ten reports in the literature of mothers who have suffered four to 12 attacks with successive pregnancies; some have recovered promptly at delivery, and there is an association with puerperal mania (Brockington 1996).

The frequency of suicide during pregnancy is a vexed question: there are problems about the accuracy of the data – not all suicides are reported to the coroner, not all have necropsies and not all necropsies include an examination of the uterus. Both suicide and pregnancy are often concealed. Nevertheless, there seems to be a difference between studies which appeared at the beginning and end of this century, with higher rates in the first quarter. These showed that about 13% of female suicides were pregnant – rather a high figure, suggesting that pregnancy was a risk factor when illegitimate pregnancy was considered so shameful. More recent studies show lower rates. The study by Weir (1984) outweighs all others in its thoroughness. Weir studied all the depositions for 1696 suicides in central London between 1943 and 1962: 3.9% were pregnant; 40 out of 66 pregnancies were conceived outside marriage – about six times the illegitimacy rate in London at the time. His data suggest that the crisis of adjustment to an unwanted pregnancy was still a factor in suicide in the mid-twentieth century.

Severe prenatal depression is sometimes left untreated, because of fears for the effect of drugs on the foetus. These have been exaggerated. No antidepressive drug is known to have teratogenic effects, nor any other effects on the growing foetus. There are occasional reports of side-effects or withdrawal symptoms in neonates, including heart failure, seizures, hypothermia, laryngeal spasm and retention of urine, so medication is more to be avoided in the last trimester. It is too early to reach conclusions about the safety of the recently introduced serotonin reuptake inhibitors. Electroconvulsive therapy (ECT) is safe, provided that the mother is competently oxygenated during the brief period of anaesthesia. Pregnant women should be screened for rare syndromes of pseudo-cholinesterase deficiency before receiving ECT.

3.2.3 Alcoholism

Pregnancy has a beneficial effect on alcohol addiction, but, if heavy abuse continues, there are severe effects on the foetus. Ethanol is teratogenic – causing the "foetal alcohol syndrome" (Lemoine et al. 1968), whose features include facial dysmorphism due to maxillary hypoplasia. Alcohol's main effect, however, is intra-uterine growth retardation (Ulleland 1972). Exposed infants may have withdrawal symptoms. Learning difficulties are a long-term consequence.

3.2.4 Other Addictions

Cannabis is commonly abused by pregnant women, but probably has no major effects on the foetus. Lysergic acid diethylamide (LSD) may have teratogenic or mutagenic effects. Phencyclidine addiction leads to withdrawal symptoms.

Narcotic addicts, like alcoholics, have multiple emotional and social problems, and most do not seek antenatal care. The infants may be affected by maternal malnutrition and infections such as venereal disease, hepatitis, endocarditis and acquired immune deficiency syndrome (AIDS). Narcotics are not teratogenic, but a high proportion of the infants are of low birth weight, partly explained by prematurity, and partly by

intra-uterine growth retardation. A withdrawal syndrome develops in most babies. The perinatal mortality rate and frequency of sudden infant death (SID) are increased.

In the management, all mothers should have social case work. Other drugs of abuse should be gradually withdrawn. If it is decided to withdraw heroin, this should be done in the second trimester, replacing it by methadone. Naloxone should only be given to save life after an overdose, because it precipitates a foetal abstinence syndrome. After birth, the infants should be kept in hospital for 14 days. Respiratory depression can be treated by naloxone, seizures and withdrawal symptoms by diazepam.

Cocaine may be teratogenic, causing genito-urinary abnormalities, but the evidence is conflicting. Its main effects are cardiovascular. It causes uterine vasoconstriction, and this can lead to abruptio placentae (separation of the placenta from the uterine wall). There is intra-uterine growth reduction, and premature labour is common – even commoner than with other addictions. The infants may suffer cerebral infarction. There is a withdrawal syndrome, but this is less severe than with narcotics. There is evidence of an increased risk of SID.

3.2.5 Eating Disorders

There are psychological and somatic reasons for an antagonism between pregnancy and anorexia nervosa; nonetheless, the majority of women with anorexia nervosa recover, with return of menstruation when the weight reaches about 80% of the standard weight. There are numerous case reports and several long-term studies establishing the fact that many women with a history of anorexia nervosa give birth to children in the normal way. The overall effect on fertility has been quantified by a 12-year Danish study: the average number of children (0.6) was about one third the usual figure (Brinch et al. 1988). A minority become pregnant while in the throes of the disease. If the mother continues to restrict her diet, the foetus may suffer from malnutrition, with reduced abdominal girth and low birth weight. Occasionally, it has been necessary to rescue the infant by elective caesarian section. When mothers are actively anorexic, their children may be affected by their attitudes to feeding, suffering stunted growth.

Bulimia nervosa is often improved by pregnancy (Morgan et al. 1999). The pressure of the enlarging uterus on the stomach may make bulimia more difficult. Some mothers report that the presence of the baby inhibits bingeing. About half relapse after delivery. Bulimic mothers also sometimes show deviant mothering, ignoring their children while overeating or vomiting, or restricting food supplies.

3.2.6 Obstetric Factitious Disorder

Self-induced illness behaviour can extend into the obstetric domain (Goodlin 1985; Jureidini 1993). Women may induce bleeding to simulate threatened miscarriage, placenta praevia or post-partum haemorrhage. They may simulate rupture of membranes in order to precipitate an early delivery. Others have been caught manipulating instruments, e.g. the external tachodynamometer. Two patients even attempted to simulate hydatidiform mole, by adding human chorionic gonadotrophin to blood samples.

3.2.7 Pre-partum Psychosis

Numerous asylum surveys have testified to the lower frequency of psychosis during pregnancy than after delivery. This has been confirmed by two large surveys. Paffenbarger, in his study of patients admitted to hospital in Ohio (Paffenbarger and McCabe 1966) found an overall rate of 0.7 per 1000 population per year, compared with 4.0 per 1000 for post-partum psychosis and 3.6 per 1.000 in non-child-bearing women. Kendell, in Edinburgh, had the advantage of linking obstetric and psychiatric case registers (Kendell et al. 1987). A study of 54,087 births showed rates for psychosis of 2.1 per month before conception, 2.0 during pregnancy, 51 in the first post-partum month, 25 in the second, 13 in the third and four thereafter; thus the risk of admission with psychosis was unchanged during pregnancy and much lower than after childbirth.

Pregnancy probably has no effect on chronic delusional states, but there is a beneficial effect on menstrual psychoses, and possibly bipolar and cycloid psychoses. Acute manic and cycloid episodes occur during pregnancy, and some seem remarkably similar to puerperal psychosis. Indeed, they have been observed in women with a history of typical puerperal psychosis (Brockington 1996).

It is, on the whole, safe to prescribe neuroleptic agents during pregnancy. Phenothiazines and butyrophenones are not teratogenic. The main (but infrequent) hazard is the appearance of sedation and extra-pyramidal symptoms in the newborn. Lithium is relatively dangerous. It may cause cardiac malformations: at least 12 cases of the rare Ebstein's anomaly have been reported. As delivery approaches, reduced renal clearance can result in toxicity on normal doses: five cases of alarming blood levels (up to 5 m*M*/l) have been reported, with coma and convulsions in the

mother. Even at normal blood levels, babies exposed to lithium have suffered lethargy, hypotonicity and other effects. Carbamazepine has been associated with rather high rates for congenital abnormality, and sodium valproate with spina bifida.

3.2.8 Obstetric Liaison Services

In view of the complexity of the psychological response to pregnancy, and the frequency of anxiety, depression and other psychiatric disorders, comprehensive obstetric and psychiatric services should include a liaison between the two specialities. In addition to the diagnosis and treatment of the pre-partum psychiatric disorders described above, the high level of supervision in the antenatal clinics can be exploited to detect those vulnerable to post-partum illness. Since many women wish to avoid the stigma of psychiatric referral, it may be best to employ specially trained midwives, whose role would be to detect and treat milder disorders and to refer those with more serious illness or vulnerability.

3.2.9 Psychopathology of Parturition

Childbirth can be one of the severest of human ordeals, and it would be surprising if its torments did not result in acts of desperation. They are fully described in the older literature. They are much rarer now, but may still occur where obstetrics is primitive, or pregnancy denied. Auto-caesarian section has been described. There are several reports of completed suicide, by defenestration, drowning, hanging or coal gas poisoning. Rage attacks are also described. They can endanger the foetus and may be a factor in some cases of neonaticide. Confusional states are rare, but well documented (Engelhard 1912; Kirchberg 1913; Sarrat 1911).

3.3 Infant Loss

The child may be lost under a wide variety of circumstances:

- Termination of pregnancy, at the behest of the mother
- Miscarriage, ectopic pregnancy and late termination of a wanted child for medical reasons
- Foetal death in utero, stillbirth, neonatal death and SIDS
- Relinquishment to adoption

3.3.1 Termination of Pregnancy

Termination of pregnancy is condemned by the Roman Catholic Church and Islam. During this century, and especially the last 30 years, there has been a general movement towards liberalisation.

The indications include the following:

- Medical grounds – to preserve the health and life of the mother
- Humanitarian grounds – when pregnancy has resulted from rape or incest
- Genetic or eugenic grounds – where there is a risk of congenital abnormality
- Psychiatric grounds – e.g. to prevent suicide
- Social grounds – because pregnancy is untimely and disruptive, to the disadvantage of the unborn child
- On demand – in some nations, it is believed that women should be free to decide on pregnancy and maternity

There has been an intense debate on the validity of psychiatric indications; this turns on the psychiatric consequences of a refusal to terminate. Suicide threats are common, but are rarely carried out; nevertheless, there can be no doubt that unwanted pregnancy is a factor in completed suicide. A history of puerperal psychosis is not an indication, because episodes are equally likely to follow abortion. There are other, arguably more serious, puerperal complications such as mother–infant relationship disorders, which are more common after unwanted pregnancy. These can be avoided by adoption, but the psychological effects of relinquishment are not negligible.

The psychological effects of termination have been thoroughly explored. Most of those who voluntarily abort suffer no adverse effects, either in the short or long term. There is often relief, even euphoria, and a reduction in anxiety, depression, anger, guilt and shame. A minority experience regret and self-reproach over the "murder" of the baby; some feel like criminals, worry about punishment or the nemesis of sterility or future congenital malformations. A few develop clinical depression (Edelberg and Galant 1925), which may ensue after the next pregnancy. There is evidence that manic or cycloid episodes, similar to puerperal psychosis, occur after abortion (David 1985).

In order to minimise the psychological risk, prudent decision-taking is of the essence, and counselling has a valuable role. The most difficult part of the experience is loneliness and isolation. Many do not inform their parents and, when they do, endure negative responses and unwelcome pressure. The attitude of the child's father is crucial. It is axiomatic that a woman should make her own decision – one of the most difficult that

she will ever take, with profound long-term consequences. It often has to be taken hastily, in an atmosphere of conflict and turmoil. The best outcomes are found when a woman makes her decision in a context of respect and support from partner, parents, friends or counsellor.

3.3.2 Miscarriage

Miscarriage is a common event and occurs in perhaps 40% of all conceptions, but only 10% after pregnancy becomes apparent through amenorrhoea or other signs. Its emotional consequences are not trivial, but comparable to foetal death in utero or perinatal death – less severe, because there has been little time for attachment to the newly conceived, but still the loss of greatly desired child. The event itself, with the clots and foetal tissue passed suddenly and painfully, may be disturbing. Some of the psychological symptoms may resemble post-traumatic stress disorder, with perseverative and intrusive re-experiencing ("flashbacks") and nightmares. There is a sense of failure, guilt and anger. The incidence of depression is four times the rate found in the general population (Friedman and Gath 1989). There may be depressive episodes at the time of expected delivery, anniversary reactions and an increased risk of post-partum emotional disorder after a later normal delivery.

Helping a mother who has suffered a miscarriage is a variant of grief therapy, in which the intense distress can be shared, and the sadness, guilt and anger ventilated with support. Mental health workers should never imply that spontaneous abortion is anything but a significant loss.

Late termination for medical reasons, although a deliberate intervention, is psychologically similar to miscarriage and to foetal death in utero. The pregnancy has advanced beyond the stage of quickening. These mothers experience intense anxiety at the time of the tests and require a more radical surgical intervention. Depression is common, and grief long-lasting (Iles and Gath 1993; Lloyd and Laurence 1985). All these women required counselling, before and after the termination.

3.3.3 Foetal Death In Utero, Stillbirth, Neonatal Death and Sudden Infant Death

Reactions to foetal death in utero, stillbirth, neonatal death and SID are generally more severe than to miscarriage, and each has its special characteristics. When the baby dies in late pregnancy, the mother carries a corpse within her and must also undergo a futile labour. If it dies during labour, the loss is sudden and shock pronounced, with a strong sense of unreality (Bruce 1962). When the child dies in the first week, the parents have to endure a period of great anxiety, with dwindling hope; they may be involved in the decision to switch off the respirator and witness the child dying. The later death of an infant, when the maternal response is fully developed – especially SID – is at the very top of the catalogue of calamities; there is no warning or preparation, and the death is followed by a forensic investigation (Cornwell et al. 1977; Smialek 1978).

When helping the parents, the principles of therapy are as follows:

- Honesty and openness in communication. The admission of errors is delicate, but the parents' guilt should not be reinforced by refusal to accept responsibility. Recrimination, litigation or querulant reactions are common. Staff should accept this as normal and try not to be defensive. Repeat interviews are recommended because the parents may be too stunned to register the first time.
- The mother will often be helped by seeing and holding the dead baby. A photograph should be taken and kept, as well as other mementoes.
- After stillbirth, most mothers prefer segregation, e.g. in a single room with their husband, and early discharge. The postnatal examination should be private.
- The mother should be visited by a member of the primary care team. A lactating mother may need bromocriptine or to donate milk to a milk bank. Hypnotics may help mothers troubled by insomnia. The doctor should be alert for secondary depression.
- The bereaved mother needs to share her distress with a sensitive and sympathetic person, with the time and interest to listen and guide. This support will often come from husband, family and friends. If they are unable to help, professionals, especially chaplains or nurses, should step in. The aim is to help the mother to grieve and accept her loss. After SID, the parents need much support during the police interrogations, forensic tests and inquest. Self-help groups and voluntary agencies are invaluable for some mothers.
- All parents want to know why the baby died. A simple, rational explanation will relieve fear, misconception and guilt. The necropsy can help, but parents should be warned that often no explanation is found. Necropsies in SID are specialised: the pathologist can play a vital psychological role and should be available for discussion.

- The dignity of naming and a burial ceremony are often helpful. The dead child remains one of the family and, in imagination, continues to grow up.
- No doctrinaire advice can be given about the timing of the next pregnancy, which is a personal decision. Increased anxiety during pregnancy and the puerperium can be expected.
- To helping the grieving sibling, the routine and rhythm of family life should be disturbed as little as possible. The parents should not be afraid to show their emotions – it is best to acknowledge how sad they feel and how much they will miss the baby. They should try to give a factual account of what happened, avoiding euphemisms. It is important to reassure the children that they are not responsible, that they will not lose the love of the parents and that neither they, nor their parents, are in imminent danger of death. The child can be helped to grieve by looking at pictures of the dead sibling, attending the funeral and visiting the graveyard.

To avoid these severe and prolonged psychological effects, a relinquishing mother needs counselling during the pregnancy (Harvey 1977). The aim is to emerge from the experience with self-respect and dignity. After delivery, the mother should be encouraged to see, hold and even breast-feed the infant; this facilitates the grieving process. Photographs of the infant should be filed. The mother should not normally be housed with mothers who are keeping their babies. Follow-up counselling should be continued for at least 6 months. The mother may wish to be put in touch with a society for relinquishing parents. Adoptive parents should accept any gift or token of the natural mother's love. Information on the outcome of the child should be available. Some birth-mothers wish to provide up-to-date information about themselves, so that the child knows that they are now respected citizens. A recent innovation is the practice of "open adoption", in which both sets of parents meet. There is even "continuing open adoption", which means that they remain in contact over the course of the child's development.

3.3.4 Relinquishment

Adoption is an ancient custom, allowing parental rights to be transferred from the birth-mother to another woman. It used to be the main way of satisfying the longing for children and dealing with accidental pregnancy; but there has been a great social change in the last 20 years. In spite of the huge increase in extra-marital births, the number of adoptions is falling steadily. This is not due to spectacular improvements in the treatment of infertility, reducing the demand, nor to the relaxation in the abortion laws, reducing the supply, but a new tolerance of single motherhood.

Relinquishment can be one of the most stressful events, comparable to bereavement and divorce, and may lead to intense depression and sadness (Rynearson 1982). Instead of understanding and support, there is often loneliness and ostracism. Time is no healer of the grief over a surrendered child. The sense of loss fluctuates – worse at anniversaries and at the birth of later children. Some harbour deep, unresolved feelings for years. There is often a fantasy of reunion or restitution. As time goes on, there is a new component: the mother knows that the adult child may seek to find her, and there is the hope that this event, which she cannot influence, may happen. Adopted children often seek out their parents and the wish for reunion is reciprocated by many birth-parents, as shown by the growth of organisations for relinquishing parents, helping them to trace their offspring.

4
Psychiatry of the Post-partum Period (I.F. Brockington)

4.1
Normal Puerperium

For many or most mothers, giving birth is a supreme moment, and euphoria or elation is common. Some may be too excited to sleep. These feelings of peace, fulfilment and accomplishment help to sustain mothers during the weeks of strain which follow. Prolonged euphoric reactions, lasting a week or more, are probably mild episodes of puerperal mania and are often followed by depression.

Newly delivered mothers have to face a number of challenges, including the following:

- Physical exhaustion and painful sequelae of pelvic trauma.
- Breast-feeding, which has many advantages, but is often difficult to establish.
- Insomnia. Sleep deprivation is a cause of irritability and increased volatility.
- The maternity blues. About half experience a brief period of dysphoria, usually lasting a few hours, occurring between the third and the fifth day. Mothers may become sensitive to minor rebuffs and are surprised and puzzled by their uncharacteristic weeping.

- Recovery of normal figure and attractiveness, which may be threatened by weight gain and stretch marks.
- Loss of libido. Episiotomy and vaginal trauma often cause dyspareunia; nevertheless, sexual relations are usually resumed within 1–3 months, though reduced in frequency and with a delayed return of orgasm. For this and other reasons (e.g. jealousy), the marriage may come under strain.
- Social privation, including loss of employment, income and leisure, confinement to the house, and boredom.

With this background of rapid biological, social and emotional transition, it is not surprising that a wide variety of psychiatric disorders occur in the puerperium. Indeed, the psychiatric complications of childbirth are more numerous and complex than in any other human situation. They fall into four main groups – psychosis, disturbances of the mother–infant relationship, various anxiety, obsessional and stress-related "neuroses" and depression.

4.2 Post-partum Psychoses

4.2.1 Classification

A variety of different psychoses can begin after childbirth. There are a group of rare organic disorders which occur soon after delivery. They include the following:

- Idiopathic confusional states, similar to those seen during parturition
- Exhaustion to the point of stupor (Tott 1844)
- Delirium tremens
- Post-eclamptic delirium – usually an acute organic syndrome followed by profound amnesia (Kutzinski 1909; Olshausen 1891), but occasionally with manic features
- Infective delirium

Other psychoses include the following:

- Psychogenic psychosis, e.g. delusional jealousy, which can arise as an understandable reaction to changing relationships and the quiescence of sexual life. Psychogenic psychoses have also been reported in fathers and adoptive mothers (Trixler and Jádi 1981).
- Manic, cycloid or depressive psychosis.

The term "puerperal psychosis" (synonyms: "puerperal insanity", "puerperal mania") usually refers to this last group. The first clear description was written in 1797 by the German obstetrician Osiander. Since Fürstner's description of *hallucinatorische Irresein der Wöchnerinnen* (Fürstner 1875), it has always been recognised that acute psychoses of cycloid type ("amentia") are associated with the puerperium.

4.2.2 Clinical Features of Puerperal Psychosis

Puerperal psychoses are acute, rapidly reaching a climax of severity. The onset is usually between 2 and 14 days after delivery. There is no specific symptom, syndrome or course, but there is evidence of a close relationship with manic depressive (bipolar) disorder:

- About 35% of episodes meet criteria for manic or schizo-manic illness (Brockington 1996).
- The bipolarity of the illness is often seen in the course of the episode or recurrences.
- There is a high puerperal attack rate in women with a history of non-puerperal mania – about 20% (Bratfos and Haug 1966; Kendell et al. 1987; Reich and Winokur 1970).
- There is often bipolar disorder in the families.

Mania is particularly severe, often with "schizo-affective" features. The delusions cover the whole gamut of morbid ideas. Verbal hallucinations, thought insertion, echo phenomena, thought broadcasting and "made" impulses may occur, accompanied by ideas of control or possession. Catatonic features are common. There is often an apparent confusion, bewilderment, perplexity or "dreamlike delirium". Occasionally this is seen in pure form – Marcé's (1858) "transitory intellectual enfeeblement".

4.2.3 Therapy and Prevention

In the days before modern therapy, the median duration was about 6 months, but since ECT and neuroleptic medication, it has been reduced to a few weeks. Neuroleptic agents should be used with caution, because of severe extra-pyramidal side effects, including the neuroleptic malignant syndrome. Lithium has been used increasingly, since the link with manic depressive psychosis was recognised. There is evidence, from retrospective studies, for its prophylactic value in women at high risk (Stewart et al. 1991). It may have adverse effects on breast-fed infants, though there is only one reported instance of a transient adverse reaction. ECT is effective in manic as well as depressive patients.

The location of treatment is an issue. Hospitalisation can have serious effects. Even in the nineteenth century, thoughtful clinicians urged that this disorder be treated at home, where the patient can maintain her role as wife, homemaker and mother and her relationship with the newborn. If hospital admission is

necessary, there are great advantages in conjoint mother and baby admission.

4.2.4 Aetiology

The frequency is about 1 per 1000 pregnancies at a threshold of hospital admission within the first trimester (Kendell et al. 1987; Terp and Mortensen 1998). The distribution is worldwide. There is no link with twin pregnancies, breast-feeding, single parenthood or stillbirth. Episodes have occurred after short gestation (22–28 months).

Puerperal psychosis belongs to a group of biological brain disorders with high heritability and an inborn tendency (diathesis) to develop psychotic episodes throughout life. Thus the problem of its causation can be broken down into three subsidiary questions:

1. What is the nature of the diathesis?
2. What determines the clinical polarity – mania, depression, cycloid features or switching from one to another?
3. What is the trigger which provokes the episode?

The first two questions are not unique, but concern the larger problem of bipolar disorder. The third is pertinent – indeed, this is where the study of reproduction-related illness can make its special contribution. The clinical facts, however, suggest not one, but several, pregnancy-related triggers (Brockington 1996):

- Abortion
- Pregnancy itself, especially the last trimester
- The early puerperium, especially the first 10 days
- Post-partum menstruation
- Menstruation in general (see menstrual psychosis)
- Weaning

These triggers can be added to the list of other biological events which trigger bipolar episodes, including surgical operations, adrenocortical steroid treatment and seasonal climatic changes. There are instances of the combination of these triggers in the life-history of individual women (Esquirol 1818).

The evidence for a link between puerperal psychosis and menstruation implicate the steroid hormones. Cortisol, oestrogen and progesterone are structurally related. Oestrogen has many interactions with neuro-transmitter systems, including dopaminergic, noradrenergic, adrenergic, serotonergic and cholinergic receptors. Progesterone also interacts with neurotransmitter amines. Many other hormones are greatly affected by pregnancy, including gonadotrophic hormones, prolactin, oxytocin and neurophysin, thyrotropin, corticotrophin and corticosteroids, renin, angiotensin and human placental lactogen. It is too early to determine which hormone and receptor is involved. Neuro-scientific studies of this psychosis are in their infancy, but this is one of the few psychoses whose onset can be predicted within a short time interval, so that it is feasible to study its pathogenesis.

4.3 Mother–Infant Relationship Disorders

Just as the emerging relationship with the foetus is an important pre-partum psychological process, so the growth of the mother–infant relationship is a vital post-partum development. In popular parlance, this is often called "bonding". The mother–infant and infant–mother relationships are two different things, the infant's attachment coming later, after 7–8 months. The mother–infant relationship consists essentially of ideas and emotions, aroused by the sight, sound, feel and memory of the infant, which find their expression in affectionate and protective behaviour. Its immense power is revealed in self-sacrifice and the pains of separation. The mother's emotional response enables her to maintain a never-ending vigilance and endure the exhausting toil of the nurture of the newborn.

There is no "critical period" in the development of the maternal response – no special magic in immediate contact with the newborn. Close proximity from the start ("rooming-in") confers confidence in mothering skills. In the development of maternal emotions, the infant plays an important part. While in the womb, hearing is acute (Peiper 1924). Within a month of birth, babies can discriminate speech. From 6–12 weeks, they react preferentially to human speech and singing (Bühler et al. 1927). At birth, they are programmed to respond to the human face. At 36 hours, they will imitate facial expressions, and by 3 weeks this is advanced. Eye-to-eye contact mediates the interaction (Robson 1967), and gazing becomes an absorbing activity on both sides. The baby's smile is another catalyst; by about 6 weeks, it is linked to the human face, and the mother will devote time to bringing it out. Even before infants can babble, videotape studies have shown how they contribute to a dialogue. At an early stage, mother and infant learn to stimulate and reinforce one another – "to dance together".

Sometimes the maternal response is immediate, primed by her pre-partum affiliation, but often there is a delay, which may worry the mother. For the first 3–4 weeks, many mothers feel bruised, tired and insecure, and their babies seem strange and distanced. This phase may end when the baby begins to smile and look at her. Thereafter there is an incremental growth, until, by the end of the third month, mothers felt pangs of conscience on leaving their babies.

The term "mother–infant relationship (or bonding) disorders" covers a spectrum of clinical states. These disorders are common in mothers referred for psychiatric help, e.g. 22% of post-partum referrals and 29% of those presenting with "postnatal depression" (Brockington 1996). They are poorly recognised by medical practitioners and psychiatrists. Yet they are the disorders most specific to childbirth and an important part of the work of peri-partum mental health teams. It is these disorders, rather than uncomplicated depression, which are likely to have long-term effects on the behaviour, mood and cognitive development of a child.

In the mildest form, the maternal emotional response is lacking. The mothers complain of a lack of warmth towards the infant and may feel estranged – the baby does not seem to be their own. Their reaction varies: some seek reassurance or counsel from their own mothers or confidantes, but others conceal their feelings. Shame is an important reason why these disorders seldom come to medical attention and impedes diagnosis. Failure to diagnose is regrettable, because treatment is easy – often explanation and reassurance suffice.

At a severe level, there is hostility and rejection of the infant (Oppenheim 1919), which may be covert or overt. The mother may try to persuade her own mother or another relative to take over or may demand that the infant is fostered or adopted. Some mothers try to escape, leaving home for hours on end, sometimes repeatedly. The most poignant manifestation of rejection is the secret wish that the baby "disappear" – be stolen or die. Rejection is often associated with pathological anger, with its risk of child abuse.

Seeking the causes, it seems certain that unwelcome pregnancy is a factor. Unfortunate events at the time of childbirth, e.g. death of a twin, previous stillbirth or a painful and unpleasant delivery may contribute. There is much evidence that the infant's social contribution is important. Sick infants, or those delayed in their social responses – handicapped or premature – may be at risk. Persistent crying, failure to sleep, feeding difficulties, vomiting and a difficult temperament put a mother under severe strain. Post-partum depression may be primary. Indeed, mothers may lose an established "bond" for the duration of an inter-current depression. In other cases, the relationship disorder precedes and seems more severe than depression.

In the diagnostic assessment, an interview is usually sufficient, but in-patient observation yields valuable information, to which each member of the multidisciplinary team contributes. The crucial assessments of maternal behaviour are made by psychiatric nurses, who keep a shift-by-shift record of salient incidents, reporting the mother's statements about the baby, her competence and skill, her affectionate behaviour and her response to crises. The long period of 24-h observation is invaluable in obtaining an accurate overall view.

This is a difficult disorder to manage successfully. Depression is sometimes severe, whether as cause or consequence. The mother experiences extreme guilt and may be exposed to criticism. Family relationships deteriorate, leading to severe, prolonged and irreconcilable conflicts. With skilled treatment, however, there can be optimism about a successful outcome. The principles of management are as follows:

- When hostility and rejection are prominent, the primary decision is whether to attempt treatment at all. The child has no need to be reared by the birth-mother. The mother must be given freedom of choice – it is dangerous for her to feel trapped in a tunnel of unwelcome motherhood. At the same time, the father also has his rights. The option of relinquishing the infant must be openly acknowledged and fully discussed with both parents. Usually the mother opts for treatment.
- Antidepressant treatment should normally be given, in the form of psychotherapy, drugs or ECT.
- The specific element of therapy is working on the dyadic relationship. It is a mistake to separate mother and baby – this merely compounds the problem by adding an element of avoidance. If there is any hint of abuse, or if she is troubled by aggressive impulses, she must never be left alone with her infant. She must be relieved of irksome burdens of infant care. When the baby is calm and content, and the mother feels at ease, she is encouraged and helped to interact with him – to cuddle, talk to him, play and bring out his smile and laughter. The aim and essence of treatment, therefore, is to create the circumstances in which mother and child can enjoy each other. Participant play therapy and baby massage are the key therapies. A panel of recovered mothers, who can instil hope in those still in trouble, is an important resource.

Treatment can take place in various settings:

- Home treatment can be successful, provided there is enough support to relieve the mother of night care and stressful duties.
- Day hospital treatment can provide individual support and group discussion, as well as specific therapies.
- With the most severe and refractory cases, the proper setting is an in-patient mother-and-baby unit.

4.4
Anxiety, Obsessional and Stress-Related Neuroses

4.4.1 Post-traumatic Stress Disorder

After long, hard and excessively painful labours, some women suffer the repetitive intrusion of images and memories, similar to those which occur after the harrowing experiences of war and natural disaster (Bydlowski and Raoul-Duval 1978). If they become pregnant, the symptoms may return, especially in the last trimester, with terrifying nightmares and insomnia. They may develop a secondary tocophobia, with avoidance of child-bearing.

This disorder can be treated by ventilation and cognitive/behavioural psychotherapy.

4.4.2 Querulant Reactions

Another reaction to a severe labour experience is pathological complaining (*querulantenwahn*). These women complain bitterly about their delivery, especially the "dehumanisation" of modern obstetrics, and their humiliation by procedures involving the most intimate parts of their bodies. This disorder can be treated by a psychotherapeutic approach which distracts the mother from her grievances and reinforces productive child-centred activity.

4.4.3 Puerperal Panic

Some mothers are incapacitated by anxiety in the early puerperium, overwhelmed by the fear that they will not be able to cope with their helpless infants (De Armond 1954). The panic and agitation seen in these extreme examples is merely an exaggeration of the feelings that many women, especially first-time mothers, experience when they confront their responsibility. This disorder can often be handled by the family, without invoking professional help. All difficulties are avoided if the mother's own mother or other members of the extended family are at hand.

4.4.4 Exaggerated Fears for the Health and Safety of the Infant

In some women, motherhood can lead to a persistent state of over-arousal, as described by Moll (1920) under the title of *maternitätsneurose*. The mother is anxious about banal tasks which put the baby at risk (e.g. bathing) and sensitive to the slightest indication of illness. These symptoms are prominent in mothers whose children have been born after years of infertility or recurrent miscarriage. In some, the over-concern about the infant's health resembles an adult Briquet's syndrome – hypochondriacal neurosis by proxy. Another cause of post-partum anxiety is the fear of SID (Weightman et al. 1999). These mothers are insomniac, because they lie awake listening to the baby's breathing; they may check the infant 20–30 times every night or even wake them to ensure they are alive. This results in excruciating tension and exhaustion.

A mother may be helped by ventilation of these fears and explanations about the rarity of SID and the infant's resistance to asphyxia. Devices to monitor the infant's breathing should be installed. The vicious cycle of insomnia and hyper-vigilance can be interrupted periodically by involving relatives or friends, so that the mother can sleep under sedation. Anxiety management and cognitive therapies may help.

4.4.5 Phobic Avoidance of the Infant

A mother who suffers from excessive anxiety, whatever the cause, may develop a phobia for her infant and become unable to approach him or her at all (Sved-Williams 1992). The same is true in maternal obsessional disorders involving intrusive filicidal thoughts.

The treatment is similar to that of other post-partum anxiety disorders. The mother may need antidepressive or sedative medication, but the main approach is behavioural, gradually desensitising her to her fear of the infant.

4.4.6 Obsessions of Child Harm

Obsessive-compulsive disorder may present in the puerperium, with extravagant thoughts, images or impulses of child harm. These mothers fear being left alone with their children and may take extraordinary precautions to avoid attacking the child. Post-partum obsessional states may also have the content of child sexual abuse.

Ventilation and explanation are an important part of the treatment, but are rarely sufficient. All these mothers require psychotropic medication, either in the form of antidepressant agents or neuroleptics. It is important to discourage avoidance of the child and to encourage interaction, strengthening positive maternal feelings. Cognitive/behavioural treatment can help her to achieve mastery over irrational impulses.

4.5 Depression

Puerperal melancholia was one of the first post-partum psychiatric disorders to be identified – in merchants' wives in Portugal (1551) and Belgium (1608) (De Castelo Branco 1551/1620; Castro 1617). Patients with up to eight attacks were later described. When, in the 1950s, attention turned to milder, more common disorders, post-partum depression was found to be common. A major contribution was made by the Gordons in New Jersey (Gordon and Gordon 1959). They reviewed the literature, studied 100 normal mothers, made a controlled study of aetiological associations, conducted a follow-through study and showed that social casework was superior to psychoanalytic explorations. Postnatal depression soon become a household word.

The scientific value of this concept must be examined with some scepticism. Depression after childbirth is similar to any other depression (see Brockington 1996, pp. 172–174) and only slightly more frequent than at other times in the lives of women during the reproductive years. The term "postnatal depression" has the danger of introducing into the minds of the unwary the mirage of a homogeneous disorder with a single cause. The causes of post-partum depression are numerous and much the same as those which cause depression in human beings of all ages.

In mothers with recurrent puerperal depression, we would expect to find specific factors. There is no consistent association with parity, dystocia or breast-feeding, but there is some evidence for a link with caesarian section, earlier miscarriage, menstrual mood disorder and the maternity blues. Unwanted pregnancy has been found to be a predictor of peri-partum depression. It has been suggested that the burden of child-rearing, rather than child-bearing, is a factor, but this has been challenged: the twin study by Malmquist and Kaij (1971) and the suicide study by Høyer and Lund (1993) showed that parous women have a lower risk of depression.

The effects of post-partum depression on family life, and on the emotional climate in which children are reared, is of great concern. These effects depend on the degree and duration of maternal depression and the extent to which it involves interactions with the child. A growing child needs not only care, but also emotional support, approbation, attention and stimulation. The mother is his or her primary environment, and her mood dominates the child's world. Even very young infants are disturbed by deviant social behaviour in the mother. In the long term, poor communication may affect the acquisition of speech, language and social skills, and through the pervasive influence of language, other educational deficits (Murray et al. 1996). Nevertheless, these are not universal and inevitable effects: some depressed mothers are sustained by the interaction with their children, and some are warm and stimulating, putting in a tremendous effort, in spite of their distress and symptoms (Pound et al. 1988).

Treatment begins with effective diagnosis. Many more mothers are depressed than ever make their way to their family doctor's surgery. The reasons for the failure to seek help are not fully understood. It can be guessed that some recover early, some do not realise they are ill and others are ashamed or apprehensive about confessing their symptoms. Primary care teams miss many cases. In such circumstances, a screening procedure with high sensitivity and specificity is required. The Edinburgh Postnatal Depression Scale (EPDS; Cox et al. 1987) meets this need. Patients identified by screening, or self-referral, require a full psychiatric examination, exploring the symptoms and course of the illness, and its setting in the context of the mother's life-history, personality and circumstances, especially this pregnancy. Her relationships with her spouse, baby, other children and family of origin must be explored and the available supports established. The diagnosis of depression is relatively easy, provided that the clinician is alert to the possibility, but it is equally important to identify the factors which contribute to vulnerability and the specific components of post-partum disorders. The initial interview should be held at home. This is partly because it is difficult for mothers to attend a clinic while fettered with the care of young children. A domiciliary assessment also has the advantage of meeting the mother in her own environment and has a quality which cannot be achieved in the office.

4.6 Services for Mentally Ill Mothers

The outline of a utopian service is slowly emerging. Its aims are to prevent disorders in those who are vulnerable, to make an early and accurate diagnosis and to intervene rapidly and effectively with minimal disruption of family life.

This requires the following:

- A specialist multidisciplinary team capable of treating severe and intractable illness and undertaking service development, treatment innovations,

training of medical, nursing and paramedical staff and research into the cause and cure of parapartum disorders.
- A community service which provides domiciliary assessment and home treatment, with minimal disruption of family life. Home visits by community nurses are an ideal method of continuing care.
- Day care, which is a rich source of support, bringing the mother into contact with other mothers with similar difficulties, with minimal family disruption. There are group discussions, one-to-one meetings with staff and specific therapies – play therapy, baby massage and anxiety management, as well as motherhood classes, occupational therapy and drama therapy.
- In-patient facilities for mothers and infants, including toddlers. Mother-and-baby conjoint admission was pioneered in 1948 by Dr. T.F. Main. There has been a dissemination of joint admission facilities throughout Britain and also in Australia, Canada, New Zealand, France, Holland, Belgium, Norway and Israel. There are a few specialised units with eight to ten beds or facilities to admit a family; these offer economies of scale, greater safety and an appropriate milieu.
- An obstetric liaison service capable of detecting illness and vulnerability during pregnancy.
- Links with other agencies providing services for mothers, especially the social services, midwifery and primary care teams.
- A network of voluntary organisations, working independently, but with close cordial ties with the professional service.
- Medico-legal expertise to deal with child abuse, infanticide and the custody of children by mentally ill mothers.

5
Psychiatry of the Abuse and Murder of Children (I.F. Brockington)

5.1
Child Abuse

Cruelty to children has long been a shameful feature of human life, which was eventually condemned. The vicious maltreatment of helpless infants has only recently come to the attention of the profession and the public. Its recognition began with French experts on legal medicine, especially Tardieu in 1860. His brilliant insights were unwisely overlooked, and it was necessary to rediscover child abuse through clues from pathology and paediatric radiology, especially bilateral subdural haematoma, retinal haemorrhages and subperiosteal haematomas (Caffey 1946). In 1955, Woolley and Evans made the crucial observation that, when the infants were removed from their homes, no new lesions developed. These discoveries were publicised by Kempe et al. in 1962, 100 years after the original discovery.

The frequency of major abuse is about 6 per 1000 births (Baldwin and Oliver 1975), but much higher in high-risk groups, such as premature infants. Abuse can be subdivided into non-accidental injury, neglect and sexual abuse (not dealt with here).

5.1.1 Non-accidental Injury

The injuries include the following:

- "Battering", resulting in widespread bruising, testifying to the rage or fanaticism of the punitive parent. It may result in rhabdomyolysis, fat embolism, visceral injuries or hypo-volaemic shock.
- Head injuries, caused by direct blows, or swinging the baby by its legs. These cause more severe damage than simple falls and result in multiple, complex fractures and brain injury.
- Fractures, due to jerking the limbs or gripping the thorax while shaking the child. There may be different lesions in several locations, at different stages of healing. They are often multiple and tend to be more severe than accidental fractures.
- Shaking, which is believed to cause subdural haematoma without direct trauma, through rupture of veins under the skull. It also causes retinal haemorrhages.
- Bites, burns and scalds, e.g. by deliberately immersing the infant in scalding water.
- Split lips, lacerations of the mucous membrane or torn frenulum, due to rough use of the feeding bottle.
- Suffocation to suppress crying. This accounts for a *small* proportion of SID (Emery 1989).
- Poisoning by drugs or table salt. Parents may also deprive the child of water or force fluids.
- Fabricating or causing illness, e.g. infections. Recurrent suffocation, poisoning and factitious illness are grouped under Munchausen's syndrome by proxy (Meadow 1977).

5.1.2 Neglect

Neglect includes the following aspects:

- Emotional neglect, resulting in infantile depression and failure to thrive

- Deliberate starvation
- Sequestration – prolonged imprisonment in confined quarters

5.1.3 Causes

The causes of child abuse and neglect include, as background, maternal youth and poverty, lack of social support and the burden of large families. There may be a history of abuse in the parent's own childhood ("the cycle of familial violence"; Oliver 1993). There is often some form of psychiatric disorder, e.g. learning difficulties, depression, substance abuse or extreme sleep deprivation, rarely psychosis. There is frequently a disorder of personality, especially explosive irritability, callousness and cruelty. Factors in the infant are important, including prematurity, illness, handicap and a difficult temperament. Excessive crying, failure to sleep, diarrhoea or refusal to feed may push a mother to the limit of her endurance.

Most of these have an indirect effect. The proximate causes are found in the interaction between mother and infant. Injury results from the mother's anger at a time of crisis, neglect from hatred of the child. The motivation for Munchausen's syndrome by proxy is enigmatic, but perhaps results from maternal anxiety and its deviant expression. The mother's behaviour is governed by the quality of her relationship with the child, her self-control, her experience and her resources in terms of maternal stratagems.

5.1.4 Treatment

The management begins with investigation. Silverman, who obtained some of the first confessions (Silverman 1953), and other pioneers (Ounsted et al. 1974) have advised on how to elicit the history of injury. Interrogation should be sympathetic and quiet, not angry or hostile. It is wise to delay until the physical and radiographic findings are conclusive; with Munchausen's syndrome by proxy, this requires much investigation. The physician interviews the parent privately and indicates his confidence that the injuries must have been inflicted by an adult, and everyone has an obligation to avoid repetition. He would welcome the parents giving him the full story. It is useful to go over each violent act in detail. He will ask about sources of worry, stress or tension and about the baby's behaviour and the parents' upbringing. He should to make clear to the parents that all of us, given adequate provocation, could batter babies. The aim is to understand – not to look back, but to look forward to the future and the positive steps which can be taken.

Although the law differs from one country to another, it is generally mandatory immediately to involve the police and the agencies with a responsibility to protect the child. There is often a duty of notification. The investigation should be handled by a stable, experienced team including a doctor, police and specialist social workers. The position should be made clear at the outset in front of a witness. The parents may deny noticing the injury or give an implausible explanation. In disputed cases, it is necessary to question the spouses independently and other relatives.

Having established the diagnosis, the first duty is to protect the child. There is a high risk of recurrence, and other children in the family are at increased risk. The paediatric hospital provides a sanctuary for a time. Often the only safe remedy is to separate the child from its parents. If it returns home, its condition must be carefully monitored. A play group or day nursery is invaluable, because some infant care can be taken over, and the infant's health monitored unobtrusively. Family centres can provide a setting for many-sided non-residential support. Supervised accommodation in mother-and-baby homes, run by the social services or voluntary agencies, provides shelter and an opportunity to learn maternal skills. The mother needs practical advice and prolonged, protective (not punitive) support from social workers, home helps or lay counsellors, assisting and directing her in her responsibilities. Limits must be set, with frequent visits to ensure that acceptable standards are reached. Child abuse occurs at moments of crisis, and an important intervention is an emergency "hot line".

The mother needs psychotherapy. Child abuse is one of those disorders for which psychotherapy is not just important and indispensable, but virtually the only treatment. It is essential that the counsellor is not involved in the supervision of the baby's health and can give all their attention to the mother. These mothers have a deep sense of helplessness and inferiority and feel worthless and devalued, as underlined by society's rejection of them now. They have intense dependency needs. Punishment and criticism are useless. Lay counsellors ("foster grandparents") may be of particular value. Many, however, have commented on the difficulty in reaching the mothers. Their treatment requires the prolonged efforts of an experienced outreach service, involving a resourceful team of therapists who support each other and are fully aware of the resistance of these mothers.

The children also need therapy. They have been physically damaged and often show obvious signs of fear. As Tardieu (1860) wrote:

One is struck by the facial appearance of these poor children, exposed to ill-treatment and privation. Their

faces breathe sadness. They are timid and fearful. Sometimes their eyes are dull, but often express a burning resentment. It is amazing how rapidly their physiognomy changes, when they are rescued and put under protection.

Ounsted et al. (1974) used the term "frozen watchfulness" to denote a hyper-vigilant child, quiet, slow to move and respond, showing reduced interest in things or people, carefully avoiding gaze fixation and crying only in extremis.

Others have stressed their depression – "a profound apathy to the point of stupor" (Baldwin and Oliver 1975). Older children may show a superficiality of relationships or an anxious reversed solicitude for the mother – the child turned therapist.

5.2 Infanticide

The term "infanticide" covers the killing of infants and children in a wide variety of circumstances, broadly divided into "neonaticide" (killing of the newborn) and "filicide" (the later murder of a child).

5.2.1 Neonaticide

Suppression of the neonate (especially female infants) has been customary in many societies, as an official policy or "grass-roots" custom to control population growth. An entirely different phenomenon is anomic (lawless) neonaticide, in which a desperate mother, who has concealed her pregnancy, kills the infant immediately after parturition. This was a major public health problem in Europe during the nineteenth century. During this century, the frequency has fallen dramatically (Wilkey et al. 1982). Contraception and a relaxation of the abortion laws have played a part, as have the development of social work and the support of single mothers.

Neonaticide is a challenge to forensic pathology. The evidence comes from three sources – signs of recent parturition, proof that the infant was born alive (i.e. had breathed at least once) and evidence of the means of death. Tardieu (1868/1880) detailed the methods of killing and their relative frequency: suffocation is by far the commonest, followed by drowning, head injury and strangulation; wounds, poisoning and neglect are infrequent. Concealment of the cadaver is a problem for the mother, especially difficult for urban dwellers.

Little is known about the mental state of mothers who kill the newborn. Deliberate planned murder of the newborn, sometimes aided and abetted by the father, is probably rare. In the great majority of cases, the explanation of the crime is to be found in the circumstances of concealed pregnancy which were fully elucidated by Jörg (1837) in his classic monograph *Die Zurechnungsfähigkeit der Schwangern und Gebärenden*. Most are immature girls, too passive to resist intercourse or terminate the pregnancy (Gummersbach 1938). All factors conspire to force them into crime – "the scoff and scorn of a taunting world" and economic necessity. There is no-one to turn to, because it is dangerous to consult. A tendency to deny the fearful reality, to retreat into wishful thinking may occupy some weeks or months. They can hope for a miscarriage, or a stillbirth, or even their own death in labour. The birth is awaited as if a miracle could avert it. But in the end, they must endure, in secret, an agonising first delivery, without the support of a midwife, making superhuman efforts to suppress the slightest murmur of complaint. No-one can imagine the extremes of depression, fear and desperation of these girls. They are eventually faced with a crying baby and respond by silencing it, killing the child out of fear and perplexity. Most of the assaults, employing suffocation or drowning, are compatible with panic. A minority, which involve head trauma or stabbing, testify to rage and hatred.

There has been much debate whether the defence of insanity can be invoked. It is probable that most of these babies die when the mother is in the grip of an emotional crisis – seized by fear, despair or fury. The emotional extremes experienced by these women, in their pain, exhaustion and lonely peril, put the definition of insanity under strain; but *emotional* crises are not generally acceptable in law as evidence of insanity, which requires a *cognitive* factor. The medico-legal problem centres on the fact that impaired consciousness undoubtedly occurs in labour, though rarely. The causes include confusional states, eclampsia, syncope due to loss of blood and exhaustion to the point of stupor. If the burden of proof is with the defence – i.e. the prosecution must prove *actus reus* and *mens rea*, but the defence must prove insanity – there can be no valid evidence in unattended deliveries, because the diagnosis of confusion and syncope require an eye-witness. Some jurists, however, have been concerned about the possibility of a miscarriage of justice – that a mother who killed her baby when her consciousness was clouded could be wrongly condemned.

We know little of the psychological consequences of neonaticide. The long-term effects are likely to resemble those of criminal abortion, but the horror of the event is of a different magnitude. Post-traumatic stress disorder and depression can be expected, and there is a lifetime of coming to terms with an atrocious crime.

5.2.2 Filicide

A parent may kill a child in a variety of different mental states. These can be grouped under sane filicides, child abuse, depression, psychosis and trance states. The majority of murdered children are killed by their parents, and the majority of female murderers kill their own child.

Occasionally, children are killed by parents who are not mentally ill. An unwanted child may be killed "in cold blood". Euthanasia is another setting – a sane mother may kill a beloved child to relieve its suffering – Perrussel's "homicide through pity" (Perrussel 1923). The "Medea complex" is the name given to vengeful child murder, which may be an element in some filicides and some neonaticides.

Child abuse is a relatively common cause of filicide. Death results from ill-tempered assaults or over-zealous punishment, without homicidal intent. In the Queensland study (Wilkey et al. 1982), 18 out of 49 child murders were due to non-accidental injury. In various surveys, the rates have been about 1 per 100,000 children. The death rate in children already subjected to abuse and returned to their families is much higher. Fathers play a major part in punitive child abuse.

Depression is the commonest cause of filicide. Studies of convicted mothers underestimate the frequency of depressive filicide, because many mothers do not survive; e.g. in Denmark, 82 out of 92 parents who poisoned their children completed suicide (Harder 1967). Depressive murder is often committed in the belief that the child's best interests are being served – delusional altruism or mercy-killing (libericide). Often the mother kills to spare the child the misfortune of being motherless after her suicide. In contrast to those who kill an unwanted child, mothers surviving depressive filicide make no attempt to conceal the crime: typically, they confess and request punishment. Mothers may kill more than one child. Unlike filicidal males, few women attack their husband or cohabitee – family murder is rare in women. Men also kill their children during an attack of depression, and family murder seems more common. Family suicide was a peculiarly Japanese institution: Komine (1938) unearthed the staggering total of 4323 cases between 1930 and 1938.

Filicide during puerperal manic or cycloid illness is rare. It may occur in chronic delusional psychosis, especially if the morbid ideas involve the child, or through command hallucinations. There are several instances of filicide in menstrual psychosis. Delirium tremens is an occasional cause. Post-epileptic delirium or epileptic automatism can result in motiveless filicide. The same is true of somnambulism.

6 Psychiatric Aspects of Hormonal Contraception (M. Lanczik)

6.1 Introduction

Contraception with "the Pill" is in widespread use and more than 99% effective. It has both positive and negative somatic effects and may also have undesired psychotropic effects, or even desired ones. Thus the physician is confronted by psychosyndromes provoked by the use of oral contraception, but may also, in other cases, use it successfully to treat mental disturbances, such as mood disorders.

A total of 15% of women who discontinue oral contraception do so because of "mental side effects" (Milsom et al. 1991), and mood alterations, usually depressive, are present in one third of all women who use oral contraception (Cullberg 1972; Clare 1985). The risk of mental side-effects from oral contraception is significantly increased in women who suffer from pre-menstrual dysphoric syndrome (Hammarbäck and Bäckström 1989). It has been stated that women with severe pre-menstrual dysphoric syndrome are more likely to develop mood disturbances when taking oral contraceptives, while women with a "mild" peri-menstrually occurring psychosyndrome are more likely to experience an improvement of mood (Bäckström et al. 1992).

Because most oral contraceptives are combined preparations containing varying amounts of androgens, oestrogens and progesterone, it is difficult to tell in individual cases which component hormone is responsible for causing a mental disorder or a possible "psychotherapeutic" effect. It is much easier to trace the somatic effects of these preparations, both desired and undesired, to their individual components.

6.2 Anxiety Disorders and Depression in the Setting of Oral Contraceptive Use

A number of relatively old studies revealed a relationship between the use of oral contraceptives (in higher doses than now usual) and depressive mood disorders, and even panic attacks (Wagner and Berenson 1994). More recent studies reveal no such relationship when low-dose contraceptives are used. Nonetheless, oral contraceptives have been reported to lead to relapses of phasic endogenous depression and pre-menstrual dysphoric syndrome in women with a past history of these disorders (Kendler et al. 1988; Bancroft and

Sartorius 1990). Triphasic preparations are clearly related to pre-menstrual dysphoric syndrome and recurrent menstrual depression (Bancroft et al. 1987b; Bancroft and Rennie 1993).

The extent to which oral contraceptives may induce milder, i.e. subclinical depressive mood disturbances is still controversial (Parry and Rush 1979; Kendler et al. 1988; Bancroft and Sartorius 1990). It is clear that they generally reduce sexual interest and that preparations with high progesterone activity in particular, including depot preparations, may lead to easy fatigability and sad mood disturbances and promote lethargic behaviour (Graham and Sherman 1992; Westhoff et al. 1995).

According to a number of anecdotal reports, oral contraceptives not only improve physical well-being, e.g. by alleviating pre-menstrual oedema and breast pain, but also improve the mental manifestations of pre-menstrual dysphoric syndrome (Graham and Sherman 1992). Prospective studies of this question, however, have yielded inconsistent findings.

Vitamin B_6 supplementation, used to treat the functional vitamin B_6 deficiency that accompanies oral contraceptive use, has also been reported to help improve mild depressive mood disturbances (Winston 1973).

In conclusion, much more research is necessary regarding the mental effects of oral contraceptives, both desired and undesired, from the point of view of both psychopathology and endocrinology.

7
Peri-menopausal Dysphoric Syndrome and Depressive Illnesses in the Peri-menopausal Period (M. Lanczik)

The term "peri-menopause" refers to the 5–7 years preceding the last menstrual period (the menopause), a time of life in which ovarian function progressively ceases and, as a result, the plasma concentrations of the oestrogens, progesterone and the androgens decline. The increasing deficiency of these hormones is paralleled by a rise of gonadotropin activity. The initial consequence of these hormonal changes is a disturbance of the menstrual cycle, i.e. the menses become rarer and less regular and ultimately stop. The beginning of the post-menopausal period is endocrinologically defined by an oestrogen level below 40 pcg/ml and a follicle-stimulating hormone (FSH) level above 25 μ/ml. Most women have multiple physical symptoms in the peri-menopausal period, and some also have mood disturbances reaching psychiatric intensity.

7.1
Clinical Features and Classification

There is controversy over whether endogenous depression appearing peri-menopausally differs in its psychopathological features from that appearing at other times (Winokur 1973; Schmidt et al. 1996). It is characteristic of peri-menopausal affective illnesses that they are always accompanied by multiple somatic manifestations.

We distinguish two characteristic types of disturbance on the basis of their psychopathologically relevant manifestations. Most patients suffer from a so-called hyperaesthetic-emotional asthaenia, which is clinically indistinguishable from post-partal dysphoria (maternity blues, post-partum blues) and from many cases of pre-menstrual dysphoric syndrome. This syndrome is characterised by sleep disturbances (50%), disturbance of concentration (53%), lability of affect (64%), increased irritability (70%), anxiety (67%), tearfulness (70%), sad mood disturbance (61%), lack of interest (56%), easy fatigability (53%; Schmidt and Rubinow 1991; Schmidt et al. 1997) and decreased libido (10%–85%; McCoy and Davidson 1985; Cutler et al. 1987).

The affective syndrome of the second group of patients is characterised by a severe, sad mood disturbance with hopelessness, life-weariness of occasionally suicidal intensity, diminished initiative, lack of interest, volitional and cognitive inhibition, self-reproach (up to a severe delusion of guilt) and, as in the first group, sleep disturbances. These manifestations are no different from those of endogenous depression appearing outside the peri-menopausal period.

It is quite obvious that two characteristic types of peri-menopausal affective disorder are to be identified diagnostically as such and that they differ not only in their clinical features, but also in their aetiology, pathogenesis and course. The first type, which may be called peri-menopausal dysphoric syndrome, is directly caused by hormonal changes. The second type, peri-menopausal depression, represents the initial manifestation, or a relapse, of endogenous depression in patients with the appropriate genetic predisposition, for which the hormonal changes of the peri-menopausal period act merely as a trigger mechanism. The lack of distinction between these two clinical entities at least partly explains the conflicting epidemiological data on this mental disorder that may be found in the literature.

A total of 80% of all peri-menopausal women experience vasomotor disturbances such as hot flushes and cold sweats, which mostly occur at night and thereby disturb sleep. The more severe these vasomotor

or hormonally induced manifestations are, the greater the risk that a depressive, dysthymic or dysphoric syndrome will develop. When the somatic manifestations subside, the psychic manifestations do also.

Because many women of this age also complain of tachycardia, vertigo and shortness of breath, hyperthyroidism and panic disorder must be included in the differential diagnosis.

7.2 Aetiology

Both neuroendocrine and psychosocial factors have been invoked as causes of these disorders.

7.2.1 Endocrinological Aspects

Baischer et al. (1995) found a negative correlation between the Hamilton Depression Score and the concentration of 17-β-estradiol in peri-menopausal women with depressive and anxious mood disturbances. Huerta et al. (1995) found an elevated FSH concentration in patients with these disorders. Thus affective disorders in the peri-menopausal period may also be considered the result of oestrogen withdrawal.

It is therefore reasonable in any case for women over age 40 suffering from an affective psychosyndrome of depressive, dysthymic or dysphoric type, or from anxiety disorders, to be screened for the possibility of peri-menopausal or menopausal manifestations by the measurement of 17-β-estradiol, thyroid parameters and FSH.

7.2.2 Psychosocial Aspects

Not only these neuroendocrine factors, but also psychosocial factors play an important role in the causation and maintenance of the disorder. Divorced, widowed or separated women are more likely to suffer from peri-menopausal affective disturbances than women with partners (Avis and McKinlay 1991). Women with lesser degrees of education also seem to have an elevated risk for these disorders (O'Connor et al. 1994; Avis and McKinlay 1995; Huerta et al. 1995). On the other hand, there is no elevated risk associated with the death of parents or with children moving away from home (Kaufert et al. 1992).

7.2.3 Risk Factors

Women who have had pre-menstrual dysphoric syndrome, endogenous depression, post-partal mood disturbances or a dysphoric syndrome related to the use of oral contraceptives have a significantly elevated risk of developing a dysphoric or depressive illness once again in the peri-menopausal period (Stewart and Boydell 1993; Hay et al. 1994; Pearlstein 1995). The longer the peri-menopausal period lasts, the more likely a mood disturbance will occur (Avis et al. 1994).

7.3 Epidemiology

It remains a matter of debate whether the peri-menopausal period is associated with an elevated risk of developing a depressive illness (Weissman et al. 1988). Kessler et al. (1993) even report a diminished prevalence of depressive syndromes severe enough to be classified as endogenous depression in this age-group. On the other hand, it is recognised that depressive syndromes may occur as a consequence of hysterectomy, with or without oophorectomy (Kritz-Silverstein et al. 1993; Oldenhave et al. 1993).

7.4 Treatment

The first distinction that must be drawn is that between depressive syndromes rising to the intensity of an endogenous depression on the one hand and dysphoric syndromes on the other. Our differential therapeutic recommendations are largely based on this dichotomy. There are essentially three strategies for the treatment of peri-menopausal affective syndromes: (1) medical antidepressant therapy, (2) hormone supplementation therapy and (3) combined therapy, i.e. both antidepressant and hormonal therapy.

7.4.1 Hormonal Therapy

Controlled studies have shown oestrogen monotherapy to be ineffective for the treatment of peri-menopausally manifesting depressive syndromes that fulfil the diagnostic criteria for endogenous depression (Schmidt et al. 1996).

Oestrogen supplementation therapy is particularly useful in patients who have not only mental, but also physical manifestations of hypo-oestrogenism, e.g. nocturnal hot flushes and sweats. Patients suffering from mood disorders after hysterectomy and oophorectomy (i.e. iatrogenic menopause) have been found to respond well to oestrogen supplementation (Sherwin 1988).

Anxiety states in the peri-menopausal period, which may be of vasomotor origin, also respond better to

hormone supplementation therapy than to anxiolytics or antidepressants.

7.4.2 Psychopharmacotherapy and Its Application in Combination with Hormonal Treatment

Thymoleptic medical therapy is the treatment of choice for patients with peri-menopausal relapses of endogenous depression, or peri-menopausal dysphoric syndromes.

A combined treatment with both thymoleptics and hormone substitution is particularly helpful in patients with peri-menopausal relapses of endogenous depression, who suffer from an oestrogen and progesterone withdrawal syndrome with both mental and physical manifestations (Zweifel and O'Brien 1997).

The increased sweating that characteristically occurs in the peri-menopausal period is further exacerbated by the use of antidepressants. Clonidine may be used to counteract this undesired effect (Feder 1995).

An increase of libido may be achieved by androgen (testosterone) supplementation (Sherwin et al. 1985).

8 References

Andersch B, Hahn L (1981) Premenstrual complaints. II. Influence of oral contraceptives. Acta Obstet Gynecol Scand 60: 579–583

Andersch B, Wenderstram G, Hahn L, Öhman R (1986) Premenstrual complaints. I. Prevalence of premenstrual symptoms in a Swedish urban population. J Psychosom Obstet Gynaecol 5: 39–49

Appelt H (1988) Ergebnisse psychoendokrinologischer Forschung in der Gynäkologie – Überblick und klinische Relevanz. In: Appelt H, Strauß B (eds) Psychoendokrinologische Gynäkologie. Enke, Stuttgart, pp 37–65

Arbeit SA (1975) A study of women during their first pregnancy. Ph.D. thesis, Yale University

Avis NE, McKinlay SM (1991) A longitudinal analysis of women's attitudes towards menopause: results from the Massachusetts Women's Health Study. Maturitas 13: 65–79

Avis NE, McKinlay SM (1995) The Massachusetts women's mental health study: an epidemiologic investigation of the menopause. JAMA 50: 45–49

Avis NE, Brambilla D, McKinlay SM, Vass K (1994) A longitudinal analysis of the association between menopause and depression. Results from the Massachusetts Women's Mental Health Study. Ann Epidemiol 4: 214–220

Bäckström T, Boyle H, Baird DT (1981) Persistence of symptoms of premenstrual tension in hysterectomized women. Br J Obstet Gynaecol 88: 530–536

Bäckström T, Sanders T, Leask R, Davidson D, Warner B, Bancroft J (1983) Mood, sexuality, hormones and the menstrual cycle. II. Hormone levels and their relationship to the premenstrual syndrome. Psychosom Med 45: 503–507

Bäckström T, Linde BÅ, Cavalli-Björkman B (1992) Effects of oral contraceptives on mood: a randomized comparison of three phasic and monophasic preparations. Contraception 46: 253–268

Baischer W, Koinig G, Hartmann B, Huber J, Langer G (1995) Hypothalamic-pituitary-gonadal axis in depressed premenopausal women: elevated blood testosterone concentrations compared to normal controls. Psychoneuroendocrinology 20: 553–559

Baldwin JA, Oliver JE (1975) Epidemiology and family characteristics of severely-abused children. Br J Prev Soc Med 29: 205–221

Bancroft J, Rennie D (1993) The impact of oral contraceptives on the experience of premenstrual mood, clumsiness, food craving and other symptoms. J Psychosom Res 37: 195–202

Bancroft J, Rennie D (1995) Perimenstrual depression: its relationship to pain, bleeding, and previous history of depression. Psychosom Med 57: 445–452

Bancroft J, Sartorius N (1990) The effects of oral contraceptives on well-being and sexuality. Oxford Rev Reprod Biol 12: 57–92

Bancroft J, Bayle H, Warner P (1987a) The use of an LHRH agonist busenelin in the long-term management of premenstrual syndromes. Clin Endocrinol 27: 171–172

Bancroft J, Sanders D, Warner P, Loudon N (1987b) The effects of oral contraceptives on mood and sexuality: a comparison of triphasic and combined preparations. J Psychosom Obstet Gynecol 7: 1–8

Bancroft J, Williamson L, Warner P, Rennie D, Smith K (1993) Perimenstrual complaints in women complaining PMS, menorrhagia, and dysmenorrhea: towards a dismantling of the premenstrual syndrome. Psychosom Med 55: 133–145

Barnhart KT, Freeman EW, Sondheimer SJ (1995) A clinician's guide to the premenstrual syndrome. Med Clin North Am 79: 1457–1472

Bergant A, Guggenberger G, Ulmer H, Richter R, Heim K, Dapunt O. Ärger und prämenstruelles Syndrom. Psychother Psychosom Med Psychol (in press)

Berger CP, Presser B (1994) Alprazolam in the treatment of two subsamples of patients with late luteal phase dysphoric disorder: a double-blind, placebo-controlled crossover study. Obstet Gynecol 84: 379–385

Biegon A, Reches A, Snyder L, McEwen BC (1983) Serotonergic and noradrenergic receptors in the rat brain: modulation by chronic exposure to ovarian hormones. Life Sci 32: 2015–2021

Binder H (1937) Psychiatrische Untersuchungen über die Folgen der operativen Sterilisierung der Frau durch partielle Tubenresektion. Schweiz Arch Neurol Psychiatr 40: 1–49, 249–276

*Bivin GD, Klinger MP (1937) Pseudocyesis. Principia, Bloomington

Bledin KD, Cooper JE, MacKenzie S, Brice B (1984) Psychological sequelae of female sterilisation: short-term outcome in a prospective controlled study. Psychol Med 14: 379–390

Bratfos O, Haug JO (1966) Puerperal mental disorders in manic-depressive females. Acta Psychiatr Scand 42: 285–294

Brezinka C, Huter O, Biebl W, Kinzl J (1994) Denial of pregnancy. J Psychosom Obstet Gynaecol 15: 1–8

Brinch M, Isager T, Tolstrup K (1988) Anorexia nervosa and motherhood: reproduction pattern and mothering behaviour of 50 women. Acta Psychiatr Scand 77: 611–617

**Brockington IF (1996) Motherhood and mental health. Oxford University Press, Oxford

Brockington IF, Kelly A, Hall P, Deakin W (1988) Premenstrual relapse of puerperal psychosis. J Affect Disord 14: 287–292

Bruce SJ (1962) Reactions of nurses and mothers to stillbirths. Nurs Outlook 10: 88–91

Brzezinski AA, Wurtman JJ, Wurtman RJ, Gleason R, Greenfield J, Nader T (1990) D-Fenfluramine suppresses the increased calorie and carbohydrate intakes and improves the mood of women with premenstrual depression. Obstet Gynecol 76: 296–301

Bühler C, Hetzer H, Tudor-Hart B (1927) Soziologische und psychologische Studien über das erste Lebensjahr. Quell Stud Jugendkd 5: 107–124

Bydlowski M, Raoul-Duval A (1978) Un avatar psychique méconnu de la puerpéralité: la nevrose traumatique post-obstétricale. Perspect Psychiatr 4: 321–328

Caffey J (1946) Multiple fractures in long bones of infants suffering from chronic subdural haematoma. Am J Roentgenol 56: 163–173

Cartwright A (1988) Unintended pregnancies that lead to babies. Soc Sci Med 27: 249–254

Casper RF, Hearn MT (1990) The effect of hysterectomy and bilateral oophorectomy in women with severe premenstrual syndrome. Am J Obstet Gynecol 162: 99–101

Casson P, Hahn PM, van Vugt DA (1990) Lasting response to ovariectomy in severe intractable premenstrual syndrome. Am J Obstet Gynecol 162: 99–101

Castro R à (1617) De Universa Mulierum Medicanovo et Antehac a Nemine Tentato Ordine. Hamburg, p 314

Chaturvedi SK, Chandra PS, Gururaj G, Pandian RD, Beena MB (1995) Suicidal ideas during premenstrual phase. J Affect Disord 34: 193–199

Chuong CJ, Colligan RC, Coulam CB, Bergstrahl EJ (1988) The MMPI as an aid in evaluating patients with premenstrual syndrome. Psychosomatics 29: 197–202

Clare AW (1985) Hormones, behaviour, and the menstrual cycle. J Psychosom Res 29: 225–233

Cohen RL (1988) Psychiatric consultation in childbirth settings. Plenum, New York

Collins A, Eneroth P, Landgren BM (1985) Psychoneuroendocrine stress responses and mood as related to the menstrual cycle. Psychosom Med 47: 512–527

Condon JT (1987) The battered foetus syndrome. J Nerv Ment Dis 175: 722–725

Condon JT (1993) The premenstrual syndrome: a twin study. Br J Psychiatry 162: 481–486

Cooper PJ, Gath D, Rose N, Fieldsent R (1982) Psychological sequelae to elective sterilisation: a prospective study. Br Med J 284: 461–464

Cooper PJ, Campbell EA, Day A, Kennerley H, Bond A (1988) Non-psychotic psychiatric disorder after childbirth. A prospective study of prevalence, incidence, course and nature. Br J Psychiatry 152: 799–806

Corney RH, Stanton R (1991) A survey of 658 women who report symptoms of premenstrual syndrome. J Psychosom Res 35: 471–482

Cornwell J, Nurcombe B, Stevens L (1977) Family response to loss of a child by sudden infant death syndrome. Med J Aust 1: 656–658

Courrier D, Duboucher H, Pouget E (1927) Pseudo-grossesse créée par la persistance d'un corps jaune périodique. Mem Soc Biol 2: 271–272

Cox JL, Holden JM, Sagovsky R (1987) Detection of postnatal depression: development of the 10-item Edinburgh Postnatal Depression Scale. Br J Psychiatry 150: 782–786

Cranley MS (1981) Development of a tool for the measurement of maternal attachment during pregnancy. Nurs Res 30: 281–284

Cullberg J (1972) Mood changes and menstrual symptoms with different gestagen/estrogen combinations. A double-blind comparison with placebo. Acta Psychiatr Scand 236 [Suppl]: 1–45

Cutler WB, Garcia CR, McCoy NL (1987) Perimenopausal sexuality. Arch Sex Behav 16: 225–235

Dalton K (1959) Menstruation and acute psychiatric illness. Br Med J 1: 148–151

Dalton K (1987) Incidence of premenstrual syndrome in twins. Br Med J 1: 148–151

David HP (1985) Post-abortion and post-partum psychiatric hospitalisation. CIBA Found Symp 115: 150–164

De Armond M (1954) A type of post partum anxiety reaction. Dis Nerv System 15: 26–29

de Castelo Branco RJ (1551/1620). Second century. Sete centúrias de curas médicos. Biblioteca de facultade de medicina de Lisboa., p 104

de Jong R, Rubinow D, Roy-Byrne P, Hoban MC, Grover GN, Post RM (1985) Premenstrual mood disorder and psychiatric illness. Am J Psychiatry 142: 1359–1361

de Leon-Jones FA, Val E, Herts C (1982) MHPG excretion and lithium treatment during premenstrual tension syndrome. Am J Psychiatry 139: 950–952

Dinane TG, O'Keane V (1991) The premenstrual syndrome: a psychoneuroendocrine perspective. Clin Endocrinol Metab 5: 143–165

Distler W, Graf M (1986) Neurochemische und endokrine Veränderungen in der Prä- und Postmenopause. Gynäkologie 19: 202–209

Dupouy R, Courtois A (1930) Des psychoses gravidiques et en particulier de la psychopolynévrite: syndrome de Korsakoff. Encephale 25: 284–301

Dye L, Warner P, Bancroft J (1995) Food craving during menstrual cycle and its relationship to stress, happiness of relationship and depression: a preliminary enquiry. J Affect Disord 34: 157–164

Edelberg H, Galant S (1925) Über psychotische Zustände nach künstlichem Abort (psychosis post abortum arteficialem). Z Ges Neurol Psychiatr 97: 106–128

Ekblad M (1963) Social-psychiatric prognosis after sterilisation of women without children. Acta Psychiatr Scand 39: 481–514

Emery JL (1989) Sudden infant death and suffocation. Br Med J 299: 456

*Endicott J (1993) The menstrual cycle and mood disorders. J Affect Disord 29: 193–200

Engelhard JLB (1912) Über Generationspsychosen und den Einfluß der Gestationsperiode auf schon bestehende psychische und neurologische Krankheiten. Z Geburtshilfe Gynäkol 70: 727–812

Eriksson E, Lisjo P, Sundblad C, Anderssen K, Andersch B, Modigh K (1990) Effect of clomipramine on premenstrual syndrome. Acta Psychiatr Scand 81: 87–88

Eriksson E, Sundblad C, Lisjö P, Modigh K, Andersch B (1992) Serum levels of androgens are higher in women with premenstrual irritability and dysphoria than in controls. Psychoneuroendocrinology 17: 195–204

Eriksson E, Hedberg MA, Andersch B, Sundblad C (1995) The serotonin reuptake inhibitor paroxetin is superior to the

noradrenaline reuptake inhibitor maprotiline in the treatment of premenstrual syndrome. Neuropsychopharmacology 12: 167–176

Esquirol JED (1818) Observations sur l'aliénation mentale à la suite de couches. J Gen Med Chir Pharm Françaises Étrangères (series 2) 1: 148–164

Ewald G (1922) Bestrahlungsergebnis bei einer menstruell rezidierenden Psychose. Monatschr Psychiatr Neurol 52: 6–21

Feder R (1995) Clonidine treatment of excessive sweating. J Clin Psychiatry 56: 35

Frank RC (1931) The hormonal causes of premenstrual tension. Arch Neurol Psychiatry 26: 1053–1057

Freeman EW, Sondheimer S, Weinbaum PJ, Rickels K (1985) Evaluating premenstrual symptoms in medical practice. Obstet Gynecol 65: 500–505

Freeman EW, Rickels K, Sondheimer SJ, Polansky M (1990) Ineffectiveness of progesterone suppository treatment for premenstrual syndrome. JAMA 264: 349–353

Freeman EW, Rickels K, Sondheimer SJ, Denis A, Pfeifer S, Weil S (1994) Nefazodone in the treatment of premenstrual syndrome: a preliminary study. J Clin Psychopharmacology 14: 180–186

*Freeman EW, Rickels K, Schweizer E, Ting T (1995) Relationship between age and symptom severity among women seeking medical treatment for premenstrual symptoms. Psychol Med 25: 309–315

Freeman EW, Rickels K, Sondheimer SJ, Wittmaack FM (1996a) Sertraline versus desipramine in the treatment of premenstrual syndrome: an open-label trial. J Clin Psychiatry 57: 7–11

Freeman EW, Rickels K, Sondheimer SJ (1996b) Fluvoxamine for premenstrual dysphoric disorder: a pilot study. J Clin Psychiatry 57 [Suppl 8]: 56–60

Friedman T, Gath D (1989) The psychiatric consequences of spontaneous abortion. Br J Psychiatry 155: 810–813

Friedmann M (1894) Über die primordiale menstruelle Psychose (die menstruale Entwicklungspsychose). Munch Med Wochenschr 41: 4–7, 27–31, 50–53, 69–71

Fürstner C (1875) Über Schwangerschafts- und Puerperalpsychosen. Arch Psychiatr Nervenkr 5: 505–543

Gath D, Rose N, Bond A, Day A, Garrod A, Hodges S (1995) Hysterectomy and psychiatric disorder: are the levels of psychiatric morbidity falling. Psychol Med 25: 277–283

Girdler SS, Pedersen CA, Light KC (1995) Thyroid axis function during the menstrual cycle in women with premenstrual syndrome. Psychoneuroendocrinology 20: 395–403

Gise LH, Lebovits AH, Paddison PL, Strain JJ (1990) Issues in the identification of premenstrual syndromes. J Nerv Ment Dis 178: 228–234

Goodlin RC (1985) Pregnant women with Munchausen syndrome. Am J Obstet Gynecol 153: 207–210

Gordon RE, Gordon KK (1959) Social factors in the prediction and treatment of emotional disorders of pregnancy. Am J Obstet Gynecol 77: 1074–1083

Graham CA, Sherwin BB (1987) The relationship between retrospective premenstrual symptom reporting present oral contraceptive use. J Psychosom Res 31: 45–53

Graham CA, Sherwin BB (1992) A prospective treatment study of premenstrual symptoms using triphasic oral contraceptive. J Psychosom Res 36: 257–266

Graze KK, Nee J, Endicott J (1990) Premenstrual depression predicts future major depressive disorder. Acta Psychiatr Scand 81: 201–205

Gummersbach H (1938) Die kriminalpsychologische Persönlichkeit der Kindesmörderinnen und ihre Wertung im gerichtsmedizinischen Gutachten. Wien Med Wochenschr 88: 1151–1155

*Halbreich U, Endicott J (1985) Relationship of dysphoric premenstrual changes to depressive disorders. Acta Psychiatr Scand 71: 331–338

Halbreich U, Smoller JW (1997) Intermittend luteal phase sertraline treatment of dysphoric premenstrual syndrome. J Clin Psychiatry 58: 399–402

Hallman J (1986) The premenstrual syndrome: an equivalent of depression? Acta Psychiatr Scand 73: 403–411

Hallman J, Orleland L, Edman G, Schalling D (1987) Thrombocyte monoamine oxidase activity and personality traits in women with severe premenstrual syndrome. Acta Psychiatr Scand 76: 225–234

Hamilton JA, Parry B, Alagna S, Blumenthal S, Herz E (1984) Premenstrual mood changes: a guide to evaluation and treatment. Psychiatric Ann 14: 426–435

Hammarbäck S, Bäckström T (1988) Induced anovulation as treatment of premenstrual tension syndrome: a double-blind cross-over study with GnRH-agonist versus placebo. Acta Obstet Gynecol Scand 67: 159–166

Hammarbäck S, Bäckström T (1989) Demographic factors in subgroups of women seeking help for premenstrual syndrome. Acta Obst Gynecol Scand 68: 247–254

Harder T (1967) The psychopathology of infanticide. Acta Psychiatr Scand 43: 196–245

Harrison WM, Rabkin JG, Endicott J (1985) Psychiatric evaluation of premenstrual changes. Psychosomatics 26: 789–792

Harrison WM, Endicott J, Rabkin JG, Nee J (1987) Treatment of premenstrual dysphoria with alprazolam and placebo. Psychopharm Bull 23: 150–153

Harrison WM, Endicott J, Nee J (1989) Treatment of premenstrual depression with nortriptyline: a pilot study. J Clin Psychiatry 50: 136–139

Harrison WM, Endicott J, Nee J (1990) Treatment of premenstrual dysphoria with alprazolam: a controlled study. Arch Gen Psychiatry 47: 270–275

Harvey K (1977) Caring perceptively for the relinquishing mother. Am J Maternal Child Nurs Jan/Feb: 24–28

Hay AG, Bancroft J, Johnstone (1994) Affective symptoms in women attending a menopause clinic. Br J Psychiatry 164: 513–516

Heidrich A, Schleyer M, Spingler H, Albert M, Knoche M, Fritze J, Lanczik M (1994) Postpartum blues: relationship between not-protein-bound steroid hormones in plasma and postpartum mood changes. J Affect Disord 30: 93–98

Herzberg BN, Coppen A (1970) Changes in psychological symptoms in women taking oral contraceptives. Br J Psychiatry 116: 161–164

Hitzig JE (1827) Mord in einem durch Eintreten des Monatsflusses herbeigeführten unfreien Zustande. Hitzigs Z Kriminalrechtspflege 6(12): 237

Hoppeler PA (1955) Nachuntersuchungen von 100 auf Grund psychiatrischer Indikation ohne vorherige Interruptio graviditatis sterilisierten Frauen. Praxis 2: 24–30, 3: 46–50

Høyer G, Lund E (1993) Suicide among women related to number of children in marriage. Arch Gen Psychiatry 50: 134–137

Huerta R, Mena A, Malacara JM, Diaz de Leon J (1995) Symptoms at perimenopausal period: its association with attitudes towards sexuality, life-style, family function, and FSH levels. Psychoneuroendocrinology 20: 135–148

Hurt SW, Schnurr PP, Severino SK, Freeman EW, Gise LH, Rivera-Tovar A, Steege JF (1992) Late luteal phase dysphoric disorder in 670 women evaluated for premenstrual complaints. Am J Psychiatry 149: 525–530

Iles S, Gath D (1993) Psychiatric outcome of termination of pregnancy for foetal abnormality. Psychol Med 23: 407–413

Jörg JCG (1837) Die Zurechnungsfähigkeit der Schwangern und Gebärenden. Gebhardt, Leipzig

Jureidini J (1993) Obstetric factitious disorder and Munchausen syndrome by proxy. J Nerv Ment Dis 181: 135–137

Kaufert PA, Gilbert P, Tate R (1992) The Manitobe Project: a reexamination of the link between menopause and depression. Maturitas 14: 143–155

Kempe CH, Silverman FN, Steele BF, Droegemueller W, Silver HK (1962) The battered-child syndrome. JAMA 181: 17–24

Kendell RE, Chalmers JC, Platz C (1987) Epidemiology of puerperal psychoses. Br J Psychiatry 150: 662–673

Kendler KS, Martin NS, Heath AC, Handelsman D, Eaves LJ (1988) A twin study of the psychiatric side effects of oral contraceptives. J Nerv Ment Dis 176: 153–160

Kessler RC, McGonagle KA, Swartz M, Blazer DG, Nelson CB (1993) Sex and depression in the National Comorbidity Survey I: lifetime prevalence, chronicity and recurrence. J Affect Disord 29: 85–96

Keye WR, Trunnell EP (1986) A psychosocial model of premenstrual syndrome. Int J Fertil 31: 259–262

Kirchberg P (1913) Psychische Störungen während der Geburt. Arch Psychiatr Nervenkr 52: 1153–1163

Kitayama I, Yamaguchi T, Harada M, Okano T, Nomura J, Hatotani N (1984) Periodic psychoses and hypothalamo-pituitary function. Mie Med J 34: 127–138

Klaiber EL, Broverman DM, Vogel W, Kobayashi Y (1979) Estrogen therapy for severe persistent depressions in women. Arch Gen Psychiatry 36: 550–554

Komine S (1938) Beobachtung über die Entstehung des Selbstmordes der Eltern mit Kindermord in Japan. Psychiatr Neurol Jpn 42: 15–16 (German summary; full version in *kanji*, pp 210–226)

Kow LM, Pfaff D (1985) Estrogen effects on neuronal responsiveness to electrical and neurotransmitter stimulation: an in vitro study on the ventromedial nucleus of the hypothalamus. Brain Res 347: 1–10

Kritz-Silverstein D, Wingard D, Barrett-Connor E, Morton D (1993) Hysterectomy, oophorectomy and depression in older women. Abstracts of the 4th Annual Meeting of the North American Menopause Society, p 83

Kutzinski A (1909) Über eklamptische Psychosen. Charité Ann 33: 216–260

Lanczik M (1995) Entstehungsbedingungen und Verlauf postpartal auftretender psychischer Erkrankungen und Störungen. Postdoctoral thesis, Würzburg

Lanczik M, Brockington IF (1999) Das postpartale dysphorische Syndrom: Psychopathologie, Diagnostik, Ätiologie. Fortschr Neurol Psychiatr 67: 60–67

Lanczik M, Spingler H, Heidrich A, Becker T, Kretzer B, Albert P, Fritze J (1992) Postpartum blues – depressive disease or pseudoneurasthenic syndrome. J Affect Disord 25: 47–52

*Leifer M (1977) Psychological changes accompanying pregnancy and motherhood. Genet Psychol Monogr 95: 55–96

*Lemoine P, Harousseau H, Borteyru J P, Menuet J C (1968) Les enfants de parents alcooliques: anomalies observées. Ouest Med 25: 476–482

Lenzinger E, Diamant K, Vytiska-Binstorfer E, Kasper S (1997) Prämenstruelle dysphorische Störung (PMDS). Ein Überblick über Diagnose, Epidemiologie und Therapieansätze. Nervenarzt 68: 708–718

Lloyd J, Laurence KM (1985) Sequelae and support after termination of pregnancy for foetal malformation. Br Med J 290: 907–909

Luine VN, Rhodes JC (1983) Gonadal hormone regulation of MAO and other enzymes in hypothalamic areas. Neuroendocrinology 36: 235–241

Luine VN, Khylchevskaya RI, McEwens B (1975) Effect of gonadal steroids on activities of monoamine oxidase and cholin acetylase in rat brain. Brain Res 86: 293–306

Maddocks S, Hahn P, Möller F, Reid RL (1986) A double blind placebo-controlled trial of progesterone vaginal suppositories in the treatment of premenstrual syndrome. Am J Obstet Gynaecol 154: 573–581

*Main TF (1948) Mothers with children in a psychiatric hospital. Lancet ii: 845–847

Malikian JE, Hurt S, Endicott J, Delaney JR (1989) Premenstrual dysphoric changes in depressed patients. Abstracts of the American Psychiatric Association, 142nd Annual Meeting, San Francisco, p 128

Malmquist A, Kaij L (1971) Motherhood and childlessness in monozygotic twins. II. The influence of motherhood on health. Br J Psychiatry 118: 22–28

Marcé LV (1858) Traité de la folie des femmes enceintes, des nouvelles accouchées et des nourrices. Baillière, Paris

Martin RL, Roberts WV, Clayton PJ (1980) Psychiatric status after hysterectomy. A one-year prospective follow-up. JAMA 244: 350–353

McCoy NL, Davidson JM (1985) A longitudinal study of the effects of menopause on sexuality. Maturitas 7: 203–209

Meadow R (1977) Munchausen syndrome by proxy. The hinterland of child abuse. Lancet 2: 343–345

Menkes DB, Taghavi E, Mason PA, Spears GFS, Howard RC (1992) Fluoxetine treatment of severe premenstrual syndrome. Br Med J 305: 27–31

Milsom I, Sundell G, Andersch B (1991) A longitudinal study of contraception and pregnancy outcome in a representative sample of young Swedish women. Contraception 43: 111–119

Moll L (1920) Die Maternitätsneurose. Wien Klin Wochenschr 33: 160–162

Moos RH (1968a) The development of menstrual distress questionnaire. Psychosom Med 30: 853–867

Moos RH (1968b) Typology of menstrual cycle symptoms. Am J Obstet Gynecol 103: 390–402

Morgan JF, Lacey JH, Sedgwick PM (1999) Impact of pregnancy on bulimia nervosa. Br J Psychiatry 174: 135–140

Mortola JF, Girton L, Fischer U (1991) Successful treatment of severe premenstrual syndrome by combined use of gonadotropin-releasing hormone agonist and estrogen/progestin. J Clin Endocrinol Metab 72: 252A

Murray L, Hipwell A, Hooper R (1996) The cognitive development of 5-year-old children of postnatally depressed mothers. J Child Psychol Psychiatry 37: 927–935

Muse KN, Cetel NS, Futterman LA, Yen SSC (1984) The premenstrual syndrome: effects of "medical ovariectomy". N Engl J Med 311: 1345–1349

O'Connor VM, del Mar CB, Sheehan M, Siskind V, Fox-Young S, Cragg (1994) Psycho-social factors contribute more symptom reporting by middle-aged women than hormonal status. Maturitas 16: 63–69

O'Hara MW (1987) Postpartum "blues", depression, and psychosis: a review. J Psychosom Obstet Gynecol 7: 205–277

Okano T (1999) Thyroid function and postpartum psychiatric disorders. Arch Womens Ment Health 1: 157–165

Oldenhave A, Jaszman LJB, Everaerd WTAM, Haspels AA (1993) Hysterectomized women with ovarian conservation report more severe climacteric complaints than do normal climacteric women of similar age. Am J Obstet Gynecol 168: 765–771

Oliver JE (1993) Intergenerational transmission of child abuse: rates, research and clinical implications. Am J Psychiatry 150: 1315–1324

Olshausen R (1891) Beitrag zu den puerperalen Psychosen, speciell den nach Eklampsie auftretenden. Z Geburtshilfe Gynäkol 21: 371–385

Oppenheim H (1919) Über Misopädie. Z Ges Neurol Psychiatrie 45: 1–18

*Osiander FB (1797) Neue Denkwürdigkeiten für Aerzte und Geburtshelfer. Rosenbusch, Göttingen

Ounsted C, Oppenheimer R, Lindsay J (1974) Aspects of bonding failure: the psychopathology and psychotherapeutic treatment of families of battered children. Dev Med Child Neurol 16: 447–456

Paffenbarger RS Jr, McCabe LJ (1966) The effect of obstetric and perinatal events on risk of mental illness in women of childbearing age. Am J Public Health 56: 400–407

Parry BL, Rush AJ (1979) Oral contraceptives and depressive symptomatology: biologic mechanism. Compr Psychiatry 20: 347–358

Parry BL, Berga SL, Mostofi N, Sependa PA, Kripke DF, Gillin JC (1989) Morning versus evening bright light treatment of late luteal dysphoric disorder. Am J Psychiatry 146: 1215–1217

Parry BL, Berga SL, Kripke DF, Klauber MR, Laughlin GA, Yen SSC, Gillin JC (1990) Altered waveform of plasma nocturnal melatonin secretion in premenstrual depression. Arch Gen Psychiatry 47: 1139–1146

Parry BL, Mahan AM, Mostofi N, Klauber MR, Lew GS, Gillin JC (1993) Light therapy of late luteal phase dysphoric disorder: an extended study. Am J Psychiatry 150: 1417–1419

Pauleikhoff B (1986) Endogene Psychosen. Pressler, Hürtgenwald

Pearlstein T (1995) Hormones and depression: what are the facts about premenstrual syndrome, menopause, and hormone replacement therapy? Am J Obstet Gynecol 173: 646–653

Pearlstein TB, Thoft J, Rubinstein D (1988) Psychiatric diagnosis and luteal variation in PMS women. Montreal, Canada, New Research Abstracts, p 10

Pearlstein TB, Frank E, Rivera-Tovar A, Thoft JS, Jacobs E, Mieczkowski TA (1990) Prevalence of axis I and axis II disorders in women with late luteal dysphoric disorder. J Affect Disord 20: 129–134

Pearlstein TB, Stone AB, Lund SA, Scheft H, Zlotnick C, Brown WA (1997) Comparison of fluoxetine, bupropion, and placebo in the treatment of premenstrual dysphoric disorder. J Clin Psychopharmacol 17: 261–266

Peiper A (1924) Sinnesempfindungen des Kindes vor seiner Geburt. Monatschr Kinderheilkd 29: 236–241

Percheval A (1911) Des manoeuvres abortives chez les femmes qui ne sont pas enceintes. Thesis, University of Paris

Perrussel G (1923) L'homicide altruiste des mélancoliques et des persécutés. Thesis, University of Paris

Pop VJM, de Rooy HAM, Vader HL, van der Heide D, van Son M, Komproe ICH, Essed GGM, de Geus CA (1991) Postpartum thyroid dysfunction and depression in an unselected population. N Engl J Med 324: 1815–1816

Pound A, Puckering C, Cox A, Mills M (1988) The impact of maternal depression on young children. Br J Psychother 4: 240–252

Reeb M (1933) Aménorrhée avec symptômes de pseudo-grossesse due à un corps jaune persistant ou à des kystes lutéiniques. Bull Soc Obstet 22: 244–245

Reich T, Winokur G (1970) Postpartum psychoses in patients with manic depressive disease. J Nerv Ment Dis 151: 60–68

Rickels K, Freeman E, Sondheimer S (1989) Buspirone in treatment of premenstrual syndrome. Lancet i: 777

Rickels K, Freeman EW, Sondheimer S, Albert J (1990) Fluoxetine in the treatment of premenstrual syndrome. Curr Ther Res 48: 161–166

Robson KS (1967) The role of eye-to-eye contact in maternal-infant attachment. J Child Psychol Psychiatry 8: 13–25

Rohde A, Marneros A, Fischer J, Diedrich K (1992) Häufigkeit und Art prämenstrueller Symptomatik unter dem Einfluß erlebter Infertilität: Eine vergleichende Studie. Geburtsh Frauenheilkd 52: 291–296

Roy-Byrne P, Rubinow DR, Hoban MC, Grover GN, Blank D (1987) TSH and prolactin responses to TRH in patients with premenstrual syndrome. Am J Psychiatry 144: 480–484

Rubinow DR, Hoban C, Grover GN, Galloway DS, Roy-Byrne PP, Andersen R, Merriam GR (1988) Changes in plasma hormones across the menstrual cycle in patients with menstrually related mood disorders and in control subjects. Am J Obstet Gynecol 158: 5–11

Rynearson EK (1982) Relinquishment and its maternal complications: a preliminary study. Am J Psychiatry 139: 338–340

Sarrat J (1911) De l'infanticide dans des rapports avec les psychoses transitories des femmes en couches. Thesis, University of Paris

Schmidt PJ, Rubinow DR (1991) Menopause-related affective disorders: justification for further study. Am J Psychiatry 148: 844–852

Schmidt PJ, Nieman LK, Grover GN, Muller KL, Merriam GR, Rubinow DR (1991) Lack of effect of induced menses on symptoms in women with premenstrual syndrome. N Engl J Med 324: 1174–1179

Schmidt PJ, Grover GN, Roy-Byrne P, Rubinow DR (1993) Thyroid function in women with premenstrual syndrome. J Clin Endocrinol Metab 76: 671–674

Schmidt PJ, Roca A, Rubinow DR (1996) Psychiatric disorders during the peri-menopause. Baillieres Clin Psychiatry 2/4: 701–711

Schmidt PJ, Roca CA, Bloch M, Rubinow DR (1997) The perimenopause and affective disorders. In: Berga SL (ed) Seminars in reproductive endocrinology. Thieme, New York, pp 297–299

Schönthal (1892) Beiträge zur Kenntnis der in frühem Lebensalter auftretenden Psychosen. Arch Psychiatr Nervenkrankh 23: 816–833

Schwarzer R, van der Ploeg HM (1987) Emotionale Veränderungen während des Menstruationszyklus – das prämenstruelle Syndrom. Psychother Psychosom Med Psychol 37: 237–243

Selitzky SA (1925) Cerebropathia et psychopathia toxica gravidarum. Zentralbl Gynäkol 37: 2070–2073

Sherwin BB (1988) Affective changes with estrogen and androgen replacement therapy in surgically menopausal women. J Affect Disord 14: 177–188

Sherwin BB (1990) Up-regulatory effect of estrogen on platelet 3H-imipramine binding sites in surgically menopausal women. Biol Psychiatry 28: 339–348

Sherwin BB, Gelfand MM, Brender W (1985) Androgen enhances sexual motivation in females: a prospective, crossover study of sex steroid administration in the surgical menopause. Psychosom Med 47: 339–351

Silverman FN (1953) The roentgen manifestations of unrecognised skeletal trauma in infants. Am J Roentgenol 69: 413–427

Singer K, Cheng R, Schou M (1974) A controlled evaluation of lithium in premenstrual tension syndrome. Br J Psychiatry 124: 50–51

Sletten IW, Gershon S (1966) The premenstrual syndrome: a discussion of its pathophysiology and treatment with lithium ion. Compr Psychiatry 7: 197–206

Smialek Z (1978) Observations on immediate reactions of families to sudden infant death. Pediatrics 62: 160–165

Smith S, Rinehart JS, Ruddock VE, Schiff I (1987) Treatment of premenstrual syndrome with alprazolam: results of double.blind, placebo-controlled, randomized crossover clinical trial. Obstet Gynecol 70: 37–43

Spitzer RL, Severino SK, Williams JBW, Parry BL (1989) Late luteal phase dysphoric disorder and DSM-III-R. Am J Psychiatry 146: 892–897

Splett T (1998) Die wissenschaftliche Forschung zu den menstruationsabhängigen psychischen Erkrankungen in der deutschen Psychiatrie bis 1945 unter besonderer Berücksichtigung der Arbeiten des Freiherrn Richard von Krafft-Ebings (unpublished)

Stahl GE (1702) De affectibus periodicis. Dissertation, Halle an der Saale

Steiner M (1987) The effects of gonadal hormones on brain and behaviour. Prog Neuropsychopharmacol Biol Psychiatr 11: 115–119

Steiner M, Haskett RF, Osmun JN, Carroll BJ (1980) Treatment of premenstrual tension with lithium carbonate – a pilot study. Acta Psychiatr Scand 61: 96–102

*Steiner M, Steinberg S, Stewart D, Carter D, Berger C, Reid R, Grover D, Streiner D (1995) Fluoxetine in the treatment of premenstrual dysphoria. Canadian Fluoxetin/Premenstrual Dysphoria Collaborative Study Group. N Engl J Med 332: 1529–1534

Steiner M, Korzewka M, Lamont J, Wilkins A (1997) Intermittent fluoxetine in the treatment of women with premenstrual dysphoria. Psychopharmacol Bull 33: 771–774

Stewart DE, Boydell KM (1993) Psychological distress during menopause: association across the reproductive life cycle. Int J Psychiatry Med 23: 157–162

Stewart DE, Klompenhouwer JL, Kendell RE, Van Hulst AM (1991) Prophylactic lithium in puerperal psychosis. The experience of three centres. Br J Psychiatry 158: 393–397

Stone AB, Pearlstein TB, Brown WA (1991) Fluoxetine in the treatment of late luteal dysphoric disorder. J Clin Psychiatry 52: 290–293

Su T-P, Schmidt PJ, Danaceau M, Murphy DL, Rubinow DR (1997) Effect of menstrual cycle phase on neuroendocrine and behavioral responses to the serotonin agonist m-chlorophenylpiperazine in women with premenstrual syndrome and controls. J Clin Endocrinol Metab 82: 1220–1228

Sundblad C, Modigh K, Andersch B, Eriksson E (1992) Clomipramine reduces premenstrual irritability and dysphoria: a placebo controlled study. Acta Psychiatr Scand 85: 39–47

Sundblad C, Hedberg MA, Eriksson E (1993) Clomipramine administered during the late luteal phase reduces the symptoms of premenstrual syndromes: a placebo-controlled trial. Neuropsychopharmacology 9: 133–145

Sundblad C, Wikander I, Andersch B, Eriksson E (1997) A naturalistic study of paroxetine in premenstrual syndrome: efficacy and side-effects during ten cycles of treatment. Eur Neuropsychopharmacol 7: 201–206

Sved-Williams AE (1992) Phobic reactions of mothers to their own babies. Aust NZ J Psychiatry 26: 631–638

*Tardieu A (1860) Étude médico-légale sur les sévices et mauvais traitements exercés sur des enfants. Ann Hygiene 15: 361–398

*Tardieu A (1868/1880) Étude médico-légale sur l'infanticide. Baillière, Paris

Tellenbach H (1983) Melancholie. Problemgeschichte, Endogenität, Typologie, Pathogenese, Klinik, 4th edn. Springer, Berlin Heidelberg New York

*Terp IM, Mortensen PB (1998) Post-partum psychoses: clinical diagnoses and relative risk of admission after parturition. Br J Psychiatry 172: 521–526

Tott CA (1844) Fälle von melancholia attonita bei Neuentbundenen. Neue Z Geburtskd 17: 187–190

Trixler M, Jádi F (1981) Adoptáció utáni 'post partum' pszichózisok. Orvosi Hetilap 122: 3071–3074

Ulleland CN (1972) The offspring of alcoholic mothers. Ann NY Acad Sci 197: 167–169

Vié J, Bobé J (1932) Les idées délirantes de grossesse. Étude sémiologique et pathogénique. Encephale 28: 468–502

von Krafft-Ebing R (1902) Psychosis Menstrualis. Eine klinisch-forensische Studie. Enke, Stuttgart

von Neugebauer F (1912) Tentamen abortus provocandi deficiente graviditate. Zentralbl Gynäkol 36: 100–104

Wagner KD, Berenson AB (1994) Norplant-associated major depression and panic disorder. J Clin Psychiatry 55: 478–480

Walker A, Bancroft J (1990) Relationship between premenstrual symptoms and oral contraceptive use: a controlled study. Psychosom Med 52: 86–96

Walton J, Youngkin E (1986) The effect of a support group on self-esteem of women with premenstrual syndrome. J Obstet Gynecol Neonatal Nurs 16: 174–178

Warner P, Bancroft J, Dixson A, Hampson M (1991) The relationship between perimenstrual depressive mood and depressive illness. J Affect Disord 23: 9–23

Watson NR, Studd JWW (1990) Use of estrogen in treatment of the premenstrual syndrome: a comparison of the routes of administration. Contemp Rev Obstet Gynaecol 2: 117–123

Weightman J, Dalal BM, Brockington IF (1999) Pathological fear of cot death. Psychopathology 167: 246–249

Weir JG (1984) Suicide during pregnancy in London 1943–1962. In: Kleiner GJ, Greston WM (eds) Suicide in pregnancy. Wright, Boston

Weissman MM, Leaf PJ, Tischler GL, Blazer DG, Karno M, Bruce ML, Florio LP (1988) Affective disorders in five United States communities. Psychol Med 18: 141–153

Westhoff C, Wieland D, Tiezzi L (1995) Depression in users of depot-medroxyprogesterone acetate. Contraception 51: 351–354

Wilkey I, Pearn J, Petrie G, Nixon J (1982) Neonaticide, infanticide and child homicide. Med Sci Law 22: 31–34

Winokur G (1973) Depression in the menopause. Am J Psychiatry 130: 92–93

Winston F (1973) Oral contraceptives, pyridoxin, and depression. Am J Psychiatry 130: 1217–1221

Wollenberg R (1891) Drei Fälle von periodisch auftretender Geistesstörung. Charité Ann 16: 427–476

Wood N, Most A, Dery GK (1982) Prevalence of premenstrual symptoms. Am J Public Health 72: 1257

Wood SH, Mortola JF, Chan YF, Moossazadeh F, Yen SS (1992) Treatment of premenstrual syndrome with fluoxetine: a doubleblind, placebo-controlled, crossover study. Obstet Gynecol 80: 339–344

Woolley PV, Evans WA (1955) Significance of skeletal lesions in infants resembling those of traumatic origin. JAMA 158: 539–543

Yonkers KA, Halbreich U, Freeman E, Brown C, Pearlstein T (1996) Sertraline in the treatment of premenstrual dysphoric disorder. Psychopharmacol Bull 32: 41–46

Yonkers KA, Halbreich U, Freeman E, Brown C, Endicott J, Frank E, Parry B, Pearlstein T, Severino S, Stout A, Stone A, Harrison W (1997) Symptomatic improvement of premenstrual dysphoric disorder with sertraline treatment. JAMA 278: 983–988

Zweifel JE, O'Brien H (1997) A metaanalysis of the effect of hormone replacement therapy upon depressed mood. Psychoneuroendocrinology 22: 189–212

CHAPTER

14

W.G. Jilek, L. Jilek-Aall

Culture-Specific Mental Disorders

Translator: E. Taub

1 Introduction

1.1 Comparative Cultural Psychiatry

At the beginning of the twentieth century, Emil Kraepelin announced the birth of a new scientific discipline that he named "comparative psychiatry" (*vergleichende Psychiatrie*), which he predicted would play an important role in both general psychopathology and ethnopsychology (Kraepelin 1904). We now know that Kraepelin, in his later years, attached special importance to sociocultural factors in the generation, manifestations, and distribution of mental illnesses (Jilek 1995). He died suddenly in 1926 before he could work out the definitive formulation of the discipline he had inaugurated, which would best be called "comparative cultural psychiatry," but is now generally known as "transcultural psychiatry," "cross-cultural psychiatry," or "ethnopsychiatry."

In accordance with Kraepelin's conception, H.B.M. Murphy, in his basic text on this discipline (Murphy 1982), identified its principal concerns as the identification, verification, and explanation of the links that exist between mental disorders and the different psychosocial characteristics of peoples and cultures. The term "culture" here comprises "the ideas, values, habits, and other patterns of behaviour which a human group consciously or unconsciously transmits from one generation to another" (Murphy 1982, p. 10).

1.2 Cultural Relativism and Universalism

The diverse views of what constitutes psychopathology have extended, from earliest times, over a continuum ranging from extreme cultural relativism to extreme universalism. Radical psychiatric universalism maintains that mental disorders are essentially the same all over the world, while radical cultural relativism is summed up by the motto: *on est fou par rapport à une société donnée*, i.e. "one is mad in relation to a given society" (Beguin 1952). The debate over cultural relativism and universalism in relation to psychopathology has been stimulated, since 1977, by the proclamation of a so-called new transcultural psychiatry, whose proponents accuse the supposedly old transcultural psychiatry of having subjugated itself to biologically and organically oriented psychiatry (Kleinman 1977, 1988; Littlewood 1990; Lewis-Fernandez and Kleinman 1994). The champions of this so-called new transcultural psychiatry criticize well-known comparative psychiatric research projects, including those of the World Health Organization (WHO), for allegedly having been designed from their outset to find greater similarity than diversity in the cultural comparison of psychopathological phenomena, so that, beneath the obscuring layers of cultural particularity, the postulated biological core might be found. A further error of the "old" transcultural psychiatry is said to be the application of "Western" psychiatric categories in areas of non-Western culture, which supposedly constitutes a category fallacy. In contrast, the "new" transcultural psychiatry stresses the specific cultural meaning of psychopathological phenomena, which must always be understood in the context of the culture in question, for which reason ethnographic methods are preferred to epidemiological methods in the area of research. The "new" transcultural psychiatry demands special consideration of indigenous notions of causality and of explanatory models based on folk beliefs, i.e. the culturally immanent or "emic" standpoint is given priority over the culturally overreaching or "etic" perspective of the scientific observer (these current terms of American anthropology were formed in vague analogy to the linguistic terms "phonemic" and "phonetic"). It is hardly of assistance to cross-cultural psychiatric research, however, when such apparent antagonisms are carried to the extreme on theoretical grounds; instead, the culturally immanent and culturally overreaching perspectives ought to complement each other (Pfeiffer 1994). These tendencies of the so-called new transcultural psychiatry reflect the pronounced cultural relativism that exerted great influence on twentieth-century Anglo-American anthropology, while French anthropology, represented by the works of Claude Lévi-Strauss, strove to extract the basic and universal structures of human thought processes out of cultural diversity.

Some radical cultural relativists characterize psychiatric diagnoses as European-American folk categories and consider it futile to design a generally applicable diagnostic classification scheme (Patel and Winston 1994) or else define modern psychiatry as a culturally constructed Western folk system (Gaines 1992). The great majority of psychiatrists involved in transcultural research reject the ideological position of extreme cultural relativism and recognize that the pioneers of transcultural psychiatry always tried to discover differences in psychopathological phenomena across cultures and to explain these by means of sociocultural factors – without, however, overlooking the universal elements of human behavior that transcend individual and ethnic/cultural differences.

2 Cultural Influence and Psychopathology

Sociocultural factors influence, to different degrees, all aspects of mental disorders, including their generation, symptom profile, course, and prevalence (Marsella 1988; Fabrega 1993). The same is true of normal psychological functions, even those that are known with certainty to have a neurophysiological basis: with respect to color perception, for example, we now know that the spectrum is divided in entirely different ways by different cultural/linguistic groups (Berlin and Kay 1969). Castillo (1997, pp. 268–270) summarizes neurobiological research findings of relevance to cultural learning and comes to the conclusion that the sociocultural environment of the individual becomes physically structured in the neural network of the brain over the course of his or her enculturation.

A *pathoplastic* effect of sociocultural factors, i.e. an effect that forms the clinical picture, is generally accepted, no doubt because of the extremely large number of unambiguous examples. Recent evidence was provided by the findings of the Determinants of Outcome of Severe Mental Disorders (DOSMED) projects of the WHO: the relative frequency of certain symptoms in schizophrenic patients in highly developed and developing countries was very different (Sartorius et al. 1986; Jablensky et al. 1992); significant differences in schizophrenic manifestations were also found when probands in Agra, India, and Ibadan, Nigeria, were compared, and the local researchers found that the content of psychotic symptoms identified critical concerns of the culture in question (Katz et al. 1988). An analysis of psychiatric symptom formation in members of three ethnic minority groups in British Columbia (Salish Indians, Russian Doukhobortsi, and Low German Mennonites) revealed that cultural factors were the most important distinguishing criteria, more so than gender or socioeconomic status (Jilek-Aall et al 1978). The pathoplastic influence of ethnic and cultural identity is also greater than the influences of geographical proximity, historical relations, ethnolinguistic relatedness, and religious connection: recent comparative psychiatric studies show that there are significant differences in schizophrenic manifestations between Malta and Libya (Maslowski 1986), Japan and China (Fujimori et al. 1987), Korea and China (Kim et al. 1993), and Pakistan and Saudi Arabia (Ahmed and Naeem 1984). The International Pilot Study of Schizophrenia (IPSS) and DOSMED global research projects, carried out in international cooperation by the WHO Mental Health Division, revealed an extremely important finding for comparative cultural psychiatry: the course and outcome of schizophrenic psychoses were clearly and uniformly better in project sites in non-Western countries than in the centers of the highly technologically developed Western countries (Jablensky et al. 1994). Cultural influence was listed as an important determinant of this finding, but could not be defined any more precisely (Jablensky et al. 1992). In this connection, longitudinal studies of schizophrenic psychoses in Japan, Hong Kong, and Singapore appear to be relevant, as they revealed favorable prognoses comparable to those seen in the developing countries in the WHO projects (Ogawa et al. 1987; P.W.H. Lee et al. 1991; Tsoi and Wong 1991). These research findings come from highly technologically developed countries, but from cultures that are still very different from those of modern Western societies. The critical difference in relation to the course and outcome of schizophrenic disorders, and probably other psychiatric disorders as well, thus seems not to be that between more and less technologically developed countries, but between modern Western societies and non-Western populations that have largely succeeded in preserving important elements of their traditional culture.

A *pathogenic*, i.e. disease-causing, effect of sociocultural factors is most easily demonstrable in psychopathological conditions of organic/toxic origin that occur in the context of certain sociocultural factors (Jilek 1982). These include, for example, neuropsychiatric disorders of genetic origin, promoted by culture-typical marriages between cousins or perinatal brain injury through certain customary childbirth practices, and especially the sanctioning of alcohol consumption in societies such as Western society, inevitably opening the way to alcohol abuse, with its well-known psychosocial and psychiatric consequences. The rapid and overwhelming westernization ("modernization") of small, tribal societies is accompanied by pathological psychosocial phenomena, as Murphy (1961) was able to show as early as 1959 in a review of data available at the time. Since then, an accumulation of research findings has confirmed the pathogenic psychosocial effects of the forced and rapid transformation of traditionally grown communities to the modern mass society. The consequences of such a social upheaval are mostly not a general acculturation or new cultural creation, but rather marginalization and deculturation of broad segments of the population, associated with critical levels of alcohol and drug abuse and their psychiatric complications, as well as youth suicide. Such behavior after the loss of culture is often a manifestation of *anomic depression* based on anomie, confusion of cultural identity, and relative deprivation (Jilek 1974, 1981). Psychopathological sequelae following excessively rapid culture change have been documented for indigenous populations in North America

(Jilek 1974; Jilek-Aall 1974, 1988; Kraus and Buffler 1979; Hochkirchen and Jilek 1985; Kirmayer 1994), Greenland (Thorslund 1990), and Oceania (Jilek 1987a; Rubinstein 1992; Hezel 1993). In Africa, similar societal changes also cause an increased frequency of transient psychoses and psychosis-like conditions (see below).

3 "Culture-Specific Disorders" or "Culture-Bound Syndromes"

3.1 Development of the Concept

Kraepelin was convinced that the characteristics of an ethnic group also find their expression in mental illness. The eighth edition of his textbook (Kraepelin 1909, 1913) is probably the first comprehensive psychiatric text mentioning disorders known today as "culture-bound syndromes," namely *latah*, *amok*, and *koro*, with which he became familiar during his travels in Asia (Jilek 1995). He did not assume, however, that entirely new types of mental illness would be found in other cultures, and he considered the conditions mentioned to be variants of well-known clinical pictures. In the *American Handbook of Psychiatry*, Arieti and Meth (1959) were the first to address culture-bound syndromes under the heading of "exotic psychotic syndromes," mainly relying upon a survey by Yap (1951). It was also Pow Meng Yap who introduced the concept of "culture-bound" clinical conditions to the psychiatric literature, first in the form of "atypical culture-bound psychogenic psychoses" (Yap 1962) and ultimately as "culture-bound reactive syndromes" (Yap 1967, 1969). According to his definition, which is still the clearest to date, culture-bound syndromes are culture bound only in the sense that their symptom patterns are unusual and are determined in form and frequency by cultural factors (Yap 1974). Yap (1969) was the first to point out that a culture-bound syndrome need not be "exotic," but may also arise under the cultural conditions of modern Western society, e.g. anorexia nervosa – the first, that is, if we set Devereux aside, who construed schizophrenia as an "ethnic psychosis of the Occident." The term introduced by Yap was at the focus of the debate in the 1980s and 1990s between "cultural relativists" and "universalists" in transcultural psychiatry, and this influenced the manner in which it was interpreted. The interpretation of Simons (1985a) may serve as an example of a universalistic interpretation. He considers culture-bound syndromes to be indigenously defined folk illnesses that are strongly influenced by the social and cultural realities of the society in which they occur, although this in no way implies that biological factors are irrelevant to the formation of these syndromes. In the same multi-author text on the subject (Simons and Hughes 1985), Carr (1985) defines culture-bound syndromes from the culture-relativistic point of view as the result of a social learning process that is legitimated as illness within the indigenous system. Littlewood and Lipsedge (1985) prefer to restrict the application of the term to certain locally occurring, temporally limited, and nonbiologically caused patterns of behavior for which the affected individuals are not considered responsible as usual. Decidedly culture-relativistic interpretations were proffered mainly by North American ethnologists. Ritenbaugh (1982), for example, states that culture-bound syndromes can be understood only in a specific cultural context, and the indigenous etiological theory, which symbolizes the essential meaning and behavioral norms of the culture, must be taken into account; diagnosis and successful treatment should be based on culture-specific ideology and technology. Other researchers have aimed at a more neutral definition of terms. Tseng and McDermott (1981) speak of culture-related specific psychiatric conditions that occur with some frequency in certain culture areas, possess special and unusual characteristics, and have manifestations that are closely related to that particular culture.

Prince and Tcheng-Laroche (1987) prefer to keep indigenous explanatory models out of the definition; they see culture-bound syndromes as groupings of clinical signs and symptoms that, because of their special psychosocial characteristics, appear in only a limited number of cultures. In view of the often ideologically motivated disputes over these definitions, some authors have proposed less controversial designations, such as "folk diagnostic categories" (Levine and Gaw 1995) or "folk psychiatric disorders" Hughes (1985a). Hughes (1985b) compiled a long list of 168 "culture-bound or folk psychiatric syndromes," containing a great number of locally used, indigenous names of generally occurring conditions. H.B.M. Murphy (1982), in his basic text on comparative psychiatry, discusses the known culture-bound syndromes without subsuming them under this title. Meanwhile, the term coined by Yap has been sanctioned, so to speak, by the official diagnostic manual of American psychiatry, DSM-IV (APA 1994).

3.2 Delimitation of the Concept

In Annex *2* of the Research Edition of the International Classification of Mental and Behavioral Disorders,

ICD-10 (WHO 1993), it is correctly stated that the culture-specific disorders have two common features:

1. They are not easily accommodated in an international psychiatric classification system.
2. They were first described in, and then considered in close relation to, a certain population group or cultural region.

For this Annex, 12 examples of widely varying importance were drawn from the great number of culture-specific disorders and briefly discussed. The American diagnostic manual DSM-IV contains a "glossary of culture-bound syndromes" with 25 entries, of which, however, only a few concern distinct syndromes. The DSM-IV glossary also contains wholly nonspecific expressions such as *locura*, the Spanish colloquial term for insanity or folly, or *mal de ojo*, Spanish for the "evil eye," to which many tradition-directed populations ascribe a general, magical, disease- and misfortune-producing effect (Hauschild 1982). Possession by the *zar*, which is listed in the same glossary, is not a pathological phenomenon, but a healing cult (Jilek 1993). The DSM-IV glossary also includes general pathological conditions that are regionally known under certain folk names. In the present chapter, the term "culture-specific disorders" refers to locally recognized, specific symptom patterns that can be plausibly brought into relation with the sociocultural emphases and stresses typical of a given population.

3.3
Attempts at Nosological Classification

In view of the bewildering profusion of various local and regional disease names, it is no wonder that early attempts were made to classify culture-bound syndromes in a generally recognized diagnostic system on the basis of the symptoms that the investigator held to be most important. Yap provided, at the same time as his new term, a diagnostic classification system for the culture-bound reactive syndromes. He divided these into three main categories (Yap 1967, 1969): paranoid syndromes, emotional syndromes, and syndromes with disordered consciousness. These categories were further subdivided into primary fear reactions, rage reactions, culture-specific nosophobia, and trance dissociation and later (Yap 1974) modified to include fear reactions, rage reactions, and reactions based on "mental obfuscation." The first English-language textbook of transcultural psychiatry (Kiev 1972) contained a division of the culture-bound mental disorders into anxiety states, obsessional compulsive neuroses, hysterical disorders, phobic states, depressive reactions, and dissociative states. With all due respect to these attempts to seek "unity in the diversity" of culture-bound syndromes (Sartorius 1979), it can only be concluded, from the perspective of transcultural psychiatry, that harmonization of indigenous ethnic terminology with a psychiatric diagnostic system is not possible (Pfeiffer 1980, 1994; Littlewood and Lipsedge 1985). A nosologic classification is neither scientifically nor clinically satisfactory, as the diagnosis-determining manifestations of most culture-bound syndromes overlap, and the indigenous disease terms are semantically overlaid with traditional meanings and interpretations of folk medicine that significantly influence both the illness behavior of the affected person and the reactions of others. It thus seems somewhat artificial when Simons (1985a) sorts the culture-bound syndromes into "taxa," for which he postulates the same or similar neurophysiologically based behavior patterns; for most of these syndromes, no neurophysiological basis can be demonstrated. Pfeiffer's method (1980, 1994) of considering culture-specific stress zones, behavioral forms, and coping styles, while taking into account the interpretations of folk medicine, seems to be of greater heuristic value and clinical utility. This chapter contains an exposition of culture-typical emphases and stress situations that may provoke specific mental disorders in affected individuals. To satisfy the statistical demands of clinical practice and research, the local disease terms will be supplemented, wherever possible, with the corresponding ICD-10 classifications. As may be inferred from the above discussion, an exact agreement between psychiatric diagnostic categories and indigenous concepts of illness cannot be achieved.

4
Folk Idioms for General Emotional Distress

Susto, Espanto **(ICD-10: F43; ?F45.1)**

Heterogeneous combinations of symptoms are regarded by some authors as a culture-bound syndrome only because a single, commonly accepted popular diagnosis is applied, even though the patients so designated have diverse sociocultural identities and lifestyles and their complaints may be quite different. A good example of this phenomenon is *susto*, also called *espanto*, *espasmo*, or *miedo* – Spanish terms that all connote a frightening experience. These disease designations are found among Spanish-speaking populations almost everywhere in the Western hemi-

sphere, and individuals are diagnosed in the city and in the countryside without regard to gender, age, social class, or ethnic origin – European, Amerindian, or African (Rubel 1964; Rubel et al. 1985; Hollweg 1997). Diverse symptoms that often resemble neurotic and somatoform syndromes, but may also include definitely organic diseases (Sal y Rosas 1958; Tousignant 1979), are given the folk diagnosis of *susto* or one of its equivalents under the following conditions:

1. The symptoms are related by their sufferers to a frightening experience, generally in the manner of a rationalization (O'Nell 1975).
2. This frightening experience is assumed to lead to a separation of the soul from the body, either spontaneously or through the action of chthonic or demonic powers (Sal y Rosas 1958; Logan 1979; Tousignant 1979), which decidedly weakens the resistance of the sufferers to disease-inducing influences.
3. The diagnosis of *susto* or equivalent is confirmed by a folk healer; this does not, however, rule out the consultation of a modern physician as well for treatment of the individual symptoms.

The folk-etiologic theory of loss of the soul is reminiscent of shamanic ideas of Paleolithic origin, which the Latin Americans adopted from the indigenous Amerindian population – an example of "reverse acculturation." The authors encountered the same conceptions in healing ceremonies of North American Indians (Jilek 1990) and Southeast Asian mountain tribes (Jilek and Jilek-Aall 1990).

Nervios, **Nerves (ICD-10: F32.11; ?F48.0)**
The psychosocial stress placed on postmenopausal women by marital conflicts, family problems, loneliness, and the suffering of emotional and material deprivations and losses may lead, in many, very diverse cultures, to a chronic, dysphoric mood state with somatic manifestations. The Hispanic population of the Caribbean, Central America, and North America and rural Anglo-American groups refer to such symptoms, in accordance with outdated academic medical conceptions, by the Spanish term *nervios* and the English term *nerves* (Low 1985; Nations et al. 1988).

Ataque de Nervios **(ICD-10: F43; ?F44)**
A direct, acute reaction to traumatic emotion-laden experiences in Hispanic women of the Caribbean and Central American area, and also in North America, is known as an *ataque de nervios* (Guarnaccia 1993; Guarnaccia et al. 1989; Oquendo 1994; Oquendo et al. 1992), an ictally appearing, uncontrolled behavior of brief duration and generally favorable prognosis, frequently a dissociated state that may resemble a *grande ataque hystérique*, in other cases an event resembling a panic attack (Liebowitz et al. 1994).

Hwa-byung **(ICD-10: F34.8)**
Korean women of mature age in Korea and North America, as well as women in other cultures, may react to continuous stresses in marriage, family life, and social situation, and to life situations perceived as unbearable, with somatic symptoms in an anxious and depressive mood state similar to *nervios* or *nerves*, but in the Korean case accompanied by pronounced feelings of anger and, frequently, by the sensation of a (nonexistent) epigastric mass. This chronic depressive condition is designated *hwa-byung*, "fire disease," in accordance with the conceptions of traditional Chinese and Korean medicine, and is correspondingly treated by specialized, shamanic healers, most of whom are women (S.H. Lee 1977; Lin 1983; Lin et al 1992; Min 1989; Min et al 1990; Pang 1990).

5 Culture-Typical Reactions to Extreme Environmental Conditions

"Arctic Hysteria"
"Arctic hysteria" is a collective designation for several conditions, some of them with dramatic manifestations, that have been observed and described among the indigenous population of arctic and subarctic regions, first by explorers and later by ethnologists, physicians, and psychiatrists. These brief reactions are provoked by a stressful situation in which the affected individuals, while experiencing acute anxiety, develop a temporary state of dissociated consciousness. The extreme climatic conditions of the arctic winter expose the inhabitants of these areas to physical deprivations and mental stresses that provide the ecological background for the various patterns of "arctic hysteria" (Novakovsky 1924). The survival of the small group, living together in a very crowded space, requires frictionless cooperation and thus demands of the individual, from earliest youth onward, the strict control of aggressive, competitive, and sexual impulses and emotions. In the traditional Eskimo culture, shamanic séances, in which the entire group actively participated, often provided an opportunity for the indirect expression of intense emotions in dissociated states of consciousness, e.g. in prolonged, ecstatic drum songs leading to a *pivdlerorneq* (Inuit, "drum dance fit") (Holtved 1967). These practices facilitated the occurrence of dissociative states of so-called arctic

hysteria in anxiety-producing situations (Foulks 1972). The best-known phenomena of this type are *pibloktoq* (Inuit, "crazy") and *kayak-svimmel* (Danish, "kayak dizziness").

Pibloktoq **(ICD-10: F44.7; F44.88)**
Pibloktoq-type attacks were observed among Eskimo groups in Greenland, Canada, Alaska, and Siberia and described by several authors in essentially similar terms (Brill 1913; Czaplicka 1914; Gussow 1960; Wallace and Ackerman 1960; Holtved 1967; Foulks 1972): the affected individual, usually a fairly young Eskimo woman, seems to be anxious and irritated, withdraws from group interaction, and starts to sing or to imitate bird or animal calls. Suddenly, she lapses into a state of wild agitation, tears the clothes from her body, and rushes – while screaming loudly – into the snow or ice, sometimes even into icy water. In most cases, however, the affected person remains within a certain range of safety, so that she can be easily saved. Violent actions directed against others or herself are rare. The usually brief stage of motor arousal ends with collapse, sometimes also with convulsive movements, followed by a deep sleep from which the affected person awakens with no clear memory of the event and without any sequelae whatsoever. The traditional culture generally ascribes the attack to possession by an evil spirit. In addition to the above-mentioned combination of ecological and psychosocial factors, the hypothesis has been advanced that *pibloktoq*-like conditions are caused by inadequate and unbalanced nutrition, leading to deficits of calcium and vitamin D_3, or to excessive amounts of vitamin A (Wallace and Ackerman 1960; Wallace 1972; Landy 1985). In the last few decades, in the course of rapid westernization, the small tribal societies of the arctic, subarctic, and Pacific region have been flooded with imported alcoholic beverages (Jilek 1987a). Under the influence of alcohol, the phenomenology of stress-induced reactions in the indigenous population of northernmost North America has changed: the traditional manifestations of arctic hysteria have now given way, under the influence of acculturation, to much more dangerous episodes of acting-out while intoxicated, in which, however, certain traditional behavior patterns can still be discerned (Foulks 1985).

Kayak-svimmel **(ICD-10: F41.0)**
In traditional seal hunts, the Eskimo often has to spend hours in a state of social isolation and sensory deprivation, sitting motionless in his kayak, surrounded by the monotonous loneliness of the Arctic, waiting for a seal to appear. Approximately 10% of seal hunters suffer an attack of *kayak-svimmel*, also known as kayak-angst, in such situations (Meldorf 1900; Bertelsen 1940; Ehrstrom 1951; Gussow 1963, 1970). Attacks of kayak-angst begin with dizziness, lightheadedness, and the feeling of a disturbance of balance. A sensation of cold rising from below makes the hunter fear that the kayak is filling with water. Anxiety increasingly grips him: he trembles and breaks out in sweat, and he may sustain a panic attack with inability to move and illusional perceptions. If help arrives in time, the affected individual may reach shore again, by himself or with assistance. Attacks of kayak-angst are accompanied by vegetative symptoms and may return with increasing intensity until the affected person gives up hunting in a kayak altogether. Now that the Eskimos have largely abandoned their traditional way of life, kayak-angst has become rare.

Windigo, Witiko, Witigo **(ICD 10: F32, in the Current Indigenous Meaning)**
The so-called *windigo* psychosis of the northern Algonquin tribes of Canada, especially the Ojibwa and Cree, has been mentioned under various diagnostic labels in practically all discussions of the culture-specific disorders ever since Linton (1956) referred to "cannibalistic insanity" as an example of a so-called primitive psychosis. Typically described as an obsessive and impulsive, murderous desire for human flesh as the result of delusional possession by, or transformation into, a mythical cannibalistic monster (*windigo*), the *windigo* phenomenon was long a favorite subject of psychodynamically inclined American anthropologists (Hallowell 1934; Landes 1938; Linton 1956; Parker 1960; Hay 1971). The periods of famine in winter which, when extreme, may have led to cannibalism out of necessity, despite the disgust it generally produces, were the ecological and historical background of the Indian myth of the *windigo* – the cannibalistic hibernal monster with a skeleton and a heart of ice. More recently, *windigo* behavior has been regarded as symptomatic of "classical depression" (Kiev 1972) or as a melancholic delusion, the content of which is based on tribal mythology and the struggle for survival in a subarctic climate (Shore and Manson 1981). The expression *windigo* is used today by the northern Ojibwa more generally to denote a state of sorrow, deep worry, or hopelessness (Marano 1982). Teicher (1960) reviews the literature on the *windigo* myths and phenomena in aboriginal North America as far back as the seventeenth century and identifies, on the basis of mostly anecdotal reports, 70 cases of *windigo* psychosis. Critical examination of these reports (Honigmann 1967; Marano 1982) could not confirm a single case of cannibalistic compulsion; in addition to sporadic earlier cases of cannibalism out of necessity, many of the individuals suspected of *windigo* were, in fact, victims of an epidemic of witch fear that struck several northern Canadian Indian

tribes during the first traumatic culture change after the European intrusion (Honigmann 1947; Marano 1982). The *windigo* phenomenon is now mainly of historical interest.

6
Mental Disorders as the Expression of Culture-Typical Emphases and Culture-Typical or Acculturative Stress Situations

6.1
Acculturative Stress Situation: Studying in a Culture-Alien Milieu

Brain Fag (ICD-10: F43.2; F48.0)
Mental stress and crises occasioned by the demands of academic study, particularly in connection with examinations, are frequently encountered in Western and East Asian cultures, despite centuries-old literary traditions and general familiarity with a school system based on reading and writing and typified by formal, competitive testing. Such academic demands place a much greater burden on schoolchildren and university students in developing countries, which have often made a rapid transition, in a single generation of "modernization," from a preindustrial, communalistic tribal society with oral traditions and cooperative, practical initiation learning to an individualistic, competitive education system imported from the West, with an emphasis on theoretical knowledge acquired from books. This type of knowledge often bears no relation to the daily life of the older generation, which, however, places heavy pressure on the younger generation to produce the academic successes which are expected to result in socioeconomic advancement in the new society – and this, in turn, further reinforces the students' mental stress. This situation is particularly typical of African students under acculturation pressure and is often associated with the development of a symptom complex known, since its first description and later delineation by Prince (1959, 1960, 1985, 1989), by the term "brain fag," coined by Nigerian university students. The clinical picture is characterized by the following:

- Unpleasant sensations in the head and body, especially of heat, burning, and crawling
- Eye and visual symptoms, especially blurry vision and tears while reading
- Inability to concentrate on books and lectures and to understand them
- Impairment of retention and memory
- Daytime fatigue and feeling of weakness, with frequent disturbance of night-time sleep

English-language abilities may suffer a transient decline (Boroffka and Marinho 1963). The intellectual capacity of affected individuals is not below average – indeed, the opposite is often true, as was mentioned by several observers and confirmed in a controlled study (Morakinyo 1985). Intercurrent organic illness of the affected students can be excluded as a cause by reason of the high prevalence of the disease; 50%–70% of Nigerian high-school and university students complained of typical symptoms when interrogated by questionnaire (Prince 1962; Jegede 1983). The brain fag syndrome has also been described in schoolchildren and university students in other African countries, including Uganda (German and Arya 1969), Ivory Coast (Lehmann 1972), Liberia (Wintrob 1973), Tanzania (Harris 1981), Malawi (Peltzer 1987; 1995), and Swaziland (Guinness 1992). These studies confirmed the relatively high frequency of the syndrome as well as the aspects in which brain fag differs from the anxiety states and depressive reactions evoked by academic stress in the West:

- Symptoms caused by the mere effort of reading
- The physical sensations of burning heat and crawling, often experienced as if there were insects crawling on – or worms beneath – the skin

In Africa, these symptoms form a part of the common, popular conception of being ill. Worms or reptiles in the head and body are often considered to be causes of disease in African folk medicine, whether the problem consists of stereotyped psychophysiological complaints (Makanjuola 1987) or an unambiguously organic condition, e.g. epileptic seizures (Orley 1970; Jilek-Aall 1979). Parasitic infestation is quite common in tropical Africa; in cases of helminthic infestation, the worms may be activated by fever, cause skin irritation, and emerge from bodily openings such as the nose and mouth (ascaris) or the conjunctiva (filaria).

6.2
Culture-Typical Stress Situation: Fear of Magical Persecution

***Bouffée Délirante*-Type Reactions (ICD-10: F23)**
Bouffée délirante, a term introduced in French psychiatry by Magnan (1893), was applied by French-speaking clinicians in West Africa (Aubin 1939; Vyncke 1957; Rainaut 1958; Salles 1961; Collomb 1965) and the Caribbean (Bustamante 1969a,b; Constant 1972) to acute transient psychotic or psychosis-like reactions evoked by a frightening experience, in the context of culturally accepted beliefs in magic, witchcraft, and the ever-present possibility of magical persecution or retribution. The symptoms consist of intense fear, dream-like confusion, unsystematized

paranoid delusions, and, frequently, visual and auditory hallucinations, usually accompanied by emotionally charged, dramatic, and sometimes dangerous acting-out behavior. A complete remission follows after a relatively short time, either spontaneously or after sedation with medications or medicinal herbs. The experience is usually sealed off by "amnesia" or denial of the event, for which the sufferer, in traditional cultures, is not held responsible. In a small number of cases, the episode may be repeated under similar conditions.

Reactions corresponding to *bouffée délirante* in English-speaking Africa are known under various diagnostic terms designating acute paranoid reactions. Lambo (1960) described an acute, agitated anxiety reaction, similar to a hysterical twilight state, that often occurs in the context of fear of bewitchment, under the name of "frenzied anxiety" and equated this reaction to *bouffée délirante* (for a review of transient psychoses in Africans, see Jilek and Jilek-Aall 1970). In the last decade, such psychotic reactions were also reported in Zimbabwe (J. Stevens 1987), Swaziland (Guinness 1992), and Egypt (Okasha et al. 1993). As several authors have stressed, most cases of *bouffée délirante* have no organic or infectious cause. Transient psychotic or psychosis-like reactions evoked by the acute fear of being affected by magic are in no way a specific phenomenon of African cultures; they have been described with some frequency in other parts of the world, including Indonesia (Pfeiffer 1994) and Tonga (Jilek 1988) as well as in southern European migrant workers (Labhardt 1963; Risso and Böker 1964). In their manifestations and course, *bouffée délirante*-type conditions are reminiscent of two nineteenth-century disease entities: the "transient amentia" of the Vienna School (Meynert 1889) and *folie hystérique*, the "hysterical psychosis" first described by Morel (1860) in Paris, a diagnostic concept revived by American authors in the 1960s with special reference to transient psychotic reactions in non-Western cultures (Hollender and Hirsch 1964; Langness 1967; Hirsch and Hollender 1969). The special significance of *bouffée délirante*-like conditions for transcultural psychiatry lies in their particularly frequent occurrence in previously tradition-directed indigenous societies undergoing a rapid transformation to modernity, in which the individual is exposed to loss of status and role confusion. This was stated as early as 1965 by Collomb (1965), repeated by later authors, and, most recently, confirmed by thorough analyses in Swaziland (Guinness 1992). The common feature of affected individuals coming under acculturation pressure in various lands is that they are still adhering to a traditionally derived, premodern world view to such an extent that they view magic and witchcraft as real threats, while at the same time having lost the protective, compensating resources of their old culture.

6.3 Culture-Typical Emphasis: Procreation and Fertility

Koro, Suo-yang (ICD-10: F48.8)

Psychoanalytical authors thought that they had found a real-life manifestation of oedipal castration anxiety in the form of the "genital disappearance" syndrome, *koro* (Kobler 1948; Devereux 1954). Yap (1965) elevated *suo-yang/koro*, which he viewed as a depersonalization, to the status of a paradigmatic example of a culture-bound syndrome (a term he had coined; see above). Among the many culture-bound syndromes, *koro/suo-yang* is the only one to have occurred in mass epidemics – indeed, on several occasions. We shall thus discuss this phenomenon in greater detail.

The Malay word *koro* is generally used today for an acute state of anxiety, accompanied by vegetative symptoms, in which the affected person believes that he or she can perceive, if he is male, the shrinking and disappearance of his penis or, if female, the shrinking of her breasts and/or labia; it is assumed not only that this process leads to impotence and/or sterility, but also that complete retraction of the external genitalia will inexorably be followed by death. This explains the "life-saving" emergency treatment practiced by folk healers in Asia when confronted by the frequently panic-like *koro* attacks, consisting of holding the penis in place, either manually or with special instruments (see the figure in Van Wulfften Palthe 1934). The term *lasa koro*, "penis shrinking," first appeared in print in Sulawesi in 1874, in Matthes's dictionary of the Buginese language. Indonesian colleagues have informed the authors that, in Sulawesi, the retraction of the head of a tortoise (Malay *kuro*) is viewed as analogous to the *koro* syndrome. Western medicine became aware of this syndrome at the turn of the century through the scientific reports of Dutch physicians in what is now Indonesia (Blonk 1895; van Brero 1897b; Voustman 1897), followed, in the 1930s, by ethnomedical analyses (van Wulfften Palthe 1934, 1935, 1936; Slot 1935; Mulder 1935). In the meantime, Western interest in classical Chinese medical theory had also made *suo yang* known to the West, specifically meaning "shrinking of the penis," or more generally "reduction of the male principle *yang*." The resulting excess of the female principle *yin* disturbs the health- and life-sustaining *yin–yang* equilibrium of the organism. Deficiency of *yang* and excess of *yin* lead to a state of "cold" that calls for the vitalizing remedies of traditional Chinese medicine; thus *suo-yang* is an indication for *yang*-strengthening remedies (Rin 1965). *Suo-yang* was already mentioned in the *Canon of*

Internal Medicine of the Yellow Emperor, Huang Di, compiled approximately 600–400 B.C., a classic work in which the *yin-yang* theory is explicated (Tseng 1973). A precise description of the *suo-yang* syndrome from the viewpoint of Chinese medicine was provided by the *New Collection of Remedies of Value* compiled by Bao in 1834 (Gwee 1968). Individual cases studied by Chinese psychiatrists in Taiwan (Rin 1965), Singapore (Gwee 1963), and Hong Kong (Yap 1965) mostly concerned insecure younger men whose *koro*-like anxiety attacks arose in connection with acute or chronic sexual problems and, in a few cases, in the context of a paranoid or depressive psychosis. Yap attempted to show that the popular conception of *suo-yang,* based on traditional Chinese medical theory, itself causes these *koro*-like symptoms to arise, under certain conditions, in susceptible individuals in the Chinese cultural sphere. He assumed that the syndrome was bound to Chinese culture and that it was exported to Indonesia by Chinese immigrants. This argument overlooks the fact that *koro* has long been endemic not only in Moslem, Christian, and animist ethnic groups of Sulawesi and Kalimantan, but also in populations that are not under Chinese cultural influence, such as the aboriginal tribes of the island of Flores (El Fakharani 1980) and of the interior of Mindanao, in the Philippines (J.W. Edwards 1985), not to mention its epidemic occurrence in non-Chinese cultures.

Kraepelin (1913) had already mentioned the occurrence of *koro*-like delusions in manic-depressive patients in Germany. With the growing interest of European and American clinicians in transcultural psychiatric literature, there are now more reports of isolated *koro*-like cases in Western countries (J.W. Edwards 1970; Dow and Silver 1973; Ang and Weller 1984; Berrios and Morley 1984; Smyth and Dean 1992; Adeniran and Jones 1994). What distinguishes these *koro*-like cases of Western patients from culture-typical *koro* or *suo-yang* in Asia is, first of all, that the *koro*-like symptoms are merely a part of a larger neurotic, psychotic, or toxic disorder; second, more importantly, that the patients develop the idea of genital disappearance on their own, and this idea is not shared by their cultural group, i.e. the idea is autistic rather than culture specific. These differential diagnostic features were pointed out years ago (Jilek and Jilek-Aall 1985) and later recommended for inclusion in the DSM-IV classification (Bernstein and Gaw 1990).

In Africa, *koro*-like conditions have been described chiefly in Nigeria. In these conditions, the affected individuals – usually men – were convinced that they had been robbed of their genitals by magical means, typically by strangers whom they had chanced to meet in the market, usually members of another tribe whom members of their own tribe generally distrust. Even when the "robbing of genitals" was belied by inspection, many affected individuals nonetheless believed that their sexual function and reproductive ability had been damaged by magical powers (Ilechukwu 1988). In Nigeria, in addition to sporadic cases of *koro*-like conditions (Ifabumuyi and Rwegellera 1979), there have also been small-scale epidemics of "magical robbing of genitals." These were stirred up by rumors and media reports to the effect that unscrupulous politicians or bankers had organized magical robbing of genitals in order to exploit the forces residing in the genitals, according to cultural tradition, for the magical enhancement of their own power or wealth (Ilewchukwu 1992). Modern Europeans to whom such ideas appear bizarre should be reminded that common popular opinion in the late Middle Ages assumed that a man was continuously liable to lose his virility, and his *membrum virile,* as the result of witches' magical attacks. The Dominican scholars Sprenger and Kraemer Institor (1487), devoted an entire chapter of their *Witches' Hammer* to this complaint, which was not infrequently voiced by young men at that time.

Koro Epidemics

The typical *koro/suo-yang* syndrome, consisting of a subjective perception of the shrinking or disappearance of the genitals or breasts with accompanying fear of death, has occurred in several times and places in epidemic fashion. The first European report of a *koro* epidemic in China was written by a French physician who observed it among young students in Szechuan in 1907; he remarked that this "Asiatic psychosis" could arise either individually or collectively, was of short duration, and was associated with the fear of being unable to procreate, "terrifying to the Chinese" (Legendre 1908). The principal site of *koro* epidemics in China is the southern part of the province of Guangdong (Leizhou and Hainan), where the mass occurrence of *suo-yang* is documented as early as 1865 and, in this century, in 1948, 1955, 1966, and 1974 – all years of particular political tension and socioeconomic turmoil in China. The last major epidemic, in 1984–1985, was thoroughly investigated by psychiatrists (Jilek 1986; Tseng et al. 1988, 1992); it struck several thousand people with typical *suo-yang* attacks, most of them male adolescents and young men, but also girls and young women. The authors had the opportunity to visit the region toward the end of the epidemic and, with the aid of Chinese colleagues, to interview affected individuals, their families, and their healers (Jilek 1986). Common among the sufferers were the feeling that their genitals or breasts were shrinking or retracting and the accompanying fear of death, with the corresponding psychomotor and autonomic/vegetative concomitants. Despite decades of Marxist indoctrination, the traditional belief still prevailed that

possession by female fox spirits caused disappearance of the genitalia. Fox spirits transforming themselves into *femmes fatales* are a classic theme of Chinese literature; the fox symbolically unites the dark domain of *yin* with the bright world of *yang* (Buber 1946). The traditional treatment of *suo-yang* victims consisted of holding the "shrinking" organ firmly and of exorcism rituals in which the fox spirits were to be "knocked out" of the body; images of the classical demon expeller Zhong Kui, or sometimes of Mao Zedong, who was identified with him, were used as prophylactic measures.

In 1967, the male Chinese population of Singapore was stricken by a *koro* epidemic affecting approximately 500 people. Because the symptoms followed the classical *suo-yang* pattern, the investigators (Koro Study Team 1969; Ngui 1969) concluded that indoctrination with the traditional Chinese medical theory of *suo-yang* had been an essential causal factor, although only one quarter of the affected individuals had prior knowledge of it. In fact, the epidemic was provoked by rumors of the deliberate poisoning of pork. The epidemic unfolded during a historical period of intense political and ethnic conflict between the Moslem Malays, who eat no pork, and the pork-eating Chinese, who felt threatened by the Malays after the British occupation forces withdrew (Murphy 1982). Psychiatric observers found no Chinese cultural influence in the *koro* epidemic in Thailand in 1976 (Jilek and Jilek-Aall 1977a,b; Suwanlert and Coates 1979). This epidemic of typical *koro* attacks began in provinces bordering on Vietnam, moved to other parts of the country, and ultimately affected more than 2000 individuals, most of them peasants, about a third of them women and children; all were Thai, and not a single one belonged to the Chinese minority. In a time of widespread fear of aggressive actions of the then victorious communist Vietnam, the Vietnamese were suspected of a malevolent attack on the reproductive capability of the Thai people by means of mass poisoning of food, drink, and tobacco. The rapid spread of the epidemic was aided by sensational news reports and stopped only by a vigorous explanatory campaign by Thai colleagues.

A few sporadic cases of psychosis with *koro* symptoms have been reported in India (Shukla and Misra 1981; Chowdhury 1992). In the summer of 1982, the "genital disappearance syndrome" attracted a great deal of attention in India through its rapid epidemic spread through Assam, West Bengal, and Meghalaya. The condition was known locally under the Assamese name *jhinjhini bimari*, "prickle disease," because a prickling feeling over the entire body preceded the *koro* symptoms. More than a thousand people of diverse ethnic, religious, and social origin – Hindus of all castes, Moslems, and animists – were involved, though not a single person of Chinese origin or under Chinese cultural influence. Indian investigators studied several hundred patients and identified them as having classic *koro* syndrome (Dutta et al. 1982; Nandi et al. 1983; Chakraborty 1984; Sachdev 1985). The affected individuals were mostly young men and male adolescents; it was estimated that one quarter of the sufferers were women. Another massive *koro* epidemic occurred in West Bengal in 1985 (Chowdhury et al. 1988; Chowdhury 1991), in the northern region of the state, an area whose ethnic and religious diversity is continually reinforced by immigration. The affected individuals this time included Hindus of Indian and Nepalese origin, Moslems, and mountain tribespeople. Women were represented in all affected ethnic groups and were the subject of a first special study (Chowdhury 1994); they complained mainly of retraction of the nipples and flattening of the breasts. The Indian *koro* epidemics came at a time of heightened social and interethnic tension in the stricken regions and were brought about by the collective fear of the rural population that they would be inundated with land-hungry immigrants (Chakraborty 1990).

In conclusion, the *koro/suo-yang* syndrome can be considered neither as a paradigm of pathogenic cultural indoctrination, nor as a manifest example of an oedipal castration complex. It is, rather, an expression of individual or collective fears, provoked by an assumed threat to reproductive ability, in cultures in which reproductive ability is a major determinant of a young person's worth.

Shen-k'uei, Dhat, Jiryan, Sukra Prameha
(ICD-10: F48.8; F45.34)

In traditional Chinese medicine, the "semen loss" syndrome carries the name *shen-k'uei*, "vital weakness of the kidneys" (Tseng 1973; Wen and Wang 1981); in Ayurvedic medicine, it is known by Sanskrit-derived designations, thus in India *dhat* (Wig 1960) or *jiryan* (Carstairs 1956), in Sri Lanka *sukra prameha* (Obeyesekere 1976). The "semen loss" syndrome is widespread on the Indian subcontinent, but has only been a subject of scientific investigation in recent decades (Malhotra and Wig 1975; Chadda and Ahuja 1990; Bhatia and Malik 1991; Paris 1992; Chadda 1995). This syndrome, in which the person supposedly suffering semen loss fears for his virility and reproductive ability, is closely related to *koro/suo-yang*, with which it may occur in combination (J.W. Edwards 1983; Bhatia et al. 1992). The patients, mostly young men, complain of a general decline of strength, loss of energy, exhaustion, disturbances of sexual function, vague bodily complaints, difficulty concentrating, and anxious-dysphoric mood. They attribute these symp-

toms to the loss of sperm in the urine or in nocturnal pollutions. Despite the hypochondriacal quality of the complaints, the patients' fears of loss of sperm must be taken seriously, because sperm, in both Chinese and Indian culture, is considered a precious, vital elixir that the organism must laboriously distill from the blood, just as the blood is distilled from food and drink. Ayurvedic tradition holds that sperm is the most valuable of the seven *dhatu*, or essential body substances. There are many Western parallels to the supposed consequences of sperm loss, ranging from Greco-Roman medical philosophy to the academic medical theories of the eighteenth and nineteenth centuries (Bottero 1991; Raguram et al.1994). Even today, true spermatorrhea would certainly arouse concern anywhere, but the *dhat* syndrome and its equivalents are distinguished by the fear of losing one's life-giving elixir in a society that places supreme value on male generative power.

6.4 Culture-Typical Emphasis: Making a Pleasant Impression on the Other Person in Interpersonal Relations

Taijin Kyôfu (ICD-10: F40.1)

Western psychiatry uses the terms "anthropophobia" or "social phobia" (*phobie sociale*, Janet 1903) for *taijin kyôfu*. There are a number of features common to Western social phobia and *taijin kyôfu*, particularly with respect to the manifestations of anxiety that are evoked by certain actual or anticipated contact situations. The Japanese conception of social phobia, however, is different from the Western one, as will be shown below. *Taijin kyôfu*, described by Morita (1921) as a subtype of the "constitutional nervous temperament" (*shinkeishitsu*), is in no way a general fear of human contact, but rather "a fear of the other person, the person one is facing, a fear of a particular interpersonal situation" (Kimura 1995, p. 135). Accordingly, Kasahara (1987) proposes the term "phobia of interpersonal relations" and shows that the interpersonal situation that arouses intense fear is one in which the other person is neither a complete stranger nor a member of one's intimate circle, but a person whom one knows to some extent. *Taijin kyôfu* is not a uniform disease type, but rather a group of structurally similar phobic anxiety conditions, including the following (Kimura 1972, 1995; Yamashita 1993):

- Fear of blushing in front of others (erythrophobia)
- Fear that others will be bothered by being looked at, resulting in avoidance of all eye contact (eye contact phobia)
- Fear that one's own, supposedly offensive body odor will disturb others (own-body-odor phobia)
- Fear that one has a particularly unattractive face or a repulsive physical defect (dysmorphophobia)
- Fear that one will tremble or make a face

The sufferers' conviction that they possess defects that they actually do not sometimes makes these phobias hard to distinguish from delusions that may also appear as temporary components of a schizophrenic mental disorder. The diagnosis of *taijin kyôfu* therefore requires the exclusion of the presence of other psychotic symptoms. The anthropophobe's relation to his human environment is fundamentally different from that seen in schizophrenic delusions of reference.

Anthropophobia of the Japanese type is a fear of presenting oneself (Iwai 1982). In severe cases, the affected individuals insist on the surgical correction of their supposed physical defect (Aoki 1981). If they have a certain degree of insight into their condition, those suffering from *taijin kyôfu* believe they are unable to master interpersonal relationships because of a "nervous weakness," as in Morita's concept of *shinkeishitsu*; they then seek help in one of the clinics that practice Morita therapy, often successfully. The theory of neuroses proposed by Morita and the Morita psychotherapy based on it are discussed in an extensive specialized literature in Japanese, which is summarized in several reviews (Suzuki and Takemura 1966; Kondo 1976; Kitanishi 1990; Goddard 1991).

According to a number of statistical determinations, social phobias are more common in Japan than in Western countries; they account for 6%–19% of neurotic patients in Japanese university outpatient clinics and a much higher percentage in private clinics (Kitanishi and Tseng 1988; Takahashi 1989). In earlier times, most *taijin kyôfu* patients were male adolescents or young men, but, in recent years, the number of female and older patients has risen, as has the number of serious cases, particularly of own-body-odor phobia and eye contact phobia. A number of Japanese authors attribute these changes in symptom prevalence and distribution to sociocultural change in the wake of industrial expansion, which influences interpersonal relationships negatively through emphasis on aggressive competition (Russell 1989).

Although *taijin kyôfu* shows parallels with the social phobia (DSM-IV: 300.23) diagnosed in North America, there is at least one major difference: the Japanese *taijin kyôfu* patient fears that the interpersonal encounter will disturb or insult the other person, while the American social phobia patient fears that such an encounter will be humiliating or embarrassing to him- or herself, i.e. although they have similar symptoms, the East Asian type of phobic anxiety is "allocentric," the American type "egocentric," in accordance with the characteristic features of each culture (Chang

1997). In their clinical analyses of *taijin kyôfu*, several authors adduce ethnographic data on the specifically Japanese cultural background of these anxiety states (Kasahara 1974; Kimura 1995; Takahashi 1989; Kirmayer 1991; Yamashita 1993). In fact, the *Tale of Genji*, written by Lady Murasaki Shikibu in 1001–1015 (Murasaki 1970), already tells of a young girl with symptoms similar to those of *taijin kyôfu*, as well as of the great importance generally attached to a pleasant body odor.

Doi (1982) analyzes the Japanese need for *amae*, the confident, imposing feeling of trust that a child has in his or her parents, and the desire to experience *amae* in interpersonal relationships with significant others; he attributes the phobic anxiety conditions discussed here to frustration of the need for *amae*. Kimura, in his semantic-phenomenological discussion of what he calls Japanese social phobias, comes to the conclusion that the self of these patients consists exclusively of the self perceived by other people. He proposed that social phobias are particularly common in Japan because "among the Japanese, the self does not contain the basis of its existence within itself" and because "in the structures of Japanese ego consciousness, the question of who one is is not answered within oneself" (Kimura 1995, p. 137).

Kirmayer (1991) stresses that *taijin kyôfu* is based, not only on specifically Japanese beliefs about the self, but also on cultural norms and rules of social interaction. His claim that there would be no *taijin kyôfu* without the medium of Japanese culture seems excessive, however, considering the relatively frequent occurrence of typical *taijin kyôfu* symptoms in Korea, particularly in young people of either sex (S.H. Lee 1987; S.H. Lee et al. 1994). The characteristic features in Korea are the same as in Japan, as is the allocentric orientation of the patient's anxiety. The first cases of "social phobia" reported in China, unlike the typical Japanese and Korean cases, show an egocentric orientation, as in North American social phobia (Yu and Prince 1991). New comparative psychiatric studies of university students in Japan, South Korea, and China show that *taijin kyôfu*-like anxieties in interpersonal situations are much more commonly found in Japanese and Korean students than in their Chinese counterparts, whose anxiety also has other contents (Kitanishi et al. 1995). Despite the shared Buddhist and Confucianist cultural heritage, the *taijin kyôfu* anthropophobias in Japan and Korea are different in type and frequency from social phobias in China, perhaps because China has undergone a different type of historical and political development.

6.5 Culture-Typical Emphasis: Propensity to Dissociation and Imitative Learning

Latah-Type Reactions (ICD-10: F48.88)

The peculiar *latah*-type reactions will be dealt with in some detail in this chapter because of their special importance for psychopathology (alternative interpretations of identical manifestations as exceptional behavior in normal psychology, a phylogenetic mechanism, an organic brain disorder, or a psychotic or neurotic disorder) and because of the central role that *latah*, like *koro*, plays in the controversy over the culture-relativistic or biopsychological orientation of transcultural psychiatry. The complete *latah* symptom complex consists of the following types of behavior:

1. Abnormal startle response with elevated general sensitivity to being startled, depending also on the individual- and culture-specific nature of the tactile, acoustic, and/or visual stimulus, and the status of the person provoking the startle response.
2. The startle response then induces the typical hypnoid alteration of consciousness, or dissociation, which can also be induced by fascination or other affective stimuli.
3. Next, there is decomposition of speech – production of unformed sounds or inarticulate words; impulsive, repeated utterance of sexual terms (coprolalia or, more accurately, pornolalia), sometimes in abbreviated or euphemistic fashion, occasionally progressing to obscene and aggressive verbal abuse.
4. Echo symptoms: echopraxia and echomimia, i.e. imitation of movements and facial expressions of the provoking person; echolalia, exact repetition of forms of address and greeting, even of words and sentences in languages of which the affected person is ignorant; and, sometimes, imitation of noises.
5. If the hypnoid alteration of consciousness is advanced, automatic obedience to commands may occur, often providing a source of merriment at the expense of the hapless *latah* sufferer.
6. In some cases, there is also cataleptic rigidity.

Areas in which the full *latah* reaction occurs include:

- Southeast Asia, the Malay-Indonesian culture area (van Brero 1895; Ellis 1897; Gimlette 1897; Abraham 1912; van Loon 1927; Yap 1952; Pfeiffer 1968; 1994; Geertz 1968; Kenny 1978; Winzeler 1995), including the Dayaks of Sarawak and the Rungus of Sabah (Chiu et al. 1972; Doolittle 1991; Winzeler 1991; Simons 1996), also Thailand (Suwanlert 1972),

Burma-Myanmar (Still 1940; Yap 1974), and the Philippines (Musgrave and Sison 1910; Simons 1996).
- Northern Eurasia, Ainu in Hokkaido and Sakhalin (Sakaki 1905; Uchimura 1935; Winiarz and Wielawski 1936; Ohnuki-Tierney 1980); Mongols (Aberle 1952), aboriginal peoples of Siberia (Hammond 1884; Jochelson 1908, 1910; Czaplicka 1914; Shirokogoroff 1935).

Recently, Minevitsch (1993) observed *latah*-type reactions among the Buryats of the Baikal region, as did Korolenko in 1996 during religious ceremonies of the Church of Christ in Novosibirsk (Professor C. Korolenko, personal communication, 1997). In the past, *latah*-type reactions, so-called Lapp panic, were described among the Sami and Skolt Lapps in northern Russia and northern Scandinavia (Collinder 1949). Cultures to which *latah* reactions are endemic have specific indigenous names for it: among Malay peoples *latah*, "nervous, ticklish," from the Sanskrit *lata*, "fool;" among the Thai *bahtschi*, "tickle-crazy;" among the Burmese *yaun*, *young-dah-te*, "ticklish;" among Filipinos *mali-mali*, and among the Ainu of Hokkaido and Sakhalin *imu*, "startle fright." In Siberia, *latah*-type reactions (Russian *miryachit*) occur among the Tungus (*olon*), Yukaghirs (*irkunji*), and Koryaks (*meryak*).

In the *latah* zones of northern Eurasia and Southeast Asia, *latah*-type reactions occur mostly in women. These peoples of northern Eurasia and Southeast Asia were not in historical contact with each other, have different types of social organization, and display no linguistic relatedness. What they have in common, however, is a cultural emphasis on the propensity to dissociation, practiced from youth onward in religious, social, and therapeutic ceremonies, shamanic séances, or trance rituals. These cultures also emphasize learning by imitation, often also by identification with the role model, and mastery of situations by copying a person behaving in a superior way.

Among the factors that may contribute to the formation of latah reactions in Malay culture, Murphy (1976) named, above all, child-rearing practices promoting dissociation and suggestibility; these result in rapid, but mechanical learning of behaviors and capabilities through passive-submissive obedience to instruction. Geertz (1968) wrote similarly about the upbringing of Javanese children. In Java, many women have an "initial dream" of sexual content some time before their first *latah* reaction (van Loon 1927; Pfeiffer 1968). This dream is culturally stereotyped and is popularly considered to provide an etiological explanation for *latah*, in addition to the action of spirits and vague notions of soul loss. In the Malay-Indonesian culture area, *latah* is regarded more as a personal trait than as an illness and is generally not thought of as an indication for traditional healing treatment. However, thorough history-taking often reveals that the first occurrence of *latah* was at a time of particular mental stress and/or was associated with a depressive reaction (Pfeiffer 1994).

The occurrence of *latah*-type reactions outside the *latah* zones of northern Eurasia and Southeast Asia has been documented for the following countries and peoples: South Africa, among the Bantus (Gilmour 1902); Eritrea and Yemen, among Yemenite Arabs (Infurna 1928; Sarnelli 1934), Libya, among the Senussi (Natoli 1937); Southern and Western Sahara among various ethnic groups (Repond 1940). Apart from the South African cases, these cases of *latah*-type reactions all occur among male members of the Arab-Islamic culture area, in which the *latah* syndrome is only partly developed, with startle and echo symptoms, but without coprolalia or automatic obedience to commands. In fact, since the studies by Landis and Hunt (1939) and later studies in human ethology, the startle response has been shown to be a universal human reaction pattern. Apparently no less universal is the imitative behavior of indigenous populations coming into close contact with Europeans for the first time, who, upon being surprised by the strangers' unfamiliar behavior, may fall into a trance and, without any sign of psychopathology, manifest extreme and prolonged echopraxia and echolalia. This was already observed by Darwin (1840) among the Indians of Tierra del Fuego; more recently, Gajdusek (1970) described similar experiences during the first contact with indigenous groups in South America and Melanesia, and Jilek-Aall (1979) during her encounter with Bantu women in a remote bush village in Tanganyika. Echo symptoms and compulsive imitation have been observed in practically all *latah*-type cases; it is thus not without reason that French psychiatry designates *latah* reactions as *névroses d'imitation*. The common occurrence of startle patterns, imitative behavior, and dissociation led Simons (1980, 1985b) to designate *latah* as *startle matching taxon* or *startle matching syndrome* (Simons 1996) and to define it as a culture-specific exploitation of a neurophysiological potential common to all human beings, indeed to all mammals.

The sporadic occurrence of *latah*-type manifestations in Europeans is occasionally mentioned in the old reports from Siberia; it affected Russian settlers who had close contact with aborigines suffering from *miryachit*. No European *latah* cases are known from Southeast Asia, where, in the colonial era, Europeans sometimes entertained themselves with the *latah* attacks of their servants. A European anthropologist who married into a Malay family with 37 *latah* cases in

1990, and has experienced this "family habit" regularly ever since, has so far remained immune to it (Bartholomew 1994). A Dutch woman in the Netherlands did, however, display *latah* symptoms, albeit without imitative behavior; she had never had any contact with the Malay-Indonesian culture (Jenner 1990). The best-known Euro-American example of a *latah*-like reaction is the so-called jumping syndrome among the rural French-speaking population of certain areas of Maine, New Hampshire, and Quebec. The jumping syndrome was described in the nineteenth century by Beard (1878, 1880) at meetings of the American Neurological Association as a trance state of brief duration induced by a sudden startle, in which the affected person jumps up, drops any object he happens to be holding, throws it away, or strikes out blindly with it, repeats words and sentences spoken to him, even in unfamiliar languages, and obeys any command given, even if nonsensical or dangerous. The jumping syndrome affects lumbermen of French Canadian descent, begins in early childhood, and runs in families. As in the Arab-Islamic cases mentioned above, these Catholic "jumpers" do not manifest pornolalia. Sporadic cases of jumping have been described in modern times as well (H. Stevens 1964, 1965; Kunkle 1967). Even though Beard (1878) stressed that this was a psychological trance state rather than a neurological disease, Gilles de la Tourette (1884) cited Beard's lectures on jumping in a *revue critique* together with the reports by O'Brien (1883, 1884) on *latah* and by Hammond (1884) on *miryachit* and defined these conditions as exotic variants of the syndrome that Charcot had named after him. The similarity of Gilles de la Tourette syndrome, an autosomal dominant neurological disease characterized by multiple motor and vocal tics, as well as coprolalia and complex mental symptoms, to *latah*-type reactions has been pointed out time and again thereafter (Mazur 1953). Recently, an attempt was made to explain this similarity by the hypothesis of common neurophysiological and neurochemical mechanisms (Howard and Ford 1992). Nonetheless, the different geographical distribution and sex-based prevalence of *latah* and Gilles de la Tourette syndrome are strong evidence against a common origin. Like Gilles de la Tourette, the earliest medical writers on the subject interpreted *latah*-type reactions as neurological or neuropsychiatric illnesses of a hereditary nature, such as petit mal epilepsy (Ellis 1897) or "provoked imitatory impulsive myospasia" (van Brero 1895) or else as a hypnotic reflex phenomenon (Gimlette 1897). The similarity of hypnotic and *latah* states was apparently demonstrated by the hypnosis experiments performed by Fletcher (1908), in which uncharacteristically aggressive behavior was produced in a female *latah* patient. The Siberian *miryachit*, too, was first regarded as an organic nervous disease (Hammond 1884) and evaluated as such when found, in indigenous recruits, by military physicians of the Tsar's army (there was a disciplinary case, originally reported by a Dr. Kaschin, among the Trans-Baikal Cossacks, who continually repeated the commands and curses of their colonel in loud echolalia; Czaplicka 1914). Nonetheless, the well-known Russian neurologist Tokarsky (1890) diagnosed *meryachenie* as a nonorganic condition. Kraepelin (1904), during his travels in Asia, felt that *latah* reminded him of observations in "hysterics." Later authors, too, regarded *latah* reactions as belonging to the sphere of hysterical phenomena, but interpreted them psychodynamically and related them to the repression of sexual impulses (Galloway 1922; van Loon 1927; Murphy 1976, 1982). In his comprehensive treatise, Yap (1952) characterized *latah* as essentially a fright neurosis, hysteriform only in the sense of Kretschmer's biological conception. Kretschmer (1958, pp. 20–24) explained his "set hypobulic reaction" with reference to the *imu* syndrome of Ainu women on Hokkaido (Uchimura 1956), in whom a culture-typical key stimulus, a snake or merely the word "snake," releases the *latah*-like "primal phylogenetic instinctual reaction." In contrast to these phylogenetically and biologically based explanations of the *latah* phenomenon, Murphy (1973) tried to show, in a comparative historical study, that *latah* first appeared in the mid-nineteenth century at a time of consolidation of the European colonial system in the Malay-Indonesian area and spread rapidly among the populations then under direct European influence. More recently, according to his thesis, the *latah* phenomenon spread from these centers to remote areas, the male cases that had appeared earlier stopped appearing, and the intensity of the attacks abated. Murphy thus considers *latah* not as an expression of the Malay cultural tradition, but as a transitional product of the interaction between this tradition and modernizing influences. In his main opus, Murphy (1982) extended this sociological hypothesis to the aboriginal Siberians, to the Francophone jumpers of Maine, and to *latah*-type behavior in general: a rapid increase in *latah* behavior always occurs when relatively powerless and indigent peoples come in contact with more powerful and wealthy groups and then seek to acquire the advantages of the latter by imitating those apparently superior and obeying their orders, without, at first, understanding them. Later, when it is realized that automatic obedience and imitative learning are less productive than understanding the principles underlying the model, the method of learning changes, and *latah* behavior becomes rarer. In the last two decades, *latah*-type reactions have been at the center of the discussion concerning the priority of universal

biopsychological or culture-specific factors in psychiatric syndromes in general, and in so-called culture-bound syndromes in particular. The debate was mainly led, on the cultural relativists' side, by Kenny (1978, 1983, 1990) and, on the side of the biopsychological universalists, by Simons (1980, 1983, 1985b, 1996). Pfeiffer (1994) recognizes the importance of both the general biopsychological and the culture-immanent perspectives; he considers the argument over priority to be futile and argues for consideration of the personal and family situation in each individual case.

6.6 Culture-Typical Emphasis: Propensity to Dissociation and Male Conflict Resolution Through Discharge of Aggression

Amok-Type Reactions (ICD-10: F68.8)

Today the word *amok*, with local variations, may be found in the dictionaries of all major languages and is generally used to denote an episode of apparently unprovoked, randomly aggressive, and homicidal behavior, ultimately ending either in suicide or in complete exhaustion and limited recollection of the event. Along these lines, *amok* is defined in the ICD-10 Annex *2* as a culture-specific disorder. In the American DSM-IV glossary, *amok* is referred to as a culture-bound syndrome; the classification "culture-specific explosive behavioral disorder" was originally suggested (Gaw and Bernstein 1992). The word *amok* comes from the Portuguese version, *amuco*, of the designation commonly used in the Hindu states of India for warriors who, like the *morituri* of ancient Rome, swore to attack and annihilate the enemy while holding their own death in contempt. Malay and Javanese warriors took over the Indian expression and the intimidating battle cry, "amok! amok!" (Shaw 1972). The heroic deeds of the *amok* warriors were exalted in popular legends such as the Malay epic *Hikayat Hang Tuah* (German translation by Overbeck 1986), in which individual *amok*-runners are glorified who heroically overcome shame or slights by carrying out mass murder with the traditional *kris* and then, ultimately, seeking and finding their own death (Pfeiffer 1994).

The word *berserkr*, occurring in the Old Norse sagas with the meaning "warrior clad in bearskin," underwent a similar semantic shift to that of *amok*. It originally designated the swordsmen in the service of medieval Scandinavian princes, who put themselves into an ecstatic state of furor in order to fight with superhuman strength. Some also wore wolfskin and were called *ulfhednar*, "wolf-heads," later "werewolves;" the shamanic idea of possession by wild animals, through donning bear- or wolfskin, plays a role here. The sagas tell not only of heroic troops of Scandinavian berserker warriors (who, incidentally, were also employed as mercenaries by Byzantine emperors), but also of individual *berserksgangr*, ecstatic fits of berserker rage, in which the storming berserker murdered at random (Weiser 1927; Laiblin 1965). In modern English, the expression "going berserk" generally denotes any type of aggressive frenzy.

In the Malay-Indonesian cultural sphere, the introduction of Islam in the fourteenth century converted running *amok* against the "infidels" to an act of religious fanaticism, and death suffered through such an act, unlike suicide (forbidden to Moslems), was considered pleasing to Allah. This tradition found its continuation in the so-called "Aceh murders" of Dutch colonists by Moslem fanatics among the Aceh of Sumatra until the end of the colonial period (Pfeiffer 1994) and also in the *amok*-like indirect suicide ritual (*juramentado*) of the Moslem Moros of the southern Philippines in the nineteenth century (Ewing 1955). Another motivation of *amok*-like behavior in the Malay-Indonesian tradition was social protest; thus enslaved debtors, for example, sought an honorable death through random murderous attacks as early as the fifteenth century (Murphy 1973). The culturally sanctioned threat of running *amok* in response to gross injustice served as a certain check on the abuse of power by rulers and wealthy men (Gullick 1958).

From the seventeenth to the nineteenth century, *amok* attracted much attention in the West through the reports of European observers and was identified with Malay-Indonesian culture, especially because no lesser figure than Captain Cook wrote from Java on the "prevailing practice" of running *amok* (Teoh 1972), and an exiled Javanese ran *amok* in Cape Town in 1786 (Prince 1991). *Amok* was often assumed to be a mental disorder caused by opium or other drugs (Pfeiffer 1994). This was claimed of *amok* by mid-nineteenth century physicians as well, along with somatic causes such as gastrointestinal diseases, febrile delirium, or dementia (Murphy 1973). Later, psychiatric diagnoses were increasingly applied: melancholy and other mental disorders (Swettenham 1889) or "epilepsia larvata" (Ellis 1893); notably, no explanation was offered for the lack of female cases. Van Brero (1897a) conceded the possibility of a "transient mental illness" with subsequent amnesia, or the claim of having seen animals or devils in the *amok* victims; nonetheless, he cautioned against calling an act insane simply because one thinks only a madman would be capable of it. Gimlette (1901) emphasized four diagnostic criteria that are still recognized as valid today:

1. Prodromal depressive brooding over actual or imagined slights.
2. An explosive homicidal outburst.

3. Persistent homicidal activity without apparent motivation.
4. Claim of amnesia for the event.

Van Loon (1927) reverted to the previous thesis of infectious delirium, stating, however, that the "peculiar psychic nature" of the Malays – pathoplastic factors – made the symptoms manifest themselves in an entirely different form than in Europe and, in general, favored the appearance of confusional states. A decade later, Amir (1939) found that most hospitalized patients were suffering not from infectious diseases, but from chronic organic brain disease or schizophrenia. Van Wulfften-Palthe (1933), who collected information about *amok*-runners among the rural population of Java and Sumatra, saw *amok* as a twilight state with sudden, unmotivated, and undirected acts of violence, which can also occur in otherwise normal individuals if the emotional stimulus is strong enough. In the 1960s, the few hospitalized patients suffering from *amok* in West Malaysia all had schizophrenic symptoms (E.S. Tan 1965, quoted in Teoh 1972). In Sarawak, East Malaysia, Schmidt et al. (1977) studied 24 *amok*-runners, including, besides Malays, also Dayaks and other aborigines, who apart from *amok* were all psychiatrically diagnosable as having organic psychosyndromes, endogenous psychoses, or neurotic disorders. Murphy (1973) documented the change of significance and function of *amok* behavior in the Malay-Indonesia area, from a glorified warrior ethos to a sanctioned means of social control and social protest which, after the establishment of the European colonial system and a modern justice system, became criminalized and unsuited to its purpose. *Amok* behavior was then transformed from a consciously motivated act to a dissociative reaction, and at present *amok* behavior is found in the Malay culture area usually as an episode in the course of a chronic mental disorder. After the Second World War, *amok* appeared on the Malay peninsula among Chinese as well as Malays (Teoh 1972). Pfeiffer (1994, p. 135) characterizes *amok* in Indonesia as an exceptional dissociative state and distinguishes the following phases:

1. Dysphoric-neurasthenic preliminary phase in the context of difficulties with the social environment, chronic illnesses, loss of social rank, or lessening of personal prestige.
2. Triggering by a sometimes trivial event that leads to decompensation of the state of heightened tension.
3. Meditative brooding, alteration of the state of consciousness, and changes in visual perception – the outside world becomes dark (*mata gelap*, "eye dark") or red, the affected person feels surrounded by threatening figures and experiences fear or rage.
4. The affected person is suddenly seized by a storm of hypermotility, the *amok*-run begins and leads to random aggression, killing, and destruction.
5. This finally ends in suicide or self-mutilation; if the affected person survives, the attack ends in deep sleep or stupor, followed by depressed mood. Amnesia is claimed for the *amok* deeds.

Yap (1951) argued for a distinction between the "true *amok*" of the Malays from other paranoid-homicidal reactions; he later classified *amok* among his culture-bound reactive syndromes as "disordered consciousness" and "rage reaction" (Yap 1967, 1974). Carr (1985) takes an unequivocally culture-relativistic position. He discusses how the major characteristics of the traditional Malay conceptual system relating to notions of value, status, and health contribute to the *amok* phenomenon, and he concludes that a significant prevalence of *amok* can only be found in a population with Malay concepts and behavioral norms. Earlier still, Carr and Tan (1976) had concluded, after study of 21 hospitalized patients with "true *amok*" in Malaysia, that *amok* is not a disease per se, but rather a "prescribed form of violent behavior" in Malay culture, released by many different factors, traditionally sanctioned as an appropriate behavior under specific conditions. Two arguments against this are that *amok*-runners are usually marginal figures, and not typical representatives of Malay-Indonesian culture, and that *amok* is extremely rare in areas of this culture that are still highly traditionally oriented (Pfeiffer 1994). The thesis of the specific Malay character of the *amok* reaction is cast into further doubt by the not infrequent observation of the *amok* syndrome among non-Malay populations. There was only one Malay among the Indonesian patients extensively documented by Pfeiffer (1994, p. 137); one can, of course, view the non-Malay ethnic groups of Indonesia as being under Malay cultural influence. With certain restrictions, this may also be true of the Philippines, where 25 soldiers were arrested between 1946 and 1956 for *amok* acts committed with firearms (Zaguirre 1957). These cases were very similar in their psychosocial aspect to the "grenade *amok*" seen in Laos during the Indochinese war between 1959 and 1970. Westermeyer (1972, 1985) reported 18 cases of *amok* attacks by young soldiers in Laos, who, in an impulsive anger reaction after a personal experience of loss or of loss of prestige, threw a hand grenade into a crowd. These soldiers killed an average of ten wholly innocent people; the majority had drunk alcohol, and half of them simultaneously committed suicide. In his attempt at a transcultural epidemiological analysis, Westermeyer (1973) comes to the conclusion that *amok* homicide waxes and wanes periodically, becoming particularly frequent at times of internal armed

conflicts and rapid social change; *amok* behavior can be "transmitted" from one ethnic group to another, contiguous one. It is typically found in alienated, relatively uneducated young men. Aubin (1939) described analogous *amok*-like attacks by Senegalese soldiers in West Africa as *état de fureur*, and Carothers (1947) described them in Kenya als "frenzied anxiety." Sir B.G. Burton-Bradley, who served as a psychiatrist in Papua New Guinea for many years, often had occasion to study surviving *amok*-runners forensically and to interview eyewitnesses to multiple homicides. His detailed description of the *amok* behavior of probands of highly differing ethnolinguistic groups, some of them still on the developmental level of the Stone Age, are entirely in agreement with the classical descriptions from Malaysia and Indonesia, except as regards visual symptoms (Burton-Bradley 1968, 1975, 1989). *Amok* must be distinguished from pseudo-*amok*, which was widespread among the population of the Highlands of Papua New Guinea. Rage mimicking that of an authentic *amok* attack, with dramatically threatening, but not murderous behavior, reminiscent of theatrical bravado, has been described in different tribes and variously interpreted as acute hysterical psychosis (Rodrigue 1963; Langness 1965), "wild man behavior" (Newman 1964), or a state of possession (Koch 1968). Pseudo-*amok* has also been attributed to the consumption of supposedly hallucinogenic mushrooms in accordance with native tradition (Reay 1960, 1965), but this was soon refuted by the eminent mycologists Heim and Wasson (Reay 1977). Pseudo-*amok* in New Guinea also serves the functions of individual catharsis and collective entertainment (Clarke 1973); the participation of costumed female "*amok*-runners" in such displays (Aufenanger 1973) offers an almost parodic contrast to the overwhelmingly male type of authentic, murderous *amok* behavior seen in New Guinea and elsewhere in the world.

In a historical study, Kon (1994) arrived at the conclusion that *amok* is not a culture-bound syndrome. It is certainly no longer possible to speak of *amok* as a specifically Malay-Indonesian phenomenon, in view of the steadily accumulating reports, in recent decades, of *amok* behavior by men of European extraction in the Western cultural sphere. In his comparative analysis of such cases, Arboleda-Florez (1979) concludes that typical *amok* behavior can arise anywhere under conditions of a rapidly changing society among individuals with feelings of social alienation and a simultaneous need for demonstrative self-assertiveness. In this connection, mention may be made of the current Western zeitgeist, propagated by mass media, which contributes to the popularization and secret admiration of macho Rambo and other "killer" types (Tan 1989). An increased prevalence of dissociative alterations of consciousness and widespread use of psychoactive substances are also features of the contemporary North American scene. This is of prognostic importance, because dangerous *amok* behavior is largely determined by an uncontrolled propensity to dissociation and a tendency toward disinhibited violence.

6.7
Culture-Typical Emphasis: Thinness in Women

Anorexia Nervosa (ICD-10: F50.0)

Immediately after formulating his concept of the culture-bound reactive syndromes, Yap (1969) proposed classifying anorexia nervosa as a Western culture-bound syndrome. Religiously motivated fasting has, of course, been seen as legitimate from time immemorial in many cultures, particularly in the Judaeo-Christian and Hindu-Buddhist spheres, and should not be interpreted as psychopathological (Banks 1992), nor should fasting for the accomplishment of political or social aims (Mahatma Gandhi!) been seen as such. Refusal of food because of lovesickness is known from Greek culture as early as the fourth century B.C. and is a classic theme in world literature, as in the Persian epic of Leila and Majnun (Pfeiffer 1994). Fasting cures for therapeutic or general "health-promoting" purposes are often recommended by both academic and "alternative" medicine. The phenomenon of refusal of food that arose in the nineteenth century among the daughters of the Western bourgeoisie and is today known as anorexia nervosa had nothing to do with religious, political, or even romantic motives. This clinical picture was first mentioned in North America by Chipley (1859) under the name *sitomania*, "dread of eating" (Vandereycken and Lowenkopf 1990). It was then described in Europe by Gull (1874) as anorexia nervosa and by Lasègue (1873) als *anorexie hystérique*, with essentially the same symptomatology now familiar to every clinician. These pioneers of psychiatry showed an impressive insight into the psychosocial causation of the syndrome, in contrast to some later authors, who postulated an organic endocrine etiology and occasionally confused cause and effect. The fascinating history of the anorexia nervosa phenomenon is discussed in detail by Brumberg (1988) and in summary form by DiNicola (1990a). Analysis of currently available epidemiological data clearly shows the decisive role of sociocultural factors in the genesis of this phenomenon (Garner and Garfinkel 1980; Prince 1985; DiNicola 1990b; Gordon 1990; Diedrichsen 1991; Garner 1991; Iancu et al. 1994; Pfeiffer 1994; Toro et al. 1994; Weiss 1995). On the basis of these data, anorexia nervosa has been shown to be a disease of modernity, without any bio-organic cause, predominantly affecting young

female members of relatively well-to-do families under the influence of idealized fashion images. Anorexia nervosa occurs primarily, and with steadily increasing incidence, in the Western cultural sphere, and recently also in upper-class girls of non-Western cultures heavily exposed to modern Western influences. The rise of anorexia nervosa (and the often associated "other side of the coin," bulimia) to a "social epidemic" (Gordon 1990) among female adolescents of the Western world was made possible by the economic and social consequences of full industrialization and of the modern consumer society, which led not only to widespread material affluence, but also to nuclearization of the family, with increasing family conflicts. In Europe, the number of cases of anorexia nervosa declined dramatically during the war years and the immediate postwar period, in which food was scarce; many chronically anorexic patients abandoned their symptoms. In today's consumer society, however, extreme slimness is the dominant ideal of feminine beauty, created by fashion, in paradoxical contrast to the enormous supply and consumption of foodstuffs. With the rise of the general standard of living, and under the influence of the visual mass media, anorexia nervosa spread from the upper and middle classes to the working class, and from western and northern to southern Europe; the disappearance of traditional family relationships played an important role in the latter development (Selvini-Palazzoli 1985). In recent decades, North America, too, has seen a remarkable increase in the incidence of anorexia nervosa, which has doubled or trebled in some areas since 1960. It has been shown that this disease became dramatically more common when the mass media promulgated the fashionable ideal type of the thin, infantilized female body type (Garner and Garfinkel 1980; Jones et al. 1980; Garner 1991; Lucas et al. 1991). Curiously, the weight tables used by American physicians from 1960 onward also show a progressive decline in female "ideal weight," so that it appears that supposedly culture-free biomedical standards have been adapted to fit popular trends in modern Western society (Ritenbaugh 1982). In Spain, where modern Western lifestyles were not adopted by most people until the 1980s, the same pathogenic effect of fashionable thinness, propagated by the advertising media, can be observed today (Toro et al. 1994; Caballero 1995). Alongside the diet-and-slimming industry, which earns high profits from countless proffered publications and therapies, a new medical and paramedical eating disorder industry (Shorter 1994) has now arisen in the Western world that also does its part to keep this fashionable illness (in both senses) in the public eye. In many traditional cultures of Asia, Oceania, and Africa, as in Europe, too, in the more distant past, good health, high social rank, and even attractiveness in young women was generally identified with a more adipose nutritional state. Since the 1980s, there have been several reported cases of anorexia nervosa in non-Western populations (DiNicola 1990b; Davis and Yager 1992; S. Lee 1996), e.g. in Japan, Hong Kong, Taiwan, Singapore, China, India, Malaysia (Chinese and Indian ethnic groups), and even in Nigeria. With the exception of Japan (Mizushima and Ishii 1983; Suematsu et al. 1985), classical anorexia nervosa in non-Western countries is still relatively rare compared to Europe and North America, predominantly arising in female students in large cities, as distinct from symptomatic anorexia in the course of a physical illness or depression (Nigeria: Ebigbo and Okunna 1986/87). Many of the non-Western patients probably lack the specifically Western symptom of fear of obesity and the obsessive idea of being fat, which was already recognized in 1880 by Charcot in Paris (Habermas 1991). The majority of Chinese anorexia nervosa patients in Hong Kong, for example, had no "fat phobia" of this type. An increasing frequency of this diagnostic criterion is, however, to be expected as non-Western societies become increasingly westernized (S. Lee et al. 1993); questionnaires given to female students in Hong Kong already show a break with traditional Chinese ideals of beauty in favor of the modern fashion for thinness (S. Lee 1993). The occurrence of anorexia nervosa among young women belonging to all classes of Western society, and the appearance of this pathological phenomenon in non-Western population groups under the influence of westernization, as well as its frequent occurrence among young female immigrants under Western acculturation pressure, are all compatible with the characterization of anorexia nervosa as a "culture change syndrome" (DiNicola 1985). This syndrome is to be understood as an individual reaction to modern Western culture's typical demand for feminine thinness, which, since the nineteenth century, has spread from the wealthy upper class to all social strata of the West and has now also begun to affect non-Western cultures.

7 Culture-Appropriate Diagnosis: Differentiation of Culture-Specific Disorders from Institutionalized Culture-Congenial States

In the preceding discussion, culture-specific mental disorders were presented as expressions of culture-typical emphases or as a result of stressful situations that may arise in a given culture or in a situation of culture change.

It is important for clinical practice to distinguish these culture-specific disorders from exceptional states that are institutionalized in a particular culture. To make such a differential diagnosis, the clinician must possess certain general prerequisites of transcultural psychiatric assessment (Tseng and Streltzer 1997), especially sensitivity for cultural particularities, understanding of the patient's culture and religion, and awareness of cultural variations of psychopathology.

One may generally state of the culture-specific disorders that occur in non-Western cultures that they (a) do not appear in the framework of ritualized behavior and (b) are recognized by indigenous healers as abnormal or morbid, and can often be treated successfully through traditional practices which have been found to be therapeutically effective in many cases of reactions caused largely or entirely by psychosocial factors (Jilek 1993).

7.1 Altered States of Consciousness

Altered states of consciousness are dissociative phenomena that also appear in a number of culture-specific mental disorders. Such states, in other forms, are institutionalized in tradition-directed cultures as nonpathological dissociations and are considered acceptable behavior when tied to culturally defined ritual conditions, in which they are induced and manifested. Altered states of consciousness are associated with characteristic changes of attention, sensory perception, affectivity, and the experience of time, and also with hypersuggestibility and culturally or religiously conditioned feelings (Ludwig 1966). In these states of consciousness, the everyday feeling of personal identity is experienced as specifically changed or as replaced by another entity. In the religious systems of tradition-directed cultures, such altered states of consciousness are thought of as extraordinary modalities in which the affected individual (a) believes himself or herself to be in special contact with a perceived being of variable provenance, identified as divine or demonic, as the spirit of a dead person, protective spirit from the animal world, etc., or (b) believes that he or she is possessed by a supernatural being whose power operates through the possessed individual; the acts of the possessed thus imply no personal responsibility and are sealed off by amnesia. These two fundamental variants of altered states of consciousness were designated by Bourguignon (1973) as (a) "trance" and (b) "possession trance" and have been reported to occur in many different cultures. The differences between these states are culturally defined and not founded in psychology or neurophysiology.

7.2 Religious Ecstasy and Demonic Possession

In traditional Christianity, trance states conforming to the belief system, such as religious ecstasy and experiences of revelation, are sanctioned at any given time according to the preference of the particular church in question, today not uncommonly in the framework of charismatic movements and fundamentalist religious communities. In the religious exercises that often precede ecstatic experiences, one may note the presence of suggestion and other somatopsychic factors that, in general, facilitate the induction of altered states of consciousness, such as fasting, dehydration, sleep deprivation, sensory deprivation, and monotonous auditory stimulation. In Judaeo-Christian tradition, the expression "possession" is reserved for possession by demonic spirits and is diagnosed and treated (by exorcism) by some ecclesiastical authorities even today (Lhermitte 1963; Rodewyk 1963; van Dam 1970).

7.3 Trance Rituals and Possession Cults

There are institutionalized trance rituals and possession cults in many non-Western cultures, because "possession," in cultures not influenced by Christian tradition, may also be thought of as positive. Ritualized trance and possession states are learned and practiced in the framework of a ceremonial for religious and/or therapeutic purposes. This happens, for example, in Amerindian initiation rites, in which altered states of consciousness are induced in motivated candidates without the effect of drugs by means of direct and indirect suggestion and rhythmic acoustic and kinetic stimulation alternating with sensory deprivation, sleep deprivation, fasting, dehydration, hyperventilation, and other induction methods (Jilek 1987b). Many of the collective trance and possession cults practiced in non-Western cultures have demonstrable psychohygienic, therapeutic, and social integrative functions. (For a global overview of these ritualized group therapies, see Jilek 1993, 1994.) A few examples may be named here: The *zar* cult is a ritual therapy for female patients with affective and psychosomatic symptoms in northeastern Africa and the Near East. The female possession cult *zebola* has a similar function in the Congo region. West African trance and possession rituals such as *orisa*, *rab*, and *lup* have historical connections to the Afro-Caribbean (*vodun*, *santeria*, *shango*) cults and to the Brazilian *umbanda*, currently the numerically largest form of religious group therapy, which syncretistically unites African

traditions, Western spiritism, and Amerindian culture elements. Indonesian trance rituals and the ecstatic self-imposed ordeals of the Hindu *taipusam* similarly have religious and psychohygienic significance, as do the revived cult dance rituals of North American Indians.

7.4
Eurocentric and Positivistic Fallacies

Western or Western-trained observers risk committing a eurocentric fallacy through the pathology labeling of institutionalized and ritualized conditions in non-Western cultures, without paying due regard to the behavioral norms and traditional explanatory models of the culture in question.

A positivistic fallacy is committed when one considers ideas and behaviors to be irrational and abnormal because they do not fit into the framework of logico-experimental science. Ritualized trance and possession states, like other religious acts, are primarily "manifestations of sentiments" (Pareto 1935), and not generally indicators of psychopathology.

8
References

Aberle DF (1952) "Arctic hysteria" and latah in Mongolia. Trans New York Academy Sci Series II 14: 291–297

Abraham JJ (1912) Latah and amok. BMJ 24: 439

Adeniran RA, Jones JR (1994) Koro: culture-bound disorder or universal symptom? Br J Psychiatry 164: 559–561

Ahmed H, Naeem S (1984) First rank symptoms and diagnosis of schizophrenia in developing countries. Psychopathology 17: 275–279

Amir M (1939) Over eenige gevallen van amok uit Noord-Sumatra. Geneeskundig Tijdschrift voor Nederlandsch-Indie 79: 2786–2798

Ang PC, Weller MPI (1984) Koro and psychosis. Br J Psychiatry 145: 335

Aoki M (1981) Jusho taijin kyôfu. In: Iida S, Iwai H, Yoshimatsu K (eds) Taijin kyôfu. Yuhikaku, Tokyo, pp 35–40

APA (1994) Diagnostic and statistical manual of mental disorders, 4th edn. American Psychiatric Association, Washington DC

Arboleda-Florez J (1979) Amok. Bull Am Acad Psychiatry Law 7: 286–295

Arieti S, Meth JM (1959) Rare, unclassifiable, collective and exotic psychotic syndromes. In: Arieti S (ed) American handbook of psychiatry, vol 1. Basic, New York, pp 546–563

Aubin H (1939) Introduction à l'étude de la psychiatrie chez les noirs. Ann Med Psychol 97: 1–29,181–213

Aufenanger H (1973) Krankheiten und Heilmittel im Bismarck Gebirge und am Wahgi-Fluss im Hochland von Neu-Guinea. Ethnomedizin 2: 329–360

Banks CG (1992) 'Culture' in culture-bound syndromes: the case of anorexia nervosa. Soc Sci Med 34: 867–884

Bartholomew RE (1994) Disease, disorder or deception? Latah as habit in a Malay extended family. J Nerv Ment Dis 182: 331–341

Beard GM (1878) Report to the American Neurological Association. J Nerv Ment Dis 5: 526

Beard GM (1880) Experiments with the "jumpers" or "jumping Frenchmen" of Maine. J Nerv Ment Dis 7: 487–490

Beguin A (1952) Qui est fou? Esprit 20: 777–788

Berlin B, Kay P (1969) Basic color terms: their universality and evolution. University of California Press, Berkeley

Bernstein RL, Gaw AC (1990) Koro: proposed classification for DSM-IV. Am J Psychiatry 147: 1670–1674

Berrios GE, Morley SJ (1984) Koro-like symptom in a non-Chinese subject. Br J Psychiatry 145: 331–334

Bertelsen A (1940) Grønlandsk medicinsk statistik og nosografi. Meddelelser om Grønland, vol 117, no 3. Reitzel, Copenhagen, pp 176–190

Bhatia MS, Malik SC (1991) Dhat syndrome – a useful diagnostic entity in Indian culture. Br J Psychiatry 159: 691–695

Bhatia MS, Thakur KN, Chadda RK, Shome S (1992) Koro in dhat syndrome. Indian J Soc Psychiatry 8: 74–75

Blonk JC (1895) Koro. Geneeskundig Tijdschrift voor Nederlandsch-Indie 35: 150–168

Boroffka A, Marinho AA (1963) Psychoneurotic symptoms in urbanized Nigerians. Transcult Psychiatr Res Rev Newslett 15: 44–46

Bottero A (1991) Consumption by semen loss in India and elsewhere. Cult Med Psychiatry 15: 303–320

*Bourguignon E (1973) Introduction: a framework for the comparative study of altered states of consciousness. In: Bourguignon E (ed) Religion, altered states of consciousness, and social change. Ohio State University Press, Columbus, pp 3–35

Brill AA (1913) Piblokto or hysteria among Peary's Eskimos. J Nerv Ment Dis 40: 514–520

Brumberg JJ (1988) Fasting girls: the emergence of anorexia nervosa as a modern disease. Harvard University Press, Cambridge, MA

Buber M (1946) Introduction. In: Quong R (ed) Chinese ghost and love stories, a selection from the Liao Chai stories by Pu Sung-Ling. Pantheon, New York, pp 9–13

Burton-Bradley BG (1968) The amok syndrome in Papua and New Guinea. Med J Aust 1: 252–256

Burton-Bradley BG (1975) Stone age crisis: a psychiatric appraisal. Vanderbilt University Press, Nashville

Burton-Bradley BG (1989) Das Amok-Syndrom in Papua und Neu Guinea. Curare 12: 177–182

Bustamante JA (1969a) El bouffée délirante en nuestro medio. Psiquiatria Transcultural (Habana) 3: 5–20

Bustamante JA (1969b) La réaction psychotique aigue, la transculturation, le sous-développement et les changements sociaux. Psychopathol Afr 5: 223–233

Caballero L (1995) Anorexia nerviosa como trastorno etnico. Psico 1: 5–7

Carothers JC (1947) A study of mental derangement in Africans, and an attempt to explain its peculiarities more especially in relation to the African attitude to life. J Ment Sci 93: 548–597

Carr JE (1985) Ethno-behaviorism and the culture-bound syndromes: the case of amok. In: Simons RC, Hughes CC (eds) The culture-bound syndromes. Reidel, Dordrecht, pp 199–223

Carr JE, Tan EK (1976) In search of the true amok: amok as viewed within the Malay culture. Am J Psychiatry 133: 1295–1299
Carstairs GM (1956) Hinjra and jiryan. Br J Med Psychol 29: 128–138
Castillo RJ (1997) Culture and mental illness: a client-centered approach. Cole, Pacific Grove
Chadda RK (1995) Dhat syndrome: is it a distinct clinical entity? Acta Psychiatr Scand 91: 136–139
*Chadda RK, Ahuja N (1990) Dhat syndrome, a sex neurosis of the Indian subcontinent. Br J Psychiatry 156: 577–579
Chakraborty A (1984) An epidemic of koro in West Bengal, India. Transcult Psychiatr Res Rev 21: 59–61
Chakraborty A (1990) Identity, land and sex. In: Stefanis CN, Soldatos CR, Rabavilas AD (eds) Psychiatry: a world perspective, vol 4. Elsevier, Amsterdam, pp 222–226
Chang SC (1977) Social anxiety (phobia) and East Asian culture. Depression Anxiety 5: 115–120
Chipley WS (1859) Sitomania: its causes and treatment. Am J Insanity 16: 1–42
Chiu TL, Tong JE, Schmidt KE (1972) A clinical and survey study of latah in Sarawak, Malaysia. Psychol Med 2: 155–165
Chowdhury AN (1991) Medico-cultural cognition of koro epidemic: an ethnographic study. J Indian Anthropol Soc 26: 155-170
Chowdhury AN (1992) Koro in schizophrenia with a case report from industrial setup. Industrial Psychiatry J (India) 5: 47–59
Chowdhury AN (1994) Koro in females: an analysis of 48 cases. Transcult Psychiatr Res Rev 31: 369–380
Chowdhury AN, Pal P, Chatterjee A, Roy M, Das Chowdhury BB (1988) Analysis of North Bengal koro epidemic with three year follow up. Indian J Psychiatry 30: 69–72
Clarke WC (1973) Temporal madness as theatre: wild-man behaviour in New Guinea. Oceania 43: 198–214
Collinder B (1949) The Lapps. Princeton University Press, Princeton
*Collomb H (1965) Bouffées délirantes en psychiatrie africaine. Psychopathol Afr 1: 167–239
Constant J (1972) Les bouffées délirantes en Guadeloupe. Psychopathol Afr 8: 169–199
Czaplicka MA (1914) Aboriginal Siberia – a study in social anthropology. Clarendon, Oxford
Darwin C (1840) The voyage of the Beagle. Dutton, New York (reprinted 1955)
Davis C, Yager J (1992) Transcultural aspects of eating disorders: a critical literature review. Cult Med Psychiatry 16: 377–394
Devereux G (1954) Primitive genital mutilations in a neurotic's dream. J Am Psychoanal Assoc 2: 484–493
Diedrichsen I (1991) Gesellschaftliche Entstehungsbedingungen bei psychogenen Eßstörungen. MMG 16: 229–237
DiNicola VF (1985) Family therapy and transcultural psychiatry: an emerging synthesis. II. Portability and culture change. Transcult Psychiatr Res Rev 22: 151–180
*DiNicola VF (1990a) Anorexia multiforme: self-starvation in historical and cultural context. I. Self-starvation as a historical chameleon. Transcult Psychiatr Res Rev 27: 165–196
*DiNicola VF (1990b) Anorexia multiforme: self-starvation inhistorical and cultural context. II. Anorexia nervosa as a culture-reactive syndrome. Transcult Psychiatr Res Rev 27: 245–286
Doi T (1982) Amae, Freiheit in Geborgenheit. Gerhards, Frankfurt am Main
Doolittle AA (1991) Latah behavior by females among the Rungus of Sabah. In: Sutlive V, Appell GN (eds) Female and male in Borneo. Borneo Research Council, Williamsburg VA, pp 121–152 (Borneo Research Council Monograph Series vol I)
Dow TW, Silver D (1973) A drug induced koro syndrome. J Fla Med Assoc 60: 32–33
Dutta D, Phookan HR, Das PD (1982) The koro epidemic in lower Assam. Indian J Psychiatry 24: 370–374
Ebigbo PO, Okunna EN (1986/87) Anorexia nervosa resulting from family rejection. Psychopathol Afr 21: 171–183
Edwards JG (1970) The koro pattern of depersonalization in an American schizophrenic patient. Am J Psychiatry 126: 1171–1173
Edwards JW (1983) Semen anxiety in South Asian cultures; cultural and transcultural significance. Med Anthropol 7: 51–67
Edwards JW (1985) Indigenous koro, a genital retraction syndrome of insular south-east Asia. In: Simons RC, Hughes CC (eds) The culture-bound syndromes. Reidel, Dordrecht, pp 169–191
Ehrstrom M (1951) Medical investigations in north Greenland 1948–1949, II. Acta Med Scand 140: 254–264
El Fakharani M (1980) Koro – ein Syndrom im Kulturwandel? Beobachtungen auf der ostindonesischen Insel Flores. Curare 3: 241–244
Ellis WG (1893) The amok of the Malays. J Ment Sci 39: 325–338
Ellis WG (1897) Latah, a mental malady of the Malays. J Ment Sci 43: 32–40
Ewing JF (1955) Juramentado: institutionalized suicide among the Moros of the Philippines. Anthropol Q 28: 148–155
Fabrega H (1993) A cultural analysis of human behavioral breakdowns: an approach to the ontology and epistemology of psychiatric phenomena. Cult Med Psychiatry 17: 99–132
Fletcher W (1908) Latah and crime. Lancet 2: 254–255
*Foulks EF (1972) The arctic hysterias of the north Alaskan Eskimo. American Anthropological Association, Washington DC (Anthropological Studies no 10)
Foulks EF (1985) The transformation of arctic hysteria. In: Simons RC, Hughes CC (eds) The culture-bound syndromes. Reidel, Dordrecht, pp 307–324
Fujimori H, Zhan-Pei Z, Kizaki Y, Zheng-Ji C (1987) Wahn und Gesellschaft in Japan und China in transkulturell-psychiatrischer Sicht. Fortschr Neurol Psychiatr 55: 323–334
Gaines AD (1992) Ethnopsychiatry: the cultural construction of psychiatries. In: Gaines AD (ed) Ethnopsychiatry. State University of New York Press, Albany, pp 3–49
Gajdusek DC (1970) Physiological and psychological characteristics of stone age man. Engineer Science 33: 26–33, 56–62
Galloway DJ (1922) A contribution to the study of "latah". J Straits Branch R Asia Soc 85: 140–150
Garner DM (1991) Soziokulturelle Aspekte bei Eßstörungen. In: Jacobi C, Paul T (eds) Bulimia und Anorexia nervosa: Ursachen und Therapie. Springer, Berlin Heidelberg New York, pp 11–23
Garner DM, Garfinkel PE (1980) Socio-cultural factors in the development of anorexia nervosa. Psychol Med 10: 647–656
Gaw AC, Bernstein RL (1992) Classification of amok in DSM-IV. Hosp Community Psychiatry 43: 789–793
Geertz H (1968) Latah in Java: a theoretical paradox. Indonesia 5: 93–104
German GA, Arya OP (1969) Psychiatric morbidity amongst a Uganda student population. Br J Psychiatry 115: 1323–1329

Gilles de la Tourette G (1884) Jumping, latah, myriachit. Arch Neurol 8: 68–74
Gilmour A (1902) "Latah" among South African natives. Scott Med J 10: 18–19
Gimlette JD (1897) Remarks on the etiology, symptoms and treatment of latah. BMJ 21: 455–457
Gimlette JD (1901) Notes on a case of amok. J Trop Med Hyg 4: 195–199
Goddard K (1991) Morita therapy: a literature review. Transcult Psychiatr Res Rev 28: 93–115
*Gordon RA (1990) Anorexia and bulemia: anatomy of a social epidemic. Blackwell, Cambridge
Guarnaccia PJ (1993) Ataques de nervios in Puerto Rico: culture-bound syndrome or popular illness? Med Anthropol 15: 157–170
Guarnaccia PJ, DeLaCancela V, Carrillo E (1989) The multiple meanings of ataques de nervios in the Latino community. Med Anthropol 11: 47–62
Guinness EA (1992) Patterns of mental illness in the early stages of urbanization.Br J Psychiatry 160 [Suppl 16]: 12–72
Gull W (1874) Anorexia nervosa (apepsia hysterica, anorexia hysterica). Trans Clin Soc Lond 7: 22–28
Gullick JM (1958) Indigenous political systems of western Malaya. Humanities, London (London University – London School of Economics and Political Science Monographs on Social Anthropology no 17)
Gussow Z (1960) Pibloktoq (hysteria) among the polar Eskimo: an ethnopsychiatric study. In: Muensterberger W, Axelrod S (eds) The psychoanalytic study of society, vol 1. International Universities Press, New York, pp 218–235
Gussow Z (1963) A preliminary report of kayak-angst among the Eskimo of West Greenland. Int J Soc Psychiatry 9: 18–26
Gwee AL (1963) Koro – a cultural disease. Singapore Med J 4: 119–122
Gwee AL (1968) Koro – its origin and nature as a disease entity. Singapore Med J 9: 3–6
Habermas H (1991) The role of psychiatric and medical traditions in the discovery and description of anorexia nervosa in France, Germany and Italy, 1873–1918. J Nerv Ment Dis 170: 360–366
Hallowell AI (1934) Culture and mental disorder. J Abnorm Soc Psychol 29: 1–9
Hammond WA (1884) Miryachit, a newly described disease of the nervous system. New York Med J 39: 191–192
Harris B (1981) A case of brain fag in East Africa. Br J Psychiatry 138: 162–163
Hauschild T (1982) Der böse Blick. Mensch und Leben, Berlin
Hay TH (1971) The windigo psychosis: psychodynamic, cultural and social factors in aberrant behavior. Am Anthropol 73: 1–19
Hezel FX (1993) Psychosis in Micronesia: an epidemiological survey. Transcult Psychiatr Res Rev 30: 66–71
Hirsch SJ, Hollender MH (1969) Hysterical psychosis: clarification of the concept. Am J Psychiatry 125: 909–915
Hochkirchen BH, Jilek WG (1985) Psychosocial dimensions of suicide and parasuicide in Amerindians of the Pacific Northwest. J Operation Psychiatry 16: 24–28
Hollender MH, Hirsch SJ (1964) Hysterical psychosis. Am J Psychiatry 120: 1066–1074
Hollweg MG (1997) Main culture bound syndromes in Bolivia. Curare 20: 23–28
Holtved E (1967) Contributions to polar Eskimo ethnography. Meddelelser om Grønland, vol 182, no 2. Reitzel, Copenhagen
Honigmann JJ (1947) Witch-fear in post-contact Kaska society. Am Anthropol 49: 222–243
Honigmann JJ (1967) Personality in culture. Harper Row, New York
Howard R, Ford R (1992) From the jumping Frenchmen of Maine to post-traumatic stress disorder: the startle response in neuropsychiatry. Psychol Med 22: 695–707
Hughes CC (1985a) Culture-bound or construct-bound? In: Simons RC, Hughes CC (eds) The culture-bound syndromes. Reidel, Dordrecht, pp 3–24
Hughes CC (1985b) Glossary of 'culture-bound' or folk psychiatric syndromes. In: Simons RC, Hughes CC (eds) The culture-bound syndromes. Reidel, Dordrecht, pp 469–505
Iancu I, Spivak B, Ratzoni G, Apter A, Weizman A (1994) The sociocultural theory in the development of anorexia nervosa. Psychopathology 27: 29–36
Ifabumuyi OI, Rwegellera GGC (1979) Koro in a Nigerian male patient. Afr J Psychiatry 5: 103–105
Ilechukwu STC (1988) Koro-like syndromes in Nigeria. Transcult Psychiatr Res Rev 25: 310–314
Ilechukwu STC (1992) Magical penis loss in Nigeria: report of a recent epidemic of a koro-like syndrome. Transcult Psychiatr Res Rev 29: 91–108
Infurna G (1928) Un piccolo foculaio di latah in Eritrea. Arch Ital Sci Med Colonial 8: 560–562
Iwai H (1982) Yugamerareta kyozo: nihonjin no taijin kyôfu. Asahi Shuppansha, Tokyo
*Jablensky A, Sartorius N, Ernberg G, Anker M, Korten A, Cooper JE, Day R, Bertelsen A (1992) Schizophrenia: manifestations, incidence and course in different countries – a World Health Organization ten country study. Cambridge University Press, Cambridge (Psychol Med Monogr Suppl 20)
Jablensky A, Sartorius N, Cooper JE, Anker M, Korten A, Bertelsen A (1994) Culture and schizophrenia. Br J Psychiatry 165: 434–436
Janet P (1903) Les obsessions et la psychasthénie, vol I. Alcan, Paris
Jegede RO (1983) Psychiatric illness in African students: "brain fag" syndrome revisited. Can J Psychiatry 28: 188–192
Jenner JA (1990) Latah as coping: a case study offering a new paradox to solve the old one. Int J Soc Psychiatry 36: 194–199
Jilek WG (1974) Salish Indian mental health and culture change. Holt Rinehart Winston, Toronto
Jilek WG (1981) Anomic depression, alcoholism, and a culture-congenial Indian response. J Stud Alcohol Suppl 9: 159–170
Jilek WG (1982) Culture – "pathoplastic" or "pathogenic"? – a key question of comparative psychiatry. Curare 5: 57–68
*Jilek WG (1986) Epidemics of "genital shrinking" (koro): historical review and report of a recent outbreak in South China. Curare 9: 269–282
Jilek WG (1987a) The impact of alcohol on small-scale societies in the circum-Pacific region. Curare 10: 151–168
Jilek WG (1987b) Veränderte Wachbewußtseinszustände in Heiltanzritualen nordamerikanischer Indianer. In: Dittrich A, Scharfetter C (eds) Ethnopsychotherapie. Enke, Stuttgart, pp 135–149
Jilek WG (1988) Mental health, ethnopsychiatry and traditional medicine in the Kingdom of Tonga. Curare 11: 161–176

Jilek WG (1990) Schamanische Symbolik in den wiederbelebten Zeremonien der Salish-Indianer der Nordwestküste. Mittlg Berl Ges Anthropol Ethnol Urgeschichte 11: 21–37

*Jilek WG (1993) Traditional medicine relevant to psychiatry. In: Sartorius N, De Girolamo G, Andrews G, German GA (eds) Treatment of mental disorders. WHO/American Psychiatric Press, Washington DC, pp 341–390

Jilek WG (1994) Modified states of consciousness: their role in ethnopsychiatry. Transcult Psychiatry Newslett (WPA) 12: 5–8

Jilek WG (1995) Emil Kraepelin and comparative sociocultural psychiatry. Eur Arch Psychiatry Clin Neurosci 245: 231–238

**Jilek WG, Jilek-Aall L (1970) Transient psychoses in Africans. Psychiatr Clin 3: 337–364

Jilek WG, Jilek-Aall L (1977a) Massenhysterie mit Koro-Symptomatik in Thailand. Schweiz Arch Neurol Psychiatr 120: 257–259

Jilek WG, Jilek-Aall L (1977b) A koro epidemic in Thailand. Transcult Psychiatr Res Rev 14: 57–59

*Jilek WG, Jilek-Aall (1985) The metamorphosis of "culture-bound" syndromes. Soc Sci Med 21: 205–210

Jilek WG, Jilek-Aall L (1990) The mental health relevance of traditional medicine and shamanism in refugee camps of northern Thailand. Curare 13: 217–224

Jilek-Aall L (1974) Psychosocial aspects of drinking among Coast Salish Indians. Can Psychiatr Assoc J 19: 357–361

Jilek-Aall L (1979) Call mama doctor – African notes of a young woman doctor. Hancock House, Saanichton/Seattle

Jilek-Aall L (1988) Suicidal behaviour among youth: a cross-cultural comparison. Transcult Psychiatr Res Rev 25: 87–105

Jilek-Aall L, Jilek WG, Flynn F (1978) Sex role, culture and psychopathology: a comparative study of three ethnic groups in western Canada. J Psychol Anthropol 1: 473–488

Jochelson W (1908) The Koryak. In: Boas F (ed) The Jesup North Pacific expedition, vol VI, part 2: Brill, Leiden (Memoirs of the American Museum of Natural History)

Jochelson W (1910) The Yukaghir and the Yukaghirized Tungus. In: Boas F (ed) The Jesup North Pacific expedition, vol IX, part 1. Brill, Leiden (Memoirs of the American Museum of Natural History)

Jones DJ, Fox JMM, Babigan HM, Hutton HE (1980) Epidemiology of anorexia nervosa in Monroe County, New York. Psychosom Med 42: 551–558

Kasahara Y (1974) Fear of eye-to-eye confrontation among neurotic patients in Japan. In: Lebra TS, Lebra WP (eds) Japanese culture and behavior. University Press of Hawaii, Honolulu, pp 396–406

Kasahara Y (1987) Social phobia in Japan. In: Social phobia in Japan and Korea. Proceedings of the first cultural psychiatry symposium between Japan and Korea. East Asian Academy of Cultural Psychiatry, Seoul, pp 3–14

Katz MM, Marsella A, Dube KC, Olatawura M, Takahashi R, Nakane Y, Wynne LC, Gift T, Brennan J, Sartorius N, Jablensky A (1988) On the expression of psychosis in different cultures: schizophrenia in an Indian and in an Nigerian community. Cult Med Psychiatry 12: 331–355

Kenny MG (1978) Latah: the symbolism of a putative mental disorder. Cult Med Psychiatry 2: 209–231

Kenny MG (1983) Paradox lost: the latah problem revisited. J Nerv Ment Dis 171: 159–167

Kenny MG (1990) Latah, the logic of fear. In: Karim WJ (ed) The emotions of culture: a Malay perspective. Oxford University Press, Singapore, pp 23–141

Kiev A (1972) Transcultural psychiatry. Free Press, New York

Kim K-I, Li D, Jiang Z, Cui X, Lin L, Kang JJ, Park KK, Chung EK, Kim CK (1993) Schizophrenic delusions among Koreans, Korean-Chinese and Chinese: a transcultural study. Int J Soc Psychiatry 39: 190–199

Kimura B (1972) Struktur des Selbstbewußtseins beim Japaner im Spiegel der sogenannten "Anthropophobien". In: Erhardt HE (ed) Perspektiven der heutigen Psychiatrie. Gerhards, Frankfurt am Main, pp 322–326

*Kimura B (1995) Zwischen Mensch und Mensch: Strukturen japanischer Subjektivität. Wissenschaftliche Buchgesellschaft, Darmstadt

*Kirmayer LJ (1991) The place of culture in psychiatric nosology: taijin kyofusho and DSM-III-R. J Nerv Ment Dis 179: 19-28

Kirmayer LJ (1994) Suicide among Canadian aboriginal peoples. Transcult Psychiatr Res Rev 31: 3–58

Kitanishi K (1990) Morita therapy from a transcultural psychiatric view. J Morita Ther 1: 190–194

Kitanishi K, Tseng W-S (1988) Social phobia among Japanese: clinical, family and cultural exploration. Paper presented at 4th Scientific Meeting, Pacific Rim College of Psychiatry, December 5, 1988, Hong Kong (abstracted in: Transcult Psychiatr Res Rev (1989), 26: 144–146)

Kitanishi K, Miyake Y, Kim KI, Liu X (1995) A comparative study of taijin kyofusho (TKS) tendencies among college students in Japan, Korea, and the People's Republic of China. Jikeikai Med J 42: 231–243

Kleinman AM (1977) Depression, somatization, and the 'new cross-cultural psychiatry'. Soc Sci Med 11: 3–10

*Kleinman AM (1988) Rethinking psychiatry. Free Press/Collier Macmillan, New York

Kobler F (1948) Description of an acute castration fear based on superstition. Psychoanal Rev 35: 285–289

Koch K-F (1968) On "possession" behaviour in New Guinea. J Polynesian Soc 77: 135–146

Kon Y (1994) Amok. Br J Psychiatry 165: 685–689

Kondo K (1976) The origin of Morita therapy. In: Lebra WP (ed) Culture-bound syndromes, ethnopsychiatry, and alternate therapies. University Press of Hawaii, Honolulu, pp 250–258

Koro Study Team (1969) The koro "epidemic" in Singapore. Singapore Med J 10: 234–242

*Kraepelin E (1904) Vergleichende Psychiatrie. Centralbl Nervenheilk Psychiatr 27: 433–437

Kraepelin E (1909) Psychiatrie – Ein Lehrbuch für Studierende und Ärzte, 8th edn, vol 1. Barth, Leipzig

Kraepelin E (1913) Psychiatrie – Ein Lehrbuch für Studierende und Ärzte, 8th edn, vol 3. Barth, Leipzig

Kretschmer E (1958) Hysterie, Reflex und Instinkt, 6th edn. Thieme, Stuttgart

Kraus RF, Buffler PA (1979) Sociocultural stress and the American native in Alaska: an analysis of changing patterns of psychiatric illness and alcohol abuse among Alaska natives. Cult Med Psychiatry 3: 111–151

Kunkle CE (1967) The "jumpers" of Maine: a reappraisal. Arch Intern Med 119: 355–358

Labhardt F (1963) Die schizophrenieähnlichen Emotionspsychosen. Springer, Berlin Göttingen Heidelberg

Laiblin W (1965) Ekstatische Männerbünde. In: Bitter W (ed) Massenwahn in Geschichte und Gegenwart. Klett, Stuttgart, pp 118–132

Lambo TA (1960) Further neuropsychiatric observations in Nigeria. BMJ 2: 1696–1704

Landes R (1938) The abnormal among the Ojibwa Indians. J Abnorm Soc Psychol 33: 14–33
Landis C, Hunt WA (1939) The startle pattern. Rinehart, New York
Landy D (1985) Pibloktoq (hysteria) and Inuit nutrition: possible implication of hypervitaminosis A. Soc Sci Med 21: 173–185
Langness LL (1965) Hysterical psychosis in the New Guinea highlands: a Bena Bena example. Psychiatry 28: 258–277
Langness LL (1967) Hysterical psychosis: the cross-cultural evidence. Am J Psychiatry 124: 143–152
Lasègue C (1873) De l'anorexie hystérique. Arch Gen Med 1: 385–403
Lee PWH, Lieh Mak F, Yu KK, Spinks JA (1991) Pattern of outcome in schizophrenia in Hong Kong. Acta Psychiatr Scand 84: 346–352
Lee S (1993) How abnormal is the desire for slimness? A survey of eating attitudes and behaviour among Chinese undergraduates in Hong Kong. Psychol Med 23: 437–440
Lee S (1996) Reconsidering the status of anorexia nervosa as a western culture-bound syndrome. Soc Sci Med 42: 21–34
Lee S, Ho TP, Hsu LKG (1993) Fat phobic and non-fat phobic anorexia nervosa: a comparative study of 70 Chinese patients in Hong Kong. Psychol Med 23: 999–1017
Lee SH (1977) A study of "hwa-byung" (anger syndrome). J Korea Gen Hosp Seoul 1: 63–69
Lee SH (1987) Social phobia in Korea. In: Rhi BY (ed) Social phobia in Japan and Korea. Proceedings of the first cultural psychiatry symposium between Japan and Korea. East Asian Academy of Cultural Psychiatry, Department of Psychiatry, Seoul National University, Seoul, pp 24–52
Lee SH, Shin YC, Oh KS (1994) A clinical study of social phobia for 10 years. J Korean Neuropsychiatr Assoc 33: 305–312
Legendre J (1908) Une curieuse épidémie. Ann Med Colonial 2: 280
Lehmann JP (1972) Le vecu corporel et ses interpretations en pathologie africaine, a propos des inhibitions corporelles en milieu scolaire. Rev Med Psychosom 14: 43–67
Levine RE, Gaw AC (1995) Culture-bound syndromes. Psychiatr Clin North Am 18: 523–536
Lewis-Fernandez R, Kleinman AM (1994) Culture, personality and psychopathology. J Abnorm Psychol 103: 67–71
Lhermitte J (1963) Diabolical possession, true and false. Burns Oates, London
Liebowitz MR, Salman E, Jusino CM, Garfinkel R, Street L, Cardenas DL, Silvestre J, Fyer AJ, Carrasco JL, Davies S, Guarnaccia P, Klein DF (1994) Ataque de nervios and panic disorder. Am J Psychiatry 151: 871–875
Lin K-M (1983) Hwa-byung: a Korean culture-bound syndrome? Am J Psychiatry 140: 105–107
Lin K-M, Lau JKC, Yamamoto J, Zheng Y-P, Kim H-S, Cho K-H, Nakasaki G (1992) Hwa-byung: a community study of Korean Americans. J Nerv Ment Dis 180: 386–391
Linton R (1956) Culture and mental disorders. Thomas, Springfield
Littlewood R (1990) From categories to contexts: a decade of the 'new cross-cultural psychiatry'. Br J Psychiatry 156: 308-327
*Littlewood R, Lipsedge M (1985) Culture-bound syndromes. In: Granville-Grossman K (ed) Recent advances in clinical psychiatry. Churchill Livingstone, Edinburgh, pp 105–142
Logan MH (1979) Variations regarding susto causality among the Cakchiquel of Guatemala. Cult Med Psychiatry 3: 153–166
Low SM (1985) Culturally interpreted symptoms or culture-bound syndromes: a cross-cultural review of nerves. Soc Sci Med 21: 187–196
Lucas AR, Beard CM, O'Fallon WM, Leonard TK (1991) 50-year trends in the incidence of anorexia nervosa in Rochester Minnesota: a population-based study. Am J Psychiatry 148: 917–922
*Ludwig AM (1966) Altered states of consciousness. Arch Gen Psychiatry 15: 225–234
Magnan V (1893) Leçons cliniques sur les maladies mentales, 2nd edn. Bataille, Paris
Makanjuola ROA (1987) "Ode ori": a culture-bound disorder with prominent somatic features in Yoruba Nigerian patients. Acta Psychiatr Scand 75: 231–236
Malhotra HK, Wig NN (1975) Dhat syndrome: a culture-bound sex neurosis of the orient. Arch Sex Behav 4: 519–528
Marano L (1982) Windigo-psychosis: the anatomy of an emic-etic confusion. Curr Anthropol 23: 385–412
Marsella AJ (1988) Cross-cultural research on severe mental disorders: issues and findings. Acta Psychiatr Scand Suppl 344: 7–22
Maslowski J (1986) Pathoplastic influences on symptoms of schizophrenia: a comparative study in Libya and Malta. Acta Psychiatr Scand 73: 618–623
Mazur WP (1953) Gilles de la Tourette 's syndrome. Can Med Assoc J 69: 520–522
Meldorf G (1900) Om Kajaksvimmelheden: Grønland og dens forhold til brugen af nydelsesmidler. Bibliotek Laeger (Kobenhavn) 1: 524–539
Meynert T (1889) Amentia, die Verwirrtheit. Jahrb Psychiatr 9: 1–112
Min SK (1989) A study of the concept of hwabyung. J Korean Neuropsychiatr Assoc 28: 604–616
Min SK, Namkoong K, Lee HY (1990) An epidemiological study on hwabyung. J Korean Neuropsychiatr Assoc 29: 867–874
Mizushima N, Ishii Y (1983) The epidemiology of anorexia nervosa in junior and senior high school students in Ishikawa Prefecture. Jpn J Psychosom Med 23: 311–319
Morakinyo O (1985) The brain-fag syndrome in Nigeria: cognitive deficits in an illness associated with study. Br J Psychiatry 146: 209–210
Morel BA (1860) Traité des maladies mentales. Masson, Paris
Morita S (1921) Shinkeishitsu no ryoho. Hakuyo-sha, Tokyo
Mulder JGA (1935) Over koro. Geneeskundig Tijdschrift voor Nederlandsch-Indie 75: 837–838
Murasaki S (1970) The tale of Genji, vol 2, part 5. Tuttle, Rutland
Murphy HBM (1961) Social change and mental health. In: Milbank Memorial Fund (ed) Causes of mental disorders: a review of epidemiological knowledge 1959. Milbank Memorial Fund, New York, pp 280–329
*Murphy HBM (1973) History and the evolution of syndromes: the striking case of latah and amok. In: Hammer M, Salzinger K, Sutton S (eds) Psychopathology: contributions from the biological, behavioral and social sciences. Wiley, New York, pp 33–55
Murphy HBM (1976) Notes for a theory on latah. In: Lebra WP (ed) Culture-bound syndromes, ethnopsychiatry, and alternate therapies. University of Honolulu Press, Honolulu, pp 3–21
**Murphy HBM (1982) Comparative psychiatry. Springer, Berlin Heidelberg New York
Musgrave WE, Sison AG (1910) Mali-mali, a mimic psychosis in the Philippine islands. Philippine J Sci B Med Sci 5: 335–339

Nandi DN, Banerjee G, Saha H, Boral GC (1983) Epidemic koro in West Bengal, India. Int J Soc Psychiatry 29: 265–268
Nations MK, Camino LA, Walker FB (1988) 'Nerves': folk idiom for anxiety and depression? Soc Sci Med 26: 1245–1259
Natoli A (1937) Casi di "latah" riscontrati fra indigeni a Bengasi. Giornal Ital Clin Trop 1: 20–24
Newman PL (1964) "Wild man" behavior in a New Guinea highlands community. Am Anthropol 66: 1–19
Ngui PW (1969) The koro epidemic in Singapore. Aust NZ J Psychiatry 3: 263–266
Novakovsky S (1924) Arctic or Siberian hysteria as a reflex of the geographic environment. Ecology 5: 113–127
O'Brien HA (1883) Latah. J Straits Branch R Asia Soc 11: 143–153
O'Brien HA (1884) Latah. J Straits Branch R Asia Soc 12: 283–284
Obeyesekere G (1976) The impact of ayurvedic ideas on the culture and the individual in Ceylon. In: Leslie C (ed) Asian medical systems: a comparative study. University of California Press, Berkeley, pp 201–226
Ogawa K, Miya M, Watarai A, Nakazawa KM, Yuasa S, Utena H (1987) A long-term follow-up study of schizophrenia in Japan with special reference to the course of social adjustment. Br J Psychiatry 151: 758–765
Ohnuki-Tierney E (1980) Shamans and imu: among two Ainu groups. Ethos 8: 204–228
Okasha A, Seif El Dawla A, Khalil AH, Saad A (1993) Presentation of acute psychosis in an Egyptian sample: a transcultural comparison. Compr Psychiatry 34: 4–9
O'Nell CW (1975) An investigation of reported "fright" as a factor in the etiology of susto, "magical fright". Ethos 3: 41–63
Oquendo M (1994) Differential diagnosis of ataque de nervios. Am J Orthopsychiatry 65: 60–65
Oquendo M, Horwath E, Martinez A (1992) Ataques de nervios: proposed diagnostic criteria for a culture specific syndrome. Cult Med Psychiatry 16: 367–376
Orley JH (1970) Culture and mental illness – a study from Uganda. East African Publishing house, Nairobi
Overbeck H (1986) Die Geschichte von Hang Tuah. Beck, Munich
Pang KYC (1990) Hwabyung: the construction of a Korean popular illness among Korean elderly immigrant women in the United States. Cult Med Psychiatry 14: 495–512
Pareto V (1935) The mind and society, vol 1. Non-logical conduct. Harcourt Brace, New York
*Paris J (1992) Dhat: the semen loss anxiety syndrome. Transcult Psychiatr Res Rev 29: 109–118
Parker S (1960) The witiko psychosis in the context of Ojibwa personality and culture. Am Anthropol 62: 603–623
Patel V, Winston M (1994) 'Universality of mental illness' revisited: assumptions, artefacts, and new directions. Br J Psychiatry 165: 437–440
Peltzer K (1987) Some contributions of traditional healing practices towards psychosocial health care in Malawi. Fachbuch Psychologie, Eschborn (Frankfurt/Main)
Peltzer K (1995) Psychology and health in African cultures. IKO Verlag für Interkulturelle Kommunikation, Frankfurt/Main
Pfeiffer WM (1968) Neuere Untersuchungsergebnisse bei Latah. Transcult Psychiatr Res Rev 5: 34–38
*Pfeiffer WM (1980) Kulturgebundene Syndrome. In: Pfeiffer WM, Schoene W (eds) Psychopathologie im Kulturvergleich. Enke, Stuttgart, pp 156–170
**Pfeiffer WM (1994) Transkulturelle Psychiatrie, 2nd edn. Thieme, Stuttgart
Prince R (1959) Report from Nigeria. Rev Newslett Transcult Res Ment Health Probl 6: 40–41
Prince R (1960) The "brain fag" syndrome in Nigerian students. J Ment Sci 106: 559–570
Prince R (1962) Functional symptoms associated with study in Nigerian students. West Afr Med J 11: 198–206
*Prince R (1985) The concept of culture-bound syndromes: anorexia nervosa and brain-fag.Soc Sci Med 21: 197–203
Prince R (1989) The brain-fag syndrome. In: Peltzer K, Ebigbo PO (eds) Text book of clinical psychology in Africa. University of Nigeria, Enugu, pp 276–287
*Prince R (1991) Amok then and now. Transcult Psychiatr Res Rev 28: 219–229
Prince R, Tcheng-Laroche F (1987) Culture-bound syndromes and international disease classifications. Cult Med Psychiatry 11: 3–19
Raguram R, Jadhav S, Weiss M (1994) Historical perspectives on dhat syndrome. J Nat Inst Ment Health Neurosci (India) 12: 117–124
Rainaut J (1958) Un aspect des psychoses transitoires en milieu africain: la bouffée aigue confusionelle et anxieuse. In: CSA (Scientific Council for Africa South of the Sahara) (ed) Mental disorders and mental health in Africa south of the sahara. CCTA/CSA-WFMH-WHO Meeting of Specialists on Mental Health, Bukavu, pp 193–214
Reay M (1960) "Mushroom madness" in the New Guinea highlands. Oceania 31: 137–139
Reay M (1965) Mushrooms and collective hysteria. Aust Territories 5: 18–28
Reay M (1977) Ritual madness observed: a discarded pattern of fate in Papua New Guinea. J Pacific Hist 12: 55–79
Repond A (1940) Le lattah: une psycho-névrose exotique. Ann Med Psychol 98: 311–324
Rin H (1965) A study of the aetiology of koro in respect to the Chinese concept of illness. Int J Soc Psychiatry 11: 7–13
*Risso M, Böker W (1964) Verhexungswahn. Karger, Basel (Bibliotheca Psychiatrica et Neurologica 124)
Ritenbaugh C (1982) Obesity as a culture-bound syndrome. Cult Med Psychiatry 6: 347–361
Rodewyk A (1963) Die dämonische Besessenheit in der Sicht des Rituale Romanum. Pattloch, Aschaffenburg
Rodrigue RB (1963) A report on a widespread psychological disorder called lulu seen among the Huli linguistic group in Papua. Oceania 33: 274–279
Rubel AJ (1964) The epidemiology of a folk illness: susto in hispanic America. Ethnology 3: 268–283
Rubel AJ, O'Nell CW, Collado R (1985) The folk illness called susto. In: Simons RC, Hughes CC (eds) The culture-bound syndromes. Reidel, Dordrecht, pp 333–350
Rubinstein DH (1992) Suicidal behaviour in Micronesia. In: Peng KL, Tseng W-S (eds) Suicidal behaviour in the Asia-Pacific region. Singapore University Press, Singapore, pp 199–230
Russell JG (1989) Anxiety disorders in Japan: a review of the Japanese literature on shinkeishitsu and taijin kyofusho. Cult Med Psychiatry 13: 391–403
Sachdev PS (1985) Koro epidemic in north-east India. Aust NZ J Psychiatry 19: 433–438
Sakaki J (1905) Imubacco. Mitteilungen der medizinischen Fakultät der kaiserlich-japanischen Universität Tokyo 6: 147–201
Sal y Rosas F (1958) El mito del jani o susto de la medicina indigena del Peru. Rev Sanidad Pol Lima 19: 167–210
Salles P (1961) Aspects pratiques des psychoses aigues transitoires chez le noir d'Afrique centrale. Med Trop 21: 1–10

Sarnelli T (1934) Primi casi di "latah" osservati nell'alto Yemen, Arabia S.O. Arch Ital Sci Med Colonial 15: 750–759

*Sartorius N (1979) Crosscultural psychiatry. In: Kisker KP (ed) Psychiatrie der Gegenwart, vol 1. Grundlagen und Methoden der Psychiatrie, part 1. Springer, Berlin Heidelberg New York, pp 711–737

Sartorius N, Jablensky A, Korten A, Ernberg G, Anker M, Cooper JE, Day R (1986) Early manifestations and first-contact incidence of schizophrenia in different cultures. Psychol Med 16: 909–928

Schmidt K, Hill L, Guthrie G (1977) Running amok. Int J Soc Psychiatry 23: 264–275

Selvini Palazzoli M (1985) Anorexia nervosa; a syndrome of the affluent society. Transcult Psychiatr Res Rev 22: 199–205

Shaw W (1972) Amuk. Federation Museums J (Malaysia) 17: 1–30

Shirokogoroff SM (1935) The psychomental complex of the Tungus. Kegan Paul Trench Trubner, London

Shore JH, Manson SM (1981) Cross-cultural studies of depression among American Indians and Alaska natives. White Cloud J 2: 5–12

Shorter E (1994) From the mind into the body: the cultural origins of psychosomatic symptoms. Free Press, New York

Shukla GD, Mishra DN (1981) Koro-like syndrome: a case report. Indian J Psychiatry 23: 96–97

Simons RC (1980) The resolution of the latah paradox. J Nerv Ment Disease 168: 195–206

Simons RC (1983) Latah II – problems with a purely symbolic interpretation. J Nerv Ment Dis 171: 168–181

Simons RC (1985a) Sorting the culture-bound syndromes. In: Simons RC, Hughes CC (eds) The culture-bound syndromes. Reidel, Dordrecht, pp 25–38

Simons RC (1985b) Introduction: the startle matching taxon. In: Simons RC, Hughes CC (eds) The culture-bound syndromes. Reidel, Dordrecht, pp 41–42

*Simons RC (1996) Boo! Culture, experience, and the startle reflex. Oxford University Press, New York

**Simons RC, Hughes CC (eds) (1985) The culture-bound syndromes. Reidel, Dordrecht

Slot JA (1935) Koro in Zuid-Celebes. Geneeskundig Tijdschrift voor Nederlandsch-Indie 75: 811–820

Smyth MG, Dean C (1992) Capgras and koro. Br J Psychiatry 161: 121–123

Sprenger J, Kraemer (Institor) H (1487) Malleus maleficarum. Frankfurt (English translation published in 1951: "The witches' hammer" with an introduction, bibliography and notes by Rev. Montague Summers. Pushkin, London)

Stevens H (1964) Jumping Frenchmen of Maine. Trans Am Neurol Assoc 89: 65–67

Stevens H (1965) Jumping Frenchmen of Maine. Arch Neurol 12: 311–314

Stevens J (1987) Brief psychoses: do they contribute to the good prognosis and equal prevalence of schizophrenia in developing countries? Br J Psychiatry 151: 393–396

Still RML (1940) Remarks on the aetiology and symptoms of young-dah-the with a report of four cases and its medico-legal significance. Indian Med Gazette February: 88–91

Suematsu H, Ishikawa H, Kuboki T, Ito T (1985) Statistical studies on anorexia nervosa in Japan: detailed clinical data on 1,011 patients. Psychother Psychosom 43: 96–103

Suwanlert S (1972) Psychiatric study of bahtsche (latah). J Psychiatr Assoc Thailand 17: 380–398

Suwanlert S, Coates D (1979) Epidemic koro in Thailand – clinical and social aspects. Transcult Psychiatr Res Rev 16: 64–66

Suzuki T, Takemura S (1966) Morita-Therapie – eine für Japan eigentümliche Psychotherapie und ihre sozial-psychiatrische Bedeutung. Z Psychother Med Psychol 16(5): 161–172

Swettenham FA (1889) The real Malay. Lane-Bodley Head, London

Takahashi T (1989) Social phobia syndrome in Japan. Compr Psychiatry 30: 45–52

Tan E-S (1989) Amok – its world-wide occurrence. Transcult Psychiatr Res Rev 26: 137–140

Teicher MI (1960) Windigo psychosis, a study of a relationship between belief and behavior among the Indians of northeastern Canada. American Ethnological Society, University of Washington, Seattle

Teoh J-I (1972) The changing psychopathology of amok. Psychiatry 35: 345–351

Thorslund J (1990) Inuit suicides in Greenland. Arctic Med Res 49: 25–33

Tokarsky A (1890) "Merjatschenie". Neurol Centralbl 9: 662–663

Toro J, Salamero M, Martinez E (1994) Assessment of sociocultural influences on the aesthetic body shape model in anorexia nervosa. Acta Psychiatr Scand 89: 147–151

Tousignant M (1979) Espanto: a dialogue with the gods. Cult Med Psychiatry 3: 347–361

Tseng W-S (1973) The development of psychiatric concepts in traditional Chinese medicine. Arch Gen Psychiatry 29: 569–575

Tseng W-S, McDermott (1981) Culture, mind and therapy. Brunner/Mazel, New York

Tseng W-S, Streltzer J (1997) Integration and conclusions. In: Tseng WS, Streltzer J (eds) Culture and psychopathology. Brunner/Mazel, New York, pp 241–252

*Tseng W-S, Mo K-M, Hsu J, Li L-S, Ou L-W, Chen G-Q, Jiang D-W (1988) A sociocultural study of koro epidemics in Guangdong, China. Am J Psychiatry 145: 1538–1543

Tseng W-S, Mo K-M, Li L-S, Chen G-Q, Ou L-W, Zheng H-B (1992) Koro epidemics in Guangdong, China. J Nerv Ment Dis 180: 117–123

Tsoi WF, Wong KE (1991) A 15 year follow up study of Chinese schizophrenic patients. Acta Psychiatr Scand 84: 217–220

Uchimura Y (1935) "Imu"; a malady of the Ainu. Lancet i: 1272–1273

Uchimura Y (1956) Imu, eine psychoreaktive Erscheinung der Ainu-Frauen. Nervenarzt 27: 535–540

van Brero PCJ (1895) Über das sogenannte Latah, eine in Niederländisch-Ostindien vorkommende Neurose. Allg Z Psychiatrie Psychisch-Gerichtliche Medicin 51: 939–948

van Brero PCJ (1897a) Einiges über die Geisteskrankheiten der Bevölkerung des malaiischen Archipels. Allg Z Psychiatrie Psychisch-Gerichtliche Medicin 53: 25–78

van Brero PCJ (1897b) Koro, eine eigenthümliche Zwangsvorstellung. Allg Z Psychiatrie Psychisch-Gerichtliche Medicin 53: 569–573

van Dam WC (1970) Dämonen und Besessene. Pattloch, Aschaffenburg

Vandereycken W, Lowenkopf EL (1990) Anorexia nervosa in 19th century America. J Nerv Ment Dis 178: 531–535

van Loon FHG (1927) Amok and lattah. J Abnorm Soc Psychol 21: 434–444

van Wulfften Palthe PM (1933) Amok. Nederlandsch Tijdschrift voor Geneeskunde 77: 983–991

van Wulfften Palthe PM (1934) Koro – een eigenaardige angstneurose. Geneeskundig Tijdschrift voor Nederlandsch-Indie 74: 1713–1720

van Wulfften Palthe PM (1935) Koro – eine merkwürdige Angsthysterie. Int Z Psychoanal 21: 249–257

van Wulfften Palthe PM (1936) Psychiatry and neurology in the tropics. In: DeLangen C, Lichtenstein A (eds) A clinical textbook of tropical medicine. Kolff, Batavia, pp 525–547

Voustman AH (1897) "Koro" in de westerafdeeling van Borneo. Geneeskundig Tijdschrift voor Nederlandsch-Indie 37: 499–505

Vyncke JC (1957) Psychoses et névroses en Afrique centrale. Académie royale des sciences coloniales, Classe des sciences naturelles et médicales, Brussels

Wallace AFC (1972) Mental illness, biology, and culture. In: Hsu F (ed) Psychological anthropology. Shenkman, Cambridge, MA, pp 363–402

Wallace AFC, Ackerman RE (1960) An interdisciplinary approach to mental disorder among the polar Eskimo of northwest Greenland. Anthropologica 11: 1–12

Weiser L (1927) Altgermanische Jünglingsweihen und Männerbünde. Konkordia, Bühl (Baden)

*Weiss MG (1995) Eating disorders and disordered eating in different cultures. Psychiatr Clin North Am 18: 537–553

Wen J-K, Wang C-L (1981) Shen-k'uei syndrome: a culture-specific sexual neurosis in Taiwan. In: Kleinman A, Lin T-Y (eds) Normal and abnormal behavior in Chinese culture. Reidel, Dordrecht, pp 357–369

Westermeyer J (1972) A comparison of amok and other homicide in Laos. Am J Psychiatry 129: 703–709

Westermeyer J (1973) On the epidemicity of amok violence. Arch Gen Psychiatry 28: 873–876

Westermeyer J (1985) Sudden mass assault with grenade: an epidemic amok form from Laos. In: Simons RC, Hughes CC (eds) The culture bound syndromes. Reidel, Dordrecht, pp 225–235

WHO (1993) The ICD-10 international classification of mental and behavioural disorders – diagnostic criteria for research. World Health Organization, Geneva

Wig NN (1960) Problems of mental health in India. J Clin Soc Psychiatry (India) 17: 48–53

Winiarz W, Wielawski J (1936) Imu – a psychoneurosis occuring among Ainus. Psychoanal Rev 23: 181–186

Wintrob RM (1973) The cultural dynamics of student anxiety: a report from Liberia. Psychopathol Afric 9: 267–294

Winzeler RL (1991) Latah in Sarawak, with special reference to the Iban. In: Sutlive VH, Appell GN (eds) Female and male in Borneo. Borneo Research Council, Willamsburg VA, pp 317–333 (Borneo Research Council Monograph Series vol I)

*Winzeler RL (1995) Latah in Southeast Asia: the history and ethnography of a culture-bound syndrome. Cambridge University Press, Cambridge

Yamashita I (1993) Taijin kyofu or delusional social phobia. Hokkaido University Press, Sapporo

Yap PM (1951) Mental diseases peculiar to certain cultures: a survey of comparative psychiatry. J Ment Sci 97: 313–327

Yap PM (1952) The latah reaction: its pathodynamics and nosological position. J Ment Sci 98: 515–564

Yap PM (1962) Words and things in comparative psychiatry, with special reference to the exotic psychoses. Acta Psychiatr Scand 38: 163–169

Yap PM (1965) Koro – a culture-bound depersonalization syndrome. Br J Psychiatry 3: 43–50

Yap PM (1967) Classification of the culture-bound reactive syndromes. Aust NZ J Psychiatry 1: 172–179

Yap PM (1969) The culture-bound reactive syndromes. In: Caudill W, Lin T-Y (eds) Mental health research in Asia and the Pacific. East-West Center Press, Honolulu, pp 33–53

*Yap PM (1974) Comparative psychiatry: a theoretical framework. University of Toronto Press, Toronto

Yu Z, Prince R (1991) Social phobias in China. Transcult Psychiatr Res Rev 28: 240–246

Zaguirre JC (1957) Amuck. J Philippine Fed Priv Med Pract 6: 1138–1149

CHAPTER

15

N. Sartorius

Psychiatry in Developing Countries

1 Introduction

The decision as to whether a country is to be called a developing country depends on a number of criteria. Some of them seem straightforward: in general, such countries have a low per capita income, their communication systems are deficient (thus preventing the efficient transport of people, goods and information), health indicators testify to an unnecessary loss of life (e.g. as a result of high child mortality) and high rates of disease, and the level of education of their populations as a whole is low. When all of these and a number of other signs of the same kind are present, it is easy to reach agreement that the country is lagging behind in development or – if most of the indicators show that the country is in a very difficult situation – to decide that it belongs to the category of "least developed countries". At present, there are some 40 countries in the latter category. However, their populations are relatively small and together they represent only a modest proportion of the total population of the world or of the developing countries. The remaining countries outside Europe, the United States, Canada, Australia, New Zealand and Japan are less easy to place within a category, and the United Nations recently proposed splitting this group into two categories (rapidly developing countries and developing countries) and placing Eastern European countries and the countries of the Community of Independent States into a separate group because it has characteristics that distinguish it from the rest of the European countries.

However, this apparently convenient division of the world into developing and developed countries does not hold up under scrutiny.[1] Firstly, in many of the so-called developed countries, there are parts (or population groups) that in all respects resemble the countries of the developing world. Secondly, in most developing countries, there are now parts (or population groups) that have achieved a high standard of living and can be favourably compared with the average of the developed world. Thirdly, some of the developing countries are lagging behind in some aspects (e.g. health criteria), while, in terms of other criteria (e.g. per capita income), they are close to or better than some of the developed countries. Fourthly, countries are continually changing, and what could be called a developed country today might tomorrow, in most respects, become close to a developing country, and vice versa. Fifthly, economic development is not necessarily the final goal of human existence, and if the world is categorized according to non-economic criteria (e.g. in terms of tolerance, non-aggression, and human interdependence), it becomes necessary to redistribute countries among the developed and the developing categories.

In view of all this, it would seem wiser to avoid dividing the world into a small number of rigidly limited categories and to consider instead smaller groups of countries (or even groups of one), recognizing that each nation has its own cultural, economic, intellectual and emotional traditions and characteristics that make it to a significant degree unique. It is possible and more desirable to describe countries in terms of a number of relevant criteria, even if this takes a little more space and time or, if grouping is necessary, to do this bearing in mind that there are many differences between the countries in the group and that most countries in the world are mosaics of developed and developing regions, of modern and traditional parts and of noble and ignoble components. Just how much of each of these is present will vary with time and from one country to another.

Medicine and science reflect the above situation. Even in very poor countries, there are institutions engaged in research and service of outstanding quality (e.g. the Institute of Psychiatry in Mexico City); on the other hand, in the developed world, there are areas in which health care is poor (e.g. in slums or remote areas), a situation made worse by the fact that there is enough money available and that a lack of governmental resources cannot be blamed for the scenes which can be observed.

The differences within and between countries are part of the reason why it is illusory to hope that a comprehensive description of mental health programmes in the developing countries can be produced in a single book chapter. Other reasons that reduce the chances of producing a detailed and well-documented description include the scarcity of data resulting from well-conducted studies of the mental health situation in many of the developing countries and the rapidity of sociocultural, economic and other changes that are present in most of them.

2 The Problems

Four groups of problems beset psychiatry in developing countries: (1) the low value given to mental health by individuals and by society, (2) the high prevalence of mental and neurological problems, (3) the disre-

[1] At various times, other terms were proposed to distinguish these two groups of countries: Third World countries, non-industrialized countries, the countries of the South. None of these designations is convincingly better than others. In this chapter, the simplest terms – developing and developed – will be used, despite their imperfection.

gard for psychosocial aspects of various health and development projects and (4) the chronic lack of resources.

2.1 Low Value Given to Mental Health

The value which individuals and societies give to mental health and the normal functioning of the mind is of decisive importance for the development of mental health programmes. Where mental health lies low on the scale of values, it is difficult if not impossible to initiate or maintain mental health programmes. The low value of mental health finds its reflection in the low priority given to mental health activities within general health programmes, in the difficulty of recruiting the best postgraduate students to psychiatry, in the feeble, if any, participation of the population in mental health activities and campaigns and in the continuing disregard for opportunities to enhance mental health programmes. At the individual level, the low value of mental health finds its expression in people's willingness to put their mental health and sanity at risk by taking drugs, in the neglect of measures that would protect or enhance the mental health of children and in a variety of other ways. In most developing countries, mental health is not seen as being without any value: rather, it is considered as less precious than many other things or acquisitions ranging from material wealth to sexual attractiveness. Countries are not monolithic in this respect; however, although population groups vary in their appreciation of mental health, few place mental health and mental life among their top priorities.

There may be many reasons for this, probably including the vagueness of the definition of mental health and mental illness. Although mental health can, at best, be a component of health (the other equally non-independent component being physical health), the term is used to describe the state of feeling well, the absence of mental illness, normal mental functioning, equanimity, coping capacity and spiritual equilibrium. The vagueness of the concept makes it difficult to rally support for it as an objective of health service programmes and makes educational programmes supposed to help in the promotion of mental health nebulous; this in turn leads to mental health slipping even lower on the scale of values. The loose boundaries of the term also allow various disciplines – ranging from theology to marketing science – and various organizations to produce and promote definitions of mental health to suit their own purposes. Groups of population at large accept one or the other of these definitions (or create another one of their own), which diminishes or removes the pressure on health services to do something about mental health.

Another reason for the low position of mental health on the scale of values is the uncertainty about the limits of mental illness. In industrialized countries, the boundaries between mental illness and simulation, possession states, laziness, personality variants, artistic creativity, ideological originality and numerous other forms of behaviour and states of mind are somewhat clearer because of the efforts of the medical profession and because of the pressure of insurance systems to make a distinction between what is a medical condition (i.e. for which treatment has to be paid for by the insurance) and a non-medical condition (for which the health insurance does not pay for treatment).

In developing countries, neither of these two types of pressure are at work. Mental health care workers are often not numerous and often tend to adhere to illness categories created on the basis of experience in other countries. If these do not fit, vague, general terms ("psychotic state") are used. Insurance systems cover only a minute proportion of the population and do not have much weight or interest in precise definitions.

While there is now no doubt about the ubiquity and similarity of the incidence of severe mental disorders such as schizophrenia, there are other psychological disorders whose incidence and prevalence seem to be different in different sociocultural settings. Most of those conditions can be placed into the categories of major classifications of mental disorders (e.g. the International Classification of Diseases, ICD), but there are some that cannot: people who suffer from them, however, seek help from both traditional healers and modern doctors. These differences in frequency and the so-called culture-specific disorders contribute to shaping the image of mental illness in the setting in which they occur and can contribute to the notion that psychiatry is a branch of medicine that has a blurred focus, that the diseases which it is supposed to treat are not real diseases (there are no culture-specific forms of cardiovascular disease and these "real" non-communicable diseases are thought to respond to the same treatment in the same way, everywhere) and that the practitioners of psychiatry are closer to spiritual advisors (or traditional healers) than to real doctors.

Culture-specific ways of describing problems, and the abundance of influential traditional health practitioners who contribute their own terms and explanations, further blur the boundaries and concepts of mental disorder. This is good and bad: good because, in many instances, such vagueness of borders diminishes the probability of rejection and stigma because of mental illness, but bad because some of those who have a recognizable and treatable mental illness do not come to seek help from qualified services even if they are provided at a cost that the patient and the family could well afford.

It could be argued that, in some particularly disadvantaged areas, it might be better for the population to continue to consider some of the diseases from which it suffers as an inevitable part of the human condition rather than as disease states for which they could obtain treatment if only their government were to provide them with a better health service or with more money that would allow them to purchase good health care. The limitations of this kind of reasoning are dangerously close to a thoroughly unethical position in which ignorance is regarded as an ally diminishing the probability of revolt and postponing changes in the prevailing socio-economic order. This is particularly so in situations in which active involvement of the population (regardless of the state of the country's health services) could prevent the occurrence of diseases or diminish the problems of those who have them. Psychiatry is an excellent example of a field of work in which much could be done even under conditions of poverty and poor health services; it is all the more deplorable that this happens so rarely.

Yet another reason for the low priority given to mental health is probably the fact that, for a very long time, there were no effective methods of medical treatment of mental disorders. While the rest of medicine achieved victory after victory and produced or declared revolutionary advances in its practice, psychiatry still used a variety of treatment methods that did not yield convincingly good results. Some of the methods used were logical in the light of the limited knowledge about mental functioning, while others were innovative; still others were closer to folk healing and religious rites than to medicine. The impression that psychiatry does not have a solid scientific basis and that it is inefficient (and that it should therefore not be given any support except insofar as it can serve to control socially unacceptable behaviour due to mental disorder and impairment) have prevailed in many countries; in the developing world, however, that image of psychiatry in the face of continuously higher demands of the efficient medical disciplines and diminishing resources at the disposal of the health care system have pushed psychiatry lower and lower on the list of priorities. Lack of support has made mental health care facilities deteriorate, has decreased the quality of psychiatric training and has increased the attractiveness of other disciplines that are more lucrative and more interesting for students of health professions.

2.2 Frequency of Mental Disorders in Developing Countries

The notion that mental disorders in developing countries are less frequent than in the industrialized world was popular among psychiatrists and other medical specialists until relatively recently. Various reasons were put forward for this contention, including the apparently more leisurely pace of life thought to characterize developing countries, the absence of those stresses seen in highly organized industrialized countries, stable and strong family relationships, healthy air and nutrition without artificial additives. For certain mental disorders, the difference was explained by physiological differences: Carothers (1947, 1953), for example, thought that depression would not affect Africans because they have smaller frontal lobes than the white population. Some of these convictions had to do with prejudice and prevailing social order: Prince (1967) examined publications about the frequency of depression in African countries and reported that, before the early 1960s, there were very few articles stating that African patients had depressive disorders as often as people elsewhere: the majority of papers stated the opposite. In later years – after African countries became independent – the findings changed: depressive disorders were reported to be seen as frequently in Africans as in other citizens of the world. Prince and others felt that this could be related to the change in the way in which doctors were viewing patients (and patients were viewing themselves) before and after gaining political independence. The diagnostic labels that were previously used only for diseases of colonial officers or citizens then became available to everyone.

Gradually, the myth of differences in the incidence of mental disorders between developed and developing countries begun to lose strength. It became clear and admitted that stresses exist in all cultural settings, villages as well as towns, and that mental disorders are frequent regardless of the level of industrial development (e.g. Leighton et al. 1963; Sethi et al. 1972). The studies coordinated by the World Health Organization (WHO) – the International Pilot Study of Schizophrenia (WHO 1975b), the Study of Determinants of Severe Mental Disorders (Jablensky et al. 1992), the studies of depression in different countries (Sartorius 1983) – all seemed to indicate that differences in the frequency of mental disorders, whenever they are reported to exist, are more probably due to methodological reasons rather than to a real difference in incidence.

While mental disorders do not seem to vary in incidence among countries, it would not be surprising to find that there are differences in the prevalence of mental disorder. This difference could result from lesser survival chances of the mentally ill in many developing countries or from other factors, e.g. the better prognosis of mental illness in the developing than in the developed countries (WHO 1979) that has been reported on several occasions. Some of the epidemiological studies of severe mental disorders in

developing countries have shown a lower prevalence of these disorders (e.g. Cooper and Sartorius 1996), but it should be noted that the size of the differences between the findings of the studies in developed countries and those in developing countries often seems to be similar to that found between studies done in the same (developed) country. The differences in the clinical syndromes of diseases such as schizophrenia have been described as well as culture-specific disorders (see Chap. 14, Vol. 2, Part 1) but, on the whole, the differences between mental disorders in different cultures seem to be less pronounced than the similarities, particularly in the instance of severe forms of mental illness. Differences in patterns of alcohol and drug abuse (e.g. between Islamic and other countries) that were found can most probably be attributed to reasons other than the level of economic development (e.g. religion). In general, it could be said that poverty and its consequences (e.g. lower education) influences the form and course of mental illness and that the numbers of poor people in developing countries is higher than in the developed world.

The poor foundations of the myth of the happy savage – of the lower frequency of mental disorders in developing countries – crumble even more convincingly when the conditions due to early or later brain damage because of poor perinatal care, frequent infections, malnutrition and under-stimulation are also considered in the assessment of the burden that mental and neurological problems might present in developing countries. Mental retardation related to iodine deficiency disorders, for example, still appears frequently in many areas of the developing world. Conditions such as short, acute psychotic states due to high fever and psychological impairment due to cerebral malaria and encephalitis are also more frequent in the developing world, but often go unreported. Mild degrees of cognitive impairment are probably higher in those developing countries in which the control and correction of sensory deficits such as poor vision is sporadic or non-existent (see Chap. 8, Vol. 1, Part 2). The all too frequent comorbidity of mental and physical diseases unfortunately continues to obscure the real numbers of people with mental illness because of the tendency to report only one (usually physical) illness per contact and because of the higher mortality of the mentally ill due to co-existent physical illness and malnutrition.

The separation of psychiatry from neurology that happened in many countries in the past four decades – useful as it may have been in many instances – has also been detrimental to the assessment of the importance of psychiatry (and of neurology) in developing countries and to the determination of the priority that mental health programmes should have. The notion that psychiatry's field of action are the functional psychoses such as schizophrenia, manic depressive illness and various neurotic states, while various types of damage of the central nervous system (regardless of the presence of psychiatric symptoms), mental retardation, learning difficulties and behavioural disorders should be the domain of the neurologists and other specialists (often not existing in the developing countries) has decreased the estimates of the magnitude of mental and neurological problems in the eyes of the health system authorities and diminished the priority that might have been given to these disorders had they been grouped together. A second, more subtle negative consequence of the division between neurology and psychiatry was that the image psychiatry might have had by dealing with mental disorders due to such organic brain damage has been lost and that the prevention of these disorders did not become an obligatory part of mental health programmes. A third consequence is related to the image of psychiatry that is portrayed as a discipline dealing with states without pathological substrates and is therefore a discipline that should not be given the same respect (or resources) as other medical disciplines.

For many health decision-makers in developing countries, the findings that mental health problems are frequent in general health care services were surprising. The notion that the frequency of communicable diseases and consequences of trauma are so high in health services that mental disorders will constitute only a small proportion of those who come forward seeking help proved to be wrong. Study after study found that as many as 10%–20% of those contacting primary health care services in developing countries have psychological problems of varying severity. In rural Senegal, for example, 17% of the 545 children and 16% of the 933 adults who were examined had such problems, although these were not necessarily recognized as such by the general health care service workers. In other African countries, the situation was similar (see, e.g. German 1987; Ndetei and Muhangi 1979). Other parts of the developing world have reported similar findings. In South India, general practitioners estimated that 10% of their patients came to them primarily because of psychological problems (Gautam et al. 1980) and a multi-centric study coordinated by the WHO also confirmed these results (Sartorius and Harding 1983; Harding et al. 1983a,b), as did a number of other studies (Sartorius et al. 1990). More recent studies have shown that the situation is similar today and that mental health problems represent a major part of the burden of general health services (Üstün and Sartorius 1995).

Unfortunately, this is not all. There are good reasons to believe that the prevalence of mental disorders will continue to increase. The main reasons for this include demographic changes and longer life expectancy: as

more and more people survive into their early twenties, the group at higher risk for schizophrenia will increase. Similarly, survival into the age of heightened risk for depressive disorders will increase the prevalence of depressive disorders in the population of developing countries, survival into old age will increase the prevalence of dementia and so on. It might of course happen that better conditions of life and better health services will decrease the numbers of preventable neurological and psychiatric problems due to early trauma, malnutrition and other ills besetting developing countries today; however, it is not easy to predict which will be quicker – the decrease in the incidence of mental and neurological disorders due to brain damage or its increase due to better survival chances.

Another reason for the probable increase of prevalence of mental disorders is the improvement of chances of survival of people suffering from chronic diseases likely to be accompanied by mental disorders such as depression. The dictum that each of our successes in terms of better medical care and saving lives has to be paid for by an increase in risks for other diseases and problems (Gruenberg 1977) was true when first said and might, regrettably, remain true in the future. The longer life expectancy of people who have a mental disorder may have the same effect on increasing the prevalence of the disorders even if incidence stays the same or slightly diminishes.

Whether the rampant pollution that accompanies industrialization in developing countries will have neuropsychiatric consequences that will increase the incidence of mental disorders is also difficult to foresee; what is more likely, however, is that man-made accidents will happen more often in developing countries and will lead to an increase in the numbers of people with mental disorders (see Chap. 17, Vol. 2, Part 1). By way of illustration, of the approximately 3 million people affected by disasters in the 1967–1991 period, more than 95% were living in the developing countries in Asia and Africa (Desjarlais et al. 1995).

Although it is likely that some of the communicable diseases that can be accompanied by mental disorders will diminish with the improvement of health care and of the health education of populations, other diseases with mental health consequences might appear and be difficult to control for years to come. If this happens, developing countries will suffer more than their developed counterparts because they cannot easily deploy additional resources to counter the new perils. A good example is infection with the human immunodeficiency virus (HIV): if the current estimates of the frequency of psychological problems associated with HIV infection and acquired immunodeficiency syndrome (AIDS) are correct, within two decades there will be more people affected by AIDS-related dementia in Africa than there are hospital beds for all inpatient care on the continent. In many developed countries, the AIDS epidemic is almost under control, but in developing countries it is still on the rise.[2] In the instance of AIDS, the psychological problems are not only the dementias appearing late in the course of the disease, but also the early reactions to the infection (Maj et al. 1993; see also Chap. 16, Vol. 2, Part 2) and the problems arising in orphaned children and devastated families.

A further contribution to the burden of mental disorders in the developing world are the mental health consequences of violence on a mass scale (e.g. wars, insurgencies, civil strife) and on an individual level (e.g. crimes), the latter particularly in urban settings. Both forms of violence are incomparably more frequent in the developing world and have taken on epidemic proportions in recent years. Although suicidal behaviour is not necessarily a psychiatric problem, it is necessary to note here that suicide is more frequent in some developing countries than in most developed countries. Sri Lanka, for example, had the second highest suicide rate in the early 1990s, and suicide among young rural women in China is among the highest in the world (Desjarlais et al. 1995).

2.3 Disregard for Psychosocial Aspects of Health and Other Development Programmes

A third obstacle to the development of mental health programmes in developing countries is the restrictive definition of psychiatry as a discipline whose competence and task is to deal only or mainly with the most severely mentally ill. Since psychiatrists often accept this definition and they are the people who run mental health programmes, such programmes will also concentrate on the most severely ill and impaired (and on their inpatient care). The education of medical students – future health care decision-makers – will foster the same image of psychiatry and make changes towards a more enlightened view of mental health and psychiatry difficult for years to come.

The restriction of mental health programmes to activities related to the most severe forms of mental disorders also decreases the importance and relevance of mental health programmes. If defined in this way, mental health programmes deal with a small proportion of the population, and their significance for the

[2]For example, a recent report summarising the findings of the annual survey in the Republic of South Africa gives the figure of 50,000 new cases of HIV infection per month.

improvement of the health of the population is minor. If, on the other hand, mental health programmes were to be seen as having to deal with human behaviour, it would be logical that they assume a position in the mainstream of public health efforts and receive significantly more support and emphasis.[3]

Psychiatrists are often reluctant to assume responsibility for the development of broadly conceived mental health programmes which encompass the prevention of mental disorders, the promotion of mental health, the delivery of contributions that psychiatry and behavioural sciences can make to the resolution of psychosocial problems and the treatment and rehabilitation of people with mental illness and impairment. Postgraduate psychiatric training does not prepare psychiatrists for such an array of tasks, and they are rarely willing to devote all their time to the development of mental health programmes. They are also reluctant to see other professionals running such programmes: as a consequence, in most developing countries mental health programmes are narrowly conceived, and the people running them are either reluctant to see them be expanded to include the areas outlined above or find it difficult to gather sufficient support to put broad mental health programmes into operation.

And yet it is logical that the role of psychiatrists in most developing countries should be different from that of their counterparts in the industrialized world. In a country such as Ethiopia, there is one psychiatrist for every 5 million population; even if that person were to spend all of his or her time treating the mentally ill people, he or she would be able to help only a minute proportion of the mentally ill in the country. Initiating and building a mental health programme that will involve a wide variety of professional and non-professional helpers might have an immeasurably more useful effect for the same investment of time and energy. It can be argued that this is not what psychiatrists are for, but the counter-argument is that the identity and tasks of any professional depend on the context, on the setting in which they exercise their profession, and that the role of the psychiatrist in a developing country therefore has to be different. Top officials in developing countries do recognize the need to develop programmes that deal with psychosocial aspects of health and development. They might be inclined to support such programmes if there were qualified individuals willing to commit their time and effort to the development of such programmes.

[3]The bad effects of the separation of psychiatry from neurology at the organic end of the spectrum described above are also significant to these considerations.

The two most important consequences of the broad conception of mental health programmes are that, firstly, these programmes would gain in importance because they would deal with problems of major importance for health – such as the promotion of healthy behaviour – and secondly, that postgraduate training in psychiatry would have to include a section that would be fundamentally different from what is currently included in postgraduate courses in most developing and developed countries. The reorientation of postgraduate education in psychiatry in order to produce sufficient numbers of leaders of national and regional mental health programmes in developing countries does not mean that the need for psychiatrists with a predominant interest in clinical psychiatry would disappear; the investment in training, however, would have to be differently weighted to allow for the development of mental health programmes and for the growth of services for people suffering from mental illness.

The WHO has promoted the notion that mental health programmes have to be comprehensive and broad if they are to be seen as an important part of the public health enterprise (Sartorius 1978; WHO 1981, 1992a). The arguments put forward were that, firstly, many mental health problems can only be resolved if the health sector as a whole is involved, secondly that many of the public health problems could be resolved more easily if technology developed by mental health programmes (e.g. concerning change of attitudes, humanization of medicine) were to be used in general health care, and thirdly, that health programmes cannot make a contribution to the overall socio-economic development unless they incorporate a mental health component. Broad mental health programmes in the form recommended by the WHO have come into existence in a number of developing countries, but are not yet the model programme accepted by all.

2.4 Scarcity of Resources

The fourth major reason for the slow development of mental health programmes in developing countries is their chronic lack of resources. Annual expenditure for health in some of the least developed countries is a few dollars – as much as one thousand times less than in countries in the industrialized world. The scarcity concerns most of the resources that are necessary for health programmes – qualified staff, facilities, medicaments, laboratories, means of transport. Salaries, if paid at all, are so low that staff have to seek other sources of income – work abroad, private practice or other occupations – often of the most surprising

nature. The scarcity of resources is often combined with an almost total neglect of peripheral health care workers, resulting in burn-out, which is increasingly seen as the chief reason for the continuous deterioration of health care in developing countries. Inpatient facilities (antiquated mental hospitals) still represent most of the resources that mental health programmes have. These institutions, which are usually in a dire state, provide a vastly insufficient number of beds for the country; in many African countries, the number of psychiatric beds is in the region of 0.5–5 per 100,000 population. In Asia, the numbers of psychiatric beds are somewhat higher, but still low in comparison with richer countries (e.g. Vietnam, 0.78/100,000; China, 0.73/100,000; the Philippines, 1.13/100,000). In the richest of the developing countries, the situation is somewhat better: even there, however, the low position that psychiatry has in the eyes of decision-makers results in a disproportionately low amount of resources for mental health care and other mental health activities. The numbers of psychiatrists vary from 1:5,000,000 in Ethiopia (Alem 1997) to 1:200,000 in Korea (Shinfuku 1993), but their distribution is uneven and most of them are practising in urban areas. The number of psychiatric nurses is also low, although here the situation is less distant from numbers in developed countries. Psychiatric social workers are rarely available. Psychologists are more numerous and often face the difficulty of finding employment in mental health care services which traditionally have not included posts for psychologists. In Latin America, the numbers of psychologists are particularly high, and many among them will seek to open a private practice or work in sectors that have nothing to do with health.

Funds for the treatment of those with mental illness are scarce and their availability irregular. The resources necessary for the maintenance of inpatient institutions, patients' food, clothing and other necessities are often not available. Distances between health service centres and the population groups are often large in developing countries, and effective transportation is of essential importance, yet vehicles are few and there is no money for fuel and maintenance and no replacement parts. The distribution of funds between mental health care services and other services is not even, and the low priority of mental health programmes is reflected in a variety of ways when it comes to budgetary provisions, new posts, fellowships, library supplies and other material. The unjust distribution of resources does not only affect the care of patients; it also has negative effects on the morale of the staff working in mental health care services. Many leave the service, and those who can be recruited are often not of the best quality.

An indicator of the state of mental health programmes is the legislation. A recent survey of legislation in 45 countries demonstrated that, in 12 of them, there was no special mental health law and the admission of patients, their treatment and other aspects of the mental health care services were governed by some informal system; a number of these countries belonged to the developing country category. The same survey established that mental health laws in the poorer countries had not been updated as recently as the laws in richer countries. In another analysis, the survey showed that, in a number of countries, mental health laws were incompatible with a policy of integration of people with mental illness into the community; all of the six countries that belonged to that category were in the developing world (see Chap. 15, Vol. 1, Part 2).

3 Possible Solutions

The difficulties alluded to above have not prevented the establishment of mental health programmes in the developing world. Often, such programmes were built around an individual excelling in energy and charisma; there are, however, also programmes that have grown in the absence of such individuals.

Although none of the developing countries has established, consistently maintained and evaluated a comprehensive mental health programme, many have adopted progressive programmes and are implementing them. The elements such programmes comprise are described below.

3.1 Promotion of the Value of Mental Health

The promotion of mental health is best understood as the process of moving mental health up on the scale of values of individuals and societies (Sartorius 1992, 1998). Mental health promotion should thus not be equated with the improvement of the mental health status of the population; improvement will be achieved by programmes of prevention and treatment of mental disorder and by various other means, e.g. better nutrition. Placing mental health high on the scale of values is, however, important for the improvement of mental health because it provides the motivation to undertake preventive and curative measures which will result in better mental health and functioning.

Since the promotion of mental health is a process of changing values, it cannot be the exclusive task of mental health care services or indeed of health services in general. Values are shaped during the upbringing of

children, at home and in school, through the media, by examples and models and in other ways. Mental health programmes in the developing countries should see the promotion of mental health (understood in this sense) as their highest priority, because success in placing mental health higher on the scale of values is essential for all other activities in the mental health field. Action concerning promotion will require multiple links and alliances with those who have a role to play in this process; the consequence of this is that mental health programmes must consider partnership with many groups with whom they have not had contact before.

Psychiatrists will usually not be in charge of mental health promotion programmes; it is important, however, that they should participate in them and advocate them, because the success of such programmes will help in the development of programmes of care and rehabilitation of people with mental illness. Sometimes, they can initiate such programmes; in Pakistan, for example, the department of psychiatry of the University of Islamabad has been the driving force behind a programme designed to promote mental health in schools (Mubbashar 1999). In other settings, mental health promotion activities have been embedded in programmes of general health promotion and community development (Badura and Kickbusch 1991; Trent 1992).

3.2
Primary Prevention of Mental Disorders

Primary prevention of mental disorders has long been considered an impossible task. When the Director-General of the WHO presented the results of an analysis of possibilities for the prevention of mental, neurological and psychosocial disorders (WHO 1986) at the 39th World Health Assembly, the assembled ministers and other top health officials were both astonished and pleased. A resolution was adopted requesting countries to take action and urging the Director-General to assist countries by preparing appropriate guidelines for action. Most of the measures presented in the report of the Director-General of the WHO have to be undertaken by sectors other than the health care sector, some have to be carried out by the health care sector but not by mental health care services (e.g. the improvement of perinatal care) and very few by mental health care workers. The latter, however, have a central role in bringing about preventive action; firstly, they have to advocate the development of preventive programmes; secondly, they have to continue to stimulate general health services and other social services to do what is necessary; and thirdly, they have to monitor the effects of the interventions undertaken.

Possibilities for the primary prevention of mental disorders are mentioned above, and more details can be found in the chapter by Eisenberg (see Chap. 8, Vol. 1, Part 2). In general, it should be stressed that, in developing countries, possibilities for the primary prevention of mental disorders (e.g. due to early damage of the central nervous system) are particularly numerous but are usually neglected (Sartorius and Henderson 1992).

3.3
Treatment of Mental Disorders

Using data from the Global Burden of Disease publication (Murray and Lopez 1996), M. Phillips (personal communication) has recently estimated that, of the 80% of people with depressive disorders living in developing countries, only 5% receive treatment. The situation is probably similar for other types of mental illness. The report by Murray and Lopez and a number of other papers have provided new ammunition for the request to introduce mental health components into primary health care – a recommendation often repeated by the WHO and numerous other bodies (WHO 1975a, 1990).

There are a number of examples of countries or areas in which mental health care has been provided through general and primary health care agents (Murthy 1996; WHO 1984, 1990). The effects have, on the whole, been rewarding. The results have indicated an increased coverage of care for people with serious mental illness and an increased satisfaction of general health care workers trained to deal with mental illness. In some developing countries, e.g. Thailand and Uganda, the treatment of mental disorders and the promotion of mental health have been included in national health care plans as one of the priorities for action. A large number of manuals enabling general practitioners and primary health care workers to deal with mental disorders have been produced and used in training programmes in many countries (WHO 1992b). The WHO has even complemented its the tenth revision of ICD (ICD-10) by a classification for use in primary health care (ICD-10 PHC). This document has been produced in order to facilitate reporting about mental illness in primary health care facilities. The PHC version of ICD-10 has a sharply reduced number of categories – only 22 categories versus more than 600 categories (see Table 1) and in some even further simplified versions only nine categories. These were selected on the basis of the frequency of diagnosis made at the primary health care level, paying attention to including only those categories which primary health care workers will be able to use with minimal instructions (and for most of which treatment is available).

Table 1. Mental health categories in the primary health care version of ICD-10 (ICD-10 PHC)

F00	Dementia
F05	Delirium
F10	Alcohol use disorders
F11	Drug use disorders
F17.1	Tobacco use
F20	Chronic psychotic disorders
F23	Acute psychotic disorders
F31	Bipolar disorder
F32	Depression
F40	Phobic disorders
F41.0	Panic disorder
F41.1	Generalized anxiety
F41.2	Mixed anxiety and depression
F43	Adjustment disorder
F44	Dissociative disorder
F45	Unexplained somatic complaints
F48.0	Neurasthenia
F50	Eating disorders
F51	Sleep problems
F52	Sexual disorders
F70	Mental retardation
F90	Hyperkinetic disorder
F91	Conduct disorder
F98.0	Enuresis

Classifications are often not used; to increase the attractiveness of ICD-10 PHC and to make it more widely used, the WHO, in collaboration with non-governmental organizations (e.g. the organizations representing general practitioners and psychiatrists), has also produced specific instructions on how to recognize the conditions to be classified and how to treat people who have them once they are recognized. The classification was produced in the late 1980s and released for test use in a number of countries. It was finalized once the results of these tests became available and published in 1996 (WHO 1996).

The introduction of mental health components into primary health care has not always been fully supported by psychiatrists. Some of them feared that the treatment of mental illness in general health care might have untoward consequences for the patient, others that the devolvement of responsibility might decrease the numbers of patients who would come forward as their clients. Neither of the two fears proved justified. General practitioners and primary health care workers can provide effective treatment to a significant proportion of patients, particularly if there are possibilities for referral and if general health care workers receive short practical training. In developing countries, other needs also have to be satisfied, including the need to ensure a regular supply of psychotropic medication and possibilities for the transportation of health care workers to remote areas and of patients for referral, for example. In both developing and developed countries, the introduction of primary health care concepts could actually have a positive effect on psychiatry, increasing its importance in the framework of public health action (Sartorius 1997) and improving its performance.

In developed countries, the treatment of mental disorders in the general health care system is, to a certain degree, a matter of choice of those concerned – the patients, their families, the general practitioner and the psychiatrist. In developing countries, it is not: if mental health care is not provided through the general and primary health care worker, it will not – for an immense proportion of the population – be provided at all.

The treatment of mental disorders, particularly those that are likely to produce a lasting impairment, has to be linked to a process of rehabilitation. In developing countries, a very small proportion of the most severely ill are inpatients in the mental hospitals, mainly built in the colonial times. Most of the remaining patients who were not rejected by their families (because the family could not afford to feed them or could not tolerate their socially unacceptable behaviour) used to stay with them and were partly integrated into the life of the family and the community. Families have thus played a central role in the life of patients during and after the disease. Current processes of urbanization and the nuclearization of families diminish the capacity of families to provide long-term care to their sick members; at present, most of the developing countries have not found an answer to the challenge of such care under conditions of the dwindling capacities of families to help.

The life of those suffering from a severe mental illness is difficult in all types of countries. In developing countries, the situation is the same or worse. While the families and communities of people who suffer from mental illness show compassion and understanding at the beginning of the illness (unless the behaviour of the mentally ill is very disturbing, bringing shame to the family or community or endangering the life of others), their capacity to care and look after these individuals is usually soon exhausted. Sometimes, those with a severe mental illness are ejected from their setting and become vagrant psychotics; sometimes, they remain hidden in some dark corner of the house for many years; often, they die a premature death because of the increased risks to their health,

their diminished capacity to cope and their lesser ability to request or obtain medical help. The introduction of psychiatric services, in collaboration with and through general health services, is therefore a life-saving operation in addition to being a process of bringing people with mental illness relief and support.

4 Conclusion

The forms of the most frequent mental disorders are similar in developing and developed countries. There are culture-specific forms of mental illness, but these represent only a small proportion of all mental illness. Most of the differences between developed and developing countries result from poverty and the many ways in which it impacts on people, on health care and on diseases.

Poverty also affects the chances for the application of knowledge to help people with mental health problems; nevertheless, the development of mental health programmes is possible and, particularly in poor countries, highly desirable. At the same time, however, it is important to realize that the reluctance of the mental health care professions to change and accept new definitions of their roles often figures highly among the obstacles to creating or improving mental health programmes.

Mental health programmes must be an integral part of general health care programmes, which, in turn, are inextricably linked to socio-economic development; however, this does not mean that change within the field of health or mental health can only happen if developing countries catch up with the industrialized world. Much can be done for mental health in developing countries today, and too little of it is undertaken. In this respect, psychiatry in developed countries has an important, often neglected role to play. Many psychiatrists from the developing world are trained in industrialized countries, and much of the technology and new knowledge is generated in these settings; it would be immensely useful if the differences that exist between countries were to be systematically considered in postgraduate medical education everywhere. This would be one way of exchanging information about ways of organizing services and providing care, but all other occasions and opportunities should also be used so that experience from countries at different stages of development can be shared and judiciously used in educating mental health care professionals and in building relevant and useful mental health programmes.

5 References

Alem A (1997) Mental health in rural Ethiopia. University of Umea, Umea

Badura B, Kickbusch I (1991) Health promotion research. World Health Organization, Copenhagen (regional publications)

Carothers JC (1947). A study of mental derangement in Africans and an attempt to explain its peculiarities, more especially in relation to the African attitude to life. J Ment Sci 93: 548–596

Carothers JC (1953) The African mind in health and disease. World Health Organization, Geneva (WHO monograph series no. 17)

Cooper JE, Sartorius N (eds) (1996) Mental disorders in China. Gaskell, London

Desjarlais R, Eisenberg L, Good B, Kleinman A (eds) (1995) World mental health: problems and priorities in low-income countries. Oxford University Press, Oxford

Gautam S, Kapur RI, Shamasundar C (1980) Psychiatric morbidity and referral in general practice – a survey of general practitioners in Bangalore City. Indian J Psychiatry 22: 295–297

German GAG (1987) The extent of mental health problems in Africa today. An update of epidemiological knowledge. Br J Psychiatry 151: 435–439

Gruenberg EM (1977) The failures of success. Milbank Memorial Fund Q 55: 3–24

Harding TW, Climent CE, Diop M, Giel R, Ibrahim HHA, Mrthy RS, Suleiman MA, Wig NN (1983a) The WHO Collaborative Study on Strategies for Extending Mental Health Care. II. The development of new research methods. Am J Psychiatry 140: 1474–1480

Harding TW, D'Arrigo Busnello E, Climent CE, Diop MB, El-Hakim A, Giel R, Ibrahim HHA, Ladrido Ignacio L, Wig NN (1983b) The WHO Collaborative Study on Strategies for Extending Mental Health Care. III. Evaluative design and illustrative results. Am J Psychiatry 140: 1481–1485

*Jablensky A, Sartorius N, Ernberg G, Anker M, Korten A, Cooper JE, Day R, Bertelsen A (1992) Schizophrenia: manifestations, incidence and course in different cultures. A World Health Organization Ten Countries Study. Psychol Med Monogr Suppl 20: 1–97

Leighton AH, Lambo TA, Hughes CG, Leighton DC, Murphy JM, Cornell BM (1963) Psychiatric disorders among the Yoruba. Cornell University Press, New York

Maj M, Starace F, Sartorius N (1993) Mental disorders in HIV-1 infection and AIDS. Hogrefe and Huber, Seattle (WHO expert series of biological psychiatry, vol V)

Mubbashar MH (1999) Mental health services in rural Pakistan. In: Tansella M, Thornicroft G (eds) Common mental disorders in primary mental health care. Routledge, London

Murray CJL, Lopez AD (1996) The global burden of disease. Published by the Harvard School of Public Health on behalf of the WHO and the World Bank. Distributed by Harvard University Press, Cambridge, MA

Murthy RS (1996) Economics of mental health care in developing countries. In: Lieh Mak F, Nadelson CC (eds) International review of psychiatry. American Psychiatric Press, Washington, pp 43–56

Ndetei DM, Muhangi J (1979) The prevalence and clinical presentation of psychiatric illness in a rural setting in Kenya. Br J Psychiatry 135: 269–272

Prince R (1967) The changing picture of depressive syndromes in Africa: is it fact or diagnostic fashion? Can J Afr Studies 1(2): 177–192

Sartorius N (1973) Culture and epidemiology of depression. Psychiatr Neurol Neuro Chir 76: 479

Sartorius N (1978) The new mental health programme of WHO. Interdiscipl Sci Rev 3: 202–206

Sartorius N (1992) The promotion of mental health: meaning and tasks. In: Trent DR (ed) Promotion of mental health, vol 1. Avebury, Aldershot, pp 17–23

*Sartorius N (1997) Psychiatry in the framework of primary health care: a threat or a boost to psychiatry? Am J Psychiatry 154 [Suppl]: 67–72

Sartorius N (1998) Universal strategies for the prevention of mental illness and the promotion of mental health. In: Jenkins R, Üstün TB (eds) Preventing mental illness. Wiley, Chichester

Sartorius N, Harding TW (1983) The WHO Collaborative Study on Strategies for Extending Mental Health Care. I. The genesis of the study. Am J Psychiatry 140: 1470–1473

Sartorius N, Henderson AS (1992) The neglect of prevention in psychiatry. Aust NZ J Psychiatry 26(4): 550–553

Sartorius N, Davidian H, Ernberg G, Fenton FR, Fujii I, Gastpar M, Gulbinat W, Jablensky A, Kielholz P, Lehmann HE, Naraghi M, Shimzo M, Shinfuku N, Takahashi (eds) (1983) Depressive disorders in different cultures. World Health Organization, Geneva

**Sartorius N, Goldberg D, de Girolamo G, Costa e Silva J, Lecrubier Y, Wittchen HU (eds) (1990) Psychological disorders in general medical settings. Hogrefe and Huber, Toronto

Sethi BB, Gupta SC, Kumar P (1972) A psychiatric survey of 500 rural families. Indian J Psychiatry 94: 183–196

Shinfuku N (1993) A public health approach to mental health in the Western Pacific Region. Int J Mental Health 22(1): 3–21

Trent D (ed) (1992) Promotion of mental health. Avebury, Aldershot

**Üstün TB, Sartorius N (eds) (1995) Mental illness in general health care. An international study. Wiley, Chichester

*WHO (1975a) Organization of mental health services in developing countries. World Health Organization, Geneva (Technical report series 564)

WHO (1975b) Schizophrenia: a multinational study. World Health Organization, Geneva (Public health papers 63)

*WHO (1979) Schizophrenia: an international follow-up study. Wiley, Chichester

WHO (1981) Social dimensions of mental health. World Health Organization, Geneva

WHO (1984) Mental health care in developing countries: a critical appraisal of research findings. Report of a WHO study group. World Health Organization, Geneva (WHO technical report series 698)

WHO (1986) Prevention of mental, neurological and psychosocial disorders. World Health Organization, Geneva (document A30.9)

*WHO (1990) The Introduction of a mental health component into primary health care. World Health Organization, Geneva

*WHO (1992a) Mental health programmes: concepts and principles. Division of Mental Health, World Health Organization, Geneva (document WHO/MNH/92.11)

WHO (1992b) An annotated directory of mental health training manuals, 4th edn. World Health Organization, Geneva

*WHO (1996) Diagnostic and management guidelines for mental disorders in primary care (ICD-10 chapter 5. Primary care version). Hogrefe and Humber, Seattle

CHAPTER

16

V. Krasnov, V. Lebedev, E. West

Psychiatric Problems Arising in Extreme Environmental Circumstances

1 Introduction

Human exposure and mastery of extreme environmental conditions inevitably resulted in a number of newly described psychiatric phenomena for which our mind and body remain phylogenetically and ontogenetically unprepared. As we continue to master harsh environments of the Arctic, Antarctic, mountain peaks, ocean depths, and space, we come upon new discoveries about human behaviour and the psyche in the process of forced adaptation to these altered, extreme conditions.

We continue to challenge the human body and mind with various professional activities under unfavourable conditions which ultimately stress human psychophysiology. For example, some professional activities require sustained maintenance of the highest degree of visual and auditory acuity in a setting of emotional and cognitive stress. It is well known that psychological and physiological factors have a direct effect on the overall functional state of an individual. There are several examples of this. Desynchronosis is a sleep irregularity due to a change in the usual circadian rhythm and is seen in shift workers and participants of lengthy expeditions. These subjects frequently demonstrate emotional instability and pronounced physical and emotional fatigue. Likewise, prolonged work under the conditions of the polar night commonly results not only on psychological stress, but also in a disruption of normal neurovegetative functions in the body. Lengthy submarine travel and work also require a special artificial illumination schedule to enable sailors to work.

There are other environmental stressors which we associate with disruption of normal bodily functions. These include chronic exposure to vibration, noise and electromagnetic radiation from power plants. As a further example, modern-age space travel and work require adaptation to zero gravity conditions for which humans are clearly unprepared by evolution on Earth.

2 Mental Health Problems

Isolated exposure to any one of these extreme environments may not necessarily produce a mental disorder. However, the amalgamated action of the untoward environmental effects presents new issues to add to our current psychiatric phenomenology, diagnosis and treatment. One such issue is the problem of psychological compatibility between members of very small teams who perform shift work or participate in lengthy expeditions in a confined space isolated from their habitual social environment and support.

We continue to face problems with professional selection for such missions, as methods are borrowed from the field of psychology that have been developed to address human behaviour under conditions that are more familiar to us. For example, there are only few data about latent interpersonal conflicts among the members of small groups of explorers who avoid meeting one another after they return home (Lebedev 1989).

It worth mentioning another known phenomenon observed in the setting of environmental isolation and extreme stress. Similar to the illusion of experiencing something double, a person may visualize his or her own double and experience other auditory, visual and tactile hallucinations and distortions of reality. Such brief and transient conditions are fully reversible and appear to be related to the attempt to adapt to an extreme situation in which reality testing and functional ability remain preserved; the individual quickly returns to an asymptomatic state.

Extreme environmental conditions may also have their effect on large populations. Ecological psychiatry is concerned with the relationship between the physical, chemical and other environmental pollutants and mental health in large groups of people. Unlike an individual's adaptation to extreme conditions, a larger group responds to ecological stressors in a more complex way and with numerous intertwined, biopsychosocial and cultural pathogenic factors.

3 Ecological Psychiatry

The new field of ecological psychiatry continues to develop along with the demand for its clinical and research applications in the recent past. However, the future areas of development in ecological psychiatry remain unclear and will require future study. Overall, ecological psychiatry may be defined as an area of complex scientific inquiry into mental health and illness in the setting of unusual, anthropogenically produced or naturally occurring environmental conditions (Krasnov et al. 1993a; Krasnov 1995).

Areas of interest include individual and public health issues related to the various effects of the psychological and social environment of mental health and illness. An area of major importance is thus the mental health of individuals who work in prolonged sensory and psychosocial isolation with forced disruption of their circadian rhythm in unusual geoclimatic

situations. Of special interest is the study of individual adaptive reactions within a small group during prolonged isolation (see above).

Epidemiological vectors address the specific features and variants in mental health and psychosomatic diseases as a result of living and working in large, new technological and industrial settings. These may include newly populated areas in the Far North of Russia and Canada, where native and non-native populations may significantly differ in their social and physical adaptation styles (Korolenko et al. 1991; Semke 1992; Krasnov 1995).

Moreover, it is important to continue research and clinical efforts related to the problems of refugees, forced migrant populations and victims of wars, torture, violence and imprisonment throughout the world (see Chaps. 18–21, this volume).

Finally, ecological psychiatry continues to promote areas of research into immediate and delayed consequences of great natural and man-made environmental disasters. For example, the Chernobyl disaster of 1986 dramatically promoted the development of ecological awareness in Russian medicine. The Chernobyl harmful exogenous factors affected millions of people and continue to remain a major health hazard. The sequel of the disaster called attention to other previously ignored large-scale technological accidents and calamitous environments. It is precisely because of the immediate and specific consequences of Chernobyl that the field of psychiatry had to re-evaluate its traditional and narrow nosography and to look for new approaches to diagnosis, evaluation, treatment and rehabilitation of the victims (Krasnov et al. 1993a,b, 1995; Krasnov et al. 1995; Viinamäki et al. 1995; Nyagu and Loganovsky 1997; Havenaar et al. 1997). This may present an area of joint inquiry for industrial and ecological specialists within psychiatry.

However, the methodology of ecological psychiatry remains largely undeveloped. It is clear that it will require biopsychosocial and cross-cultural approaches to study the role of environmental factors in human health and social functioning. The future areas of methodological development will most likely evolve around individual and group psychobiological responses to various negative and positive environments. In addition, new methods in appropriate psychiatric rehabilitation must be considered.

In recent years, a number of publications have addressed psychosomatic and non-psychotic mental disorders which are developing with increasing frequency in many hazardous industrial settings (Simon et al. 1990; Krasnov 1991, 1995; Andresen et al. 1993). These include sleep–wake cycle disturbances, severe emotional and cognitive stress, perceptual disturbances and similar conditions. The afflicted workers are found among air traffic controllers, submarine sailors, space workers and astronauts (Weybrew 1963; Hauty and Adams 1966; Tenfiord 1966; Serxner 1968; Offerlend and Roos 1969; Berry 1970; Bennet 1972; Lindsley 1972; Takla et al. 1994). Current scientific literature appears to group these conditions with other psychogenic neurotic disorders. However, this seems partly unjustified. Unlike neurotic conditions, the phenomena of adaptation appear to preserve a healthy motivation for work and do not demonstrate maladaptive interpersonal symptomatology. It is obvious that unusual, brief, reactive psychological phenomena may occur under extreme conditions. However, these experiences appear to result from the process of adaptation to the extreme conditions and do not necessarily cause particular problems or hinder work. A number of vivid descriptions of such interpersonal adaptation phenomena and latent conflicts during lengthy polar expeditions and other travel in isolation can be found in current (mainly in non-professional) literature (Byrd 1932; Slocum 1956; Law and Bechervaise 1957; Bombard 1958; Heyerdahl 1959; Mullin and Connery 1959; Serxner 1968; Leonov and Lebedev 1971; Gunderson 1973; Rivolier et al. 1988; Sandal et al. 1996).

Occasional interpersonal difficulties and conflicts appear to be related to asthenic, affective and neurovegetative conditions which tend to develop under prolonged stress and in total or partial isolation. These asthenic neurovegetative symptoms frequently include progressively increased fatigability, irritability, sleep disorders, dysphoria, anxiety, dizziness, headaches, tachycardia, unpleasant somatic sensations, vegetovascular instability, hyperaesthesia and hypervigilance. The above-mentioned phenomena are highly nonspecific and tend to occur in small expedition groups and during prolonged isolation and sensory deprivation. When the extreme conditions persist unremittingly for an extended period of time, bona fide psychiatric disorders do tend to develop as well, as documented by a number of authors (Ritter 1900; Byrd 1932; Slocum 1956; Bombard 1958; Burns and Kimura 1963; Lebedev 1989; Cooper 1993).

Various sensory and perceptual disturbances, illusions and the specific phenomena of depersonalization and derealization can frequently be observed under specific extraordinary environmental conditions. Depersonalization and derealization appear to be specifically provoked by desynchronosis, sleep deprivation, prolonged isolation and monotony under experimental conditions. Unlike similar known psychiatric conditions, these phenomena are transient, situation specific and do not affect the individual's sense of self-control.

For example, in the setting of zero gravity, depersonalization is a common experience. Proprioception and the sense of one's own position in space become impossible, since there are none of the usual terrestrial

gravity cues available. Under such conditions, the process of adaptation requires an astronaut to rely almost exclusively on visual perception in order to estimate his or her position in space. Similar adaptation may take place during a space walk outside the spacecraft, where orientation in space is aided by the Earth, the nearby spaceship, the sun and the stars (Leonov and Lebedev 1971). Nevertheless, depersonalization and derealization in space are experiences with total preservation of reality testing and without feelings of morbidity.

Numerous experimental designs have been utilized in order to study sensory and cognitive deprivation and isolation. The studies have been carried out in the laboratory and in space in zero gravity. Common findings included asthenia, somatoform conditions, brief and transient illusions, hallucinations and cognitive mistakes. At the same time, reality testing always remained intact and allowed for necessary corrections while performing specific tasks and assessing unusual situations (Heron et al. 1956; Burns and Kimura 1963; Sharpe 1969; Leonov and Lebedev 1971; Daniel and Vining 1983; Lebedev 1989).

Similar psychological phenomena have been reviewed in detail in the literature dealing with sensory deprivation (Bexton et al. 1954; Heron et al. 1956; Davis et al. 1960; Mendelson et al. 1961). Sensory deprivation and monotony appear to stimulate the human need for external stimuli and emotionally rich responses also known as "sensory hunger". Thus perception and mental images acquire additional vividness, but do not affect one's ability to distinguish them from images of real objects. This type of compensatory imagination is inseparably connected to the adaptation activity of the mind and body. The longer the period of solitude and isolation, the greater the need for social contact becomes. The psychological phenomena of this social need may be manifested by assigning personal meaning to inanimate objects, talking out loud to oneself or to one's diary and responding to internal stimuli. In solitude and isolation, a human being sometimes visualizes his or her own double and communicates with it. However, a clear distinction between reality and fantasy is always preserved. In addition to such pseudoautism, prolonged solitude and isolation produce peculiar patterns of overvalued ideas of inventiveness, reformation and religious and philosophical contemplation about the world and the purpose of existence. Ideation of this nature completely remits upon discontinuation of isolation and return to normal life and does not recur.

At the same time, comparable adaptation mechanisms may take place under monotonous and isolated conditions of some asthenizing and exhausting industrial labour environments. Not much is known about the subject except for some documented occupational hazard factors (Burns and Kimura 1963; Serxner 1968; Lindsley 1972; Sandal et al. 1996). Although sanitary and hygienic norms have been established for most physical and chemical work environments, most researchers have not studied the complex interaction between occupational hazard factors and their cumulative effect on an individual and his or her ability to adapt. Moreover, some hazardous occupational settings already exist within unsafe ecological environments. Such conditions arise at several plants in the oil industry, mining and non-ferrous metallurgy north of the polar circle, in the north of Russia and in Siberia, for example. Therefore, we face a complex research problem of resolving the interplay of multiple variables involved in the mental hygiene of industrial labour. These variables include technological factors, the industrial environment, individual and group emotional and cognitive reactions and the overall ecological background.

Human adaptation to certain specific conditions of professional activity has become a real problem for modern medicine. Frequent, lengthy geographic and military expeditions, space exploration and aeronautics, disruptive shift labour on submarines and similar work environments increasingly require humans to perform complex functions in an artificial space–time environment and isolation. Under such conditions, the perception of space and time is modulated by complex high technology, but humans are expected to perform with the highest degree of concentration, attention and alertness. Space perception and information related to operations in space and time come from the instrument panels and not from the usual perceptual cues. Operation of modern machinery, such as airplanes and ships, is made possible by indirect measurements by the instruments. Frequently, there is a discordance between individual's subjective perception and the information provided by the technological devices. For example, an object may appear to be far away when viewed through a porthole, but it is much closer according to the instruments on a boat or a spaceship. More recent errors in rendezvous between inhabited space modules and a supply aircraft and accidents with in-flight refuelling are examples of such mistakes.

4 Dynamics of Subclinical and Clinical Adjustment Reactions

However, mental disorders do arise as a result of work under extreme conditions. These disorders are typically limited to asthenic syndromes with emotional lability and neurovegetative disturbances accompanied

by inappropriate aggressive and demonstrative behaviours. Recent research efforts have enabled us to isolate and describe some of the common features of psychiatric sequelae after exposure to extreme conditions. Our early work dealt with mental disorders in members of marine vessels on long-term self-contained missions (Kokhanov and Krasnov 1986). The factors under consideration included vibration, noise, excessive humidity, change in temperature, desynchronosis, excessive cognitive demands on vision and hearing and everyday life inconveniences, isolation from the family, information deficiency and a restricted circle of personal and professional contacts. The duration of the working period and observation was up to 6 months. All individuals in the study were male and aged from 20 to 46 years old. A total of 149 individuals were studied, and a standardized structured interview was used to assess and monitor the mental state changes. Clinical findings were quantified using a specially developed psychiatric symptoms assessment scale with 60 clinical entities, including asthenic, affective and neurovegetative symptoms and signs. In addition, changes in vital signs and orthostatic blood pressure were also monitored.

The study allowed us to understand changes in mental status over the period of observation and to reveal certain patterns of behaviour associated with the gradual development of affective and neurovegetative symptoms. It became apparent that some symptoms such as asthenia and dysthymia eventually developed into stable clinical conditions. During the first month of work, 39 men (26.2%) revealed episodic disturbances, asthenia and neurovegetative signs. These findings were revealed only in the setting of a structured interview and had no influence on observed behaviour and work productivity. By the end of the study, after 6 months, 76 individuals (51%) revealed similar symptoms with asthenia and mild depression. Transient initial dysphoric reactions were excluded from these figures, as such reactions were clearly due to the initial adjustment to a new work routine and adaptation to a new social environment.

The asthenic disturbances were characterized by a gradual increase in psychological and cognitive deterioration, irritability, hyperaesthesia and sleep problems. The latter were limited to initial insomnia, insufficient depth of sleep and waking up frequently. Later sleep disturbances were marked by hypersomnia and dysphoria upon awakening. The end of the working day was typically marked by low mood and dysphoric responses to others. Subjects reported physical and emotional fatigue, headaches, excessive sweating, tachycardia, tremor and dyspnea. Vegetative findings were accompanied by aches, paraesthesias and general physical discomfort with hypochondriac and phobic ideation.

During the final 2 months of observation, mental status findings were more consistent with the development of depressive disorders with a complex association with asthenia. Overall deterioration in activity was noted during the day and especially in the morning, when most subjects had difficulty initiating most activities. Psychomotor retardation and speech slowing and poverty were also increasingly noticeable. There were significant problems with attention and memory tasks on cognitive mental status examination.

The development of emotional problems acquired a predictable pattern of earlier tension and irritability progressing to frank apathy and social withdrawal, which did not appear to affect work-related activities. Apathy was complicated by increasing signs of depression and episodic anxiety. Emotional instability was marked by intrapunitive reactions and introversion. Occasional external expression of dysphoria was accompanied by self-criticism and feelings of regret. A tendency towards low self-esteem was also discovered. Somatoform symptoms became more clearly localized to the chest, epigastric and head areas and became stable. Individuals also woke up early in the morning.

Thus a complex of depressive signs and symptoms was observed to evolve according to a specific pattern. Initial occasional asthenic and neurovegetative symptoms gave a way to dysphoria and apathy and eventually to a stable depressive disorder.

The clinical nature of mental disorders seen in long-term sailing appears to be truly multifactorial, with psychosocial and physical hardship stressors being of equal importance in symptom formation. This can be seen in the phenomenology of the depressive disorders, which include both psychological and biological manifestations such as hyperaesthesia and dysphoric mood, memory and attention problems and somatic complaints, which are usually intertwined in one complex observable pattern. At the same time, the above-mentioned depressive disorders are strikingly similar to neurotic depression and atypical depression with an organic substrate. Diurnal mood variations with early morning dysphoria, dysphoric reactions, somatoform features and neurovegetative signs seen in sailors after 6 months are non-specific symptoms of depression which are most likely related to underlying physiological and chronobiological disturbances.

It is important to note that subjects' ability to perform work-related functions remained relatively unaffected, and depression did not make it any more difficult to follow routine operations that did not demand independent decision-making. In spite of the obvious decrease in motivation, most subjects continued work activity as directed. This is most likely due to innately high motivation to preserve the upkeep of the

inner resources of an individual. Moreover, most symptoms completely remitted after the period of convalescence at the end of the 6 months. Depressed mood remitted first, followed by a brief period of euphoria and remission of the asthenic phenomena.

The vital signs were consistent with increased sympathetic tone during the period of initial asthenia and with vagal parasympathetic tone during the late stage of depressive disorder. Heart rate values were 78 ± 2.1 and 63 ± 1.8, respectively ($p < 0.001$).

Other significant studies have been conducted to investigate environmental effects on the human psychological state. Thus, in 1991, Krasnov and his colleagues studied changes in mental state in the personnel of a large airport. In this study, 402 individuals were examined using an anonymous questionnaire in order to determine the prevalence of psychiatric disorders. Stable and recurrent asthenic and psychovegetative disorders were identified in 207 subjects (51.5%). The most common symptoms included headaches, irritability, problems with concentration and attention, eye strain, daytime sleepiness, increased meteorotropic sensitivity, episodic tachycardia, arterial blood pressure changes and chest pain. These findings do not necessarily indicate separate clinical entities, but seem to be the result of work stress and psychophysiological adaptation. Most interviewed subjects revealed no family conflict, felt good about their work and the workplace and had no other apparent life stressors. At the same time, the significant symptoms appeared to be considerably more prevalent among air traffic controllers who are exposed to chronic noise and desynchronosis compared to the rest of the group.

Another example of a protracted and persistent mental disorder related to a specific environmental exposure is mental and psychosomatic syndrome in rescue team workers who dealt with the Chernobyl disaster aftermath in 1986 (Krasnov 1993a,b). The extreme psychological stress was combined with exogenous toxins and produced a peculiar clinical syndrome. The syndrome of the so-called liquidators is characterized by its symptomatic heterogeneity and multiple patho-aetiologies with the final amalgamated clinical presentation of asthenia, somatoform symptoms, dysthymia and cognitive dysfunction combined with vegetative dysregulations, psychosomatic disorders (e.g. gastritis, ulcer, asthmatic bronchitis, arthritis, arterial hypertension) in conjunction with vegetative-vascular dysregulation (varying from arterial hypotonia to hypertension), and not infrequently also with combined somatic dysfunctions (cardiovascular, gastrointestinal, osteomuscular, metabolic) and non-specific neurological disturbances of the dyscirculatory encephalopathy type. This syndrome developed a few years after the disaster, and affected individuals usually present with non-psychotic symptoms but have a tendency to develop a psycho-organic syndrome. Primary radiation disease may be ruled out, since most individuals were exposed to relatively low radiation doses (0.1–0.3 Gy), while radiation disease develops at irradiation in doses of not less than 1 Gy for a short period of time. Nevertheless, in the general context of harmful effects (e.g. aggressive components of decontamination agents, fluctuations in the external temperature around miners digging a tunnel under the destroyed reactor, desynchronosis, psychological stress), we cannot ignore the role of low and moderate doses of radiation in the development of the conditions discussed above.

5 Conclusion

In conclusion, psychiatric phenomena arising under extreme environmental conditions can be either normative pseudopathological adaptive mechanisms or bona fide psychiatric disorders with complicated long-term biopsychosocial sequelae. Pseudopathological normative adaptation phenomena may include sensory distortions, personification of inanimate objects, vivid mental images, illusions and even visual hallucinations of one's double, whereas reality testing is always completely reserved. Such phenomena appear to play an adaptive role in accommodating an individual to an extreme condition and tend to remit under normal circumstances. Pre-flight training of astronauts in Russia shows that presence of such phenomena during the training period usually prognosticate a swift adaptation to real flight conditions (Lebedev 1989).

On the other hand, psychiatric disorders do develop under unusually severe environmental conditions. According to ICD-10 (World Health Organization 1993) and DSM-IV (American Psychiatric Association 1995), such disorders are usually adjustment disorders, acute stress disorders and post-traumatic stress disorders. The new complex polymorphous disorders of multiple aetiologies such as the Chernobyl disaster victims syndrome have yet to be formally classified (Krasnov 1995).

6 References

American Psychiatric Association (1995) Diagnostic and statistical manual of mental disorders, 4th edn (DSM-IV). International version with ICD-10 codes. American Psychiatric Association, Washington, DC

Andresen B, Stark FM, Gross J (eds) (1993) Psychiatrie und Zivilisation. Humanistische Psychologie, Cologne

Bennet G (1972) Pilot incapacitation. Flight Intern 102: 569–771

Berry CA (1970) Summary of medical experience in the Apollo VII through II spaceflights. Aerospace Med 5: 500–519

Bexton WN, Heron W, Scott TH (1954) Effects of decreased variation in the sensory environment. Can J Psychol 8: 70–76

Bombard A (1958) Naufrage volontaire. Edition de Paris, Paris

Burns NM, Kimura D (1963) Isolation and sensory deprivation. In: Burns NM et al (ed) Unusual environments and human behavior. Physiological and psychological problems of man in space. Macmillan, London, 167–192

Byrd RE (1932) Alone. Putnam, London

Cooper B (1993) Single spies and batallions: the clinical epidemiology of mental disorders. Psychol Med 23(4): 891–907

Daniel TC, Vining J (1983) Methodological issues in the assessment of landscape quality. In: Altman I, Wohlwill JF (eds) Human behaviour and environment. Advances in theory and research. 6. Behaviour and the natural environment. Plenum, New York

Davis JM, McCourt WE, Solomon P (1960) The effect of visual stimulation on hallucination and other mental experiences during sensory deprivation. Am J Psychiatry 20:889–892

Gunderson EK (1973) Psychological studies in Antarctica. In: Edholm OG, Gunderson EK (eds) Polar human biology. Heinemann, London, pp 352–361

Hauty GT, Adams T (1966) Phase shifts of the human cicadian system and performance deficit during the periods of transition. Aerospace Med 37: 668–674

Havenaar J, Rumyantzeva G, van den Brink W et al (1997) Long-term mental health effects of the Chernobyl disaster: an epidemiologic survey in two former Soviet regions. Am J Psychiatry 154(10): 1605–1607

Heron W, Doane BK, Scott TH (1956) Visual disturbances after prolonged perceptual isolation. Can J Psychol 10: 13–18

Heyerdahl T (1959) The Kon-Tiki expedition. Allen and Unwin, London

Kokhanov V, Krasnov V (1986) The course of neuropsychic disorders developing in conditions of occupational activity (in Russian). Zhurn Nevropat Psychiatr 86: 1681–1684

Korolenko C, Segal B, Donskikh T (1991) Ecological psychiatry. The need for the formation of a new direction in science (in Russian). In: Piven B (ed) Ecological psychiatry. Altai Medical Institute, Barnaul, pp 4–12

Krasnov V (1991) Role of factors of industrial environment in development of some neuropsychiatric disorders (in Russian). In: Piven B (ed) Ecological psychiatry. Altai Medical Institute, Barnaul, pp 13–17

Krasnov V (1995) Ecological psychiatry (in Russian). In: Krasnov V, Gurovitch I (eds) Proceedings of the XII congress of Russian psychiatry. Russian Society of Psychiatrists, Moscow, pp 158–160

Krasnov V, Yurkin M, Voitsekh V et al (1993a) Mental disturbances in liquidators of the aftermath of the Chernobyl disaster (in Russian). Soc Clin Psychiatry 3(1): 5–10

Krasnov V, Petrenko B, Voitsekh V et al (1993b) Mental disturbances in liquidators of the aftermath of the Chernobyl disaster. 2. Clinical-pathogenetic and pathoplastic relationships (in Russian). Soc Clin Psychiatry 3(4): 6–20

Krasnov V, Yurkin M, Petrenko B et al (1995) Mental disorders in the participants cleaning up the consequences of the Chernobyl nuclear power plant accident: structure, pathogenesis and approaches to therapy. In: Nyagu A (ed) Mental health consequences of the Chernobyl disaster: current state and future prospects. Proceedings of the international conference in Kiev. Association of Physicians of Chernobyl, Kiev, pp 102–103

Law P, Bechervaise J (1957) Australia's Antarctic outposts. Oxford University Press, Melbourne

Lebedev V (1989) Personality in extreme conditions (in Russian). Politizdat, Moscow

Leonov AA, Lebedev VI (1971) Space and perception by the cosmonaut. Mir, Moscow

Lindsley DB (1972) Human factors in long-duration space flight. National Academy of Sciences, Washington

Mendelson JH, Kubazansky P (1961) Physiological and psychological aspects of sensory deprivation. In: Solomon P (ed) Sensory deprivation. Harvard University Press, Harvard, pp 91–113

Mullin CS, Connery JM (1959) Psychological study at an Antarctic IGJ station. US Armed Forces Med J 19: 290–296

Nyagu A, Loganovsky K (1997) Neuropsychiatric effects of ionizing radiation. Chernobylinterinform, Kiev

Offerlend A, Roos B (1969) The human factor. Shipwrecks and other accidents to ships. Br J Prev Soc Med 14: 49–54

Ritter C (1900) A woman in the polar night. Century, New York

Rivolier J, Goldsmith R, Lugg D, Taylor A (eds) (1988) Man in Antarctic. The scientific work of the international biomedical expedition on the Antarctic. Taylor and Frances, London

Sandal GM, Vaernes R, Bergan T (1996) Psychological reactions during polar expedition and isolation in hyperbar chambers. Aviat Space Environ Med 67(3): 227–233

Semke V (1992) Ecological psychiatry: the present and the future (in Russian). Soc Clin Psychiatry 2(3): 5–13

Serxner JL (1968) An experience in submarine psychiatr. Am J Psychiatry I: 25–30

Sharpe M (1969) Living in space. Astronaut and his environment. Doubleday, New York

Simon GE, Katon WS, Sparks PJ (1990) Allergic to life: psychological factors in environmental illness. Am J Psychiatry 7: 901–906

Slocum J (1956) Sailing alone around the world. Dover, New York

Takla NK, Koffman R, Bailey DA (1994) Combat stress, combat fatigue and psychiatric disability in aircrew. Aviat Space Environ Med 65(9): 858–865

Tenfiord OW (1966) Mental diseases among Norwegian seamen. Bull Inst Marit Trop Med Gdynia 17: 373–382

Viinamäki H, Kumpusalo E, Myllykangas M et al (1995) The Chernobyl accident and mental wellbeing – a population study. Acta Psychiatr Scand 91: 396–401

Weybrew BB (1963) Psychological problems of prolonged marine submergence In: Burns N, Chambers R, Hendler E (eds) Unusual environments and human behavior. Free Press of Glencoe, Collier-Maemill, London, pp 87–125

World Health Organization (1993) The ICD-10 classification of mental and behavioural disorders, diagnostic criteria for research. World Health Organization, Geneva

CHAPTER

17

E.J. Bromet

Psychiatric Problems Related to Natural and Human-Made Disasters

1 Introduction

Extreme events, such as natural and human-made disasters, often have short- and long-term psychological impacts that far exceed the degree of medical morbidity and mortality that ensues. Indeed, Lechat (1979) defined a disaster as a "disruption exceeding the adjustment capacity of the affected community" (p. 11). Since World War II, a number of studies have assessed the emotional consequences of a variety of natural disasters, such as hurricanes, tornadoes, floods, volcanic eruptions, and earthquakes, as well as technological and human-made catastrophes, such as the nuclear power plant accidents at Three Mile Island (TMI) in the United States and Chernobyl in Ukraine (Bromet and Dew 1995; Havenaar et al. 1996; Raphael 1986; Weisæth 1993), the Nazi Holocaust (Robinson et al. 1994), and the Cambodian massacre (Kinzie et al. 1989). Studies aimed at describing the prevalence of psychological impairment or disorders in populations exposed to disasters are best conceptualized as epidemiologic in nature. Within this rubric, they represent a special case of the classic cohort paradigm, focusing on incidence, prevalence rates, risk factors associated with onset, and short- and long-term impairment.

Before 1980, beliefs about the effects of disasters on mental health were based primarily on case reports or data from surveys of nonrepresentative and/or small samples. The study groups were typically samples of convenience, such as disaster victims seeking monetary compensation, injured persons, litigants, or volunteers. Such research was often criticized for having poor response rates, no control group, nonstandardized assessment tools, and no conceptual framework guiding the design, measurement, and analysis. With the 1980 publication of the third edition of the *Diagnostic and Statistical Manual* (DSM-III; American Psychiatric Association 1980), psychiatric epidemiology began to adopt new methodologies to assess mental disorders and symptom complexes, and disaster research soon followed suit. This entailed the widespread use of standardized diagnostic and subclinical measures in disaster research, such as the Diagnostic Interview Schedule (DIS; Robins et al. 1981), the General Health Questionnaire (Goldberg 1972), the Symptom Checklist-90 (Derogatis 1977), and the Impact of Events Scale (Horowitz 1990). For many of these instruments, normative population data are available. In addition, with the introduction of operational criteria for post-traumatic stress disorder (PTSD) and structured interview guides, a number of studies began to examine the occurrence and persistence of PTSD and post-traumatic stress symptoms in disaster-exposed compared to control populations. Greater attention has also been given to including representative samples and a comprehensive array of risk and protective factors. However, with some notable exceptions, such as the TMI nuclear power plant accident (Dew and Bromet 1993) and the Buffalo Creek flood (Green et al. 1990, 1994), the majority of disaster studies continue to be cross-sectional, assessing adjustment at a single, often arbitrary, point in time within 1 month to 3 years postdisaster. We therefore have a somewhat limited understanding of the evolution of symptoms over time and the extended long-term effects of natural and human-made disasters.

While epidemiologic designs provide a more rigorous methodology for understanding community-wide stress, pre-event baseline data are unavailable in most studies. The exceptions can be thought of as "natural experiments." Such exceptions have occurred in situations in which a population was studied for some other purpose prior to a catastrophic occurrence, and researchers were able to conduct a prospective follow-up and identify predictors of postdisaster adjustment. For example, data were collected for a psychiatric epidemiologic study modeled after the Epidemiologic Catchment Area (ECA) study (Robins and Regier 1991) in Puerto Rico in 1984 (Canino et al. 1987). In 1985, torrential rains hit the island, causing extensive mudslides and leaving 180 people dead, 4000 in shelters, and 19,000 with serious property damage. In 1987, the investigators reevaluated a group of disaster survivors ($n = 77$) and controls ($n = 298$) using a Spanish-language version of the DIS Disaster Supplement (Robins and Smith 1983). The results showed that, after adjusting for demographic variables and previous symptoms, depressive, somatic, and PTSD symptoms were significantly, albeit modestly, increased after the disaster (Bravo et al. 1990; Canino et al. 1990; Escobar et al. 1992). A methodologically different example of a "natural experiment" occurred in conjunction with a major earthquake in Italy. Trevisan et al. (1992) were in the midst of conducting a 5-year follow-up of workers participating in a coronary heart disease risk factor study when the earthquake struck. Compared to workers whose evaluations were completed before the earthquake, those whose examinations were performed afterward had significantly higher heart rates, serum cholesterol levels, and triglyceride levels. However, these elevations were not seen at the next follow-up assessment 7 years later.

Because of the uniqueness of each disaster and the different methodologies employed in each study, it is difficult to summarize the findings across the diverse events that have occurred. Moreover, the methodolo-

gies contain numerous sources of bias that constrain the inferences that can be drawn about symptoms and behaviors. The sources of bias include selective mortality (only survivors are assessed), difficulties assembling a representative sample when the disaster victims are scattered and subsequently relocated, difficulty gaining cooperation when populations are under stress and/or angry about delays or inadequacies in relief programs (Frederick 1980), and recall bias or faulty memory. Disaster studies are also subject to interviewer bias when the interviewers themselves reside in the affected areas. Finally, whether consciously or not, investigator bias can arise in analyzing data and interpreting findings either because investigators are influenced by how they themselves coped with their own life-threatening experiences or because of constraints imposed by the agencies sponsoring the research. The sources of variability in disaster research can be summarized as follows:

1. Disaster characteristics
   a) Natural versus human-made
   b) Suddenness of onset
   c) Duration of impact
   d) Extent of damage
   e) Number of deaths
   f) Evacuation
   g) Relocation
   h) Potential for recurrence
   i) Organizational response
   j) Sociocultural context
2. Methodological issues
   a) Sample differences (general population, litigants, help-seekers)
   b) Low response rates
   c) Interviewer bias
   d) Recall bias
   e) Timing of assessments
   f) Choice of measures
   g) Cross-sectional versus longitudinal

This chapter emphasizes findings on mental health effects from epidemiologic research, drawing mostly from studies using systematic methods for sampling and assessment. Three issues are discussed: the psychological problems that have been identified in adults and children; the risk factors for adverse outcomes in these populations; and the personal, situational, disaster-related, and treatment intervention factors that can mediate potential deleterious effects. As in a previous review (Bromet and Dew 1995), this chapter gives more weight to studies using representative samples of 50 or more survivors, because multivariate analyses of such samples are statistically more reliable. Although the review is limited to studies published in English, the disasters themselves occurred all over the world.

2 Rates and Types of Psychological Impairment

The most detailed and statistically sophisticated analysis of postdisaster prevalence of psychological morbidity was performed by Rubonis and Bickman (1991) in a meta-analysis of 52 mental health studies connected with natural and technological disasters. They calculated that the average excess psychological morbidity was 17%. Similarly, Weisæth (1993) estimated that the 1-year prevalence of psychological morbidity following a disaster was about 20%. However, he noted that the rate could range to as high as 50%. Indeed, by treating all disasters as equal, the average rates obscure the rates of suffering of survivors of prolonged, intense disasters, such as the Chowchilla bus kidnapping (Terr 1983) and the Nazi Holocaust, which approach 100%.

The 20% figure given in the reviews published before 1995 reflect research conducted primarily in North America and Europe (Bromet and Dew 1995). In fact, many of the world's worst disasters occur in developing countries (Lima et al. 1990b; Weisæth 1993) and in the former Soviet Union. Recent studies published after these reviews appeared suggest that the average rate may be considerably higher. For example, the rates of psychiatric morbidity reported in recent studies of natural disasters in Sri Lanka, Colombia, and India were 75%, 55%, and 59%, respectively (Sharan et al. 1996).

The specific rate of impairment or disorder varies widely with the type and magnitude of the disaster, the measures chosen for study, the populations evaluated, and the timing of the data collection. Four studies of earthquake survivors illustrate this issue. In the first study, de la Fuente (1990) administered a questionnaire containing the DSM-III criteria for PTSD, generalized anxiety, and depression to victims housed in shelters after the 1985 earthquake in Mexico. The rates for men ($n = 163$) were 18% for PTSD, 9% for anxiety, and 7% for depression; the rates for women ($n = 395$) were 38%, 24%, and 16% for each disorder, respectively. In the second study, Wood et al. (1992) determined the rate of nightmares among students at Stanford University and San Jose State University after the 1989 San Francisco Bay Area earthquake by asking them to keep written records every morning for the 3-week period after the earthquake; 40% recorded one or more nightmares, compared to 5% of controls in Arizona. The third study, by Maj et al. (1989), assessed survivors from three Italian communities who had contacted their general practitioners for physical health problems after an earthquake. They used a two-stage design, in which the 30-item Italian version of the General Health Questionnaire (Goldberg 1972)

was administered in the first stage, and the Comprehensive Psychopathological Rating Scale and ICD-9 were completed on the basis of a psychiatric evaluation in the second stage. The reported prevalence rates of psychiatric disorder (mostly neuroses) for the three communities ranged from 41.9% to 54.4%. The last study, by Goenjian (1993), assessed the effects of the 1988 earthquake in Armenia. Psychiatric evaluations were conducted on 582 individuals who had sought mental health treatment 3–6 months after the earthquake. Based on DSM-III-R (revised) criteria, 74% were diagnosed with PTSD and 22% with major depressive disorder. Thus, although all four studies focused on victims of major earthquakes, they differed substantially with respect to the magnitude, socioeconomic, and sociocultural contexts of each event, the study design (one-stage versus two-stage, types of samples, timing of data collection), and measurement (type of symptoms, approach to measurement). While the rates of psychological morbidity were clearly elevated across all the studies, it is very difficult to assign a precise numerical value, or even a reasonable range, from studies with such extensive methodological variation.

The most frequently reported symptoms in adults in the aftermath of natural disasters are somatic complaints, depression, anxiety, and post-traumatic stress symptoms, particularly intrusive and avoidant symptoms. The clustering of such symptoms is sometimes referred to as the "disaster syndrome." Escobar et al. (1992) used the phrase "disaster-reactive psychopathological repertoire" (p. 966) to describe such symptoms. It has been argued that these symptoms do not reflect separate disorders (Deering et al. 1996) but rather represent "complex somatic, cognitive, affective and behavioral effects of psychological trauma" (van der Kolk et al. 1996, p. 90).

Significant elevations of these symptom domains have been reported in studies of adult survivors of earthquakes (e.g. Cardena and Spiegel 1993; Carr et al. 1997; de la Fuente 1990; Freedy et al. 1994; Goenjian 1993; Goenjian et al. 1994; Karanci and Rustemli 1995; Maj et al. 1989; Murphy 1988; Nolen-Hoeksema and Morrow 1991; Sharan et al. 1996; Wood et al. 1992), floods (e.g. Cook and Bickman 1990; Kaniasty and Norris 1993; Ollendick and Hoffmann 1982), hurricanes (Norris and Uhl 1993), volcanic eruptions (Shore et al. 1986a,b), mudslides (Canino et al. 1990; Escobar et al. 1992), cyclones and tornadoes (Madakasira and O'Brien 1987; Milne 1977a; Parker 1977; Patrick and Patrick 1981; Penick et al. 1976), and a mass food-poisoning epidemic (toxic rape-seed oil; Lopez-Ibor et al. 1985). Elevations in these same symptom domains, as well as aggressive behavior and enuresis, were also reported in studies of children exposed to floods (Durkin et al. 1993), hurricanes (e.g. Garrison et al. 1995; Hardin et al. 1994; Lonigan et al. 1994; Shannon et al. 1994), cyclones and tornadoes (Bloch et al. 1956; Milne 1977b), a bushfire (McFarlane et al. 1987), and a blizzard (Burke et al. 1982; girls only: Burke et al. 1986).

These same symptom domains have also been shown to be elevated following human-made catastrophes (Bromet 1989). In fact, one of the earliest descriptions of post-traumatic stress syndrome in a civilian population was the detailed report of Hiroshima survivors by Lifton (1967) based on a series of in-depth interviews he conducted in Japan in 1962. Strictly speaking, the generalizability of the findings is open to question because of both the unstructured format of the interview and the selected nature of the sample, i.e. survivors from an unspecified population pool willing to talk at length about highly personal matters with a psychoanalytically oriented American psychiatrist. Nevertheless, Lifton's descriptions of psychic numbing (originally referred to as "psychological closure"), i.e. the loss of feeling from the extreme agony that resulted from witnessing mass death and dying and being unable to respond to calls for help, as well as other cardinal symptoms of intrusive and avoidant thinking, became criterion symptoms for the classification of PTSD in current nosological systems.

There are two unique mental health findings from studies of technological events that distinguish their aftermath from that of natural disasters: the protracted nature of the mental health problems and the fear of contracting an exposure-induced disease. With respect to its enduring effects, long-term follow-up studies of Nazi concentration camp survivors show persistent psychiatric and somatic symptoms enduring through the survivors' lifetime in spite of adequate occupational and family role functioning (Robinson et al. 1994). Even relatively lower impact events, such as the accident at TMI and the Buffalo Creek dam collapse (Green et al. 1990), led to long-term sequelae. For example, Baum and Fleming (1993) conducted a 6-year, longitudinal panel study of a carefully selected sample from the 5-mile radius of the TMI nuclear power plant in central Pennsylvania. The comparison groups included residents living near another nuclear reactor, near a coal-fired plant, and more than 5 miles from any power plant. The TMI group had significantly higher rates of a very wide range of symptoms, including somatic complaints, anxiety, depression, hyperarousal, intrusive thoughts about the accident, avoidance of reminders of it, elevated blood pressure, and elevated levels of urinary norepinephrine, epinephrine, and cortisol. In a similar vein, the Bromet study of mothers of young children living within 10 miles of TMI found that depression and anxiety symptoms were persistently elevated up to 10 years

after the accident, with 35% of the women exhibiting consistently high levels of distress over the period of observation (Dew and Bromet 1993).

In addition, technological events, particularly those that involve radiation exposure (Slovic 1991), lead to another enduring symptom, fear of acquiring cancer or another exposure-related disease. This fear is reinforced when cancers and other illnesses occurring among survivors are attributed, rightly or wrongly, to the exposure. For example, Misao and colleagues (1961) showed that the level of anxiety about developing "A-bomb" diseases was 75% higher in two large samples of Nagasaki survivors compared to controls. It is interesting to note that the persistent fears of radiation disease in people affected by Chernobyl have been given the disparaging label of "radiophobia" (Ginzburg and Reis 1991).

In sum, at least 20% of an exposed population will exhibit symptoms of anxiety, depression, somatization, and/or PTSD, and in some cases these symptoms may persist for years. The rates of disaster-related psychopathology range from 20% to 50%. However, there are some important risk factors associated with its expression. The next section describes both personal and disaster-related risk factors.

3 Risk Factors

Three sets of factors affect adjustment following natural or human-made disasters (Davidson and Foa 1993; Freedy et al. 1993; Table 1): (1) predisaster factors, including gender, psychiatric history, and prior traumatic stressors, (2) within-disaster factors, such as degree of exposure to horror and death, and cognitive appraisal of control, and (3) postdisaster experiences occurring during the recovery process, including informal and formal crisis support.

Table 1. Psychiatric risk factors

Period	Risk factor
Predisaster	Female sex
	Prior psychiatric history
	Prior traumatic experiences
During disaster	Severity of exposure
	Death of loved one or close friend
	Physical threat
	Perceived lack of control
Postdisaster	Inadequate practical support
	Inadequate emotional support
	Inadequate professional intervention

With respect to predisaster variables, a number of studies have shown that women have higher rates of disaster-related psychopathology than men. For example, Shore et al. (1986b) reported that, after the Mount St. Helens (northwestern United States) volcanic eruption, the 1-year onset rates of generalized anxiety, major depression, and/or PTSD were approximately twice as high in exposed women (20.9%) as in men (11.1%). Garrison et al. (1995) also found higher rates of PTSD in adolescent girls during the first few months after Hurricane Andrew (in the southeastern United States) compared to boys (9.2% versus 2.9%). It should be emphasized, however, that gender differences have not been reported consistently across disaster studies (Gibbs 1989), even though in epidemiologic research PTSD, depression, and anxiety are more prevalent in women than in men (e.g. Gruen 1993; Kessler et al. 1994, 1995). To some extent, the gender differences in disaster studies may also be due to the types of measures administered. Although women are more likely to endorse affective and anxiety symptoms, men have higher rates of substance-related problems and disorders (Kessler et al. 1994). Substance problems are often not recorded in disaster studies, even though it is well established that, in men, PTSD is comorbid with substance use disorders. Their consideration might provide a more balanced view of how men and women react to disasters.

A life history of mental disorder, or predisaster mental health status, represents a well-documented risk factor in disaster as well as other types of outcome research and constitutes a crucial vulnerability factor in most stress models (Bromet and Dew 1995; Bromet and Schulberg 1987; Solomon 1992; Gruen 1993; Rabkin 1993). For example, Nolen-Hoeksema and Morrow (1991) studied depressive and post-traumatic stress symptoms in Stanford University students 10 days and again 7 weeks after the Bay Area earthquake. A subsample of them had completed a similar battery of questionnaires 2 weeks before the earthquake. Pre-event symptoms proved to be among the strongest predictors of postevent symptomatology. In our TMI research, preaccident depression and/or anxiety disorder based on the Schedule for Affective Disorders and Schizophrenia–Lifetime (SADS-L) interviews and research diagnostic criteria (RDC) was the most important predictor of depression and anxiety disorders over the 1 year following the accident. In a follow-up of victims of a paint factory fire, Weisæth (1989) showed that, while acute PTSD was linked to the initial intensity of the exposure, outcome 4 years later was influenced more by pre-exposure psychological functioning. Similarly, McFarlane (1988) found a significant association between a history of psychiatric disorder and chronic PTSD in a large group of firefighters in Australia assessed after a major

bushfire. Finally, studies of PTSD in the general population have also found that prior history of psychopathology is a significant risk factor (Breslau et al. 1991; Bromet et al. 1998; Davidson et al. 1991).

Another important set of predispositional risk factors is prior exposure to traumatic life experiences and enduring high levels of strain at the time of the event. The evidence suggests that individuals with prior traumas are more vulnerable to the impact of a natural or human-made disaster (Carr et al. 1997). Indeed, the general population studies of PTSD have also consistently reported that prior traumatic experiences are associated with increased rates of PTSD following subsequent trauma exposure (Breslau et al. 1991; Bromet et al. 1998; Davidson et al. 1991). In general, epidemiologic studies have shown that individuals exposed to both ongoing strain and an acute stressor will be at increased risk for adverse mental health outcomes (Turner et al. 1995). It should be noted that a debate exists as to whether prior exposure to trauma is a risk factor or a protective factor (stress inoculation theory). The weight of the evidence thus far implicates prior exposure as a vulnerability factor.

Far less is known about the effects of predisaster risk factors on the psychological morbidity of children (March and Amaya-Jackson 1993). Recent reviews of the available evidence suggest that the level of distress increases with the child's age and with pre-event mental and physical health problems of the child or his or her parents (Aptekar and Boore 1990; Saylor 1993; Vogel and Vernberg 1993), although the results have not been entirely consistent. For example, in a study of American children exposed to Hurricane Hugo in the southeast United States, Lonigan et al. (1994) found that younger rather than older children were more likely to develop PTSD-like reactions. In contrast, March et al. (1997) did not find that age was a risk factor for post-traumatic symptomatology after the industrial fire in the same region.

It has also been hypothesized that mothers' response to a disaster is not only an important risk factor for the child (Bloch et al. 1956), but for some events, it may be a better predictor of post-traumatic phenomena in young children than direct exposure itself. The extent to which exposure factors, rather than parental reaction, directly predict the psychological response may depend on the age of the children and on the degree of their direct involvement in the disaster. For example, in the study by Pynoos et al. (1987) of elementary school children in the Los Angeles area exposed to a fatal sniper attack on their playground, proximity to the violence was the most important predictor of type and number of PTSD symptoms. In a subsequent study of the Armenian earthquake, Pynoos et al. (1993) demonstrated a similar correlation between proximity to the epicenter of the quake and severity of children's post-traumatic stress reactions. March et al. (1997) showed a clear dose–response effect, with the greatest PTSD symptoms reported by children who witnessed an industrial fire and had a relative or friend hurt or killed, followed by those with a relative or friend who was hurt or killed (only), those who witnessed the event (only), and the lowest level among those with neither exposure. On the other hand, several other disaster studies (Vogel and Vernberg 1993), including our TMI work on very young children (Cornely and Bromet 1986) and school-age children (Bromet et al. 1984) as well as the study by Laor et al. (1997) of preschool children after the Scud missile attack on Tel Aviv, found that mothers' response was the more significant factor. Thus the age of the sample and the nature of the event will influence the findings on risk factors in children.

Turning to the characteristics of the disaster itself, the severity of the exposure is the primary risk factor for the development of postdisaster psychiatric morbidity (Bromet and Schulberg 1987; Green 1994). The intensity derives from the magnitude of the destruction (realized or anticipated as in the case of a nuclear power plant accident), stressors incurred during the evacuation and/or relocation (Milne 1977a), sudden death of a loved one, severe physical harm, cognitive appraisal of lack of control, and predictability and personal threat or worry over future similar events (Freedy et al. 1993; Green 1994). This pattern is illustrated by findings from a study of survivors of a rail accident in England, showing that those who were trapped, witnessed death, or felt at risk of death experienced more PTSD symptoms than other survivors (Selley et al. 1997). In general, the more severe the exposure, the greater the likelihood of psychological impairment in the wake of such events.

Perhaps the most powerful and enduring of the stressors occurring in the immediate aftermath of a disaster is the untimely death of relatives and close friends. Kohn and Levav (1990) commented that, in this circumstance, grieving may be complicated by the other aspects of personal chaos with which disaster victims must cope. As noted earlier, perhaps the most graphic and poignant description of survivors' coping with death and dying is Lifton's (1967) descriptions of the survivors of Hiroshima. Specifically, he vividly depicted the loss of feeling survivors experienced from the extreme agony of witnessing mass death and dying and being unable to respond to calls for help, a phenomenon referred to currently as psychic numbing. This phenomenon and other cardinal symptoms Lifton described have also been observed in other disaster victims and, as noted above, ultimately became criterion symptoms for the classification of DSM-III PTSD. The concepts of intrusive and avoidant thinking also were operationalized in the Impact of

Events Scale (e.g. Horowitz 1990), which has been widely used in disaster research to capture aspects of the disaster syndrome.

In the postdisaster recovery period, practical and perceived support have important relationships to mental health outcomes (Fleming et al. 1982). The theory from stress research, based primarily on studies of individual life event experiences, is that social support should serve to buffer the adverse effects of stress (in this case, from a disaster) on mental health. However, social support can also be considered an aspect of the recovery environment (e.g. Green 1995) or a dynamic variable that can itself be affected by a disaster (Kaniasty and Norris 1993). It is interesting to note that, in many studies, social support is treated as a trait measure (predispositional variable) that presumably is stable over time (Gibbs 1989). Adding to the conceptual confusion is the fact that the findings on the precise role of social support in disaster research (as well as life events research) have been inconsistent. In other words, a few studies have reported that social support buffered some aspects of postdisaster morbidity (e.g. Cook and Bickman 1990; Fleming et al. 1982), while others found direct effects rather than a buffering effect (e.g. Bromet et al. 1982; Murphy 1988). In addition, within a given study, the findings on social support have also been inconsistent across demographically defined risk groups. For example, in a study of the Exxon Valdez oil spill in Alaska, Palinkas et al. (1992) found that perceived family support buffered the effects of exposure on depressive symptoms (assessed by the Center for Epidemiologic Studies Depression scale; Radloff 1977) in Euro-Americans but not in Native Alaskans. In a similar vein, Solomon et al. (1987) reported that spouse support reduced disaster-related symptoms in exposed men but not in women. In our TMI research, we found no evidence of a buffering role for social support in the sample of mothers, but found that a positive family milieu buffered the effects of TMI stress on the current mental health of school-age children (Bromet et al. 1984).

The importance of understanding how different aspects of social support operate in various demographically defined groups cannot be underestimated given its potential role in alleviating the deleterious effects of stress caused by disasters. There is growing evidence that "crisis support" (e.g. people who listen and provide practical and emotional support) promotes aspects of psychological recovery after disasters. Both naturalistic follow-up studies of disaster victims and clinical studies of traumatized patients support the therapeutic value of crisis support. The theory is that repeated discussion about the traumatic memories lessens the feelings of fear and anxiety attached to them. Consistent with this formulation, Dalgleish et al. (1996) showed that greater crisis support when the Herald of Free Enterprise ferry sank predicted fewer avoidance symptoms at 6-year follow-up.

Another consequence of many disasters is the creation of temporary shelters. Far less is known about the impact of temporary settlements on mental health, even though many recent disasters have produced large refugee populations interned in camps without adequate sanitary conditions, nutrition, privacy, or safety (Bromet 1996; Jackson 1991). Survivors of Hiroshima and Nagasaki and those evacuated from the area around Chernobyl were also subjected to social stigma by both health care professionals and lay members of the community. This aspect of the postdisaster experience has not been adequately studied or discussed, although existing data from some interned populations would suggest that the rate of PTSD stemming from this experience might be substantial (Bromet 1996).

4 Implications and Conclusions

This review makes it clear that psychiatric sequelae occur among significant proportions of affected populations and that these clinical manifestations may endure for long periods of time. While some have argued that the psychological effects of disasters are negligible (Perry and Lindell 1978; Quarantelli 1985) or short-lived (e.g. Milne 1977b), the majority of studies conducted since the introduction of DSM-III show that disasters engender substantial morbidity, particularly during the first few years (e.g. McFarlane 1993; Rubonis and Bickman 1991). Moreover, in some instances, such as disasters involving prolonged persecution or radiation exposure, the level of morbidity either did not decrease or indeed increased with time (Chamberlin 1980; Lima et al. 1990a; McFarlane et al. 1987). Logue et al. (1981a,b) lamented more than 15 years ago that few empirical studies of disasters utilized epidemiologic methods. As we have seen, there has been considerable progress in the design of psychiatric disaster research, and the findings on high levels of morbidity have been confirmed in a number of disaster populations.

Unfortunately, there is a dearth of disaster studies with longitudinal designs or extended, long-term follow-up. Thus, while high-risk groups in need of short-term services can be identified, we still have little knowledge of the relationship between short-term and long-term adjustment. Specifically, we have little knowledge of the evolution of disaster-related symptoms and syndromes and the risk factors associated

with course (as opposed to onset) of these disorders in men and women of different ages and different socioeconomic and national backgrounds. Conversely, it is imperative that we learn more about the variables that promote health and protect against adverse mental health outcomes after disasters. Such knowledge can then be used in the formulation of potentially successful interventions.

Several types of intervention strategies following disasters have been implemented. These include public information, community education, crisis counseling, individual and family outreach, and recovery counseling (Lebedun and Wilson 1989). The World Health Organization's Division of Mental Health provides training material and professional services when disasters strike, with the long-term goal of increasing self-reliance in coping with the needs of disaster victims (World Health Organization 1992). Other specific materials are also available from international agencies (see Danieli et al. 1996 for available resources).

The role of the mental health professional in disaster management programs has traditional and nontraditional elements. Such roles can range from assisting with public education and media responses to bolstering indigenous support networks, supporting family members who have lost a loved one, and working with the Red Cross and other relief organizations to provide psychological first aid (Green and Lindy 1994). In addition, new forms of psychological and pharmacotherapeutic treatments are being developed and tested with survivors of trauma (Brady 1997). It should be emphasized that the complex nature of disasters (disasters involve multiple factors acting simultaneously, including physical exhaustion and physical illness, dislocation, unavailability of basic resources) and disaster-induced psychopathology (e.g. comorbidity of PTSD, depression, anxiety, substance use disorders, and somatization, along with the chronicity of these disorders) challenge the design of clinical interventions. Research is currently being conducted on a number of issues. For example, should symptom-specific treatments be administered sequentially or simultaneously? Should medications be combined with cognitive-behavioral treatment from the outset, or should they be added as needed? How effective are the different modalities for children, e.g. debriefing, pharmacological interventions, individual psychotherapy, family therapy (Vernberg and Vogel 1993)? It is clear though, based on the retrospective data from the National Comorbidity Survey (Kessler et al. 1995), that professional intervention can shorten the duration of symptoms by about 2 years.

Finally, as Vernberg and Vogel (1993) note, prevention of psychological trauma requires rapid and appropriate crisis responses from the mental health community. The U.S. military established four principles for treating psychiatric casualties: immediacy (early intervention), expectancy (the attitude that the patient will return soon to duty), simplicity (in the forms of treatment), and centrality (of facilities). Some conceptually similar, important new work on preventing the onset of PTSD is being tested in rape victims, in which they are educated about what symptoms to anticipate, taught relaxation techniques, taken through the traumatic experience, and given home work assignments (Foa 1997). Preliminary results suggest that this combination of straightforward techniques is effective in reducing the onset of PTSD symptoms. If it proves successful, this type of intervention has widespread utility for many disaster situations. While it is difficult, if not impossible, to prevent all psychiatric sequelae from occurring after disasters, our goal should be to reduce the prevalence and severity of the morbidity and the long-term adverse effects.

5 References

American Psychiatric Association (1980) Diagnostic and statistical manual of mental disorders, 3rd edn. American Psychiatric Press, Washington, DC

Aptekar L, Boore JA (1990) The emotional effects of disaster on children: a review of the literature. Int J Ment Health 19: 77–90

Baum A, Fleming I (1993) Implications of psychological research on stress and technological accidents. Am Psychol 48: 665–672

Bloch DA, Silber E, Perry SE (1956) Some factors in the emotional reaction of children to disaster. Am J Psychiatry 113: 416–422

Brady KT (1997) Posttraumatic stress disorder and comorbidity: recognizing the many faces of PTSD. J Clin Psychiatry 58 [Suppl 9]: 12–15

Bravo M, Rubio-Stipec M, Canino GJ, Woodbury MA, Ribera JC (1990) The psychological sequelae of disaster stress prospectively and retrospectively evaluated. Am J Comm Psychol 18: 661–680

Breslau N, Davis GC, Andreski P, Peterson E (1991) Traumatic events and posttraumatic stress disorder in an urban population of young adults. Arch Gen Psychiatry 48: 216–222

Bromet EJ (1989) The nature and effects of technological failures. In: Gist R, Lubin B (eds) Psychosocial aspects of disasters. Wiley, New York, pp 120–139

Bromet EJ (1996) Impact of trauma. Curr Opin Psychiatry 9: 153–157

Bromet E, Dew MA (1995) Review of psychiatric epidemiologic research on disasters. Epidemiol Rev 17: 113–119

Bromet EJ, Schulberg HC (1987) Epidemiologic findings from disaster research. In: Hales R, Frances A (eds) American Psychiatric Association annual review, vol 6. American Psychiatric Press, Washington, DC, pp 676–689

Bromet EJ, Parkinson D, Schulberg H, Dunn L, Gondek P (1982) Mental health of residents near the Three Mile Island reactor:

a comparative study of selected groups. J Prev Psychiatry 1: 225–276

Bromet EJ, Hough L, Connell M (1984) Mental health of children near the Three Mile Island reactor. J Prev Psychiatry 2: 275–301

Bromet EJ, Sonnega A, Kessler RC (1998) Risk factors for DSM-III-R posttraumatic stress disorder: findings from the National Comorbidity Survey. Am J Epidemiol 147: 353–361

Burke JD, Borus JF, Burns BJ, Millstein KH, Beasley MC (1982) Changes in children's behavior after a natural disaster. Am J Psychiatry 139: 1010–1040

Burke JD, Moccia P, Borus JF, Burns BJ (1986) Emotional distress in fifth-grade children ten months after a natural disaster. J Am Acad Child Psychiatry 25: 536–541

Canino G, Bird H, Shrout PR, Rubio-Stipec M, Bravo M, Martinez R, Sesman M, Guevara LM (1987) The prevalence of specific psychiatric disorders in Puerto Rico. Arch Gen Psychiatry 44: 727–735

Canino G, Bravo M, Rubio-Stipec M, Woodbury M (1990) The impact of disaster on mental health: prospective and retrospective analyses. Int J Ment Health 19: 51–69

Cardena E, Spiegel D (1993) Dissociative reactions to the San Francisco Bay area earthquake of 1989. Am J Psychiatry 150: 474–478

Carr VJ, Lewin TJ, Webster RA, Kenardy JA (1997) A synthesis of the findings from the Quake Impact Study: a two-year investigation of the psychosocial sequelae of the 1989 Newcastle earthquake. Soc Psychiatry Psychiatr Epidemiol 32: 123–136

Chamberlin BC (1980) MAYO seminars in psychiatry: the psychological aftermath of disaster. J Clin Psychiatry 41: 238–244

Cook JD, Bickman L (1990) Social support and psychological symptomatology following a disaster. J Trauma Stress 3: 541–555

Cornely P, Bromet E (1986) Prevalence of behavior problems in three-year-old children living near Three Mile Island: a comparative analysis. J Child Psychol Psychiatry 27: 489–49

Dalgleish T, Joseph S, Thrasher S, Tranah T, Yule W (1996) Crisis support following the Herald of Free-Enterprise disaster: a longitudinal perspective. J Trauma Stress 9: 833–845

Danieli Y, Rodley NS, Weisæth L (1996) International responses to traumatic stress. Baywood, Amityville

Davidson JRT, Foa EB (1993) Epilogue. In: Davidson JRT, Foa EB (eds) Posttraumatic stress disorder: DSM IV and beyond. American Psychiatric Press, Washington, DC, pp 229–235

Davidson JRT, Hughes D, Blazer DG, George L (1991) Posttraumatic stress disorder in the community: an epidemiological study. Psychol Med 21: 713–721

Deering CG, Glover SG, Ready D, Eddleman HC, Alarcon D (1996) Unique patterns of comorbidity in posttraumatic stress disorder from different sources of trauma. Compr Psychiatry 37: 336–346

de la Fuente R (1990) The mental health consequences of the 1985 earthquakes in Mexico. Int J Ment Health 19: 21–29

Derogatis LR (1977) The SCL-90 manual I: scoring, administration, and procedures for the SCL-90. Clinical Psychometric Research, Baltimore

Dew MA, Bromet EJ (1993) Predictors of temporal patterns of psychiatric distress during 10 years following the nuclear accident at Three Mile Island. Soc Psychiatry Psychiatr Epidemiol 28: 49–55

Durkin MS, Khan N, Davidson LL, Zaman SS, Stein ZA (1993) The effects of a natural disaster on child behavior: evidence for posttraumatic stress. Am J Public Health 83: 1549–1553

Escobar JI, Canino G, Rubio-Stipec M, Bravo M (1992) Somatic symptoms after a natural disaster: a prospective study. Am J Psychiatry 149: 965–967

Fleming R, Baum A, Gisriel MM, Gatchel RJ (1982) Mediating influences of social support on stress at Three Mile Island. J Hum Stress 8: 14–22

Foa EB (1997) Trauma and women: course, predictors, and treatment. J Clin Psychiatry 58[Suppl 9]: 25–28

Frederick CJ (1980) Effects of natural vs. human-induced violence upon victims. Evaluation Change (special issue): 71–75

Freedy JR, Kilpatrick DG, Resnick HS (1993) Natural disasters and mental health: theory, assessment, and intervention. J Soc Behav Pers 8: 49–103

Freedy JR, Saladin ME, Kilpatrick DG, Resnick HS, Saunders BE (1994) Understanding acute psychological distress following natural disaster. J Trauma Stress 7: 257–273

Garrison CZ, Bryant ES, Addy CL, Spurrier PG, Freedy JR, Kilpatrick DG (1995) Posttraumatic stress disorder in adolescents after Hurricane Andrew. J Am Acad Child Adolesc Psychiatry 34: 1193–1201

Gibbs MS (1989) Factors in the victim that mediate between disaster and psychopathology: a review. J Trauma Stress 2: 489–514

Ginzburg H, Reis E (1991) Consequences of the nuclear power plant accident at Chernobyl. Public Health Rep 106: 32–40

Goenjian AK (1993) A mental health relief programme in Armenia after the 1988 earthquake. Br J Psychiatry 163: 230–239

Goenjian AK, Najarian LM, Pynoos RS, Steinberg AM, Manoukian G, Tavosian A, Fairbanks LA (1994) Posttraumatic stress disorder in elderly and younger adults after the 1988 earthquake in Armenia. Am J Psychiatry 151: 895–901

Goldberg DP (1972) The detection of psychiatric illness by questionnaire. Oxford University Press, London

Green BL (1994) Psychosocial research in traumatic stress: an update. J Trauma Stress 7: 341–362

Green BL (1995) Long-term consequences of disasters. In: Hobfoll SE, de Vries MW (eds) Extreme stress and communities: impact and intervention. Kluwer, Dordrecht, pp 307–324

Green BL, Lindy JD (1994) Post-traumatic stress disorder in victims of disasters. Psychiatr Clin North Am 17: 301–309

Green BL, Lindy J, Grace M, Gleser GC, Leonard AC, Korol M, Winget C (1990) Buffalo Creek survivors in the second decade: stability of stress symptoms. Am J Orthopsychiatry 60: 43–54

Green BL, Grace MC, Vary MG, Kramer TL, Gleser GC, Leonard AC (1994) Children of disaster in the second decade: a 17-year follow-up of Buffalo Creek survivors. J Am Acad Child Adolesc Psychiatry 33: 71–79

Gruen RJ (1993) Stress and depression: toward the development of integrative models. In: Goldberger L, Breznitz S (eds) Handbook of stress, 2nd edn. Free Press, New York, pp 550–569

Hardin SB, Weinrich M, Weinrich S, Hardin TL, Garrison C (1994) Psychological distress of adolescents exposed to Hurricane Hugo. J Trauma Stress 7: 427–440

Havenaar JM, Van Den Brink W, Van Den Bout J, Kasyanenko AP, Poelijoe NW, Wohlfarth T, Meijler-Iljina LI (1996) Mental health problems in the Gomel region (Belarus): an analysis of risk factors in an area affected by the Chernobyl disaster. Psychol Med 26: 845–855

Horowitz M (1990) Post-traumatic stress disorders: psychosocial aspects of the diagnosis. Int J Ment Health 19: 21–36

Jackson G (1991) Psychological effects of disaster. Trop Doct Suppl 1: 61–62

Kaniasty K, Norris FH (1993) A test of the social support deterioration model in the context of natural disaster. J Pers Soc Psychol 64: 395–408

Karanci AN, Rustemli A (1995) Psychological consequences of the 1992 Erzincan (Turkey) earthquake. Disasters 19: 8–18

Kessler RC, McGonagle KA, Zhao S, Nelson CB, Hughes M, Eshleman S, Wittchen JU, Kendler KS (1994) Lifetime and 12-month prevalence of DSM-III-R psychiatric disorders in the United States: results from the National Comorbidity survey. Arch Gen Psychiatry 51: 8–19

Kessler R, Sonnega A, Bromet E, Nelson C (1995) Posttraumatic stress disorder in the National Comorbidity Survey. Arch Gen Psychiatry 52: 1048–1060

Kinzie JD, Sack W, Angell R, Clarke G, Ben R (1989) A three-year follow-up of Cambodian young people traumatized as children. J Am Acad Child Adolesc Psychiatry 28: 501–504

Kohn R, Levav I (1990) Bereavement in disaster: an overview of the research. Int J Ment Health 19: 61–76

Laor N, Wolmer L, Mayes LC, Gershon A, Weizman R, Cohen DJ (1997) Israeli preschool children under Scuds: a 30-month follow-up. J Am Acad Child Adolesc Psychiatry 36: 349–356

Lebedun M, Wilson KE (1989) Planning and integrating disaster response. In: Gist R, Lubin B (eds) Psychosocial aspects of disaster. Wiley, New York, pp 268–279

Lechat MF (1979) Disasters and public health. Bull World Health Org 57: 11–17

Lifton RJ (1967) Death in life: survivors of Hiroshima. Random House, New York

Lima BR, Pai S, Lozano J, Santacruz H (1990a) The stability of emotional symptoms among disaster victims in a developing country. J Trauma Stress 3: 497–505

Lima BR, Santacruz H, Lozano J, Chavez H, Samaniego N, Pompei MS, Pai S (1990b) Disasters and mental health: experience in Colombia and Ecuador and its relevance for primary care in mental health in Latin America. Int J Ment Health 19: 3–20

Logue JN, Hansen H, Struening E (1981a) Some indications of the long-term effects of a natural disaster. Public Health Rep 96: 67–79

Logue JN, Melick ME, Hanson H (1981b) Research issues and directions in the epidemiology of health effects of disasters. Epidemiol Rev 3: 140–162

Lonigan CJ, Shannon MP, Taylor CM, Finch AJ, Sallee FR (1994) Children exposed to disaster: risk factors for the development of post-traumatic symptomatology. J Am Acad Child Adolesc Psychiatry 33: 94–105

Lopez-Ibor JJ, Soria J, Canas F, Rodriguez-Gamazo M (1985) Psychopathological aspects of the toxic oil syndrome catastrophe. Am J Psychiatry 147: 352–365

Madakasira S, O'Brien KF (1987) Acute post-traumatic stress disorder in victims of a natural disaster. J Nerv Ment Dis 175: 286–290

Maj M, Starace F, Crepet P, Lobrace S, Veltro F, De Marco F, Kemali D (1989) Prevalence of psychiatric disorders among subjects exposed to a natural disaster. Acta Psychiatr Scand 79: 544–549

March JS, Amaya-Jackson L (1993) Post-traumatic stress disorder in children and adolescents. PTSD Res Q 4: 1–3

March JS, Amaya-Jackson L, Terry R, Costanzo P (1997) Posttraumatic symptomatology in children and adolescents after an industrial fire. J Am Acad Child Adolesc Psychiatry 36: 1080–1088

McFarlane AC (1988) The aetiology of post-traumatic stress disorders following a natural disaster. Br J Psychiatry 152: 116–121

McFarlane AC (1993) PTSD: synthesis of research and clinical studies. In: Wilson JP, Raphael B (eds) International handbook of traumatic stress syndromes. Plenum, New York, pp 421–429

McFarlane AC, Policansky SK, Irwin C (1987) A longitudinal study of the psychological morbidity in children due to a natural disaster. Psychol Med 17: 727–738

Milne G (1977a) Cyclone Tracy. I. Some consequences of the evacuation for adult victims. Aust Psychol 12: 39–54

Milne G (1977b) Cyclone Tracy. II. The effects on Darwin children. Aust Psychol 12: 55–62

Misao T, Hattori K, Shirakawa M, Suga M, Ogawa N, Ohara Y, Ohno T, Fukuta J, Hamada T, Kuwahara H, Fukamachi K, Hitsumoto S, Kamatani T (1961) Characteristics of abnormalities observed in atom-bombed survivors. J Rad Res 2: 85–97

Murphy SA (1988) Mediating effects of intrapersonal and social support on mental health 1 and 3 years after a natural disaster. J Trauma Stress 1: 155–172.

Nolen-Hoeksema S, Morrow J (1991) A prospective study of depression and posttraumatic stress symptoms after a natural disaster: the 1989 Loma Prieta earthquake. J Pers Soc Psychol 61: 115–121

Norris FH, Uhl GA (1993) Chronic stress as a mediator of acute stress: the case of Hurricane Hugo. J Appl Soc Psychol 23: 1263–1284

Ollendick DG, Hoffmann M (1982) Assessment of psychological reactions in disaster victims. J Community Psychol 10: 157–167

Palinkas LA, Russell J, Downs M, Petterson JS (1992) Ethnic differences in stress, coping, and depressive symptoms after the Exxon Valdez oil spill. J Nerv Ment Dis 180: 287–295

Parker G (1977) Cyclone Tracy and Darwin evacuees: on the restoration of the species. Br J Psychiatry 130: 548–555

Patrick V, Patrick WK (1981) Cyclone '78 in Sri Lanka – the mental health trail. Br J Psychiatry 138: 210–216

Penick EC, Powell BJ, Sieck WA (1976) Mental health problems and natural disaster: tornado victims. J Community Psychol 4: 64–67

Perry RW, Lindell MK (1978) The psychological consequences of natural disaster: a review of research on American communities. Mass Emergencies 3: 105–115

Pynoos RS, Frederick C, Nader K, Arroyo W, Steinberg A, Eth S, Nunez F, Fairbanks L (1987) Life threat and posttraumatic stress in school age children. Arch Gen Psychiatry 44: 1057–1063

Pynoos RS, Goenjian A, Tashjian M, Karakashian M, Manjikian R, Manoukian G, Steinberg AM, Fairbanks LA (1993) Posttraumatic stress reactions in children after the 1988 Armenian earthquake. Br J Psychiatry 163: 239–247

Quarantelli EL (1985) An assessment of conflicting views on mental health: the consequences of traumatic events. In: Figley CR (ed) Trauma and its wake. Brunner/Mazel, New York, pp 173–215

Rabkin J (1993) Stress and psychiatric disorders. In: Goldberger L, Breznitz S (eds) Handbook of stress, 2nd edn. Free Press, New York, pp 477–495

Radloff LS (1977) The CES-D scale: a self-report depression scale

for research in the general population. J Appl Psychol Meas 1: 385–401

Raphael B (1986) When disaster strikes. Basic, New York

Robins LN, Regier D (1991) Psychiatric disorders in America: the Epidemiologic Catchment Area Study. Free Press, New York

Robins LN, Smith E (1983) The Diagnostic Interview Schedule Disaster Supplement. Washington University School of Medicine, St. Louis

Robins LN, Helzer JE, Croughan J, Ratcliff KS (1981) National Institute of Mental Health Diagnostic Interview Schedule: its history, characteristics, and validity. Arch Gen Psychiatry 38: 381–389

Robinson S, Rapaport-Bar-Sever M, Rapaport J (1994) The present state of people who survived the holocaust as children. Acta Psychiatr Scand 89: 242–245

Rubonis A, Bickman L (1991) Psychological impairment in the wake of disaster: the disaster-psychopathology relationship. Psychol Bull 109: 384–399

Saylor C (ed) (1993) Children and disasters. Plenum, New York

Selley C, King E, Peveler R, Osola K, Martin N, Thompson C (1997) Post-traumatic stress disorder symptoms and the Clapham rail accident. Br J Psychiatry 171: 478–482

Shannon MP, Lonigan CJ, Finch AJ, Taylor CM (1994) Children exposed to disaster. I. Epidemiology of post-traumatic symptoms and symptom profiles. J Am Acad Child Adolesc Psychiatry 33: 80–93

Sharan P, Chaudhary G, Kavathekar SA, Saxena S (1996) Preliminary report of psychiatric disorders in survivors of a severe earthquake. Am J Psychiatry 153: 556–558

Shore JH, Tatum EL, Vollmer WM (1986a) Evaluation of mental health effects of disaster, Mount St. Helens eruption. Am J Public Health 76[Suppl]: 76–83

Shore JH, Tatum EL, Vollmer WM (1986b) Psychiatric reactions to disaster: the Mount St. Helens experience. Am J Psychiatry 143: 590–595

Slovik P (1991) Perception of risk from radiation. In: Ricks RC, Berger ME, O'Hara FM (eds) The medical basis for radiation preparedness. III. The psychological perspective. Elsevier, New York, pp 211–227

Solomon SD (1992) Mental health effects of natural and human-made disasters. PTSD Res Q 3: 1–7

Solomon SD, Smith EM, Robins L, Fischbach R (1987) Social involvement as a mediator of disaster-induced stress. J Appl Soc Psychol 17: 1092–1112

Terr L (1983) Chowchilla revisited: the effects of psychic trauma four years after a school-bus kidnapping. Am J Psychiatry 140: 1543–1550

Trevisan M, Jossa F, Farinaro E, Krogh V, Panico S, Giumetti D, Mancini M (1992) Earthquake and coronary heart disease risk factors: a longitudinal study. Am J Epidemiol 135: 632–637

Turner RJ, Wheaton B, Lloyd DA (1995) The epidemiology of social stress. Am Soc Rev 60: 104–125

van der Kolk BA, Pelcovitz D, Roth S, Mandel FS, McFarlane A, Herman J (1996) Dissociation, somatization, and affect dysregulation: the complexity of adaptation to trauma. Am J Psychiatry 153[July Suppl]: 83–93

Vernberg EM, Vogel JM (1993) Children's psychological responses to disasters, part 2. J Clin Child Psychol 22: 485–498

Vogel J, Vernberg EM (1993) Children's psychological responses to disasters, part 1. J Clin Child Psychol 22: 464–484

Weisaeth L (1989) The stressors and the post-traumatic stress syndrome after an industrial disaster. Acta Psychiatrica Scand Suppl 355(80): 25–37

Weisaeth L (1993) Disasters: psychological and psychiatric aspects. In: Goldberger L, Breznitz S (eds) Handbook of stress, 2nd edn. Free Press, New York, pp 591–616

Wood JM, Bootzin RR, Rosenhan D, Nolen-Hoeksema S, Jourden F (1992) Effects of the 1989 San Francisco earthquake on frequency and content of nightmares. J Abnorm Psychol 101: 219–224

World Health Organization (1992) Psychosocial consequences of disasters – prevention and management. WHO/FHE/MNH/93.1. WHO, Geneva

CHAPTER 18

J. de Jong

Psychiatric Problems Related to Persecution and Refugee Status

1
Development of a Diagnostic Concept to Measure the Impact of Persecution

Patterns of psychological distress following traumatic experiences have been described for many years prior to the inclusion of post-traumatic shock disorder (PTSD) in DSM-III in 1980 (American Psychiatric Association 1980). For example, "shell shock," "combat fatigue," formulations denoting a presumed organic basis for trauma-related psychological symptoms, were used during the two World Wars. Kardiner (1941) noted that sufferers from "traumatic neuroses" develop an enduring vigilance for and sensitivity to threat and stated that the nucleus of the neurosis is a physioneurosis. He also noted that the "pathological traumatic syndrome" consists of an altered conception of the self in relation to the world, based on being fixated on the trauma and having an atypical dream life, with chronic irritability, startle reactions, and explosive aggressive reactions. Central in Kardiner's thinking, similar to Janet and Freud, was the fact that "the subject acts as if the original traumatic situation were still in existence and engages in protective devices which failed on the original occasion. This means in effect that his conception of the outer world and his conception of himself have been permanently altered" (Van der Kolk et al. 1996).

After World War II, many psychiatrists in the United States and the United Kingdom tried to apply Kardiner's lessons in the combat theater. They discovered group stress debriefing as well as the use of group psychotherapy and the therapeutic community. Both war and disasters made mental health care professionals aware that, under extreme conditions, the group rather than the individual is the basic unit of study and treatment (Van der Kolk et al. 1996).

After World War II, the long-term effects of trauma were studied in survivors of the holocaust and the Japanese concentration camps. The concepts of "concentration camp syndrome" or "*KZ Syndrom*" were developed, which included not only the symptoms listed under PTSD, but also personality changes which are now referred to as complex enduring personality change after catastrophic experience (ICD-10) or complex PTSD (appendix, DSM-IV). The studies on concentration camp victims showed that extreme trauma had severe biological, psychological, social, and existential consequences, including a diminished capacity to cope with both psychological and biological stressors later in life.

Other diagnoses previously used for the disorder included "adjustment reactions" and "pathological" grief responses. During the process of developing DSM-III, a number of meetings eventually led to the inclusion of PTSD (American Psychiatric Association 1980). All the different syndromes – war sailor syndrome, trench foot, rape trauma syndrome, battered women syndrome, Vietnam veterans' syndrome, and abused child syndrome – were included in the new diagnosis.

Relative to the conceptualization of other psychological disorders, that of PTSD appears to represent a step forward, in that a direct linkage between a known causal agent (a traumatic event) and a resultant syndrome is assumed. Thus, at least in theory, PTSD is an ideal diagnosis; it exceeds the level of a phenomenological description of symptoms.

It has also brought rapid progress to the field of post-traumatic stress studies and interventions. However, interpreting trauma as a disorder prevents us from understanding mass trauma and community stress. We should shift our attention from the debilitating or disturbing aspects to include aspects of normal processing of traumatic events and even of potential growth (Tedeschi et al. 1998).

1.1
Long-Term Effect of Persecution

Krystal (1988) studied the long-term outcome of massive traumatization in concentration camp victims. He suggested that the core experience of being traumatized consists of "giving up" and accepting death and destruction as inevitable. He noted that the trauma response evolves from a state of hyperalert anxiety to a progressive blocking of emotions and behavioral inhibition. In his view, trauma leads to a "dedifferentiation of affects." Traumatized patients come to experience emotional reactions merely as somatic states, without being able to interpret the meaning of what they are feeling. They become prone to undifferentiated affect storms and psychosomatic reactions. Most of the initial studies on the effects of World War II were conducted by investigators who had participated in the war or had survived concentration camps (see Dohrenwend 1998).

Lomranz (1995) reviewed the research literature on the holocaust and came to the conclusion that most studies have the following limitations:

1. They are anchored in a psychoanalytic or psychodynamic theoretical orientation.
2. They are limited mainly to intrapsychic dimensions.
3. They are seldom empirical.
4. They use extremely small samples and clinical case studies.

5. They disregard the importance of sample selection procedures.
6. They have patients as their subjects.
7. They overinclude and do not specify the exact nature of the trauma (e.g. concentration camp, hiding).
8. They lack control groups.
9. As their major focus of investigation, they have themes concerning deficiency, symptomatology, or psychopathology.

Lomranz mentions that the more recent empirical studies have underscored the remarkable adaptive capacities of survivors, revealing minimal, if any, differences with nonsurvivor control groups. In his view, the studies show that survivors experience well-being and satisfaction in their work, marriage, and family and enjoy social interactions. In some aspects of adaptation and well-being, survivors sometimes score even higher than those who did not experience the holocaust. The studies with a positive outcome quoted by Lomranz date from the 1980s. Other studies from a somewhat later period are more pessimistic; they confirm the clinical impression that survivors of World War II are at risk of a delayed onset or worsening of post-traumatic symptomatology during the later phases of the life cycle, sometimes after decades of adequate coping. This pertains to survivors of Nazi persecution, military veterans, former members of resistance movements, and war sailors. For example, Op den Velde et al. (1993) studied former members of the Dutch resistance during World War II and found that the time of onset of the first symptoms of PTSD was highly variable. The first occurrence of symptoms suggestive of PTSD followed in 31.7% of the studied population immediately or almost immediately after World War II. In 50%, PTSD was first manifested more than 20 years after the end of the war. Hovens (1994) found that, 45 years after the war, 56% of these male Dutch resistance fighters appeared to have current PTSD according to the Structured Clinical Interview for DSM-III-R (SCID). Only 4% of the sample was symptom free. It was found that veterans with PTSD had longer periods of resistance work and subsequent imprisonment. This suggests that the actual duration of the stressful circumstances may have contributed substantially to the occurrence of PTSD.

Tennant et al. (1993), in a controlled follow-up study with former Australian servicemen who were interned in Japanese prisoner-of-war camps, showed a significant increase of psychiatric disorder when the target group was between 46 and 66 years of age. A similar trend was noted in several other studies (see Aarts and Op den Velde 1996).

1.2 Transgenerational Traumatization

The late sequelae of World War II are still visible today, both among patients and in art and literature related to that period. Concentration camp survivors, resistance fighters, combat veterans, and civilian victims often struggle with the prolonged consequences of the violence and abuse inflicted by the war. It may be expected that these victims and survivors could not adequately comply with the needs and wishes of their children born after World War II, the so-called second generation. Numerous clinic-based observations have suggested that some of the effects of the persecution of parents were transmitted to their offspring. The traumatization suffered by the survivors – with its prolonged psychopathological and psychosocial after-effects, including familial losses, often of spouse and children, the frequent hastily contracted marriages upon liberation, and problems arising from resettlement – among other factors, were thought to impair parenting abilities (Levav 1998). A number of clinical studies reporting on the influence of the war on survivors have been summarized along three lines (Van der Velden and Kleber 1999):

1. *Unbalanced interaction between parents and children.* This unbalanced interaction manifested itself in several ways. First, it became apparent in difficulties in separation and individuation. Parents had experienced so many losses and separations during the war that they felt threatened by the idea that their child was growing into an independent, autonomous individual. They felt as if they were losing their child. Not being able to support the child's need for independence might lead to an "anxious attachment" between parent and child (Bowlby 1981). This anxious attachment can manifest itself in difficulties with small separations (e.g. going on a short holiday, playing at friends' houses) as well as in overprotection or overinvolvement (Russel 1980). Second, it was shown in lack of emotional support. Parents were often so occupied by the reminiscences of the war that they could not be empathetic to the emotional needs of their children or provide affection and support. A third problem was "parentification." The memories and intrusions of World War II made the parents psychologically vulnerable. Children easily register the distress and vulnerability of their parents, even if these are masked. As a result, children may feel responsible for their parents' well-being, which may result in a reversal of family roles. The parent assumes the dependent role, while the child

acquires the caring and supporting role and becomes parentified (Sigal et al. 1973).
2. *Management of emotions, in particular aggression.* Feelings of rage and hostility were often difficult to handle for those who survived the ordeals of the concentration camps of World War II. Violence was omnipresent in the camp, but at the same time a prisoner had to control his or her feelings with the utmost care in order to survive. Against such a background, aggression can become a difficult emotion in postwar life. Anger and aggression therefore often manifest themselves as extremes on the poles of a continuum. They may become extremely inhibited, sometimes resulting in outbursts of anger, or manifest themselves frequently.
3. *The burdensome theme of war in the family.* War-stricken parents often hardly discuss their war memories with their children. Their reluctance can be motivated by the wish to protect their children from the atrocities and horrors of the war. This attempt may paradoxically result in the opposite. The child does not dare to ask questions and tries to fill the gaps with his or her own imagination. The silence with regard to war memories is indicated as a conspiracy of silence (Danieli 1985) or as a family secret.

The issue of transgenerational traumatization is subject to debate. Children of victims, exposed to their parents' symptomatology without themselves undergoing violence, persecution, or war, may show the characteristic traumatic stress symptoms of intrusion, hyperarousal, and enactment similar to their parents. Parental reactions to the experience play a crucial role in the child's responses, at times determining the child's reaction to the traumatic event. Trauma clearly has a contagious effect (Steinberg 1995). Nevertheless, the child's resilience and available supports may modify the impact. In a study on the long-term effects of the Vietnam war, Kulka et al. (1991) found that children with a veteran parent suffering from PTSD had significantly more behavioral problems than did children of veteran parents without PTSD. These findings may help explain the transmission of post-traumatic stress to offspring who were spared the actual traumatic experiences.

In a study of the children of U.S. prisoners of war held during the Vietnam war, McCubbin et al. (1977) found indications of the transgenerational effects of captivity. The long-term strains of internment affected parent–child relationships and children's functioning in general, as was shown, for example, in school performance and dysfunctional symptoms of the child. A longitudinal analysis of 42 families showed factors that could account for these effects:

1. The severity of the mental abuse suffered during internment
2. The severity of physical abuse
3. The wife's relationship with her parents
4. The wife's involvement in the activities of the former prisoners of war

The father's involvement in preparing the family for separation was found to be a positive factor in subsequent family relationships. Children of survivors of the Nazi holocaust are a well-documented group with regard to the indirect effects of violence on children. For example, Rakoff et al. (1965) found that Holocaust survivors exhibited negative effects due to the traumatic experiences of their parents. Mooren and Kleber (1996) did two studies, one on the children of Jewish World War II survivors and one on children whose parents survived the Japanese camps in the former Dutch Indies during World War II. These studies do not confirm previous studies in which no significant differences between survivor and control group children were found. The findings of Mooren and Kleber (1996) indicate that only a few differences exist in health and parenting styles between the second generation and the comparison groups they studied. The problems of the second generation are related to the coping disturbances of the parents due to the war. Their data confirm the results of a double-blind research design that was used in Israel (Brom et al. 1994). Both studies demonstrated a small and significant difference between the offspring of holocaust survivors and their peers. These and other empirical studies (Solkoff 1992) made clear that there are some long-term aftereffects of the war in the offspring of war survivors. However, in contrast to clinical observations, studies conducted in the community did not show evidence that the offspring of concentration camp survivors have more psychopathology than controls (Levav 1998). It appears that the concept of transgenerational traumatization is rather inappropriate (Figley and Kleber 1995). The children of concentration camp survivors have not themselves experienced traumatic events. It is their upbringing that is hindered by war memories and their parents' feelings of guilt and loss. The disturbances in the offspring of war survivors are not as much an issue of transmission of trauma as an issue of a specific socialization.

These studies have several implications for clinical workers, either in the West or in low-income countries, where man-made or natural disasters are much more prevalent than in the West. Clinicians are confronted with the subtle task of unraveling a patient's history to find out whether and to what extent the parents' war experience has an impact on the life of their patients.

1.3
Cultures of Fear

The state production of fear has led researchers to speak of "cultures of fear" or "frightened populations" (Corradi et al. 1992). Salimovich et al. (1992) have identified three core features of persistent fears among the Chilean population: (1) a sense of personal weakness, vulnerability, and a feeling of powerlessness; (2) sensory perceptions remaining in a permanent state of alert; and (3) the impossibility of testing subjective experience against reality. Cultures of fear are not only produced by states in Latin America, but also by (para)militia or police forces, e.g. in the former Yugoslavia, the Caucasian republics, or Sri Lanka, as well as by cults of violence and counterviolence, e.g. in Mozambique, Uganda, Liberia, or Sierra Leone (Jeyaraja Tambiah 1992; K.B. Wilson 1992; Allen 1991). During and after a war, the "culture of fear" may interact with the chronic sequential war traumas and with the daily difficulties of living in a devastated area, a refugee camp, or a repressive environment. This leads to a continuous traumatic stress syndrome showing similarities with complex PTSD or enduring personality change after catastrophic experience. These individual diagnoses show a striking similarity to the above-mentioned core features of the "cultures of fear." We know little about the consequences of this interaction, and on the personal level empirical research hardly exists (Shresta et al. 1998). On a community level, it is tempting to hypothesize a relation between this continuous traumatic stress and the occurrence of recurrent violence as a temporarily "effective" but in the long run adverse collective coping strategy, e.g. in the Middle East, southern Africa, Algeria, Latin America, or the former Soviet Union. It is also tempting to postulate that the violence in Israel and Palestina or in the South African townships, for example, is to some extent a rebound phenomenon of a rebellious younger generation which is no longer willing to endure the humiliations or a perceived lack of resistance in their elders. The prevailing fear and suspicion weave their way into society and affect mutual support structures, personal commitments, belief in justice, and belief in democratization and human rights.

2
Refugees and Displaced Persons

2.1
Scope of the Problem and Its Geographical Distribution

The number of refugees has grown enormously over the last decades. In 1960, there were estimated to be 1.4 million refugees worldwide. By 1978, there were 4.6 million refugees seeking (international) assistance and protection. Grant (1979) estimated a total of 13 million worldwide in 1979, and Cohon and Westermeyer mention a total of 16 million refugees in 1981 (Cohon 1981; Westermeyer 1989). By the beginning of 1993, the United Nations High Commission for Refugees (UNHCR) reported a total of 18.2 million refugees and estimated the number of internally displaced people (refugees within the borders of their own countries) conservatively at 24 million (Long 1992; UNHCR 1993). The United States Committee for Refugees (USCR) (1992) mentions a current estimate of 20 million refugees covered by the 1951 UN Convention on the Status of Refugees and 20 million internally displaced persons. Many of these refugees are destitute people from poor countries who travel within or to other impoverished countries. This displacement can destabilize the host countries, aggravate regional tensions, and increase rates of environmental degradation. In Mozambique alone, for instance, the number of internally displaced persons was estimated at four million and the number of refugees at more than one million out of a total population of 14 million. Meanwhile, established nations that have traditionally offered asylum are increasingly unwilling to open their doors to massive refugee flows (Desjarlais et al. 1995).

The total number of refugees is increasing as the number of people seeking refuge is growing faster than the number of refugees whose plight is resolved through repatriation or resettlement. For every legally defined refugee, there are many more internally displaced people who suffer no less psychosocial trauma and who may not formally be identified, but who are vulnerable to similar mental disorders as refugees. Studies of the dynamic flow of uprooted populations confirm that groups may be refugees yesterday, repatriated today, and internally displaced tomorrow. In addition to the approximately 20 million official refugees, there are countless other economic and environmental refugees in the world today. More than 70 million people around the world have left their native countries, primarily in search of work. At least 10 million of these have fled from environmental decline, land degradation, and diminishing agricultural and water resources (Desjarlais et al. 1995).

The regional distribution of the refugee problem is far from even. According to Leopold and Harrel-Bond (1994), Africa contains about 30% of the world's refugees and perhaps half of the internally displaced. Europe accounts for about 3.75% and North America account for about 10.25% (including an estimated 2 million refugees from the former Yugoslavia), and the Middle East and South Asia for about 50%. Clearly, these figures are subject to fluctuation due to the volatility of national and international relations, for

example in Burundi, Congo, Rwanda, Sri Lanka, Ethiopia and Eritrea. In Africa, 70% of the refugee burden is borne by 12 countries. All 16 of the least-developed African countries are affected by the refugee problem.

Western countries spend US$ 5–7 billion on less than one million refugees in the West, while a mere US$ 340 million is spent by the UNHCR on the 17–19 million refugees in the rest of the world (Harrel-Bond 1991).

Both disasters and wars have a differential impact in low-income countries. In the period 1967–1991, an average of 117 million people living in developing countries were affected by disasters each year, as compared to about 700,000 in developed countries (a striking ratio of 166:1). War represents the most ancient and the most important form of human-made violence in terms of the magnitude of its effects. Since World War II, there have been 127 wars and 21.8 million war-related deaths. The Red Cross estimated a total twice as high, i.e. about 40 million people killed since WW II. All but two of the 127 wars have taken place in developing countries (McFarlane and de Girolamo 1996).

2.2 Psychosocial and Mental Health Consequences

Forced migration and the refugee status are often consequences of circumstances of degradation, violence, dehumanization, and torture. As was just mentioned, this is particularly true in low-income countries, where the combination of low-intensity warfare and high-intensity lethal weaponry is one of the major causes of displacement of large numbers of people, especially women and children. The militarization of these countries, which may even nowadays be the arena for conflicts between the superpowers, is a major cause in the present-day global refugee problem (Jablensky 1994).

All aspects of the refugee's ordeal take their mental and physical toll: the events precipitating the act of relocation (war, persecution, hunger, disaster, death), the process of relocation (upheaval, a long journey on foot under mental and physical strain), and the often temporary settlement in a refugee camp (discomfort, uncertainty, oppression). The resulting long- and short-term mental health and psychosocial consequences are many and varied.

2.3 Risk Factors and Mental Disorders

Jablensky et al. (1994) discuss more specific risk factors that are implicated in the development of mental disorders in refugees. These include the following: marginalization resulting in loss of self-esteem and self-confidence; socioeconomic hardship resulting in poverty, deprivation, and unemployment; poor physical health due to poor sanitation, poor nutrition, or crowding; head trauma and other physical injuries; collapse of social networks resulting in anomie, alienation, and isolation; and mental trauma due to torture, death, loss, and fear. Sartorius (1994) mentions factors which may help to mitigate several of these risk factors: opportunity to promote cohesion (versus active dispersion); continuity of life span (versus neglecting life span); promotion of initiative (versus punishing it); clear knowledge about the future (versus an uncertain future); job opportunities (versus unneeded or obligatory work); fostering a social network (versus no social network); rules of conduct (versus rules by accident); and setting goals in life (versus living from day to day).

The general effects of trauma are varied but can be described by some combination of the following: feelings of extreme vulnerability, helplessness, or despair, intense, panic-level arousal and negative emotion, a sense of being stunned, numb, and depleted, and an altered consciousness or awareness that entails a sense of derealization and depersonalization (also referred to as dissociation) (Litz and Roemer 1996). More specific mental disorders observed in refugees and attributable to their circumstances include PTSD (Friedman and Jaranson 1994), depressive disorders (Silver and Iacono 1984; Westermeyer 1986), substance abuse (Keehn 1980; Peak et al. 1981; Westermeyer 1985), panic disorder, generalized anxiety, phobia, antisocial and other personality disorders, psychosis (Yesavage 1983), childhood developmental problems, organic brain syndrome (especially in victims of violence), and associated medical and social problems (Corcoran 1982).

2.4 Morbidity Among Adults

PTSD is currently the focus of public and scientific interest. However, few epidemiological estimates exist of the overall prevalence of PTSD (regardless of exact cause) in the general population and especially in those areas of the world where wars are highly prevalent.

The Epidemiological Catchment Area survey found a lifetime PTSD rate of 0.5% among men and 1.3% among women. The number of those who showed some symptoms after a trauma was substantially higher, i.e. 15% among men and 16% among women, with an average of 2.4 symptoms (Helzer et al. 1987).

A number of studies have assessed the prevalence of PTSD among Vietnam combat veterans. The National

Vietnam Veterans Readjustment Study (NVRRS) appears to be the most validated epidemiological study of the prevalence of PTSD among Vietnam combat veterans in the general population. Current prevalence rates (6 months prior to assessment) for men and women were 15.2% and 8.5%, respectively. Lifetime prevalence rates of PTSD for men and women were 30.6% and 26.9%, respectively (Kulka et al. 1991). A study of over 2000 monozygotic twin pairs found a prevalence of PTSD of almost 17% in twins who served in South-East Asia compared to 5% in co-twins who did not do their military service in Asia (Goldberg et al. 1990).

Kaličanin et al. (1993) studied 384 war veterans in 21 psychiatric institutions in Serbia between 1991 and 1993. The most frequent diagnoses among war veterans were adjustment disorder (34%) and PTSD (27%). Among refugees the diagnoses that were encountered most frequently were adjustment disorders (27%), mixed anxiety and depressive reaction (25%), and acute and transitory psychotic disorders associated with stress (40%). The most frequent stressors related to these problems among refugees were separation from loved ones (15%), loss of property (15%), and the combination of several stressors (45%).

Rundell et al. (1990) reviewed all controlled studies of psychiatric disorders reported to be associated with war and other traumatic events. They mention four studies of concurrent psychiatric disorders in combat veterans with PTSD. Additional psychiatric diagnoses are alcohol abuse or dependence (41%–80% of PTSD patients), depression (8%–72%), and drug abuse disorders (16%–50%). Antisocial personality disorder has been reported in 3%–40%, social phobia in up to 50%, and bipolar disorder in 10%–25%.

De Girolamo and McFarlane (1996) provide an overview of the epidemiology of PTSD. Two groups of studies in their overview pertain to the consequences of persecution. The first consists of ten studies on prisoners of war and other types of prisoners, generally those imprisoned for political reasons. These studies were conducted in the United States, Germany, Turkey, Australia, and France. In all these studies, the PTSD rate varied between 50% and 70% or sometimes more. There were patients who suffered from PTSD symptoms for the first time decades after the end of their imprisonment; many others were still suffering decades after the end of detention and the first appearance of the disorder. Some studies have found an association between the appearance and the severity of PTSD and the length and the severity of imprisonment, the latter being mainly evaluated by the experience of torture and by the percentage of body weight lost. The second group of studies in their overview consists of 12 studies assessing the rate of PTSD among samples of refugees. Nine of the 12 studies were carried out among samples of resettled refugees, mostly South-East Asians. Looking more closely at some of these studies reveals the following picture. South-East Asian refugees in the United States and Canada show high rates of psychiatric disorders. According to Westermeyer (1988), 44% of Hmong adults in a 10-year follow-up in the United States qualified for an axis I disorder (13%) or chronic adjustment troubles plus depressive symptoms (31%). Among Cambodian, Hmong, Laotian, and Vietnamese psychiatric patients in the United States, Mollica et al. (1987) found that 50% of the patients suffered from PTSD. Except for one patient, the diagnosis of PTSD was consistently associated with another psychiatric diagnosis, primarily major affective disorder. Patients with PTSD had experienced twice as many traumatic events as patients with other psychiatric diagnoses. On average, the refugee patients had experienced ten traumatic events such as lack of food, water, or shelter, imprisonment, war injury, sexual abuse, or witnessing death or murder. They also witnessed an average of 1.6 instances of torture. Kinzie and Manson (1983) and Nguyen (1984) report similar data. In a study among Somalian refugees in Europe, 38% of the refugees suffered from PTSD, and an additional 60% from partial PTSD (Roodenrijs et al. 1998).

The remaining three studies are the only studies carried out among victims of wars or refugees in the afflicted areas in low-income countries. Mollica et al. (1993) studied 993 Cambodian refugees on the Thai–Cambodian border. During the Khmer Rouge regime (1975–1979), more than 85% reported lack of food, water, shelter, and medical care, brainwashing, and forced labor; 54% reported murder of a family member or friend; and 17% reported rape or sexual abuse. During the refugee period (1980–1990), most of these figures decreased to some extent; during the same period only reports of murder, head injury, and rape/sexual abuse decreased to 5%. Fifty-five percent qualified for a depression, and 15% for PTSD. Summerfield and Toser (1991) conducted a study in Nicaragua. Since they interviewed a small sample of convenience with the General Health Questionnaire, which was not validated for Nicaragua, firm conclusions can hardly be drawn from their data. Ramsay et al. (1993) looked at a sample of torture victims that found refuge in Great Britain. This means that only one of the three aforementioned valid epidemiological surveys carried out so far was in a low-income country (Mollica et al. 1993). El Sarraj et al. (1996) found prevalence rates of 20% or more among 550 torture survivors in Gaza. Shresta et al. (1998) studied 526 Bhutanese torture survivors and a matched control group in Nepal. A diagnosis of PTSD was significantly more common in the tortured group (14%) than in the nontortured group (3%). Significantly more tortured

refugees had high anxiety scores (43% versus 34%) and high depression scores (25% versus 14%).

In Africa, a few descriptive studies have been done to assess psychosocial and psychiatric problems among various populations such as refugees and displaced persons. So far, no valid epidemiological studies have been published.

Based on these studies, it may be concluded that, in general, PTSD rates among victims of wars and persecution are 50% or higher, that these victims often suffer from additional psychiatric disorders, and that they have usually faced very difficult circumstances (e.g. torture, starvation, witnessing killings) and are a highly traumatized population.

2.5 Traumatic Stress and Its Consequences for Children, Adolescents, and Families

Tens of millions of children are victims of persecution or misfortune, becoming refugees, displaced persons, child soldiers, or casualties of war. Of the above-mentioned 43 millions refugees and displaced persons in the world, 70%–80% are mothers with children (Forbes Martin 1992). The current increase of refugee populations, both in low-income and in high-income countries, creates a causal chain of disruption of communities and families. Women, children, and the elderly are particularly vulnerable according to the literature (Desjarlais et al. 1995). However, based on our own experience among the Lhotsampa refugees from Bhutan in Nepal, Sudanese refugees in Northern Uganda, or refugees in the shelters surrounding Addis Ababa, for example, men are afflicted at least as seriously as women; both peasants and white-collar workers mostly lack the opportunity to pursue their profession, which impairs their social identity. Women on the other hand maintain a social role raising their children, doing household chores, or engaging in petty trade. However, this accumulation of tasks, their sense of responsibility for their idle husbands, and the risk of being raped do make women a vulnerable group.

The surviving elderly often carry a heavy responsibility in the transmission of cultural values. This transmission often takes place during elaborate rites of passage, which can be inhibited in the refugee camps due to high costs, the long duration of the rituals, or restrictions imposed by camp authorities.

The hardship of children is compounded by their dependence on parents and by their mental and physical vulnerability. In (post)war conditions, parents may themselves suffer from mental or psychosocial problems, which in turn affect their children. If both parents have died or are lost, as often happened in Rwanda, for example, the children may be left to fend for themselves. Crude death rates among refugees and internally displaced persons soar to levels as high as 50 times the baseline crude death rate in their home areas. Most of these deaths occur among young children. During periods of displacement, the mortality rate can be as much as 60 times the expected rates (Toole and Waldman 1993). Among displaced Kurds in northern Iraq, 63% of the deaths occurred among the 17% of the population younger than 5 years (Yip and Sharpe 1993). In addition, children are especially at risk for developmental attrition, aggressive and antisocial behavior disorders frequently associated with substance abuse, and seizure disorders (Westermeyer 1989; Desjarlais et al. 1995). Developmental attrition results from the failure to reach normal developmental landmarks, so that cumulative defect becomes even larger. It is the result of inadequate intake of nourishment vital for body and mind: proteins, calories, and micronutrients; perceptual and cognitive stimuli; and social relations and interactions. Attrition manifests itself in three ways: physically, as height and weight well below norms for age; in school settings, as learning failure and retarded mental development; and behaviorally, as psychiatric disorder and social deviance. The disadvantages to which poor children are subjected – especially when they are exposed to manmade disaster – start during pregnancy. Poor nutrition and limited health care during pregnancy increases the likelihood of poor outcomes. If children born of a complicated pregnancy are reared under adverse social conditions, they will suffer long-term retardation in cognitive development (Desjarlais et al. 1995). These children are also at risk of developing schizophrenia later in life, whereas children with pellagra have a higher risk of developing dementia in adulthood.

Children need familial support and must live in communities rather than in orphanages or resettlement camps. The wars in Mozambique, Zimbabwe, Eritrea, Rwanda, Congo, and Sudan have led to the orphaning or dislocation of thousands of children who have been witness to the bloody conflict. One of the priorities for abandoned children is family reunion. In addition to the usual ways of family tracing by means of photographs or radio messages, the present author was interested to learn of the way Sudanese children were supported in a refugee camp in North Kenya. For about 5 years, this group of nearly ten thousand children had been fleeing forced circumscription by either the northern Sudanese army or one of the guerrilla movements of southern Sudan. Each plot in the refugee camp had a large sign of one of the southern Sudanese cities so that children could try to find one of their family members back. Children who did not find a family member lived under the supervision of an adult who received a modest per

diem for his or her work. Fifty children living in ten huts in a circle cooked their own meals, visited schools, and had leisure activities under the supervision of their tutor. This kind of solution is preferable to orphanages that are often set up with the best of intentions but with questionable results. Children in orphanages in Mozambique often appear withdrawn, apathetic, regressed, and fearful. Orphanages for girls often turned into a breeding place for prostitution, whereas orphanages for boys easily resulted in banditry.

Children are not only easy victims of the violence, but they are also actively drawn into the violence. During the drafting of the Declaration of the Rights of the Child, several low- and high-income countries opposed to the minimal age of participating in a war being increased to 16 years. In a number of African, Latin American, and Asian countries, children are forced into a military training and service. Many boys are forced to serve as porters or to serve in a (rebel) army, whereas young girls are forced to serve as housemaids or are sequentially raped as "partners" of soldiers. This was a common fate in Mozambique, and it still is for children and adolescents in northern Uganda if they are abducted by the Lords Resistance Army or by one of the rebel movements. Insubordination, be it refusal to follow orders or an attempt to desert, can result in summary execution. Many children were forced by RENAMO to kill a parent or an inhabitant of their own village to make it impossible for them to return to their home village and so to transform them into willing tools of war (Minter 1989; Geffray 1990; Vines 1991). These and similar atrocities, such as cutting off body parts in Mozambique, equal those committed by the Lords Resistance Army or Holy Spirit Movement in Uganda, the Khmer Rouge regime in Cambodia, or in Sierra Leone.

Violence can become a dominant way of being in the world. Acts of violence can work their way into the practices of everyday life. Violence relates both to effective ways of acting in a community and to a reworking of the moral sensibilities that define those ways of acting.

As a consequence of trauma and violence, children may show symptoms of silent withdrawal or hyperactivity accompanied by violent behavior. Regression to earlier developmental stages is common. Preschool children may show frequent or continuous crying, pathological behavior involving clinging, bedwetting, and loss of bowel control, thumb and finger sucking, frequent nightmares and night terrors, as well as unusual fear of actual or imagined objects.

Children of early school age display similar behavior and are overtly unhappy, nervous, restless, irritable, and fearful. Self-stimulation such as rocking or head banging may be observed. In addition, refusal to eat and physical complaints with a psychosomatic basis (headache, dizziness, abdominal pain) are frequent. These children may also display behavior appropriate of much younger children, such as prolonged muteness and bed-bound incontinence or clinging behavior. They frequently have specific fears, e.g. of being left alone in a room or of situations which remind them of traumatic events. Such reactions may be limited to a couple of days or weeks or may persist over a period of months or years. Many of these children become victims of a continuous strain trauma. Such trauma can result in personality disorders and even have knock-on effects in descendants, as was mentioned before.

War-related fears weave their way into the mental lives of exposed children. A study of 3- to 9-year-olds in Lebanon indicated that war was the major topic of conversation in 96% of the children, of play in 80%, and of drawing in 80% (J. Abu-Nasr, unpublished). The drawings of Ugandan refugee children indicate their preoccupation with their experiences of violence, death, and starvation. Drawings depict soldiers shooting their mothers, infants bleeding to death, decapitations, dogs eating human corpses, and people crouching in the forests with jutting ribs and swollen bellies. In a follow-up after 1 year, these children were still producing such drawings (Harrel-Bond 1986). A study of child refugees in Mozambique showed that 77% of the children had witnessed murder and 51% had been physically abused or tortured. A total of 64% of the children were abducted, and of these 75% were forced to serve as porters and 28% were trained for combat (Boothby et al. 1991). Sudanese refugee children in North Uganda suffered significantly more from the war than the Ugandan children in the same area: 94% of the families lost their property, 81% lost a family member, 92% suffered from a lack of food or water, 62% had no medical care when illness occurred, 28% of the children had been tortured, and 25% been lost or kidnapped. Compared to the Ugandan children, the Sudanese children reported significantly more PTSD complaints such as trouble sleeping, nervousness, traumatic memories, behavioral problems, depressive symptoms, and psychosomatic complaints. Despite these adversities, the plight of these children has not come to an end, since the situation in these camps – as in many other refugee camps – is extremely difficult. The Sudanese children experience considerably more daily difficulties than the Ugandan children. This is mainly caused by the poverty in the camps. The refugee children report lack of food, lack of clothes, lack of school materials, poor sanitation, and lack of medical care. They also have to contribute to household tasks, mainly because their mothers spend a great deal of time obtaining water and firewood (Paardekooper et al. 1999).

Straker et al. (1992) carried out a 3-year follow-up study of 60 young people in South-Africa. They wanted

to find out the extent to which exposure to political violence and violent conduct within that context can become generalized in behavior in other situations. All the adolescents had been abused by the police. Despite their adverse backgrounds, their militancy, and their endorsement of violence as a political tool, they did not see violence as an end in itself. Only a minority saw violence as a means to further personal gain and revenge.

Dawes (1990) studied women and children who were involved in riots in South Africa. A total of 52% of unmarried and 69% of married women suffered from PTSD. Nine percent of children suffered from PTSD, and the symptoms were most common among children whose mothers suffered from PTSD. Macksoud (1993) studied 220 children between the age of 6 and 16 years in Greater Beirut. A total of 83% had been exposed to shelling, 63% to combat, 60% to a change of residence, 55% deprived of food, 53% had their house bombarded, 11% witnessed a killing of an extended family member or other, and 9% witnessed a panic reaction in an extended family member. She also found that the number of traumatic events experienced ranged from 0 to 20, with an average of 5.75. In their study on the effects of the Gulf war in Irak, Dyregrof and Raundalen (1993) found that 90% of children had lost an average of four friends. Pynoos and Nader (1993) found that, among 51 Kuwaiti children within the age range of 8–21 years, 86% knew people who had been captured, 76% people who had been injured, 65% saw injured or dead people, 31% reported severe post-traumatic reactions, 40% moderate, 29% mild, and 4% of children reported no reaction.

Punamäki (1982) found that 35% of a sample of Palestinian children had symptoms of fear, 22% showed withdrawal behavior, 27% difficulties sleeping, 92% nightmares, and 80% were afraid that the Israeli army would attack their homes. She later found that strong ideological commitment in children between the age of 10 and 16 had a moderating effect on anxiety, insecurity, and depression (Punamäki 1996). The more the mother is socially embedded, politically active, and ideologically (or religiously) committed, the better her mental health and that of her children (Punamäki 1989).

Baker (1990) explored the effect of violence on a sample of 796 Palestinian children. A total of 80% started out fighting each other, 25% exhibited destructive behavior, one third suffered headaches, and one third difficulties sleeping, 12% bit their nails, and 43% felt depressed. Qouta et al. (1995) studied Palestinian children aged 11–12 years and found that the more traumatic experiences they had and the more they participated in the intifada, the more concentration, attention, and memory problems they had. Rosenbaum and Ronen (1992) studied 277 Israeli children to explore the effects of violence. They found a high rate of anxiety, especially in the evening, which was very high in the first weeks, but started to decrease in the fifth week. Females were more anxious than males, and there were similarities between the perception of the child and his or her parents concerning the danger. Solomon (1994) compared Israeli children in a high-exposure area with a low-exposure area. Children from the shelled area reported more coping activities than children in a nonshelled area. Children focusing on threat reported more psychological stress than children who focused on avoidance. In a study on the psychological sequelae among Angolan adolescents 6 months after the 1993 war, Ventura (1997) studied three adolescent groups differing in their degree of war exposure. The results show that the prevalence and symptoms of PTSD were greater in the refugee group (90%), followed by the nonrefugees living in Lubango (82%), and the Angolans who at the time of the study had resided for more than 1 year in Portugal (22%). There was a relationship between increased war exposure and an increase in anxiety, depression, adjustment, and behavioral problems and a decrease in intellectual functioning and self-concept. It is praiseworthy that – in contrast to adults – so many studies on children have been carried out in the afflicted areas.

However, in comparison to adults, a limited number of studies exist on children and adolescents who became refugees in a host country. Kinzie et al. (1986, 1998) found a 50% prevalence rate of PTSD in 40 Cambodian adolescents living in the United States who had been exposed to war trauma. They found that the prevalence of PTSD and depression continued to persist for years after their initial diagnosis in these refugees who were traumatized as children. Arroyo and Eth (1985) found that one third of 30 Central American refugee children (17 years or less) met criteria for PTSD. Mghir et al. (1995) found that 34% of 38 adolescent and young adult Afghan refugees in the United States met criteria for major depression, PTSD, or both.

Mahjoub (1995) provides a summary of our knowledge about the psychosocial consequences of wars among children and adolescents. Children appear to suffer less from wars than adults for the following reasons: children forget easier, are less concerned and too young to understand the horrors, and have the future to distract them from the past. Mahjoub's second conclusion is that, in addition to the short- and medium-term effects that have been studied so far, the long-term effects of wars on children need additional attention. Third, the traumatic syndrome among children has not yet been well defined. One of the reasons for this is that the presence of symptoms depends

on the measurements used (e.g. asking the mothers instead of the children about anxiety or depression). Fourth, although the differences in symptoms between boys and girls are virtually nonexistent, certain age-groups are less vulnerable than others (e.g. until the age of 2–3 years and during the latency period). Fifth, the stress of children is mediated through their environment; the role of the family has been shown in the British studies after World War II as well as in studies in kibbutzim and among Palestinians, for example.

2.6
Reaction of the Donor Community

The position of bilateral or multilateral donors is often ambivalent with respect to mental health care among refugees. Most refugee programs are emergency oriented in material and logistic terms, emphasizing food and shelter. So far, little attention has been paid to the psychological state of refugees within their cultural context. Refugees have often been reduced to mindless or psyche-less individuals whose only recognized needs are material and physical. This attitude is partially due to the perception that the degree of trauma suffered by refugees is such that their problems are too overwhelming to address. Therefore, until recently, only a few relief and development agencies included psychosocial problems on their agendas. In this way, they unwittingly contributed to a "conspiracy of silence and denial," as it has been described in the literature on the post-holocaust era, for example. The argument is often advanced that it is impossible to do anything substantial or meaningful. This results in an avoidance of the issues of the individual psychological suffering and of the attendant economic consequences. At the same time, this reaction of the (Western) donors seems to be a repetition of its own postwar coping style.

Discussions that the present author has had with Africans from various socioeconomic backgrounds have indicated that the issues of individual and collective grief and bereavement are much more important – both in terms of personal and economic consequences – in Africa than in the West. Asians seem to take an intermediate position in this regard.

The increase in refugee populations implies a progressive disruption of the natural social life of communities or families living in camps, which easily become "total institutions" in Goffman's sense. Learned helplessness and a "dependency syndrome" may develop (von Buchwald 1994). This is especially likely to happen in camps incorporating the same type of authoritarian regimes. Dependency may be reinforced by the international donor agencies that see refugees as helpless and vulnerable people, without considering the possibility that their very survival is due to ingenious coping strategies and resilience.

3
Interventions for Survivors of Persecution

3.1
Interventions for Adults and Children

In a rational world, the development of services for both adults and children should be based on public mental health considerations. The selection of treatment priorities in a society coping with the massive consequences of human-made violence would ideally take place with the help of the public mental health criteria leading to interventions discussed below.

3.1.1 Prevalence

The first criterion is the *prevalence* of the problem, preferably determined with the help of a culturally sensitive epidemiological survey. A survey is important to distinguish individuals who, once exposed to the traumatic events, have developed a disorder from those who have not. Epidemiology is also valuable in the design of services, because it provides prevalence estimates of the affected populations. Moreover, epidemiology can provide information on chronic symptoms as a result of type II trauma or continuous traumatic stress as well as on disability. Chronicity is a critical issue in arguing for secondary and tertiary prevention. As has been mentioned before, until now only three epidemiological studies have been performed in lower-income countries where violence is highly prevalent. Since probably less than 10% of the traumatized population seeks treatment, it is difficult to determine the size and the nature of psychosocial and mental health problems that have to be addressed by service providers. Finally, epidemiological research can generate information on risk factors and protective factors by studying sociodemographic and economic variables, gender, coping style, social network, and structural and functional social support, for example. Modification of these factors is an important objective of an effective intervention and prevention program as long as primary prevention in the form of putting an end to war is seen as an unrealistic option.

Ideally, an epidemiological survey should include a study on help-seeking behavior. Most studies on traumatic stress focus on populations under treatment. However, in order to understand the dynamic process

of adaptation to traumatic stress, one must not only look at those who seek treatment. This is even more important in war-affected areas, where "treatment" in the Western sense is often absent. In other words, to develop curative and preventive interventions, we need information both on the prevalence of disorders and on the way these disorders are distributed over a certain area. Studying help-seeking behavior provides information on the indigenous and allopathic services where people try to get support or, conversely, where they do not try to do so.

3.1.2 Community Concern

Community concern can be hard to assess in postwar circumstances. Victims may have been deprived of basic human needs such as shelter, water, or food for a considerable amount of time. They may be accustomed or even conditioned to ask for material help from (non)governmental organizations (NGOs). They may also be so accustomed to the effects of traumatic stress that a distortion has taken place of the population norm with regard to normality and deviancy. For example, during focus groups discussions in a war-ridden area in East Africa, the mothers stated that their children were not affected by the war. Since the results seemed hard to believe, the focus group discussions were repeated. It turned out that all children had night terrors and that all children up to the age of 12 were wetting their bed in an area where being potty trained at the age of 1 year is not exceptional. In addition to qualitative techniques such as focus groups or key informant interviews, community concern can be assessed by looking at the type of problems that people present to community services or health services in the affected areas. Their presentation of specific problems provides insight into their concerns which is complementary to the results of the aforementioned epidemiologic study or the research on help-seeking behavior. For example, in our work among Sudanese refugees in Uganda, the main problems presented by the refugees are poverty, forced marriage, violence within the family, suicide (attempt) and homicide, mental disorder and epilepsy, magic–religious illness such as Jok-Jok (spirit possession), alcohol abuse, acquired immunodeficiency syndrome (AIDS), rape, infertility, and "misled youth." Although most of these problems are related to the war activities and the daily problems in the refugee camps, and although many of the problems concur with epidemiologic data, these presented problems give vital information on intervention priorities as seen by the community. Another indication for community concerns are the requests being forwarded by community leaders among the refugees. These requests can be transmitted by district or provincial authorities to the government and subsequently to international agencies or organizations (see Sect. 3.1.7 on "Political Acceptability").

3.1.3 Seriousness

In postwar situations, general distress and minor psychological disorders may affect almost the whole population. These problems may thus be a priority based on the first criterion of (point) prevalence. However, the management of these psychosocial problems needs to be in balance with serious psychiatric disorders that are brought forward by the population. After a war, an increase in the incidence of major psychiatric disturbances is often seen, primarily caused by the absence of services during the war period. In addition, several groups among the population may show a different response if mental health services are offered. For example, when setting up mental health care services in Cambodia, we found that in areas with high concentrations of returnees from the Thai border camps, the consumption of mental health care services was many times higher than in areas that had never known any allopathic mental health care service over and above the services that were delivered by the local healers and monks. The reason for the difference was that, although both groups regarded mental disorder as a serious predicament, only the returnees from the border camps knew from their previous experience in Thailand that such a thing as allopathic treatment for mental disorder exists (Somasundaram et al. 1999).

3.1.4 Susceptibility to Management, Treatability, or Feasibility

The criterion of susceptibility to management, treatability, or feasibility is important both with regard to the question of whether people with certain problems get support from their environment and whether there are sufficient resources in terms of personnel, time, and funds to treat specific problems. For example, in some areas of Africa and Asia, the prevalence of epilepsy is as high as 3.7%–4.9% and it is often presented to psychiatric and primary health care services (Adamolekun 1995). In addition, especially in situations of massive stress, a large number of people show symptoms of dissociation varying from individual possession as an "idiom of distress" to classical fugue states and epidemics of mass psychogenic illness with or without psychogenic fits (de Jong 1987). While setting up services, it has to be considered which health care sector is best equipped to deal with the high prevalence of all kinds of convulsions.

Offering treatment to those with epilepsy is a feasible option. A total of 95% of a sample of West African patients with generalized epileptic convulsions were correctly diagnosed and treated with phenobarbital by primary health care workers who received a couple of hours of training; the average seizure frequency decreased from 16 to 0.34 per month (de Jong 1996). On the other hand, dealing with the equally highly prevalent dissociative states often requires sophisticated and scarce psychotherapeutic skills. In many cultures, adequate management for both groups implies triage of the epileptic patients and referral of those with dissociative states to the local healers or possession cults. A similar problem exists with regard to the treatment of complex PTSD as a result of type II trauma in war-affected areas. In the West, the psychotherapy of complex PTSD requires a long-term commitment from both therapist and client. In most war-affected areas, psychotherapists are not or hardly available and long-term therapy is mostly alien to the local culture. This is one of the reasons why mental health care professionals often have to resort to short therapies, limiting themselves to the stabilization phase of the three-phase model of Janet, for example (Van der Kolk et al. 1996; Meichenbaum 1997).

3.1.5 Sustainability

The sustainability of a program will depend primarily on the institutional capacity and the creation of enough human resource capacity to continue the interventions. To guarantee continuation of the activities after repatriation, the program should train as many (para)professionals as possible from the refugee community. (It is obvious that this type of empowerment also has a preventive effect on the community.) Sustainability is also important with regard to the provision of psychotropic drugs. As mentioned before, populations of refugees, displaced persons, or returnees will often contain individuals with serious psychiatric disorder or epilepsy that present to the mental health services. When setting up a program, it has to be remembered that these patients need to continue their medication after resettlement or repatriation.

In our view, donors should abstain from engaging themselves in mental health care services without a long-term commitment. Some donors customarily fund subsequent project phases on a 6-month basis. From an ethical point of view, one wonders how these donors expect mental health care professionals to assist traumatized people with psychosocial interventions, psychotherapy, or maintenance treatment if they have to disrupt their work every 6 months hoping for new funds to arrive.

Another aspect promoting the sustainability of a program is to take a politically neutral stance while implementing the program (see Sect. 3.1.7 on "Political Acceptability").

3.1.6 Knowledge, Skills, and Availability of (Mental) Health Care Professionals

Before setting up an intervention program on an intersectoral basis, the number and types of mental health care professionals, general health workers, and other possible trainees from other sectors such as education or social services should be assessed. This assessment should answer questions regarding their ability to handle different types of psychosocial and mental health problems, their normal duties and responsibilities, and what kind of training and supervision they need in order to be able to deal with the consequences of man-made violence. It should also determine the extent to which the trainees themselves have been traumatized by the war. Training should address their traumatic experiences, and after the training a second assessment should be carried out to find out whether the trainees are able to deal with the traumas of others. Group and individual debriefing, supervision and intervision, and job rotation are useful measures to prevent burnout.

3.1.7 Political Acceptability

It is important to find out the (hidden) agenda of policy makers and their opinion on the possible implications of the work. For example, both in epidemiological research and in stress research, it is important to measure traumatic stressors and life events before, during, and after the human-made disaster in order to measure the effect of these independent variables on psychosocial well-being and psychopathology. The results of this type of research play a central role in designing culturally appropriate interventions. However, the same data can also be used for other purposes, such as human rights work or advocacy against repressive government practices at home or in a guest country. Governments may therefore be ambivalent toward this kind of psychosocial and research activity. On the one hand, the activities may stimulate democracy, respect for human rights, and psychosocial support. On the other hand, governments may be afraid of being exposed as repressive. They may try to hinder a psychosocial program or, conversely, welcome human rights activ-

ists in case they themselves are imprisoned after a coup, thus needing the benefits of human rights that their previous enemies achieved.

3.1.8 Ethical Acceptability

The possible harm that might be inflicted on others, e.g. by carrying out research that lacks cultural sensitivity or by carrying out research that does not result in provision or improvement of services, should be considered. All too often scholars are eager to collect psychodiagnostic or psychometric data to be published without questioning whether the data will help in formulating preventive or curative interventions for the affected population.

Western-style informed consent with signatures on an elaborate consent form has to be discussed thoroughly before being applied. As Bromet (1995) argued after the Chernobyl accident, such forms may be perceived with distrust and suspicion. The procedure may evoke fear, e.g. of being disowned by one's country, as we found out in countries such as Cambodia, Ethiopia, or Gaza.

Another ethical consideration is the above-mentioned sustainability of the project. Psychosocial and mental health assistance requires a long-term commitment. In lower-income countries, it may take 5–7 years before a local training of trainers (TOT) group has been trained itself and before it has subsequently trained and supervised sufficient secondary- and tertiary-level staff to ensure continuity of the work.

3.1.9 Cultural Sensitivity

Culture defines reality for its members. It defines the purpose of the life of the individual and the group and prescribes and sanctions proper behavior. The beliefs, values, and behaviors of a culture provide its members with personal and social meaning, learned through tradition and transmitted from generation to generation. Culture serves two functions. It is integrative, i.e. represents the beliefs and values that provide individuals with a sense of identity. It is also functional, i.e. furnishes the rules for behavior that enable the group to survive and provide for its welfare, while supporting an individual's sense of self-worth and belonging. These two functions are analogous to the warp and woof of a tapestry (Kagawa-Singer and Chi-Ying Chung 1994). The weaving technique is universal, but the patterns that emerge from each culture are particular. A thread can be taken out and compared cross-culturally, but its function can only be understood within the cultural fabric from which it came.

As a result, each aspect of a public mental health intervention has to be tested for its cultural assumptions and consequences. For example, Green (1993) has suggested eight generic dimensions of trauma: (1) threat to life and limb; (2) severe physical harm or injury; (3) receipt of intentional injury/harm; (4) exposure to the grotesque; (5) violent/sudden loss of a loved one; (6) witnessing or learning of violence to a loved one; (7) learning of exposure to a noxious agent; (8) causing death or severe harm to another. Some of these dimensions (e.g. 1–3) can be regarded as universal stressors, but others are perceived differently in specific cultures. For example, in Uganda we found that group rape of abducted women can be dealt with by a collective purification ritual under the aegis of the elderly, whereas in Algeria, Cambodia, Nepal, or Namibia the shame caused by rape can lead to suicide or marginalization of the victim. Violence in the family is regarded as a serious consequence of continuous traumatic stress in many cultures, but some cultures are lenient toward battery violence, whereas in other cultures it is found to be unacceptable. Loss of an older loved person (dimension 5 and 6) who has children and some accumulated wealth can be acceptable in African animist cultures, since the person will travel to the reign of the ancestors. On the other hand, the death of a child in the same culture is a disaster, even though some Westerners think that parents suffer less in cultures with high exposure to child mortality. An individual who is a perpetrator (dimension 8) may be regarded as a hero if he is a (child) soldier in some West African countries or in the Middle East. Even exposure to the grotesque can be mediated by religious convictions such as the role of karma in Buddhism in Asia. These culturally determined subjective components are important in determining subsequent psychopathology. Therefore, the development of scales to quantify the severity of traumatic exposure is a complex issue necessitating collaboration with cultural informants and social scientists. The role of culture is equally important in the design of any psychosocial, psychotherapeutic, or psychiatric intervention, be it curative or preventative. This mean that local mental health care professionals – especially if they were trained abroad – should "indigenize" their knowledge and expertise in order to provide services to the victims of violence who often have a different cultural background and reside in peripheral rural areas. In my opinion, Western mental health care professionals who are equipped with the best of intentions but without work experience in low-income countries or without extensive experience in working with refugees in their home countries should be extremely cautious in offering their Western

culture-bound expertise in conflict-ridden areas elsewhere.

3.2 Translating Public Mental Health Priorities into Program Priorities

Organizations, either governmental or nongovernmental, usually choose a model to implement psychosocial and mental health care interventions in postwar situations. Their choice is often determined by the model which was consciously or unconsciously established by the psychiatric services in the area before the disaster took place. Siegler and Osmond (1974) distinguished eight possible models: the medical, moral, invalidity, psychoanalytic, social, psychedelic, conspiracy, and family interaction models. The models implemented in a specific area are often determined by social or colonial history and often need thorough transformation to be effective in postwar circumstances. For example, the Soviet approach, with its emphasis on medical authority, hospital-based care, and the conspiracy model, is an obstacle for adequate interdisciplinary community-based interventions in countries such as the former Yugoslavia, the Caucasian republics, Cambodia, or Vietnam. The centralized custodial care in Portugal's former colonies Mozambique and Angola does not provide a model to set up mental health care services in the rural areas afflicted by the war. The French psychoanalytic model which is en vogue in Algeria does not equip its mental health care professionals to do outreach work in the communities suffering from the sequelae of the violence and human rights violations. Similar considerations play a role in the treatment of refugees with mental problems in high-income countries. Some countries prefer to set up special or category-based services for refugees, other countries opt for integrated care, and in some countries refugees use the general psychiatric services but may be referred to services with a specialization in complex cultural psychiatric and psychotherapeutic problems.

In addition to the role of local models, the international (donor) community plays an important role in the choice of a specific approach. For example, because many Westerners can identify themselves with the suffering of people in the former Yugoslavia and because it is a politically sensitive area, it seems that currently more money is spent on the psychosocial consequences of the war in ex-Yugoslavia than in the rest of the world. If more funds are available, more mental health care professionals can be hired or trained, which easily results in a more or less classical dyadic or group psychotherapeutic approach within existing institutions or facilities with the possible disadvantage that large parts of the afflicted population do not receive proper care. If funds and professional manpower are limited, as in most lower-income countries, governments or NGOs often have to resort to cheaper public mental health care solutions with the possible flaw that semiprofessionals with limited training have to support large numbers of people suffering from serious sequential traumatization. The choice of a specific model is further complicated by the preferences or capacities of professionals or organizations. For example, there are large numbers of medical doctors worldwide who are well trained in public health but have little knowledge of psychotrauma treatment, while most psychiatrists and psychotherapists find it difficult to translate their professional expertise into a public mental health care perspective. Social scientists often hear that they are still so committed to time-consuming participant observation that the results of their inquiries lag behind in the turmoil of postwar areas. The choice of a model is also determined by organizations working in the field who have their own opinions on the implementation of mental health care services. The public discourse of all kinds of organizations is often quite different from their behavior in the field. For example, one of the United Nations (UN) agencies involved in helping refugees in Uganda had limited knowledge of current views on integrating different types of services within a horizontal public health care approach. When we started establishing mental health care services in the area, we were obliged to spend almost a million dollars to set up our own logistic infrastructure, including transport and telecommunication, to convince the UN agency that we would not "use" others to implement our work. In other words, we added another vertical program to a variety of initiatives in the medical field. After this investment, we had shown that we might be serious partners. This enabled us to begin dismantling our infrastructure in order to set up a suitable horizontally integrated public health care system sharing logistics with other organizations providing medical or social services.

3.3 Outline of a Public Mental Health Model for Massive Traumatization

The international multisite program of the Transcultural Psychosocial Organization (TPO or "Peace of Mind"), a World Health Organization (WHO) Collaborative Centre, is active in Algeria, Burundi, Cambodia, Congo, Ethiopia, Gaza, India (Tibetans), Kosovo, Mozambique, Namibia, Nepal, Sri Lanka, Uganda, and the Netherlands. The outline of its approach will be

described here as an example of how mental health care services can be set up in conflict areas. The TPO program is based on an interdisciplinary approach combining public mental health care, psychology, psychiatry, social science, psychotherapy, and epidemiology. Within a comprehensive approach, each discipline contributes to finding practical solutions for the complex problems of transcultural diagnosis and to designing effective, community-oriented interventions. This outline illustrates our multimodal approach, which can be used in a rather eclectic way. In this approach, three aspects of transcultural public mental health care programs are emphasized which, if neglected, may hamper the effectiveness of interventions. First, cultural variables and the context in which behavior occurs have to be studied both on the population and on the individual level in order to understand normal and deviant behavior. Second, the use of Western quantitative research instruments which are not based on culture-specific qualitative data should be avoided. This lack of cultural validation of instruments perpetuates the so-called category fallacy, in which indigenous diagnoses are overlooked and Western categories imposed where they have no cultural validity (Kleinman and Good 1985). Third, the long-term outcome of programs is often primarily determined by the original design. This is not realistic with a view to the major changes that may occur among the target population. The TPO model is a combination of an ecological, traumatic stress, and stress diathesis model (Lazarus and Folkman 1984). It uses a methodology which captures the idiomatic description of mental health problems that fit local cultural illness experiences in order to bolster indigenous coping strategies. Each project in the different countries develops its own culture-specific approach within this general outline or blueprint.

TPO generally recruits long-term professional expertise from abroad if local expertise is scarce. Local and foreign staff form an intercultural and interdisciplinary TOT team. Local staff may have little formal training, since most of the low-income countries where the program is active only have a few psychologists, psychiatrists, or social scientists. Even if these local professionals are available, they usually need additional training in topics such as psychosocial and mental health care interventions, group and individual trauma treatment modalities, crisis intervention, prevention, and human rights violations in order to cope professionally with the consequences of war, oppression, or civil unrest. Our intervention strategy uses a combination of Western and local methods. We exchange knowledge between the local members of the TOT group and the expatriate trainers. We try to prevent neocolonization of the non-Western mind by the uncritical use of Western trauma treatment techniques such as cognitive-behavioral or psychoanalytic approaches. Instead, we carefully build on local coping strategies and on healing techniques that match local expectations and practices by healers. In this way, syncretistic therapy methods are developed combining universal "common factors" (Frank and Frank 1993) or nonspecific therapy variables (Garfield and Bergin 1986) with culture-specific healing methods. This enables technology transfer about the consequences of trauma to a wide number of local practitioners and beneficiaries. We alleviate the psychological consequences of massive organized violence through the management of identified cases – both adults and children – and through the empowerment of the community. Collaboration is sought with local and international organizations in rural development, vocational skills training, and income-generating activities. Participatory action/research is carried out in order to evaluate and monitor our programs in each country, at the same time involving local staff to ensure a future capacity to evaluate the progress of programs once TPO leaves.

The program objectives are realized through a number of activities. Psychosocial and mental health problems are identified and assessed by means of a multimethod approach combining qualitative, ethnographic methods with quantitative epidemiological and psychometric methods. Using this specific approach (Agar and Murdoch 1994), we are able to detect priority problems within the communities and the different categories of people who will be trained by the program. This helps to elicit indigenous coping strategies; explanatory models on the level of the population and the individual, local idioms of distress; folk illnesses and culture-bound syndromes; proverbs, metaphors, and somatizations related to distress; and culturally mediated protective factors. The latter are used for training and reinforcing self-management and self-help activities within the framework of a psychosocial intervention program. The research and the training sessions produce additional culture-specific material on grief, trauma, and coping. This material is used to adapt the existing training manuals, e.g. for health care workers, teachers, relief workers, and healers. Part of the training is based on the WHO/UNHCR book on the *Mental Health of Refugees* (de Jong and Clarke 1996). Culturally appropriate adaptations of this book are produced for South-East Asia and East Africa. The material is also used to design preventive interventions in the field of psychosocial assistance and mental health (de Jong 1995). One of the preventive empowering strategies is to support local

authorities and camp managers in rural development based on a primary care approach emphasizing intersectoral collaboration. In addition, the program is monitored and the effectiveness of the interventions is improved by using variables developed during the action research.

Each country or participating area tailors the program to suit local circumstances while following the outline. Both research and interventions are essential. In order to constantly monitor and improve local program implementation, each project uses a combination of qualitative and quantitative research techniques such as participant observation, focus groups, key informant interviews, and epidemiologic surveys with psychodiagnostic and psychometric instruments. The outcome of the interventions is assessed with an effectiveness methodology. The results are fed back as a cybernetic loop in the ongoing intervention projects.

Each intervention model chosen by an organization or a government has to address the restoration of shattered assumptions. Janoff-Bulman (1992) mentions that we all have a culturally determined set of assumptions – or schemata – that guide our cognitions of reality and our plans and actions. A basic aspect is the illusion of personal invulnerability, which protects us from stress and anxiety but also from adopting preventive behavior. A second assumption is that we live in a meaningful world. By following certain rules and by behaving in a worthy way, we protect ourselves against misfortune. A third assumption is that we are decent and worthy people. We possess a positive self-image and self-esteem. When overwhelming trauma occurs, our social reality is often disrupted and our personal schema is also completely shattered. This results in strong emotions such as powerlessness, anger, and anxiety or, conversely, in their denial. Coping with stress implies coming to terms with the new reality and trying to fit new assumptions about the world with its harsh facts and with ourselves. Culturally defined coping strategies may serve to come to terms with shattered assumptions about the world and ourselves. This may happen by trying to redefine the traumatic event to make it consistent with previous reality or by making sense of the event by finding a purpose in it or by restoring the belief in an orderly world with the help of religion or by involvement in a democratization process. Alternatively, a survivor may try to change his or her behavior to regain autonomy and control. A person may seek social support, set up self-help groups, or decrease his or her dependency on authorities or organizations. This chapter has tried to provide guidelines to support survivors in finding a new homeostasis in times of war and persecution.

4 References

Aarts PGH, Op den Velde W (1996) Prior traumatization and the process of ageing. Theory and clinical implications. In: Van der Kolk B, McFarlane A, Weisaeth L (eds) Traumatic stress. The effects of overwhelming experience on mind, body and society. Guilford, New York, pp 359–377

Adamolekum B (1995) The aetiologies of epilepsy in tropical Africa. Trop Geogr Med 47(3): 115–117

Agar MH, Murdoch RO (1994) Investigating recent trends in heroin use in Baltimore city: a pilot 'quantitative' research project. Centre for Substance Abuse Research, Maryland

Allen T (1991) Understanding Alice: Uganda's holy spirit movement in context. Africa 61(3): 370–400

American Psychiatric Association (APA) (1980) Diagnostic and statistical manual of mental disorders, 3rd edn. American Psychiatric Association, Washington DC

Arroyo W, Eth S (1995) Children traumatised by Central Am warfare. In: Pynoos RS, Eth S (eds) Posttraumatic stress disorder in children. American Psychiatric Association, Washington DC, pp 103–120

Baker A (1990) The psychological impact of the intifada on Palestinian children in the occupied West Bank and Gaza: an exploratory study. Am J Orthopsychiatry 60: 496–505

Boothby N, Upton P, Sultan A (1991) Children of Mozambique: the cost of survival. United States Committee for Refugees, Washington DC

Bowlby J (1981) Attachment and loss. Pelican, London

Brom D, Kfir R, Dasberg H (1994) A controlled double-blind study on the offspring of Holocaust survivors. Poster presented at the Annual Conference of the Society of Traumatic Stress Studies, Chicago, 23 November 1994

Bromet EJ (1995) Methodological issues in designing research on community-wide disasters with special reference to Chernobyl. In: Hobfoll SE, de Vries MW (eds) Extreme stress and communities: impact and intervention. Kluwer, Dordrecht, pp 267–283

Cohon JD (1981) Psychological adaptation and dysfunction among refugees. Int Migr Rev 15: 255–275

Corcoran JDT (1982) The concentration camp syndrome and USAF Vietnam prisoners of war. Am J Psychiatry 10: 991–994

Corradi J, Weiss Fagen P, Garreton M (1992). Fear at the edge: state terror and resistance in Latin America. University of California Press, Berkeley

Danieli Y (1985) The treatment and prevention of long-term effects and intergenerational transmission of victimization: a lesson from holocaust survivors and their children. In: Figley CR (ed) Trauma and its wake: The study and treatment of posttraumatic stress disorder. Brunner/Mazel, New York, pp 295–313

Dawes A (1990) The effects of political violence on children: a consideration of South African and related studies. Int J Psychol 25: 13–31

de Girolamo G, McFarlane AC (1996) The epidemiology of PTSD: a comprehensive overview of the international literature. In: Marsella AJ, Friedman MJ, Gerrity ET, Scurfield RM (1996) Ethnocultural aspects of posttraumatic stress disorder. American Psychological Association, Washington DC, pp 35–36

de Jong JTVM (1987) A descent into African psychiatry. Royal Tropical Institute, Amsterdam

de Jong JTVM (1995) Prevention of the consequences of man-made or natural disaster at the (inter)national, the community, the family and the individual level. In: Hobfoll SE, de Vries MW (eds) Extreme stress and communities: impact and intervention. Kluwer, Boston

de Jong JTVM (1996) A comprehensive public mental health programme in Guinea-Bissau: a useful model for African, Asian and Latin-American countries. Psychol Med 26: 97–108

de Jong JTVM, Clarke L (eds) (1996) Mental health of refugees. World Health Organization, Geneva

*Desjarlais R, Eisenberg L, Good B, Kleiman A (1995) World Mental health. Problems and priorities in low-income countries. Oxford University Press, New York

*Dohrenwend BP (ed) (1998) Adversity, stress and psychopathology. Oxford University Press, New York

Dyregrof A, Raundalen M (1993) A longitudinal study of war-exposed children in Iraq. International Conference on Mental Health and the Challenge of Peace, Gaza, 13–15 September 1993

El Sarraj E, Punamäki RL, Salmi S, Summerfield D (1996) Experiences of torture and ill-treatment and posttraumatic stress disorder symptoms among Palestinian political prisoners. J Traum Stress 9: 595–606

Figley CR, Kleber RJ (1995) Beyond the "victim". In: Kleber RJ, Figley CR, Gersons BPR (eds) Beyond trauma: cultural and societal dynamics. Plenum, New York, pp 75–95

Forbes Martin S (1992) Refugee women. Zed, London

Frank JD, Frank JB (1993). Persuasion and healing: a comparative study of psychotherapy. Johns Hopkins University Press, Baltimore

Friedman MJ, Jaranson J (1994)The applicability of the post-traumatic stress disorder concept to refugees. In: *Marsella AJ, Bornemann T, Ekblad S, Orley J (eds) Amidst peril and pain. American Psychological Association, Washington DC

Garfield SL, Bergin AE (eds) (1986) Handbook of psychotherapy and behavior change, 3rd edn. Wiley, New York

Geffray C (1990) La cause des armes au Mozambique. Anthropologie d'une guerre civile. Karthala, Paris

Goldberg J, True WR, Eisen SA, Henderson WG (1990) A twin study of the effects of the Vietnam war PTSD. JAMA 263: 1227–1232

Grant B (1979) The boat people and "age" investigation. Penguin, New York

Green BL (1993) Identifying survivors at risk: trauma and stressors at cross events. In: Wilson JP, Raphael B (eds) International handbook of traumatic stress syndromes. Plenum, New York, pp 135–144

Harrell-Bond B (1986) Imposing aid: emergency assistance to refugees. Oxford University Press, Oxford

Harrell-Bond B (1991) Proceedings of the International Conference on The Mental Health and Well-being of the World's Refugees and Displaced Persons, Stockholm, 8–11 October 1991

Helzer JE, Robbins LN, McEvoy L (1987) Post-traumatic stress disorder in the general population: findings of the Epidemiological Catchment Area Survey. N Engl J Med 317: 1630–1634

Hovens JE (1994) Research into the psychodiagnostics of PTSD. Eburon, Delft

Jablensky A, Marsella S, Ekblad B, Jansson L, Levi L, Bornemann T (1994) Refugee mental health and well-being: conclusions and recommendations. In: Marsella AJ, Bornemann T, Ekblad S, Orley J (eds) Amidst peril and pain. American Psychological Association, Washington DC

Janoff-Bulman R (1992) Shattered assumptions: towards a new psychology of trauma. Free Press, New York

Jeyaraja Tambiah S (1992) Buddhism betrayed? Religion, politics, and violence in Sri Lanka. University of Chicago Press, Chicago

Kagawa-Singer M, Chi-Yung Chung R (1994) A paradigm for culturally based care in ethnic minority populations. J Community Psychol 22: 192–208

Kaličanin P, Bukelić, Išpanović-Radojković, Lečic-Toševski D (eds) (1993) The stresses of war. Institute for Mental Health. Belgrade, pp 139–159

Kardiner A (1941) The traumatic neuroses of war. Hoeber, New York

Keehn RJ (1980) Follow-up studies of World War II and Korean conflict prisoners. Am J Epidemiol 111: 194–211

Kinzie JD, Manson S (1983) Five years' experience with Indochinese refugee psychiatric patients. J Operation Psychiatry 14: 105–111

Kinzie JD, Sack WH, Angell RH, Manson S, Rath B (1986) The psychological effects of trauma on Cambodian children. I. The children. J Am Acad Child Adolesc Psychiatry 25: 370–376

Kinzie JD, Sack WH, Angell RH, Clarke G, Ben R (1998) A three-year follow-up of Cambodian young people traumatised as children. J Am Acad Child Adolesc Psychiatry 28: 501–504

Kleinman A, Good B (1985) Culture and depression. University of California Press, Berkeley

Krystal H (1988) Integration and self-healing: Affect, trauma and alexithymia. Analytic Press, Hillsdale

Kulka RA, Schlenger WA, Fairbank JA, Hough RL, Jordan BK, Marmar CR, Weiss DS (1991) Trauma and the Vietnam war veteran generation. Brunner/Mazel, New York

Lazarus RS, Folkman S (1984) Stress, appraisal and coping. Springer, Berlin Heidelberg, New York

Leopold M, Harrel-Bond B (1994) An overview of the world refugee crisis. In: Marsella AJ, Borneman T, Ekblad S, Orley J (eds). Amidst peril and pain. American Psychological Association, Washington DC, pp 17–33

Levav I (1998) Individuals under conditions of maximum adversity: the holocaust. In: Dohrenwend BP (ed) Adversity, stress and psychopathology. Oxford University Press, New York

Litz BT, Roemer L (1996) Post-traumatic stress disorder: an overview. Clin Psychol Psychother 3(3): 153–168

Lomranz J (1995) Endurance and living: long-term effects of the Holocaust. In: Hobfoll SE, de Vries MW (eds) Extreme stress and communities: impact and intervention. Kluwer, Dordrecht, pp 325–353

Long LD (1992) Ban Vinai: the refugee camp. Columbia University Press, New York

Macksoud M (1993) Assessing war trauma in children. J Refugee Stud 1

McCubbin HI, Dahl BB, Lester G, Ross B (1977) The returned prisoner of war: factors in family reintegration. J Marriage Family 39: 471–478

McFarlane AC, de Girolamo G (1996) The nature of traumatic stressors and the epidemiology of posttraumatic reactions. In: Van der Kolk B, McFarlane A, Weisaeth L (eds) Traumatic stress. The effects of overwhelming experience on mind, body and society. Guilford, New York

Mahjoub A (1995) Approche psychosociale des traumatismes de guerre chez les enfants et adolescents palestiniens. Editions de la Méditerranée, Tunis

*Meichenbaum D (1997) Treating post-traumatic stress disorder. A handbook and practice manual for therapy. Wiley, New York

Mghir R, Freed W, Raskin A, Katon W (1995) Depression and posttraumatic stress disorder among a community sample of adolescent and young adult Afghan refugees. J Nerv Ment Dis 183: 124–130

Minter W (1989) The Mozambican National Resistance (Renamo) as described by ex-participants. Dev Dialog 1

Mollica R, Wyshak G, Lavelle K (1987) The psychosocial impact of war trauma and torture on South East Asian refugees. Am J Psychiatry 144: 1572–1576

Mollica R, Donelan K, Svang TOR, Lavelle J, Elias C, Frankel M, Blendon RJ (1993) The effect of trauma and confinement on functional health and mental health status of Cambodians living in Thailand-Cambodia border camps. JAMA 270(4): 581–586

MoorenGTM, Kleber RJ (1996) Late gevolgen van de Tweede Wereldoorlog onder Indische jeugdige oorlogsgetroffenen: een onderzoek naar gezondheid en verwerking van de oorlogsjaren. Gedrag Gezondheid 24: 224–232

Nguyen SD (1984) Mental health services for refugees and immigrants. Psychiatr J Univ Ottawa 9: 85–91

Op den Velde W, Hovens JE, Falger PR, de Groen JHM, van Duijn H, Lasschuit LJ, Schouten EGW (1993) PTSD in Dutch resistance veterans from World War II. In: Wilson JP, Raphael B (eds) International handbook of traumatic stress syndromes. Plenum, New York, pp 219–230

Paardekoper BP, de Jong JTVM, Hermanns JMA (1999) The psychological impact of war and the refugee situation on South Sudanese children in refugee camps in northern Uganda: an exploratory study. J Child Psychol Psychiatry (in press)

Peak WE, Robinowitz R, Roberts WR et al (1981) Adjustment differences among male substance a varying in degree of combat experience in Vietnam. J Consult Clin Psychol 49: 426–437

Punamäki RL (1982) Childhood in the shadow of the war. A psychological study on attitudes and emotional life of Israeli and Palestinian children. Curr Res Peace Violence 5: 26–41

Punamäki RL (1989) Political violence and mental health. Int J Ment Health 17(4): 3–15

Punamäki RL (1996) Can ideological commitment protect children's psychosocial wellbeing in conditions of political violence? Child Dev 67: 55–69

Pynoos RS, Nader K (1993) Issues in the treatment of posttraumatic stress in children and adolescents. In: **Wilson JP, Raphael B (eds) International handbook of traumatic stress syndromes. Plenum, New York

Qouta S, Punumäki RL, El Sarraj E (1995) The relations between traumatic experiences, activity, and cognitive and emotional responses among Palestinian children. Int J Psychol 30(3): 289–304

Rakoff V, Sigal JJ, Epstein N (1965) Children and families of concentration camp survivors. Can Ment Health 14: 24–26

Ramsay R, Gorst-Unsworth C, Turner S (1993) Psychiatric morbidity in survivors of organised state violence. Br J Psychiatry 162: 55–59

Roodenrijs TC, Scherpenzeel RP, de Jong JTVM (1998) Traumatische ervaringen onder Somalische vluchtelingen in Nederland [Traumatic experiences among Somalian refugees in the Netherlands]. Tijdschr Psychiatr 40(3): 132–142

Rosenbaum M, Ronen T (1992) How did Israeli children and their parents cope with the threat of daily attacks by scud missiles during the Gulf war? Ministry of Education Conference on the Stress Reaction of Children in the Gulf War, Rmat Gan, Israel

Rundell JR, Ursano RJ, Holloway HC et al (1990) Psychiatric responses to trauma. Hosp Community Psychiatry 40(1): 68–74

Russel A (1980) Late effects – influence on the children of the concentration camp survivors. J Contemp Psychother 4(2): 87–94

Salimovich S, Lira E, Weinstein E (1992) Victims of fear: the social psychology of repression. In: Corradi J, Weiss Fagen P, Garreton M (1992) Fear at the edge: state terror and resistance in Latin America. University of California Press, Berkeley, pp 72–89

Sartorius N (1994) Psychologisch-psychiatrische Problemen von Migranten und Flüchtlingen. Symposium of the German Psychiatric Association on Transcultural Psychiatry, 19–21 February, Reichenau, Germany

Shrestha NM, Sharma B, van Ommeren M, Regmi S, Makaju R, Komproe I, Shrestha C, de Jong JTVM (1998) Impact of torture on refugees displaced within the developing world: symptomatology among Bhutanese refugees in Nepal. JAMA 280(5): 1–6

Siegler M, Osmond H (1974) Models of madness, models of medicine. Harper and Row, New York

Sigal JJ, Silver D, Rakoff V, Ellin B (1973) Some second generation effects of survival of the Nazi persecution. Am J Orthopsychiatry 43: 320–327

Silver SM, Iacono CU (1984) Factor-analytic support for DSM-III's PTSD for Vietnam veterans. J Clinical Psychol 40: 5–14

Solkoff N (1992) Children of survivors of the Nazi Holocaust: a critical review of the literature. Am J Orthopsychiatry 62: 342–358

Solomon Z (1994) The pathogenic effects of war stress: the Israeli experience. In: Hobfoll SE, de Vries MW (eds) Extreme stress and communities: impact and intervention. Kluwer, Boston

Somasundaram D, van de Put W, Eisenbruch M, de Jong JTVM (1991) Introducing mental health services in Cambodia. Soc Sci Med 48(8): 1029–1042

Steinberg I (1995) Treating the loss of a child. In: Figley CR, Mazze N, Bride B (eds) Death and trauma. Brunner/Mazel, New York

Straker G, Moosa F, Becker R, Nkwale M (1992) Faces in the revolution. Philip, Cape Town

Summerfield D, Toser L (1991) 'Low intensity' war and mental trauma in Nicaragua: a study in a rural community. Med War 7: 84–99

*Tedeschi RD, Parc CL, Calhoun LG (eds) (1998) Post-traumatic growth. Positive changes in the aftermath of crisis. Erlbaum, London

Tennant GC, Goulsen K, Dent O (1993) Medical and psychiatric consequences of being a prisoner of war of the Japanese: an Australian follow-up study. In: **Wilson JP, Raphael B (eds) International handbook of traumatic stress syndromes. Plenum, New York, pp 231–240

Toole MJ, Waldman RJ (1993) Refugees and displaced persons: war, hunger and public health. JAMA 270: 600–605

UNHCR (1993) The state of the world's refugees: the challenge of protection. Penguin, New York

United States Committee for Refugees (USCR) (1992) World refugee survey. Author, Washington DC

**Van der Kolk B, McFarlane A, Weisaeth L (1996) Traumatic stress. The effects of overwhelming experience on mind, body and society. Guilford, New York

Van der Velden PG, Kleber RJ (1999) Children of war victims: a controlled study. J Traum Stress (submitted)

Van der Velden PG, Eland J, Kleber RJ (1994) De Indische naoorlogse generatie [The Indonesian post-war generation]. Bohn Stafleu Van Loghum, Houten

Ventura MMF (1997) O stress postraumático e suas sequelas nos adolescentes do sul de Angola. PhD thesis. Minho University, Minho, Portugal

Vines A (1991) Renamo terrorism in Mozambique. Indiana University Press, Bloomington

von Buchwald U (1994) Refugee dependency: origins and consequences. In: Marsella AJ, Bornemann T, Ekblad S, Orley J (eds) Amidst peril and pain. American Psychological Association, Washington DC

Westermeyer J (1985) A clinical guide to alcohol and drug problems. Praeger, Philadelphia

Westermeyer J (1986) Planning mental health services for refugees. In: Williams C, Westermeyer J (eds) Refugees' mental health issues in resettlement countries. Hemisphere, New York

Westermeyer J (1988) DSM-III psychiatric disorders among Hmong refugees in the United States: a point prevalence study. Am J Psychiatry 145: 197–202

Westermeyer J (1989) Psychiatric care of migrants: a clinical guide. American Psychiatric Press, Washington DC

Wilson KB (1992) Cults of violence and counter-violence in Mozambique. J S Afr Stud 18: 531–582

Yesavage JA (1983) Dangerous behaviour by Vietnam veterans with schizophrenia. Am J Psychiatry 140: 1180–1183

Yip R, Sharp TW (1993) Acute malnutrition and childhood mortality related to diarrhea: lessons from the 1991 Kurdish refugee crisis. JAMA 270: 587–590

CHAPTER 19

L. Tata Arcel, I. Genefke, M. Kastrup

Psychiatric Problems Related to Torture

1 Introduction

Torture is a widespread human rights problem practised in 115 out of 215 or in 53.5% of all countries (Amnesty international 1997). The medical and psychological after-effects on its victims and their families now constitute a public health problem demanding concerted efforts from the health systems in all parts of the world. A special concern for mental health care providers in European countries are the mental health needs of traumatised torture survivors among refugees granted asylum after wars and persecution from repressive regimes in Europe and elsewhere (e.g. the former Yugoslavia, Eastern Europe, Turkey, the Middle East, south-east Asia). Baker (1992) estimated that between 5% and 35% of the world's 14 million refugees had suffered at least one experience of torture. Research on torture and its consequences is needed for several important reasons.

Torture is the most abhorrent form of human rights violation. The aims of torture have always been to destroy the personality of the victim and at the same time set an example for the rest of the community (Genefke and Vesti 1998; Basoglu et al. 1997). Scientists can contribute to turning society's and the media's attention to the health problems connected with the issue, thus raising public awareness and bringing pressure to bear on governments and international organisations to combat the practice of torture. Studying the trauma of torture as a deliberate attack aimed at destroying an individual's identity gives us valuable insights into the effects of extreme stress and the processes of traumatisation and coping. Thus the study of torture trauma and its mental health implications contributes to developing methods of psychiatric assessment and classification of trauma-related symptoms and to refining stress theory.

Following the introduction of the diagnosis of post-traumatic stress disorder (PTSD) in DSM-III in 1980, a generic theory unifying previous approaches to the long-lasting psychological effects of trauma was proposed. According to this theory, different types of traumata such as combat experience, child abuse, sexual abuse and torture may provoke similar psychopathology. However, PTSD as diagnostic category has been shown to be inadequate to encompass the entire symptomatology of torture survivors (Shresta et al. 1998; Arcel 1998).

This chapter will present the current knowledge of psychological and psychiatric problems in torture survivors world-wide. Within this frame, it will discuss whether symptoms following torture constitute a specific syndrome qualitatively different from other stress response syndromes (Rasmussen 1990; Genefke and Vesti 1998). The chapter will furthermore give a clinical description of how torture trauma affects emotion, consciousness, memory, self and interpersonal relationships. Finally, treatment principles for torture-related psychiatric problems will briefly be reviewed.

2 Definition and Types of Torture

A widely accepted definition of governmental torture is stated in the United Nation's Convention against Torture and Other Cruel, Inhuman or Degrading Treatment or Punishment (United Nations General Assembly 1992):

> The term torture means any act by which severe pain or suffering, whether physical or mental, is intentionally inflicted on a person for such purpose as obtaining from him or a third person information or a confession, punishing him for an act he or a third person has committed, or is suspected of having committed, or intimidating him or coercing him or a third person, or for any reason based on discrimination of any kind, when such pain is inflicted by or by the instigation of, or with the consent or acquiescence of a public official or other person acting in an official capacity.

Torture affects not only the individual victim, but also the entire community by creating a culture of fear that intimidates and terrorises. In repressive cultures, the mere fact that an individual has been detained and tortured is enough to stigmatise that individual, creating mistrust for his possible disloyalty (Turner and McIvor 1997). Relationships in families that have been tortured together at home or during detention are burdened with strong and contradictory feelings of shame, guilt, aggression and humiliation on the one hand and compassion and identification on the other. Often torture is a secret that families share.

High-risk target groups as regards governmental torture are members of the political opposition, human rights activists, ethnic minorities, refugees, labour union leaders, student leaders, journalists, criminals or ordinary civilians when arrested and interrogated.

We differentiate between primary victims, i.e. those subjected to direct torture, and secondary victims, i.e. their families. The torturers are soldiers, police officers, prison officers and militia. Torturers are

trained in a variety of techniques of physical and psychological abuse.

2.1 Prevalence

As no large epidemiological studies exist in the community, the rate has been measured as the number of torture survivors in refugee populations, recognising that refugees are high-risk populations. The percentages vary between 20% and 70% according to the selection of the groups. It was 20% in a random sample of 3000 of the 10,000 asylum seekers who arrived in Denmark in 1986 (Jepsen 1988). Typically, the prevalence is higher in refugee populations referred to psychiatric clinics, e.g. 70% in males and 31% in females in a selected outpatient psychiatric clinic in Norway between 1991 and 1995 (Lavik et al. 1996). Montgomery (1998) found 30% tortured in a sample of Middle Eastern asylum seekers, and 20% were tortured in a Bosnian refugee population of 2023 included in a psychosocial programme in Croatia (Arcel et al. 1998).

2.2 Torture Methods

For practical reasons, methods have been divided into physical and psychological, although this division is artificial, as many of the psychological methods, e.g. water, food and sleep deprivation, involve not only psychological but also physiological sequelae, and many physical methods, e.g. sexual torture, involve psychological elements too.

The most common physical methods of torture are blunt violence, "falanga" (persistent beating of the soles of the feet), suspension (by the arms or legs), strapping of body or body parts with rope or handcuffs, electrical torture, suffocation, sexual torture carried out by humans or animals, insertion of foreign bodies into body openings, cuts with sharp instruments, teeth torture, immobilisation in forced positions, detention in cells smaller than the human body, burns with cigarettes or acidic liquids, "telephono" (beating on the ears with cupped hands), cold-water showers, pulling of hair and exposure to extreme heat or extreme cold.

Psychological methods include blindfolding and isolation, severe humiliation, sexual torture, forced nakedness, death threats, verbal abuse, mock executions, witnessing torture of others or family members, solitary confinement, exhaustion, deprivation of food, water and sleep and being forced to listen to high-pitched sounds or music. As the knowledge of physical torture increases, more sophisticated, mainly psychological methods of torture are being developed by torturers in order to prevent documentation through physical sequelae.

3 Physical Sequelae

Physical torture methods leave symptoms and physical signs. Certain types of torture are related to specific symptoms and signs. Chronic and late sequelae are related to the severity of the applied method. These sequelae are often the basis for a documentation of physical torture. It is important to examine the victim close to the time of torture, as traces may disappear with time. In contrast, the psychological symptoms described later in this article are persistent. (For a detailed listing of physical sequelae related to methods of torture, see Skylv 1992; Holtan 1998.) As torture is a violent attack on the body, many torture victims have poor physical health and many medical complaints and initially describe physical rather than psychological symptoms. It is therefore important to mention here a few physical sequelae.

The sequelae can be fresh or badly healed fractures, ostitis, periostitis, nerve and vessel injury, fibrosis in muscles fasciae and connective tissue, injury to tendons and ligaments, distortions and scars from beatings, whippings, cuts with a knife or other sharp instruments, burns or electrical torture. Other symptoms may include sexual dysfunction, damaged teeth, the whiplash syndrome, punctured eardrums and hearing loss. In survivors of falanga, we find smashed heels, and the rate of walking is slow and the distance limited. Alopecia is found following electrical torture. Torture survivors' chronic pain and tension influences the general tension of the musculature, resulting in fibrositis, fibromyalgia and myofascial pain. Former prisoners of war who have been subjected to torture are reported to have increased cumulative incident rates of chronic disorders of the peripheral nervous system, joints and back and an increased rate of peptic ulcer (Nice et al. 1996). There is increasing recognition of the chronic nature of many of the sequelae, in particular chronic pain.

Sexual torture may leave traces in the musculoskeletal system, structural injuries, functional disturbances and dysfunction of the pelvic joints in women. These women often have low lumbar pain, complain about pains in genitalia, menstrual disturbances and sexual problems. Sexual dysfunction and testicular atrophy in men are seen following electrical torture on genitals.

4 Medical Assessment

Important themes in the medical assessment preceding psychiatric assessment and treatment include the following:

- The medical assessment of the torture victim is an important part of the entire diagnostic process and preparation for the psychological treatment. Chronic medical conditions or distracting physical symptoms should be identified and assessed before any psychological treatment is initiated. Differences in type of injury following torture are small between different parts of the world despite the fact that particular forms of torture may prevail in certain cultures. Specialities such as traumatology, sports medicine, forensic medicine, neurophysiology and rheumatology provide the basis for the diagnosis and treatment of sequelae of physical torture.
- Torture victims have many medical complaints. Pains in the head, back, abdomen or chest are often the major complaints. Pain can recur years after torture and must be addressed. However, the complaints require physical evaluation in a slower, more deliberate and less invasive approach in torture victims than in others because the pain may be a manifestation of a primary psychiatric problem.

  In a study of 50 victims, Juhler and Smidt-Nielsen (1995) found that there was a discrepancy between the large number of physical complaints and the small number of objective medical findings, with the exception of musculoskeletal findings, as shown in Table 1.

  The survivors may feel reassured to know that their symptoms are organically unfounded, as they often fear that the torture has inflicted damage on them.

Table 1. Discrepancies between percentage of physical complaints and medical findings in 50 victims of torture

Type of complaint	Complaints (%)	Findings (%)
None	0	4
Neurological	86	22
Cardiopulmonary	74	14
Gastrointestinal	68	24
Urological	34	4
Sexual/genital	54	16
Musculoskeletal	92	92

- As some examinations may resemble the torture situation, medical procedures must be discussed with the survivor. Especially vulnerable are patients with a history of electrical torture or teeth extraction.
- The physical evaluation can document the history of torture and the methods used.
- The role of the primary care physician is less threatening for the survivor, thus creating a bridge of trust for further psychiatric care. When a patient is being treated for urgent physical symptoms or medical conditions, psychotherapy can begin.

5 Psychological Sequelae

During the last two decades, an increasing body of research and theory has been produced in many, mainly Western countries where torture victims among refugees have been investigated. Terms such as complex PTSD (Herman 1992), chronic organic psychosyndrome (Petersen et al. 1985) and torture syndrome (Genefke and Vesti 1998) have been used to denote the complexity of torture trauma. Important questions that need to be addressed include whether torture constitutes a specific syndrome, i.e. whether it evokes a predictable constellation of signs and symptoms. Initially, attempts to identify the key features of the effects of torture were subsumed under the heading of "torture syndrome" proposed by Danish doctors (Amnesty international, Danish Medical Group 1977). This proposal was based on a combination of thorough medical examinations of the physical indications of torture in most of the major organ systems and a psychiatric examination of the short-term and long-term psychological effects of both physical and psychological torture in treatment seekers.

At an early stage, it was concluded by the Rehabilitation and Research Centre for Torture Victims (RCT) in Denmark that the worst sequelae of torture were psychological, as confirmed by other international studies (Allodi 1980; Kordon et al. 1988; Rasmussen 1990). Researchers in favour of a distinct torture syndrome claim that the issue of infliction of deliberate violence on one human being by another with the aim of destruction has a decisive influence on the formation of symptoms (Somnier et al. 1992). Gelinas (1993) supports this point by differentiating between "facticity" (e.g. breaking a leg in an accident) and "agency" as causes of trauma. According to Gelinas, it is the malevolent intention of the perpetrator that causes the

most extreme trauma. So far, the answers are inconclusive, primarily due to major methodological problems in research.

Many studies do not differentiate sufficiently between symptoms connected with the difficulties of being a refugee in exile and symptoms related to the torture experience. As very few epidemiological studies exist, investigations are hampered by selection bias by investigating treatment seekers. The definition of torture varies from investigation to investigation, and the degree of severity is not adequately operationalised. Severity measured by the number of incidents is too crude a measure (Holtz 1998). Head trauma as a possible aetiological factor in symptoms is rarely examined, and impairment may thus be interpreted as psychologically founded. Personality traits, e.g. resilience, ego strength, attachment style and others important for coping with torture trauma, are very rarely investigated. The numerous health problems caused by physical torture are not clearly differentiated from psychosomatic symptoms. The importance of other traumatic events, e.g. dramatic loss of family members, divorce, unemployment, exile and separations as additive to torture stress, is often not investigated.

However, it has now been established that torture is a very important life stress event that causes numerous other stress events. The survivors may lose their physical health, work, family, status in the family and in society and finally, if exiled, their country, language and cultural environment (Mollica et al. 1987). The most commonly reported psychiatric and psychological symptoms in many different parts of the globe include anxiety, depression, irritability, emotional instability, cognitive memory and attention problems, personality changes, behavioural disturbances, neurovegetative symptoms such as lack of energy, insomnia, nightmares, and sexual dysfunction even if the person has not been sexually tortured (Mollica and Caspi-Yavin 1992; Cunningham and Cunningham 1997; Holtz 1998).

In Table 2, we present the most frequent psychiatric symptoms observed by the RCT psychotherapists. The classification of the symptoms in the DSM-IV diagnostic categories (American Psychiatric Association 1994) is shown.

Nine of the most common symptoms are encompassed by the diagnostic criteria for PTSD. It is interesting to note that symptoms "change in personality" and "survivor's guilt" are included in the revision of PTSD in DSM-IV (American Psychiatric Association 1994) under the heading of associated descriptive symptoms. They are mentioned among a number of symptoms that are more commonly seen in association with an interpersonal stressor in criterion A.

In the WHO International Classification of Diseases of 1992, it is emphasised that post-traumatic stress disorder may show a chronic course over many years and a transition to an enduring personality change. The diagnostic guidelines emphasise that the change must be present for at least 2 years and should not be attributable to a pre-existent personality disorder or to a mental disorder other than PTSD.

The personality change includes inflexible and maladaptive features with impairment in interpersonal, social and occupational functioning.

6 Post-traumatic Stress Disorder and Co-morbidity

The introduction of PTSD has shifted the focus of investigators away from demonstrating the presence of a unique syndrome and given impetus to substantial research on torture documenting in most investigations that PTSD is found in a significantly higher degree in refugees with torture experiences compared to refugees without (Shresta et al. 1998; Holtz 1998; Priebe and Esmaili 1997). In connection with tortured individuals, it is important to consider the neurobiological aspects of PTSD, as the threat to the organism leads to the activation of multiple neurobiological systems. It has been shown that, among individuals with PTSD, a higher level of adrenaline and noradrenaline is found than in non-traumatised individuals. Another finding is the increased activity of the autonomic nervous system, with abnormal sleep architecture and elevated blood pressure and heart rate (Kaplan et al. 1994). The numbing phase often seen in PTSD may be partly explained by repeated depletion of catecholamines (Saporta and van der Kolk 1992). The same authors claim that noradrenergic dysregulation might explain the difficulty in modulating aggression found in PTSD victims.

Van der Kolk and Greenberg (1987) suggest that autonomic arousal mediated by the locus coeruleus may account for the increase in flashbacks and nightmares in PTSD victims.

There is an increasing recognition of the importance of neurological factors in understanding the connections between PTSD and trauma. At the same time, PTSD has been shown to be inadequate when it comes to encompassing the entire symptomatology seen in torture trauma victims. Co-morbidity is the rule and in most investigations includes the diagnoses of anxiety disorder, major depression and somatoform disorder (Shresta et al. 1998; Holtz 1998). It has additionally been shown that different forms of torture produce different PTSD symptoms (Ramsay et al. 1993). The majority of symptoms reported by individuals who suffered isolation or blindfolding, impact torture and

Table 2. Twelve major symptoms seen by psychotherapists at the Rehabilitation Centre for Torture Victims (RCT) in Denmark and location of symptoms within diagnostic categories of DSM-IV

PTSD criterion	Symptom (DSM-IV)[a]	Description or comment
Emotional lability	D2, D4	Irritability or outbursts of anger, hypervigilance
Sleep disturbances	B2, D1	Nightmares, difficulty falling asleep or staying asleep
Disturbances in ability to concentrate/remember	C3, D3	Psychogenic amnesia, difficulty concentrating
Avoidance of thoughts or feelings associated with torture experience	C1	
Avoidance of activities or situations that arouse recollections of torture experience	C2	
Diminished ability to establish personal relationships	C5, 6	Feelings of detachment or estrangement from others, restricted range of affect
Markedly diminished interest in several significant activities	C4	
Sense of foreshortened future	C7	
Sudden acting or feeling as though torture situation were recurring	B3, 4	Flashbacks occur after exposure to events or environmental stimuli that symbolise or resemble aspect of torture experience
Change in personality		
Survivor guilt		
Anxiety		Anxiety that is specific to torture experience and is neither neurotic nor psychotic

Empty cells indicate that there is no associated DSM-IV diagnostic category or no description/comment.
PTSD, post-traumatic stress disorder.
[a]B1, recurrent and intrusive distressing recollections of the event, B2, recurrent and distressing dreams of the event; B3, acting or feeling as if the traumatic event were recurring; B4, intense psychological distress at exposure to internal or external cues that symbolise or resemble an aspect of the traumatic event; B5, physiological reactivity on exposure to internal or external cues that symbolise or resemble an aspect of the traumatic event; C1, efforts to avoid thoughts, feelings or conversations associated with the trauma; C2, efforts to avoid activities, places or people that arouse recollections of the trauma; C3, inability to recall an important aspect of the trauma; C4, markedly diminished interest or participation in significant activities; C5, feeling of detachment or estrangement from others; C6, restricted range of affect; C7, sense of a foreshortened future; D1, difficulty falling or staying asleep; D2, irritability or outbursts of anger; D3, difficulty concentrating; D4, hypervigilance, D5, exaggerated startle response.

other forms of physical torture were PTSD intrusion symptoms, whereas individuals who had been sexually tortured described more avoidance phenomena. This may suggest that PTSD is not a uniform syndrome in tortured individuals. Vesti and Kastrup (1995) concluded that a substantial proportion of survivors developed symptomatology close to the entity of PTSD, whereas others did not. Current results indicate that the strongest predictors of PTSD symptoms are the perceived severity but not objective severity of torture and the impact of the captivity experience on the family and the quality of social support in post-captivity, as shown by Basoglu et al. (1994a), who compared tortured political activists with non-political tortured victims in Turkey. Threats and humiliation and being forced to watch others being tortured were factors predicting PTSD among resettled refugees in Australia (Cunningham and Cunningham 1997).

The largest investigation of torture survivors to date world-wide, comparing tortured with non-tortured Bhutanese refugees in Nepal, found the diagnoses of PTSD, depression and anxiety to be significantly more common in the tortured group (Shresta et al. 1998). The total number of torture techniques experienced was the only significant predictor of PTSD and depression. The presence of PTSD, depression and anxiety leads the authors to propose the classification "disorders of extreme stress - not otherwise specified" as being a better way to describe the sequelae of torture than PTSD. Their investigation identified symptoms of shame, mistrust, concerns, unexplained pain, the conviction of being permanently injured, many medical complaints, loss of appetite, sleep disturbances and loss of sexual desire.

Political involvement, being prepared for torture, spirituality and an inner locus of control during torture and imprisonment appeared to play an active role in the development of protective coping methods

(Basoglu et al. 1997; Shresta et al. 1998; Holtz 1998; Qouta and el Saraj 1997). PTSD symptoms were associated with severe physical torture combined with a lack of effective support in male refugees from Iraq staying in the United Kingdom (Gorst-Unsworth and Goldenberg 1998). The prevalence of obsessional symptoms in the group was 20%. Poor social support rather than severity of trauma was a predictor of depression in the long term. Feelings of meaninglessness and "existential dilemmas" were widely reported in the study group.

In conclusion, there is now agreement that simply describing torture survivors as having PTSD is an inadequate description. PTSD in torture survivors is not inevitable, but genetic factors such as vulnerability to stress or resilience, proneness to anxiety, developmental deficits, previous psychiatric history, incapacitating physical sequelae, quality of social environment and individual coping efforts play a decisive role. It is beyond doubt that torture exposes people to the risk of developing psychiatric symptoms and thus social problems. The more prolonged, repeated, and unpredictable the experience of torture is, the more traumatic it is and the more serious the psychiatric symptoms are.

The complexity of the torture trauma means that neither PTSD nor a single torture syndrome can provide an adequate diagnostic basis. Connections between the general stress theory, head injury literature and results on torture symptoms need to be correlated in theoretical models of diagnosis and treatment.

7
Torture as Extreme Stress

Stress research has highlighted the universal human automatic response of "fight or flight" under threat accompanied by the emotions rage and fear, which have common physiological similarities (Cannon 1953). The "fight or flight" response is triggered not only in the original traumatic situation but also by memories. When the mind detects a threat, the sympathetic nervous system is activated and the individual is instantly prepared to fight back or, if he or she considers the danger to be too great, to flee. Inability to either fight back or flee before, during and after torture or when re-experiencing the traumatic memories leaves the individual in a state of extreme physiological arousal and with the emotions of fear and rage that cannot find an outlet. Mental images of the torture create arousal in the brain and other body parts, with accompanying feelings of rage and fear. The physiological activation causes anxiety despite the fact that a threat no longer exists. It is therefore more accurate to talk about traumatic reactions as psychosomatic reactions, as there is no psychological stress without involved physiological stress. There is no doubt that extreme traumatic experiences alter the central nervous system and the physiological system generally, which can explain the symptom of hyperarousal in torture victims. As Basoglu also emphasises, "no other trauma seems to come closer to the experimental paradigms used to explore the processes of traumatisation in animals and humans" (Basoglu et al. 1997).

The hypothesis from the stress theory about the "dose-response" relationship is well confirmed by torture research. However, the results on the protective factors mentioned earlier show that the subjective interpretation of the objective experience contributes to the development of trauma. The more life-threatening the torture situation is experienced as, the more traumatising it becomes, and, similarly, the more helpless a person has felt while subjected to torture, the more traumatising the effect will be. Of all the variables examined in a study comparing political victims and non-political victims of torture, it was shown that the fact that the non-political victims were less psychologically prepared for torture was by far the strongest predictor of psychological problems (Basoglu et al. 1997).

The literature on severe stress, trauma and coping gives us valuable insights into the impact of torture on attachment, emotion, consciousness, memory, self and interpersonal relationships (Allen 1995; Lazarus and Folkman 1984; Lazarus 1991). Disturbances in these areas create vicious circles and expose the individual to the risk of developing psychiatric symptoms.

7.1
Attachment

Torture disturbs attachment to other people, as it disturbs the sense of safety and trust. Behnia (1997) emphasises that distrust in torture victims makes rehabilitation and resettlement a difficult and painful process. Attachment developmentally has the function not only to protect the individual by care-giving, but also to regulate physiological arousal, as Bowlby (1973) has shown. Gradually, the regulation of the physiological functioning, e.g. through self-soothing activities, becomes internalised during normal development.

Torture disturbs not only basic trust and the sense of safety, but also physiological regulation, as it generates hyperarousal beyond the normal range. Self-soothing activities are severely affected and may be impaired. It goes without saying that torture can

disturb even sound foundations of attachment in childhood. If the tortured individual has developmental deficits in self-regulation or in attachment, it is clear that the processes of restoring physiological arousal regulation and of regaining trust in other human beings will be more lengthy and difficult and in some cases may never be achieved. An enduring and dependable relationship with other human beings – the mental health care provider being one of them – is necessary in order to restore safety and to relearn regulation through self-soothing techniques, e.g. relaxation, visualisation.

7.2 Emotion

Despite coping efforts, emotions in torture survivors are intense and prolonged and lose their signal value for adaptation (Lazarus 1991).

Chronic anxiety, for example, means that the individual is preoccupied with controlling the emotion itself and thus turns his or her focus away from solving the problems that trigger the anxiety. Uncontrollable anxiety is connected with feelings of helplessness and a loss of control, both experienced intensely during captivity and torture.

Torture evokes anger and aggression in the "fight or flight" response, which cannot be expressed during captivity and in many instances in very few other places afterwards, unless the individual feels secure. The family is usually a secure base where repressed feelings can be expressed. However, outbursts of anger counteract the restoration of attachment and mutual trust.

The individual may regain a feeling of control and power in destructive aggression in the family, but it eventually leaves him or her in a vicious circle of guilt and self-hatred.

If current situations are reminiscent of the torture trauma, hostile destructiveness may be expressed in the maltreatment of family members or in self-destructive tendencies.

7.3 Self-Image and Body Image

Self-image and body image are negatively affected by torture. The helplessness and humiliation often trigger feelings of being weak and dirty, with a damaged and distorted body. The feeling of shame is at the core of the torture trauma. Guilt is felt for surviving or for acts of disloyalty during captivity and torture, i.e. confession, not being heroic. Defensive measures against shame, caused especially by sexual torture, are withdrawal and avoidance, isolation from others.

Shame and rage are interrelated in the same spiral. Shame evokes rage. Rage means being out of control and evokes shame. This spiral also contributes to violence in the family and elsewhere.

Chronic stress can lead to emotional exhaustion, resulting in depression and in a lack of energy, anhedonism and a general disengagement in the world and previous activities.

Prolonged helplessness and hopelessness can eventually lead to learned helplessness. Loss and a sense of failure lead individuals to blame themselves for what they have experienced.

7.4 Consciousness and Memory

The ability to alter one's own consciousness differs from person to person, but trauma increases an individual's ability to dissociate him- or herself from traumatic memories as a protective mechanism. However, the dissociated experiences have a tendency to return intrusively in memory flashbacks and nightmares. Moreover, dissociation increases detachment from the outer world. Dissociation may protect the individual against painful experiences, but it also constricts the emotional life of the individual, leaving him or her with poor symbolisation. In addition, dissociation interferes with cognitive processes and learning capacity and creates gaps in self-experience. Restoring contact with reality re-associates the fragments of visual images, memories and bodily sensations (Herman 1992).

Extreme stress may impair memory such that sounds, smells and images do not occur in sequence. Tortured individuals have too strong a memory in one sense, e.g. images are brought to mind too easily, but too weak a memory in another sense, as images are fragmented and can be experienced with feelings of unreality. The blurred memory of the past can affect the memory of current things (Allen 1995).

7.5 Self and Relationships

Herman (1992) points out that, in victims of severe trauma, all structures of the self, i.e. the image of the body, the internalised images of others and self, the values and ideals that give a sense of coherence, are invaded systematically and broken down. A victim of repeated and prolonged trauma may even lose the sense of self. Traumatic experiences lead to internal

conflicts and conflicts with other people, which render the self insecure and affect the capacity to act.

8 Psychological Assessment

The effects presented above have implications for assessment, and many initial barriers may emerge between clinician and patient. Traditional clinical methods of asking torture survivors to relate their trauma in their own words are frequently ineffective (Mollica and Caspi-Yavin 1992).

The survivor may not reveal all the symptoms or tell the whole story of torture. The effect of trauma on memory, possibly impaired memory caused by a head injury, denial and avoidance of painful memories, cultural norms about revealing painful experiences, fear of contaminating the clinician with the evil of torture and language barriers are all problems that the clinician may face during the assessment. Language barriers can be overcome through specially trained interpreters, who should be regular members of the medical team. Clinicians, especially inexperienced ones, may be overwhelmed by strong emotions when interviewing a torture victim about his or her experiences. They may feel emotional resistance in hearing the horrors of torture and concern about creating severe emotional distress or re-traumatisation in the patient. Another pitfall in both the assessment and treatment of tortured individuals is over-identification with the survivor's suffering and over-idealisation of his or her former activities. Supervision on these counter-transference reactions is necessary.

Anamnesis and history taking must show sensitivity and allow the survivor to control the amount of information given, since establishment of a trusting relationship has a high priority (Jacobsen and Smidt-Nielsen 1997).

A trusting relationship is enhanced if the assessment includes not only symptoms and history of torture but also the person's more important life experiences, personal values, former and current life situation, family situation and practical problems impeding his or her reintegration in society, be it the original one or a new host society.

Concerning the exact information on torture, the application of a checklist of traumatic experiences, including types of torture and symptoms, may ease communication during assessment and is to be recommended rather than open-ended interviews on these issues. The use of structured interviews and diagnostic instruments presents additional advantages for reassessment and research purposes.

Questionnaires commonly used to measure trauma symptoms include the Allodi trauma scale (Allodi 1985), the Mississippi Combat-Related PTSD (Keane et al. 1988), the Impact of Event Scale (Zilberg et al. 1982) and the Harvard Trauma Questionnaire (Mollica et al. 1991).

9 Treatment

Psychodynamic therapy and counselling, cognitive behavioural therapy and individual and group therapy are suitable for torture survivors (Basoglu 1992). There are as yet no clear research results on the effectiveness of different interventions. None of the above-mentioned methods has proved to be better than the others. The length of treatment varies; the insight therapy carried out at the RCT involves 20 to 80 sessions, while behaviour therapy involves ten to 20 sessions.

Most rehabilitation centres for torture survivors in countries where torture has been or is practised and in countries of resettlement base their work on a multidisciplinary principle with a broad spectrum of medical, psychiatric, psychological, physiotherapeutic and social care. A total of 126 centres and programmes are currently providing services for victims of torture in 54 countries (Network Coordination Division 1998). Most centres are run by non-governmental organisations funded by a range of sources. In most countries, the centres are dependent on international funds, local donations, the United Nations's Voluntary Fund for Torture Victims and money from the European Union or national governments. The scarcity of financial support often creates disruptions in the flow of money and in the continuity of treatment, and significant professional time is taken up by fund-raising activities. There is now increasing emphasis on community-based interventions and on developing mainstream health systems in order to address the needs of tortured individuals.

The supportive elements prevail in psychotherapy, especially during the initial phases. Survival and resettlement needs may initially dominate more than torture experiences or psychiatric symptoms. However, social support for a survivor with PTSD may not be of great use if the person is unable to profit from it due to psychological disturbances (Basoglu 1992). The therapist must be prepared for great fluctuations in psychological functioning depending on the phase of trauma exploration in therapy or stressors experienced in the victim's life as an exile in a strange culture with financial, employment, housing and educational problems (Somnier et al. 1992). Reconstructing and retelling traumatic experiences will always weaken defence

mechanisms and revive disturbing and often aggressive feelings towards the therapist and family members.

It is imperative for the therapist to find a balance between denial and the "conspiracy of silence" surrounding horrendous experiences and premature pressure on the patient to tell his or her story. It is important that survivors are allowed to tell their story at their own pace in order to avoid re-traumatisation. Torture victims are very likely to experience the therapist as an interrogating torturer (Tocilj-Šimunković and Arcel 1998). As bad psychological functioning impairs the survivor's whole quality of life and family life, some therapists suggest that the main objective of therapy is not total reconstruction of the trauma or symptom reduction, but rather increased functionality in achieving personal goals. The achievement of limited personal goals will create a good basis for further trauma reconstruction.

Cultural issues prevail when a therapist works with torture survivors who are refugees. They are often quite alien to psychological treatment and insist on the treatment of physical symptoms as the main objective of therapy; such patients need a thorough explanation of the nature of psychotherapy. Expression of emotion may be a social taboo in males from many cultures, as it is understood as a loss of control or "feminine" behaviour.

The core issues in any therapy independent of its theoretical orientation are as follows:

- Establishing a working alliance in a trusting relationship in order to restore faith in humankind and to re-establish attachment
- Recollection and cognitive reconstruction of traumatic experiences under the safe conditions of therapy
- Strengthening the individual's coping mechanisms by including the current life situation and its conflicts in therapy
- Attribution of meaning to the torture experience and helping the survivor to discard the "victim role"

The victim's family may be included in some phases of therapy. Muscular tension may be treated with physiotherapy and psychotherapy and may subsequently reach levels corresponding to those of a healthy working adult by the end of the treatment. During physiotherapy, the survivor is gradually made to understand the connection between his or her symptoms and the vulnerable points at which the torturers directed the torture.

Almost all torture survivors seeking treatment, i.e. both refugees and others, will need practical assistance along with medical and psychological treatment. Focusing on survivors' current life situation, family, personal values, on existential dilemmas created by their possible ideological disillusionment or the opposite and on the advantages and disadvantages of their role in their country's political situation is just as important for torture victims as the medical assessment and treatment.

Medication has been found to be effective in major depressions, obsessive–compulsive disorders, anxiety and psychotic symptoms. In a review article, Smith et al. (1998) reached the conclusion that tricyclic antidepressants and monoamine oxidase inhibitors produce modest improvements in positive symptoms of PTSD.

10 Prevention

In the medical work for the eradication of torture, the public health model with its distinction of three levels of prevention was the first to be developed and has been widely recognised. The primary focus aims at society and tries to prevent the practice of torture. In order to achieve this, many strategies may be used, including identification of agents of torture, the legal basis and the actions to be taken. Secondary prevention focuses on high-risk groups, e.g. medical and law-enforcement personnel, and may include arranging training activities for these groups. Tertiary prevention aims to provide rehabilitation at an individual level for those who have been exposed to human rights violations and for their families (Arcel et al. 1995; Arcel 1994, 1998). The prohibition of torture from a legal point of view is regarded as an imperative rule of international law, and specific international and national mechanisms have been devised to prevent torture. The emphasis here is on the state's obligation to ensure that the prohibition of torture becomes an integral part of the training of civilian and military law-enforcement personnel and medical personnel and others involved in the custody, treatment and interrogation of detainees. During medical examinations of victims by doctors from many countries connected with the International Rehabilitation Council for Torture Victims (IRCT)/Rehabilitation and Research Centre for Torture Victims (RCT) programme, it was particularly disturbing to discover the extent of participation in torture by medical doctors and other health care professionals who were present during the administration of torture, who determined the intensity and who decided when to continue or to stop (Vesti et al. 1998).

Doctors at risk are prison doctors who are often asked to help the torturers and forensic doctors who are tempted or directed to falsify death certificates and to minimise the extent of physical and psychological sequelae. Military doctors are put under pressure to conceal the atrocious activities of their colleagues (Rasmussen 1990; Eitinger and Weisaeth 1998; Pross 1988). International professional support networks can facilitate regulation of the professional ethical problems of doctors at risk and support these physicians in their roles as witnesses of conscience.

Seen in a developmental perspective, torture and organised violence is closely related to the political, structural and cultural conditions in a society. As a consequence, preventive interventions may be linked to a change in the underlying sociopolitical causes and to the creation of the necessary conditions for human rights and development at the level of society.

11 References

Allen JG (1995) Coping with trauma: a guide to self-understanding. American Psychiatric Press, Washington DC

Allodi F (1980) The psychiatric effects in children and families of victims of political persecution and torture. Dan Med Bull 27(5): 229–232

Allodi F (1985) Physical and psychiatric effects of torture: Canadian study. In: Stover E, Nightingale EO (eds) The breaking of bodies and minds: torture, psychiatric abuses and the health professions. Freeman, New York, pp. 66–78

American Psychiatric Association (1994) Diagnostic and statistical manual of mental disorders, 4th edn (DSM-IV). American Psychiatric Association, Washington, DC

Amnesty international (1997) Report 1997. Amnesty international, London

Amnesty international, Danish Medical Group (eds) (1977) Evidence of torture: studies by the Amnesty International Danish Medical Group. Amnesty International, London

Arcel LT (ed) (1994) War victims, trauma and psycho-social care. ECTF, Nakladnistvo Lumin, Zagreb

**Arcel LT (ed) (1998) War violence, trauma and the coping process. Armed conflict in Europe and survivor responses. International Rehabilitation Council for Torture Victims (IRCT), Copenhagen

Arcel LT, Folnegović-Šmalc V, Kozarić-Kovačić D, Marušić A (eds) (1995) Psycho-social help to war victims: women refugees and their families. International Rehabilitation Council for Torture Victims (IRCT), Copenhagen

Arcel LT, Folnegović-Šmalc V, Tocilj-Šimunković G, Kozarić-Kovačić D, Ljubotina D (1998) Ethnic cleansing and post-traumatic coping – war violence, PTSD, Depression, anxiety and coping in Bosnian and Croatian refugees. A transactional approach. In: Arcel LT (ed) War violence, trauma and the coping process. Armed conflict in Europe and survivor responses. International Rehabilitation Council for Torture Victims (IRCT), Copenhagen, pp 45–78

Baker R (1992) Psychosocial consequences for tortured refugees seeking asylum and refugee status in Europe. In: Basoglu M (ed) Torture and its consequences: current treatment approaches. Cambridge University Press, Cambridge, pp 83–106

**Basoglu M (ed) (1992) Torture and its consequences. Current treatment approaches. Cambridge University Press, Cambridge

Basoglu M (1993) Prevention of torture and care of survivors. An integrated approach. JAMA 270(5): 606–611

Basoglu M, Paker M, Paker O, Ozmen E, Marks I, Incesu C, Sahin D, Sarimurat N (1994a) Psychological effects of torture: a comparison of tortured with non-tortured political activists in Turkey. Am J Psychiatry 151(1): 76–81

Basoglu M, Paker M, Özmen E, Tasdemir Ö, Sahin D (1994b) Factors related to long-term traumatic stress responses in survivors of torture in Turkey. JAMA 272(5): 357–363

Basoglu M, Mineka S, Paker M, Aker T, Livanou M, Gök S (1997) Psychological preparedness for trauma as a protective factor in survivors of torture. Psychol Med 27: 1421–1433

Basoglu M, Jaranson J, Mollica R, Kastrup M. Torture and its consequences. In: Gerrity E, Keane T, Tuma F (eds) Mental health consequences of torture and related violence and trauma. Plenum, New York (in press)

Behnia B (1997) Distrust and resettlement of survivors of war and torture. Int J Ment Health 25(4): 45–58

Bowlby J (1973) Attachment and loss. 2. Separation. Basic Books, New York

British Medical Association (1992) Medicine betrayed: the participation of doctors in human rights abuse. Zed, London

Cannon WB (1953) Bodily changes in pain, hunger, fear and rage, 2nd edn. Branford, Boston

Cunningham M, Cunningham JD (1997) Patterns of symptomatology and patterns of torture and trauma experiences in resettled refugees. Aust NZ J Psychiatry 31: 555–565

Eitinger L, Weisaeth L (1998) Torture – history, treatment, and medical complicity. In: Jaranson JM, Popkin MK (eds) Caring for victims of torture. American Psychiatric Press, Washington, pp 3–14

Gelinas DJ (1993) Relational patterns in incestuous families, malevolent variations, and specific interventions with the adult survivor. In: Paddison PL (ed) Treatment of adult survivors of incest. American Psychiatric Press, Washington, pp 1–34

Genefke I, Vesti P (1998) Diagnosis of governmental torture. In: Jaranson JM, Popkin MK (eds) Caring for victims of torture. American Psychiatric Press, Washington, pp 43–59

Gorst-Unsworth C, Goldenberg E (1998) Psychological sequelae of torture and organised violence suffered by refugees from Iraq: trauma-related factors compared with social factors in exile. Br J Psychiatry 172: 90–94

Herman JL (1992) Trauma and recovery. Basic Books, New York

Holtan NR (1998) How medical assessment of victims of torture relates to psychiatric care. In: Jaranson JM, Popkin MK (eds) Caring for victims of torture. American Psychiatric Press, Washington, pp 107–113

Holtz TH (1998) Refugee trauma versus torture trauma: a retrospective controlled cohort study of Tibetan refugees. J Nerv Ment Dis 186(1): 24–34

Jacobsen L, Smidt-Nielsen K (1997) Torture survivor – trauma and rehabilitation. International Rehabilitation Council for Torture Victims (IRCT), Copenhagen

**Jaranson JM, Popkin MK (eds) (1998) Caring for victims of torture. American Psychiatric Press, Washington DC

Jepsen S (1988) The general health of asylum seekers. The Danish experience. In: Miserez D (ed) Refugees – the trauma of exile. The humanitarian role of the Red Cross and Red Crescent. Nihoff, Boston

Juhler M, Smidt-Nielsen K (1995) Identification of torture survivors: a comparative study of medical complaints and findings in torture survivors. In: Caring for survivors of torture. Challenges for the medical and health professions. Abstracts of papers presented at the VIIth International Symposium, 15–17 November 1995, Cape Town

Kaplan H, Sadock B, Grebb J (1994) Synopsis of psychiatry, behavioural sciences, clinical psychiatry, 7th edn. Williams and Wilkins, Baltimore

Keane TM, Caddel JM, Taylor KL (1988) Mississippi Scale for combat-related posttraumatic stress disorder: three studies in reliability and validity. J Consult Clin Psychol 56(1): 85–90

Kordon DR, Edelman LI, Lagos DM et al (1988) Psychological effects of political repression. Sudamerican/Planeta, Buenos Aires

Lavik NJ, Hauff E, Skrondal A, Solberg O (1996) Mental disorder among refugees and the impact of persecution and exile: some findings from an outpatient population. Br J Psychiatry 169: 726–732

Lazarus RS (1991) Emotion and adaptation. Oxford University Press, New York

Lazarus RS, Folkman S (1984) Stress, appraisal, and coping. Springer, Berlin Heidelberg New York

Mollica RF, Caspi-Yavin Y (1992) Overview: the assessment and diagnosis of torture events and symptoms. In: Basoglu M (ed) Torture and its consequences: current treatment approaches. Cambridge University Press, Cambridge, pp 253–274

Mollica RF, Wyshak G, Lavelle J (1987) The psychosocial impact of war trauma and torture on South East Asian refugees. Am J Psychiatry 144: 1567–1572

Mollica RF, Caspi-Yavin Y, Bollini P, Truong T, Tor S, Lavelle J (1991) The Harvard trauma questionnaire: validating a cross-cultural instrument for measuring torture, trauma and posttraumatic stress disorder in Indochinese refugees. J Nerv Ment Dis 180(2): 11–16

Montgomery E (1998) Refugee children from the Middle East. Scand J Soc Med Suppl 54: 1–154

Network Coordination Division (1998) Rehabilitation of torture victims. Update on centres and programmes worldwide. International Rehabilitation Council for Torture Victims (IRCT), Copenhagen

Nice DS, Garland CF, Hilton SM, Baggett JC, Mitchell RE (1996) Long-term health outcomes and medical effects of torture among US Navy prisoners of war in Vietnam. JAMA 276(5): 375–381

Petersen HD, Abildgaard U, Daugaard G, Jess P, Marcussen H, Wallach M (1985) Psychological and physical long-term effects of torture. A follow-up examination of 22 Greek persons exposed to torture 1967–1974. Scand J Soc Med 13: 89–93

Priebe S, Esmaili S (1997) Long-term mental sequelae of torture in Iran – who seeks treatment? J Nerv Ment Dis 185(2): 74–77

*Pross C (1988) Wiedergutmachung: der Kleinkrieg gegen die Opfer. Athenäum, Frankfurt am Main

Qouta S, el Sarraj E (1997) Prison experiences and coping styles among Palestinian men. Peace and conflict. J Peace Psychol 3(1): 19–36

Ramsay R, Gorst-Unsworth C, Turner S (1993) Psychiatric morbidity in survivors of organised state violence including torture. A retrospective series. Br J Psychiatry 162: 55–59

Rasmussen OV (1990) Medical aspects of torture: torture types and their relation to symptoms and lesions in 200 victims, followed by a description of the medical profession in relation to torture. Dan Med Bull 37[Suppl 1]: 1–88

Saporta JA Jr, van der Kolk BA (1992) Psychobiological consequences of severe trauma. In: Basoglu M (ed) Torture and its consequences: current treatment approaches. Cambridge University Press, Cambridge, pp 151–181

**Shresta NM, Sharma B, van Ommeren M, Regmi S, Makaju R, Komproe I, Shrestha GB, de Jong JTVM (1998) Impact of torture on refugees displaced within the developing world. Symptomatology among Bhutanese refugees in Nepal. JAMA 280(5): 443–448

Skylv G (1992) The physical sequelae of torture. In: Baþoðlu M (ed) Torture and its consequences: current treatment approaches. Cambridge University Press, Cambridge, pp 38–55

Smith MW, Cartaya OJ, Mendoza R, Lesser IM, Lin K (1998) Conceptual models and psychopharmacological treatment of torture victims. In: Jaranson JM, Popkin MK (eds) Caring for victims of torture. American Psychiatric Press, Washington, pp 149–169

Somnier F, Vesti P, Kastrup M, Genefke IK (1992) Psycho-social consequences of torture: current knowledge and evidence. In: Basoglu M (ed) Torture and its consequences: current treatment approaches. Cambridge University Press, Cambridge, pp 56–71

Tocilj-Šimunković G, Arcel LT (1998) Group psychotherapy with victims of torture. In: Arcel LT (ed) War violence, trauma and the coping process. Armed conflict in Europe and survivor responses. International Rehabilitation Council for Torture Victims (IRCT), Copenhagen, pp 143–154

Turner SW, McIvor R (1997) Torture. In: Black D, Newman M, Harris-Hendriks J, Mesey G (eds) Psychological trauma. A developmental approach. Bell and Bain, Glasgow, pp 205–215

United Nations General Assembly (1992) The convention against torture and other cruel, inhuman or degrading treatment or punishment. Committee Against Torture. United Nations, New York (fact sheet no. 17)

**van der Kolk BA, Greenberg MS (1987) The psychobiology of the trauma response: hyperarousal, constriction, and addiction to traumatic reexposure. In: van der Kolk BA (ed) Psychological trauma. American Psychiatric Press, Washington, DC

Vesti P, Kastrup M (1995) Treatment of torture survivors: psychosocial and somatic aspects. In: Freedy JR, Hobfoll SE (eds) Traumatic stress: from theory to practice. Plenum, New York

Vesti P, Helweg-Larsen K, Kastrup M (1998) Preventing the involvement of physicians in torture. In: Jaranson JM, Popkin MK (eds) Caring for victims of torture. American Psychiatric Press, Washington, DC, pp 185–199

Zilberg NJ, Weiss DS, Horowitz MJ (1982) Impact of event scale: a cross-validation study and some empirical evidence supporting a conceptual model of stress response syndromes. J Consult Clin Psychol 50: 407–414

CHAPTER

20

V. Folnegović-Šmalc

Psychiatric Problems Related to Violence and Rape

1 Introduction

For a long time, the focus of psychiatric interest in violence and rape has been on the personality of violators and/or rapists, and the discussion has been restricted mostly to forensic papers or textbooks. As for victims, the focus was on the victimizing role in the act of violence and therefore on the description of their mental condition prior to and during the act. Only recently have the consequences of violence and rape on the mental condition of victims and their environment come into the centre of attention (Burgess and Holstrom 1974). The description of the mental condition of victims and the corresponding diagnostic categories – severe stress (F43) and enduring personality change after catastrophic experience (F62.0) – were first introduced by the World Health Organization (WHO) in the International Classification of Diseases 10 (WHO 1992). Much attention in modern psychiatry is given to children and adolescents who have been victims of violence and/or rape, since they often suffer more severe and enduring consequences than adults. Provided that the intervention is adequate and timely, the development of severe and enduring consequences in children can often be successfully prevented (Kocijan-Hercigonja et al. 1996).

2 Epidemiology

Even though violence (and rape as a subcategory of violence) has been part of society since history began to be recorded, the incidence of violence seems to have increased over the past decades. Modern wars that rage not only on the front lines, but also behind the lines, the genocide that took place during the twentieth century (e.g. Jews in the Second World War, Bosnians and Croats in the recent war in the former Yugoslavia) and increased antisocial civilian behaviour may all have contributed to the increased incidence of violence.

During the recent war in Croatia, there were 650,000 refugees and displaced persons; 8646 civilians were wounded, of whom 11% were children and 25% women. The number and nature of traumatic events were studied in some refugee camps, and the results show that 40% of civilians were physically maltreated, 10% witnessed killing and 10% were raped or witnessed rape (Folnegović-Šmalc and Folnegović 1998). Children are often victims in modern wars. In the war in Croatia alone, 268 children were killed, 971 wounded and 900 lost both and 5497 one parent (Kocijan-Hercigonja et al. 1996).

In the United States, violence increased substantially from 1958 to 1968: homicide by 52%, rape by 71% and robbery by 143%. According to the Uniform Crime Reports, 1,932,274 violent crimes were committed and 109,002 individuals were raped in the United States in 1992 (Tardiff 1995).

Post-traumatic stress disorder (PTSD) was initially studied in Vietnam veterans. The prevalence of PTSD in Vietnam veterans up to 19 years after the war was 15% (Kulka et al. 1990), with an even higher rate in wounded veterans (Helzer et al. 1987).

Some very significant epidemiological data are presented in studies of monozygotic twins. Of the twins who had served in Vietnam, 16.8% had PTSD, while only 5% of the twins who had not served had PTSD (Goldberg et al. 1987). Trauma survivors' risk of PTSD may be related to positive heredity for PTSD (Watson et al. 1995). These data point to the need for further investigation of possible genetic factors in the development of mental consequences of trauma.

3 Factors Influencing the Development of Mental Consequences

Victims of violence experience severe psychotrauma that can lead to the development of serious mental problems. The consequences of violence for the victim depend on a number of factors, such as the nature of the violence and the way it is performed, the duration and number of psychotraumatic events, who the violence was performed by, the victimizing role of the victim, the victim's personality structure, mental condition and past diagnosis (Frank and Anderson 1987) and previous traumatic or other experience, the age, gender and educational level of the victim, the general medical, social and financial consequences of violence and their reversibility, the length of time since the violence has occurred, the support provided by the family and environment and primary prevention.

3.1 Nature of Violence and the Way It Is Performed

The nature of violence and the way it is performed generally reflects the personality structure and cultural and financial background of violators and their relationship towards the victim. Physical maltreatment is most common in poor and primitive environments. If the victim is a woman, and in some cases in men as well, rape often takes place. In better-off environments, violence is often performed in a more sophisticated way (e.g. conducting biological experiments).

Rape is a very common form of violence in which sexuality is in the service of non-sexual needs and wishes, used to express power, anger and domination. Sexuality can also be the basic motive for rape, but this is rarely the case. Rapists in these cases are usually mentally disturbed people, their victims are often children or the elderly and sexual act is often incestuous.

Rape can be public or without witnesses, and there can be one or more rapists. It can be combined with other forms of mental or physical maltreatment, or it may be the only form of abuse. Trauma can be collective or single. Mass public rape is commonly used in war situations to terrify the enemy and to achieve certain goals of war strategy, such as ethnic cleansing. In public rape, there can be two different groups of victims. Direct or active victims are those who have been raped, while indirect or passive victims are those who witnessed the rape or were threatened to be raped as well. There is a significant difference between the two groups from the gynaecological point of view, but there is often not much difference from the psychiatric point of view.

Raphael (1986) found higher rates of PTSD after collective traumatic situations than after single trauma. According to the results of a study of the connection between the development of PTSD and the nature of psychotraumatic events (Folnegović-Šmalc and Folnegović 1998), PTSD developed in more than 80% of individuals who had been publicly raped, wounded or witnessed killing (Fig. 1).

3.2
Duration and Number of Psychotraumatic Events

The severity of mental symptoms is considered to depend on the duration and number of times violence and/or rape has occurred, symptoms being more severe if the duration is longer. The results of our study (Folnegović-Šmalc and Folnegović 1998) show that PTSD develops significantly more often in people who have experienced four or more psychotraumatic events (Fig. 2). The therapeutic outcome is also more favourable for people who have experienced a smaller number of traumatic events.

3.3
Violator

Parents, children, other family members, acquaintances, strangers or enemies can all be violators. In the abuse of children, the most common violators are parents, and in the abuse of elderly, the most common violators are children or other family members. Violent acts towards the mentally ill or disabled people are often performed by hospital staff. In war situations, violent acts are most commonly performed by hostile army, but in modern wars, they are often performed by former neighbours or acquaintances who have chosen the opposite side in the war. Mental suffering is more intense and the number of symptoms increases if the victim knows the violators or rapists than in cases when the violators are strangers (Kozarić-Kovačić et al. 1995).

3.4
Victimizing Role of the Victim

Victimizing is very important for violence that takes place within family, especially for physical maltreatment. Mentally ill, alcoholics and people with personality disorders exert a victimizing influence more

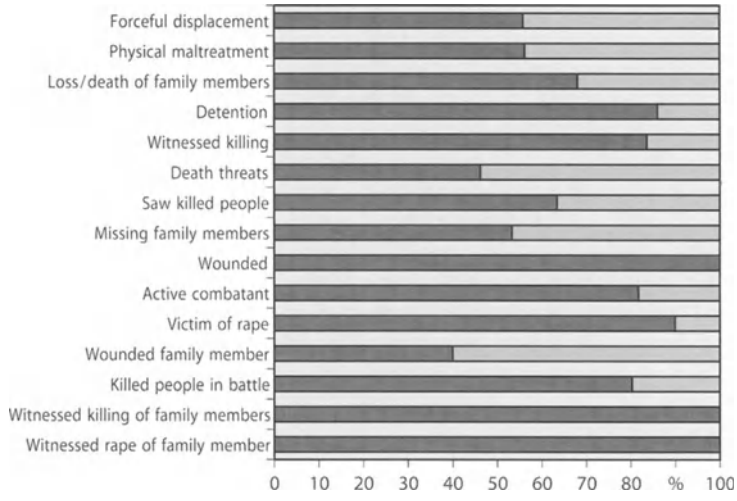


Fig. 1. Incidence of post-traumatic stress disorder (PTSD) in refugees with different traumatic events. *Shaded bars*, PTSD; *white bars*, no PTSD

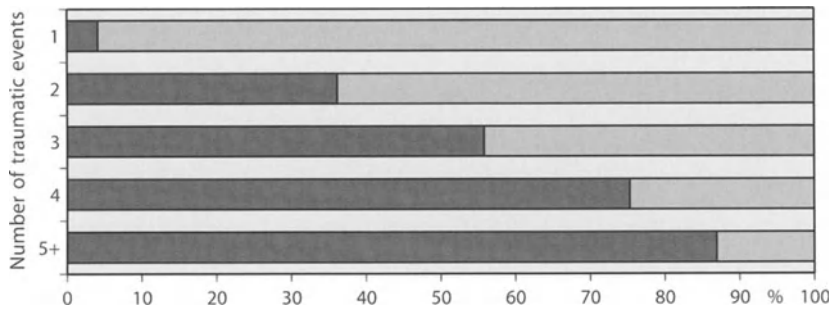


Fig. 2. Number of traumatic events in refugees with (*shaded bars*) and without (*white bars*) post-traumatic stress disorder (PTSD)

often. In war, however, there is no direct victimizing, unless opposite interests and attitudes are considered victimizing.

3.5
Victim's Personality Structure, Mental Condition and Previous Experience

In the description of acute stress reaction or PTSD (ICD-10), a traumatic event is described as one that "is likely to cause pervasive distress in almost anyone" (WHO 1992). Emphasizing that distress can be caused in *almost* anyone, this quotation indicates that the development of distress or PTSD is determined by the victim's personality structure and that individual vulnerability and coping capacity play a role in the occurrence and severity of acute stress reactions. Coping is defined as constantly changing cognitive and behavioural efforts in order to manage specific external and/or internal demands appraised as taxing or exceeding the resource of the person (Lazarus and Folkman 1984). It is well known that, in war for instance, many people have been exposed to similar traumatic experiences and yet not all of them have developed PTSD or other forms of mental decompensation. Furthermore, positive development in the form of defence mechanisms and maturity has been observed in poly-traumatized individuals who have spent months under traumatizing circumstances (Kozarić-Kovačić et al. 1998).

Earlier traumatic experience plays an important role in the way a person reacts to trauma (Breslau et al. 1991, 1997). People who have experienced severe psychotraumatic events earlier in life are more vulnerable when new trauma occurs. The recurrent traumatic event can be a precipitating factor for the development of serious mental decompensation, PTSD or even a psychotic condition. One of our patients is a good example: aged 63, married and retired; as a child during the Second World War he had to flee and leave home; his father was killed and his mother ran away to another country; he suffered severe mental and physical traumata; until 1991, he had never complained of or been treated for mental disorders. One evening during the war in 1991, he saw people on the news being forced to leave their homes in Vukovar. He became acutely psychotic and was admitted to hospital with mixed dissociative disorder (F44.7). After a few days of treatment, the psychotic symptoms withdrew and PTSD developed, the intrusive content being characterized by his early traumatic experience.

In addition to early traumatic experience, pre-existing psychiatric disorders, a family history of anxiety disorders and prolonged early separation from parents all contribute to the increased risk of mental decompensation after traumatic events (Breslau et al. 1997). The most common pre-existing psychiatric disorders are personality disorders. Alcoholism more commonly appears as a co-morbidity than as a pre-existing disorder. Co-morbidity depends partially on gender differences, depression being more prevalent in the female population and alcoholism being more prevalent in the male population (45%), followed by depression (37%), anxiety, personality disorders, opioid dependence and psychosis (Folnegović-Šmalc and Folnegović 1998). Out of veterans that had PTSD, 54% had co-morbidity, while 7% had developed enduring personality change after catastrophic experience during the 3 years after PTSD was diagnosed.

3.6
Age, Gender and Educational Level of the Victim

The victim's age is a very important factor in determining the kind and intensity of mental decompensation after violence and rape. The most vulnerable groups are children (Robinson et al. 1994; Shannon et al. 1994), adolescents and the elderly (Cheung 1994).

Most studies indicate a greater frequency of PTSD in the female population (Breslau et al. 1997; Davidson et al. 1991; Helzer et al. 1987), but some of them

indicate the same frequency for both genders (Folnegović-Šmalc and Folnegović 1998; Kozarić-Kovačić et al. 1998).

Low socio-economic status seems to be positively correlated to the development of mental difficulties after traumatic experience (Weisaeth 1996). Educational level, on the other hand, negatively correlates with the intensity and number of symptoms. The cognitive and emotional personality structure is reflected in the way a person interprets traumatic events. The level of education determines defence mechanisms and coping ability (Horowitz 1993; Myers et al. 1984) and influences cognitive levels of interpretation; it can therefore postpone or play a protective role in the development of severe mental consequences.

3.7
General Medical, Social and Financial Consequences of Violence

General medical, social or financial consequences of traumatic events can contribute to the development of mental symptoms. Taking rape in war as an example, mental symptoms and PTSD are likely to develop in victims. If general medical consequences such as infections or vaginal lesions occur, the likelihood of PTSD developing is greater. Furthermore, of those raped women who became pregnant and had a baby, almost 100% develop PTSD. They perceive the baby as "a foreign body in their womb". In this group, suicidal behaviour, psychotic symptoms and sexual disorders, most likely orgasmic disorders, dyspareunia and vaginismus, are also frequent (Kozarić-Kovačić et al. 1995).

3.8
Duration of Period After Violence Has Occurred

The length of time that has elapsed since violence had occurred is very important for an individual's assessment of the intensity of psychotrauma. The victim's assessment of the intensity of psychotrauma was studied in a group of refugees and displaced persons that had experienced the following traumatic events: exile, hunger (personal or other family members), loss of material goods, permanent life-threat and physical torture, permanent physical disability by torture (personal or other family members) or death of a family member. Immediately after arrival at the camp, i.e. within a few days of the traumatic experience, hunger was rated as the most intense traumatic experience. Two to three months later, the death of a family member was rated as the most intense traumatic experience by the same group (Folnegović-Šmalc and Folnegović 1998). These results indicate that, shortly after violence has occurred, existential problems and material losses are most significant for the individual, but as time passes, the individual experiences irreversible emotional consequences as most traumatic.

3.9
Support from Family and Environment

Psychosocial and material support, understanding from their family and environment and the ability to make autonomous decisions are significant for traumatized subjects immediately after the traumatic event has occurred. Providing help in dealing with feelings of shame and guilt and recognizing possible suicidal intentions are of extreme importance.

People that have been raped need support in organizing medical examinations, in knowing the truth about possible sexually transmitted diseases (STD) or pregnancy and, if they insist, in keeping the event secret. People who want to keep the traumatic event secret should not be included in group therapy. The family should be advised on how to provide maximum support, and the individual should return to family as soon as possible. An exception to this is when a husband cannot accept that his wife has been raped, in which case the return to the family can be postponed until they both reach a decision about their future life together.

It is important to remind caregivers and members of the family that, prior to their experience of assault, rape victims were no different from anyone else of their age and social background with respect to psychiatric symptoms or diagnoses and that, following rape, most victims experience a large number of serious psychiatric symptoms and may appropriately be considered for interventions directed at reducing the intensity of those symptoms. The interventions should include education about normative reactions to rape and about the falsehood of rape myths, exploration of the meaning of the assault to the victim, relaxation training and other anxiety-reduction techniques as well as guidance in decision-making regarding disclosure of the rape, prosecution of the assailant and possible changes in where the victim lives or other life-style changes (Frank and Anderson 1987).

3.10
Primary Prevention

Primary prevention of mental consequences of violence is focused on providing basic existential security

and support by family and environment, limiting exposure, providing education about the individual expectation of the person who has been traumatised – about the course, prognosis and consequences (e.g. pregnancy) – education of family members, providing general medical services, regulating sleep and debriefing. If co-morbid disorders occur, adequate therapy is needed.

4 Clinical Presentation

Mental response to stressful events reflects the complexity of psychotrauma, and various alternatives may predominate in the clinical presentation within this complex model. A variety of traumatic events can thus provoke various mental symptoms. Clinical experience confirms that the nature of symptoms is usually determined by the nature of the stressful event, while their intensity is determined by the pre-morbid personality structure, the intensity and duration of traumatic events and the attitude towards the victim in his or her environment.

Symptoms can meet the criteria for acute stress reaction, PTSD, other reactions to severe stress and reaction to severe stress, unspecified. Conditions that meet criteria for diagnostic categories included in sections other than "reaction to severe stress" can also develop. These include sleep disorders, eating disorders, alcoholism, other substance-related disorders, depression, anxiety and psychotic conditions. In describing PTSD, recent classifications of mental disorders have focused on symptoms in veterans or prisoners so that, despite their severity, symptoms in people who have been raped often have to be classified as other reaction to severe stress (ICD-10), which is a worse category in diagnostic terms. This indicates the need for further subclassification of the reaction to severe stress category.

People who have been raped often have following symptoms: changes in behavioural patterns, markedly diminished interest in life or significant activities, prolonged silences, social withdrawal, memory disturbances, depression, feeling guilty and anxious, feeling numb or detached, sleep disturbances or loss of appetite. Victims may experience an extremely wide range of emotions. The impact of the rape may be so severe that feelings of shock or disbelief are expressed. They can show different emotional styles: the expressed style, in which feelings of fear, anger and anxiety are shown through such behaviour as crying, sobbing, smiling, restlessness and tenseness; and the controlled style, in which feelings are masked or hidden and calm, composed or subdued affect is seen. Confused thinking, difficulty in making decisions, uncertainty and poor problem solving can also be observed. Victims can feel helpless and self-pitying, and the risk of suicidal behaviour and actual suicide is high. Aggression towards others and intrusive symptoms are rare. People who have been raped may have symptoms of sexual disorders such as orgasmic or sexual pain disorders that are rarely observed in other psychotraumatized individuals. Victims are often afraid of possible sexual contacts. Dysmenorrhoea and symptoms such as those accompanying pregnancy are frequent. Psychosomatic disorders and complaints about feeling tired, exhaustion, chest pain, headaches, nausea and vomiting, accompanied by excessive sweating, increased heart rate, elevated blood pressure and rapid breathing, are also more frequent in people who have been raped than in other psychotraumatized subjects.

5 Malingering and Negation

As emphasized in DSM-IV, when making a diagnosis of PTSD in people who have experienced some forms of physical violence, malingering must be "ruled out in those situations in which financial compensation, benefit eligibility or forensic determinations play a role" (American Psychiatric Association 1994). In contrast, in cases of rape, victims themselves often use negation and denial as defence mechanisms. Depending on the individual and cultural background, the raped person may feel guilty and the attitude towards him- or herself may be negative and rejective, even in cases when there was no victimizing influence on the part of the victim. Rape is therefore often not reported even when other forms of violence also took place. Special models of psychological, social and general medical assistance should be offered to victims in order to help them overcome these difficulties. Therapists should play an active role in discovering that the person has been raped, should offer maximum confidentiality and should provide the possibility of anonymous phone interventions and other general medical services.

6 Treatment

Debriefing, other forms of cognitive behavioural therapy and pharmacotherapy are most frequent in the treatment of PTSD (Davidson 1992). A specific approach is needed in the treatment of psychotrauma

after rape. In addition to the standard methods applied for PTSD, individual psychotherapy is recommended. In pharmacotherapy, the focus is on possible suicidal behaviour, depression and sleep disorders.

A victim of violence and/or rape should always be viewed as an individual, bearing in mind the complex and dynamic interaction of that person's past and present experience and the variety of biological, psychological and environmental influences. Mental consequences in victims of violence comprise a broad range of symptoms and intensity, whereas symptoms and defence mechanisms in victims of rape are more alike. Violence and rape affect not only the victims, but also their family, with the possibility of the development of severe mental disorders and even psychotic conditions in family members. In conclusion, violence and rape represent a serious psychiatric and public health issue, and the importance of primary prevention and timely psychiatric intervention must always be emphasized.

7 References

American Psychiatric Association (1994) DSM-IV. American Psychiatric Association, Washington DC

Breslau N, Davis GC, Andreski P, Peterson EL (1991) Traumatic events and posttraumatic stress disorder in an urban population of young adults. Arch Gen Psychiatry 48: 216–222

Breslau N, Davis GC, Andreski P, Peterson EL, Schultz LR (1997) Sex differences in posttraumatic stress disorder. Arch Gen Psychiatry 54: 1044–1048

Burgess AW, Holmstrom L (1974) Rape trauma syndrome. Am J Psychiatry 131: 981–986

Cheung P (1994) Posttraumatic stress disorder among Cambodian refugees in New Zealand. Int J Soc Psychiatry 40(1): 17–26

Davidson J (1992) Drug therapy of post-traumatic stress disorder. Br J Psychiatry 160: 309–314

Davidson JRT, Hughes D, Blazer DG (1991) Post-traumatic stress disorder in the community: an epidemiological study. Psychol Med 21: 713–721

Folnegović-Šmalc V, Folnegović Z (1998) Epidemiology of PTSD. Druš Istraž Zagreb 7

Frank E, Anderson BP (1987) Psychiatric disorders in rape victims: past history and current symptomatology. Comp Psychiatry 28: 77–82

Goldberg J, True W, Eisen S et al (1987) The Vietnam Era Twin (VET) Registry: ascertainment bias. Acta Genet Med Gemellol (Roma) 36: 67–78

Helzer JE, Robins LN, McEvoy L (1987) Post-traumatic stress disorder in the general population. N Engl J Med 317: 1630–1634

Horowitz MJ (1993) Stress and the mechanisms of defense. In: Wilson JP, Raphael B (eds) Handbook of traumatic stress syndromes. Plenum, New York

Kocijan-Hercigonja D, Rijavec M, Parry Jones W, Remeta D (1996) Psychologic problems of children wounded during the war in Croatia. Nord J Psychiat 50(6): 451–456

Kozarić-Kovačić D, Folnegović-Šmalc V, Škrinjarić J, Szajnberg N, Marušić A (1995) Rape, torture and traumatization of Bosnian and Croatian women: psychological sequelae, Am J Orthopschiatry 65(3): 428–433

Kozarić-Kovačić D, Folnegović-Šmalc, Marušić A (1998) Acute post-traumatic stress disorder in prisoners of war released from detention camps. Druš Istraž Zagreb 3(35): 485–497

Kulka RA, Schlenger WE, Fairbank JA et al (1990) Trauma and the Vietnam war generation. Bruner and Mazel, New York

Lazarus RS, Folkman S (1984) Stress, appraisal and coping. Springer, Berlin Heidelberg New York

Myers BM, Templer DI, Brown R (1984) Coping ability of women who become victims of rape. J Consul Clin Psychol 52: 73–78

Raphael B (1986) When disaster strikes. Basic, New York

Robinson S, Rappaport B, Sever M, Rappaport J (1994) The present state of people who survived the holocaust as children. Acta Psychiatr Scand 89: 242–245

Shannon MP, Lonigan CJ, Finch AJ, Taylor CM (1994) Children exposed to disaster. J Am Acad Child Adolesc Psychiatry 33: 80–93

Tardiff K (1995) Adult antisocial behaviour and criminality. In: Kaplan HI, Sadock MD (eds) Comprehensive textbook of psychiatry, vol II, 6th edn. Williams and Wilkins, Baltimore, pp 1622–1631

Watson CG, Anderson PED, Gearhard LP (1995) Posttraumatic stress disorder (PTSD) symptoms in PTSD patient's families of origin. J Nerv Mental Dis 183: 633–638

Weisaeth L (1996) PTSD: vulnerability and protective factors. In: Giller EL, Weisaeth L (eds) Clinical psychiatry. Post-traumatic stress disorder. Bailliere Tindall, London

WHO (1992) The ICD-10. Classification of mental and behavioural disorders. World Health Organization, Geneva

CHAPTER

21

N. Konrad

Psychiatry in Custody and in Prisons

Translator: E. Taub

1
Institutional Settings for Imprisonment

The commonest forms of imprisonment are on remand and under sentence following a criminal conviction; of more limited significance are detention by the police and the detention of individuals who are, or are suspected of being, illegal aliens. To be distinguished from these is detention in camps, particularly in concentration camps and as prisoners of war.

1.1
Rates of Detention

In the penal institutions of the Federal Republic of Germany, the total number of prisoners on the 31 December 1994 was 60,289. Of these, 39,291 were sentenced prisoners (of whom 1570 were women), 4265 prisoners in youth detention centres and 20,203 prisoners on remand. This gives an overall rate of detention of 73.9 per 100,000 inhabitants. Making international comparisons, there are countries with lower rates of detention, such as Sweden (50.6 per 100,000 inhabitants) or the Netherlands (50.7), but also countries with higher rates, such as England and Wales (89.9) or the USA (455) (Tomasevski 1992).

1.2
Conditions of Imprisonment

Internationally, there is great variation in the conditions under which prisoners are kept. Factors which can influence psychological state include nutritional deficits (which in niacin or vitamin B_{12} deficiency may even produce dementia), appalling hygienic conditions or maltreatment by prison staff (who may sometimes be acting out sadistic impulses). These factors should play no role in those countries which adhere to the *Standard Minimum Rules for the Treatment of Prisoners* of the United Nations or to the Council of Europe's (1987) guidelines, which to a large extent concur with those of the United Nations.

A question of significance not only for legal outcome concerns the extent of prisoners' integration into the subculture, e.g. the extent to which they have relationships with fellow prisoners which are oppressive and the role they play in the environment in which they live. The mentally ill are at increased risk of being subject to harassment (verbal threats, discrimination, physical violence) and exploitation by their fellow prisoners. Sexual abuse even extends to being raped by fellow prisoners.

Greater rates of contact and consultation with doctors are found than in the general population. The functions of this in everyday prison life are not only the provision of acute medical care, but also relief from isolation (e.g. through opportunities for contact in the waiting room/waiting cell) and the provision of treatments which at earlier stages were neglected or postponed (e.g. dental care).

1.3
Role of the Psychiatrist

The professional medical role of a psychiatrist and/or psychotherapist working in a prison has a conflict inherent in it. On the one hand, the doctor acts according to the requests and interests of his or her imprisoned patient and, following the Hippocratic oath, assigns the highest priority to the preservation and restoration of the patient's health, yet on the other hand he or she is an employee of that authority which, in carrying out the punishment required by the state, implements measures which may well damage the prisoner's health (Binswanger 1979). Unlike the surgeon or physician working in prison, who treats illnesses which have been imported from conditions of freedom or which happen to have occurred during imprisonment, the psychiatrist in prisons daily deals with large numbers of individuals with "prison reactions", which have arisen directly as a consequence of imprisonment. To a certain extent, the function of the psychiatric and psychotherapeutic treatment provided by the psychiatrist is to keep the prisoner fit for imprisonment, serving a pacifying and mollifying function (Mechler 1981). Prison psychiatrists find themselves in ethically questionable territory if they carry out psychopharmacological or other medical interventions for which there is no primary medical indication, in order to allow judicial proceedings and the penal system to run smoothly (Binswanger 1979).

Because particular stresses in imprisonment can lead to adjustment disorders, the psychiatrist is in a position to propose changes in conditions of imprisonment. These need not be confined to changes for a single individual (e.g. transfer to a protected place of imprisonment), but may also promote more general reforms in prison conditions, with the aim of creating a therapeutic atmosphere. Persisting stress from an atmosphere of violence and provocation can also lead to psychological disturbances among prison staff.

Acutely mentally ill individuals are not infrequently dealt with under conditions which seem inappropriate when compared with psychiatric treatment in the health system. Thus there are instances of mentally ill individuals being dealt with through disciplinary procedures and placed in isolation, even when they are

incapable of understanding the workings of the institutions or of following the rules (e.g. when they refuse to leave the cell when ordered to do so or violate rules about hygiene, e.g. by showering fully dressed; Hodgkins 1995). Other typical behaviours of psychiatrically disturbed prisoners include acts of self-harm, firesetting in their cells, neglect of personal hygiene and vandalising the contents of the cell (Adams 1986).

2 Selection Processes – Diversion

A psychiatrically disturbed individual who has been accused of a crime which may carry a custodial sentence passes through a series of filtering processes en route to a penal institution. In the course of these processes, indications may be identified for his or her removal from the criminal justice system (diversion) by, for example, transfer to a psychiatric facility. Whether the potential routes for diversion available in a particular country are used in practice depends on basic convictions about where a mentally disordered offender primarily belongs (Steadman et al. 1995).

2.1 Role of the Police

A key initial crossroads for the diversion process is represented by the actions of the police. In some circumstances, for an offender with conspicuous psychiatric morbidity (who may in fact be more readily identifiable as the perpetrator of a crime; Porporino and Motiuk 1995), the police may not even initiate criminal proceedings via a remanding judge, but may refer the offender to a psychiatric service. For the preliminary inquiry, decisive importance is assigned to the police report (Duff 1997).

If the police attribute no particular importance to an apparent psychiatric disorder, or if criminal prosecution is initiated in spite of this, the public prosecutor's office, as the authority responsible in Germany for criminal prosecution, may put aside the preliminary enquiry. This occurs if the accused is obviously not capable of taking responsibility for his or her criminal act and the severity of the offence committed does not suggest that he or she is generally dangerous. Internationally, there are variations between different legal systems in the definitions of responsibility under criminal law. These differences concern the capacity for being held responsible for criminal actions, fitness to plead and, in particular, the possibility of a judgement of "guilty but mentally ill". Even in Germany, the presence of a psychosis does not automatically mean that lack of responsibility for criminal actions is assumed (see Vol. 1, Part 2, Chap. 16; Konrad 1995).

2.2 Alternatives to Imprisonment

A significant question is whether there are alternatives to a custodial sentence within the criminal justice system: shortage of adequate care institutions, particularly long-term in-patient facilities, gaps in community systems of support and very narrowly formulated criteria for hospital detention under civil or public law may convey the impression of a gap in security, which in the absence of other social controls it may seem most appropriate to close using the criminal justice system (Edwards et al. 1994). Sometimes the criminal justice system is seen as operating as a facility capable of providing multiple services and, as such, as being the best place of treatment for the mentally ill (Dell et al. 1993), without account being taken of the extent to which treatments are really available in practice in particular local institutions.

However, even among those detained on remand for relatively long periods, psychiatric illnesses may not always be recognised (Teplin 1990). While staff responsible for admitting prisoners to custodial facilities generally only detect illnesses characterised by massively conspicuous behavioural abnormalities, the admission medical examination which is obligatory in some countries is not a fully reliable filter for detection of psychiatric illnesses, particularly depressive disorders, either.

2.3 Admission to a Psychiatric Hospital

Where the offence is more serious, German criminal law allows for provisional detention in a psychiatric hospital under the provisions of Art. 126a of the German code of criminal procedure (*Strafprozeßordnung*, StPO). If, in the course of remand in custody, it becomes apparent that the conditions are met for the final outcome to be a compulsory detention in a psychiatric hospital under Art. 63 of the penal code (*Strafgesetzbuch*, StGB), Art. 126a StPO may be applied to allow transfer from remand in custody. Both provisional detention under Art. 126a StPO and longer-term detention according to Art. 63 StGB generally involve admission to the official secure facilities (see Vol. 2, Chap. 17), which are obliged to accept these transfers on the orders of a judge.

In countries in which there is no legal obligation for treatment institutions within the health service to accept admissions on court orders, a psychiatrically ill individual may remain within the criminal justice system regardless of the severity of his or her disturbance. The major reasons for refusal to accept such individuals for treatment are that they appear excessively dangerous for the treatment facility (e.g. because of inadequate physical security conditions or lack of staff experience in forensic psychiatric practice), that suitability for treatment is limited (e.g. with "difficult" chronically psychotic patients) or that the indication for treatment is not sufficiently strong. Doctors' attempts to provide a route into the health system, e.g. by writing a report, may be associated with delays (time allowed for provision of reports, time for provision of a second opinion in contentious cases) and may thus even prolong the period of detention on remand (Robertson et al. 1994). The time between the issuing of the order for transfer to a psychiatric facility and the actual transfer taking place (e.g. when there is a shortage of beds) can also influence the prevalence of psychiatric disorders within penal institutions (Dell et al. 1993).

2.4
The Mentally Ill Remanded in Custody

When a mentally ill individual does end up remanded in custody, early diversion processes are important, above all because of the relatively high suicide risk. In addition to an as yet unrecognised mental disorder being overlooked (e.g. schizophrenia with predominantly negative symptoms), there is a possibility that a psychiatric disorder present at the time of the crime may have been in remission at the time of detention and may again become apparent only at a later point in time.

In some countries, such as Germany and England, there are legal provisions which allow a prisoner with a severe mental illness to be defined as unfit to plead and spared further remand or criminal proceedings. Some countries make use of the possibility of transferring mentally ill individuals in custody for (temporary) treatment of their illnesses in public hospitals, if specific provisions for this are not available within the criminal justice system (e.g. in specific forensic hospitals).

Because the presence of a psychiatric disorder may be equated with dangerousness, there may be delays compared with fellow prisoners in release with a caution (Teplin 1994) or in measures granting reduction of sentences, including early release on parole (Porporino and Motiuk 1995). This phenomenon also influences the prevalence of psychiatric disorders in custody. Finally, there may be tendencies for special conditions to be imposed when psychiatrically offenders are released on parole and for greater social surveillance to be put in place by the authorities, leading to a higher rate of recall following release on parole even where no new criminal offence has been committed (Porporino and Motiuk 1995).

3
Prevalence of Psychiatric Disorders

3.1
Methodological Problems in Empirical Studies

Empirical studies (e.g. Roesch 1996) have not confirmed the assertions which are periodically made that rates of psychiatric disorders among prisoners are increasing (e.g. Jemelka et al. 1989) and that this may be linked to deinstitutionalisation programmes. In view of the lack of clarity about diagnostic criteria which is not infrequently a characteristic of studies, and the use of selective samples and of unstandardised procedures for diagnosis, scepticism has been expressed regarding the conclusions of some of these studies. In cross-sectional studies, problems also arise from the variations in length of imprisonment prior to the survey, which can lead to confusions between illnesses present prior to imprisonment and those which have developed during imprisonment, as well as in comparisons between point and lifetime prevalence. In addition, the widespread tendency to associate psychiatric disturbance with dangerousness can encourage tendencies to conceal symptoms among prisoners who do not wish to jeopardise their chances of earlier release on parole (Smith 1987).

Variations in prevalence found in different studies may arise from the particular characteristics of the population examined as well as from the specific survey method. Thus generalisations across countries should be avoided and the prevalence figures arrived at should mainly be used for service planning (Roesch et al. 1995).

The summary presented in Table 1 only includes studies involving the examination after 1990 of a relatively large ($n > 100$) representative sample of a population of prisoners with standardised diagnostic instruments and a diagnosis which has been made on the basis of an international system of classification in which dependence syndromes have been included. Limitations arise above all from the generally limited willingness of remand prisoners to co-operate, from the reliability of the survey instruments used (e.g. Diagnostic Interview Schedule, DIS) in detecting per-

Table 1. Empirical studies on the prevalence of psychiatric disorders

Author	Country	Population	Subjects (*n*)	Survey instruments/ classification system	Diagnosis made (%)	Schizophrenia (%)[b]	Affective disorder (%)[c]	Personality disorder (%)	Alcohol dependence (%)	Drug dependence (%)
Gunn et al. (1991)	England and Wales	Sentenced men	1769	Clinical Interview Schedule/ICD-9	38	1.5	0.4	10.0	11.5	11.5
Maden et al. (1994)	England and Wales	Sentenced women	258	Clinical Interview Schedule/ICD-9	57	1.6	0	18	9	26
Brooke et al. (1996)	England and Wales	Prisoners on remand	750	Semi-structured interview SADS-L/ICD-10	62.5	2	9.5	11.2	38	
Anderson et al. (1996)	Denmark	Prisoners on remand	228	PSE-10, PCL-R, GAS, semi-structured interview/ICD-10	64	7	10	17	44	
Herrman et al. (1991)	Australia	Sentenced	189	SCID/DSM-III-R	82	3	12	Not measured	69	
Bland et al. (1990)	Canada	Sentenced men (< 2 years)	180	DIS, GHQ/DSM-III	76.7	2.2	21.1	47.8	50.6	24.4
Hodgkins and Côté (1990)	Canada	Sentenced men (> 2 years)	495	DIS/DSM-III	95.3	5.4	10.3	46.6	37.8	22.3
Arboleda-Flórez (1994)	Canada	Prisoners	1200	SCID, SCID II, PCL/DSM-III-R	60.7	1.3	4.7	5.3	46.4	

Table 1 (Continued)

Roesch (1996)	Canada	Men in prison on remand	790	DIS, BPRS, RDS, DP/DSM-III-R	93.6	4.9	10.1	64.3	77.6	63.7
Chiles et al. (1990)	USA	Sentenced men (> 2 years)	109	DIS/DSM-III-R	88	5	30	44	66	61
Teplin (1994)	USA	Men in prison (jail[a])	728	DIS/DSM-III-R	62.4	3.0	3.4	48.4	19.1	15.3
Teplin et al. (1996)	USA	Women in prison on remand	1272	DIS/DSM-III-R	70.3	1.8	13.7	13.7	23.9	52.4
Jordan et al. (1996)	USA	Sentenced women (> 1 year)	805	CIDI, DIPD-R, DIB-R/DSM-III-R	46.3	Not measured	10.8	28.0	17.1	30.3

[a]Those held in "jails" are generally prisoners on remand and those serving short sentences (< 1 year).
[b]Paranoid psychoses are also included in the "schizophrenia" diagnostic group.
[c]For classification according to ICD-9 affective psychoses, and for classification according to DSM-III(-R) major depression and bipolar disturbance are included in the "affective disorder" column.

sonality disorders and from inter-rater reliability, which is not always adequately examined.

3.2 Drug and Alcohol Use

Despite the variations which result from particular local characteristics and from differing research methods, a generalisation which may be made is that once the diagnosis of antisocial personality disorder is excluded (for which frequency estimates vary greatly because of limited reliability), drug and alcohol dependence generally constitutes the most important diagnostic group. Where more than one diagnosis has been recorded for probands, dependence syndrome plays an important part as a co-morbid disorder (Anderson et al. 1996; Brooke et al. 1996). Overlaps have been described above all between the diagnoses of antisocial personality disturbance, alcoholism, drug dependence and major depression (Abram 1990). Indeed, the use of illegal substances together with behavioural abnormalities and lack of treatment can create the route into prison (Chiles et al. 1990). Co-morbidity requires specific treatment programmes which differ from models of treatment used for "pure" disorders (Teplin et al. 1996).

Table 1 suggests a higher prevalence of psychiatric disorders among remand prisoners than among those serving sentences. Evidence supporting this conclusion comes from the finding that a transfer on the grounds of a need for in-patient treatment was judged to be indicated among 9% of remand prisoners (Brooke et al. 1996) compared with 3% of those serving sentences (Gunn et al. 1991) in England and Wales.

To the extent to which comparisons have been undertaken with the general population, a greatly raised prevalence of psychiatric illness among prisoners has been found across countries and across diagnostic groups (Arboleda-Flórez 1994; Bland et al. 1990; Brooke et al. 1996; Hodgins and Côté 1990). Considering diagnoses individually, this is particularly marked for the dependence syndrome disorders.

3.3 Sex Differences

Higher rates of morbidity are found among female than among male prisoners (Anderson et al. 1996; Maden et al. 1994), suggesting that the filters in the selection processes outlined above operate differently. Thus one would expect a greater level of need for treatment in the female penal system, related to higher prevalence of psychiatric disturbances (Maden et al. 1994). In practice, female prisoners tend to be more willing to accept offers of treatment. This may also be related to a better therapeutic atmosphere among female prisoners. Further, women's experiences of stigma are not so severe when they enter (psycho-) therapeutic treatment. Empirical evidence has not yet been provided for the hypothesis that rates of early traumatic life experiences are higher among imprisoned women than among those not in prison (Jordan et al. 1996).

4 Specific Clinical Pictures in Prison

4.1 Addiction-Related Disorders

Dependence disorders are currently the most commonly occurring psychiatric disorders in prison in most European countries and in America (Edwards et al. 1994; Teplin et al. 1996). Taking a longitudinal perspective over the past few years, a shift can be observed from alcohol dependence towards dependence on opiate-type substances and multiple substance dependence (Anderson et al. 1996), with particular national patterns of consumption and the availability of psychotropic substances having an important impact.

4.1.1 Withdrawal Syndromes

Among those with dependence disorders, psychological and physical manifestations of dependence – which vary greatly within and between individuals – may be observed in the first few days after imprisonment. The severity of drug withdrawal syndromes following forced abstinence is often overestimated. Marked withdrawal syndromes occurring in German penal institutions are often treated with medication ("supported withdrawal", e.g. with methadone and/or diazepam in gradually diminishing doses). Prolonged withdrawal syndromes and, in particular, persisting depressive syndromes after the physical manifestations of withdrawal subside may require longer-term treatment with antidepressants, e.g. doxepin.

Sometimes people with drug dependence try to conceal their addiction after imprisonment. As well as reduced general and nutritional state, alertness should be aroused by fresh injection sites and by older, hardened and dark-coloured areas of veins (not only in the elbow region) which may indicate old injection sites.

Acute withdrawal symptoms can strengthen suicidal tendencies in prison (Bogue and Power 1995). Prac-

tices which are contraindicated are iatrogenic maintenance of addictive tendencies through excessively prolonged treatment with psychotropic drugs such as chlormethiazole, which were initially indicated by psychopathological symptoms, through the use of a broad range of indications for tranquilisers or through uncritical administration of analgesics in a way that promotes dependence. Analgesics are sometimes administered by nursing staff or even prison officers without medical staff first establishing that they are indicated.

4.1.2 Therapeutic Options

Prisoners with dependence disorders constitute a particularly difficult clientele from the point of view of amenability to treatment. In countries in which treatment can be offered as an alternative to imprisonment, the patients found in the penal institutions are above all those who have failed to proceed through this channel, with many of the failures already occurring at the start of treatment, e.g. when patients fail to enter the treatment facility or leave within a few days. Those prisoners with longer sentences for whom, despite their wish for treatment, conditions of imprisonment do not allow transfer to an external treatment facility suffer because of the treatment-hostile culture of the normal penal institution become involved in the subculture and can sometimes obtain addictive substances without much effort (Chiles et al. 1990).

4.1.3 Drug Consumption in Prison

Under prison conditions, it is generally not possible to guarantee absolute freedom from drugs. Alcohol can be manufactured from yeast, fruit and bread laid down in a warm place, e.g. by the central heating. Particularly in larger, more chaotic places of imprisonment, drugs enter by a variety of routes. These are generally cannabis and heroin, and more rarely cocaine or other drugs which change hands at higher prices than heroin. Purchase of drugs via other prisoners can bring with it an increasing involvement in subcultural activities and accumulation of debts, which can sometimes make the situation in prison unbearable for an individual prisoner.

Where the supply of substances is irregular, the possibility that withdrawal symptoms may develop may sometimes not be considered, e.g. following transfer into a particularly sheltered area. Flashbacks are a rarity in everyday prison life. When prisoners who have been identified by prison staff as having psychiatric morbidity are interviewed, particular care is required in differential diagnosis, as these disorders are often categorised as belonging to the range of substance-related disorders; an important danger is that the presence of a schizophrenic illness with "secondary" drug dependence may in particular be missed on superficial interview. Of critical importance are reliable urine tests, i.e. specimens obtained under observation, preventing the use of prepared urine specimens supplied by other prisoners which may have been kept warm by the heating.

4.2 Adjustment Disorders

"Prison is by its whole nature an unnatural milieu", according to Birnbaum (1931), "and in keeping with this the state of imprisonment is an unnatural condition in which to live." The clinical pictures to be classified as adjustment disorders would not have arisen without imprisonment as a psychosocial stress or, to be precise, a critical life event.

4.2.1 Different Forms of Experience

With the reduction in the range of external stimuli and impressions, the prisoner may sink into a state empty of thoughts and feelings, sometimes extending to complete apathy, or else may experience an intensification of the emotions associated with perceptions and an increase in imaginative activity, including daydreaming (Sieverts 1979). For those with narcissistic personality traits and disorders, fantasies can include the reversal of the previous failure and the anticipation of future successes in crime.

Different individuals also cope differently with feelings of profound desolation and inner needs. Fears of maltreatment are of considerable significance for those prisoners who have maltreated or abused children, have been informers to prison or judicial staff or who cannot pay debts (from drug purchase or gambling). Tendencies to regress may be expressed through an increasingly dilapidated physical state and weight gain.

4.2.2 Symptoms

Among the many forms of experience provoked by the prison situation, fluid transitions to symptomatic pictures which meet ICD-10 criteria for adjustment disorders sometimes occur. Symptoms observed vary greatly both between individuals and over time for a single individual, and they include despondency, withdrawal into isolated ruminations in the cell,

anxiety, stupor, agitation and hostility. It has been suggested that adjustment reactions vary according to the form of imprisonment (single or communal imprisonment), but as yet there is no scientific evidence to support the generalisations which have been made about the various factors which may be significant (e.g. variations in the level of isolation, including solitary confinement, differing sizes of rooms allowing retreat to different degrees, variations in daily routine and in characteristics of fellow prisoners).

4.2.3 Basis for Explanation

Attempts to explain these "prison reactions" emphasise the triggering effects of particular stresses, such as the issuing of an arrest warrant, interviews with the police, being confronted with the statements of accomplices or witnesses, being denied visits, episodes of isolation, appointments for review of remand in custody, delivery of the indictment, end of trial and judgement being pronounced. General factors associated with imprisonment should also be taken into account: loss of social status, limitation of freedom of movement, living space and sphere of action, isolation through removal from the usual social networks, change in patterns of communication and profound alteration in previous living habits. It is difficult to gauge precisely the effects of these factors on a particular individual and the reduction in coping resources which they bring about. Stresses can also result from the uncertainty about how severe a punishment to expect, contemplation of the criminal act and the experience of lack of understanding or contemptuous treatment from the prison staff.

Ways of reacting to these stresses and the evolution of adjustment disorders are widely regarded as being related to personality (e.g. Langelüdekke and Bresser 1976), in that personality influences the significance for the individual of imprisonment, vulnerability to stress and the tendency to get involved in stress-producing situations such as arguments with fellow prisoners (Hurley and Dunne 1991). However, as various authors have described (e.g. Schleusener 1976), there are difficulties in recording reliably transient psychological changes among prisoners, perhaps because the sensitivity of the instruments thus far used to measure them is limited (Rasch 1976). "Imprisonment shock", a transient stupor-like state of total resignation, tends above all to afflict younger and very active people; the third decade is a stage of life during which intolerance of imprisonment is great (Birnbaum 1931).

4.2.4 Course

Adjustment disorders are observed less frequently among sentenced prisoners (1.9%) than among those on remand (7.6%–18.5%) (Gunn et al. 1991; Hurley and Dunne 1991; Brooke et al. 1996). As length of imprisonment increases, symptoms diminish to a degree, with adaptation to life in prison becoming greater and roles within the subculture which stabilise self-esteem perhaps being adopted (Backett 1987). Thus the proportion of remand prisoners judged to be severely disturbed in a Swiss prison fell from 57% on the tenth day after imprisonment to 43% after 60 days in prison (Harding and Zimmermann 1989).

4.2.5 Classification of "Prison Reactions"

The classification of "prison reactions" is not uniform. Langelüddeke and Bresser (1976) distinguish between panic-dominated, paranoid, aggressive, querulant, depressive, suicidal and simulated prison reactions. From the perspectives of attempting to make more reliable psychiatric diagnoses and keeping practice embedded within the framework of clinical psychiatry, there are no advantages to using terminology which deviates from the international systems of classification.

Thus the condition known as "prison rage" (a state of explosive agitation in which cell contents are smashed up and acts of violence sometimes committed towards staff or fellow prisoners) should be classified as an adjustment disorder with predominant disturbance of conduct (ICD-10 F43.24). This is observed above all following longer periods of isolation in situations in which there has been an experience of rejection, such as recall from parole or separation from a partner (Mechler 1981).

Reactive psychoses, such as those manifested by foreigners in environments where they do not speak the language, can be classified as acute and transient psychotic disorders (ICD-10 F23). In terms of differential diagnosis, strong consideration should be given to disorders related to psychotropic substances (above all delirium or psychotic symptoms) as well as to schizophrenia or delusional disorders.

4.3 Prison Psychoses

4.3.1 True Psychoses Versus Prison Psychoses

While adjustment disorders predominate in out-patient clinical practice within the criminal justice system (see Vol. 1, Part 2, Chap. 17), psychoses are the most

frequently occurring disorders in the in-patient sphere (Konrad 1997). With reference to mental illnesses with psychotic symptoms, a substantial proportion of German psychiatrists maintain the distinction which originated in the last century between "true psychoses" and "prison psychoses". The disorders regarded as "true psychoses" generally fit into the category of schizophrenic illnesses, with psychopathology which may be coloured by prison conditions in aspects such as the contents of delusions, whereas "prison psychoses" are specific reactions to imprisonment (Langelüddeke and Bresser 1976; Mechler 1981). According to this model, it is assumed that "prison psychosis" is a product of individual vulnerability factors, such as an innate disposition to become psychotic (Birnbaum 1931) or particular sensitivity to prison conditions among impulsive and opportunistic offenders (Nitsche and Wilmanns 1911), together with factors related to current circumstances, such as the prevailing social climate of the time (Wilmanns 1924) or conditions in the penal system (Knigge 1932). Blurred boundaries are described between this state and dissociative phenomena and simulation (Wilmanns 1924; Mechler 1981).

The rarely observed Ganser syndrome is located on this continuum, in which the typical symptomatic picture involves talking past the point together with a qualitatively altered state of consciousness. This state, however, appears more frequently in situations where examinations are taking place than in everyday prison life. In this condition, the symptomatic picture is reported to be relatively stable during the generally short period for which it is manifest. The syndrome is more commonly observed among people with lower than average intelligence and histrionic personality traits (Bellino 1973).

Some authors conceptualise prison psychoses as intensifications of and developments from "prison reactions", from which they differ in their persistence and in the particular symptoms which develop. Langelüddeke and Bresser (1976) characterise subtypes, differentiating between pictures characterised by paranoia and sometimes hallucinations, based on a "delusion of innocence", primarily simulated states and extreme manifestations of poorly developed personalities, with blunting and loss of initiative. Apparently unprovoked attacks on prison officers are manifestations of these states with obvious practical significance and may lead to confinement in a place of special security or other disciplinary measures.

4.3.2 Differential Diagnosis

"Prison psychoses" have not entered international classification systems (ICD-10, DSM-IV) as a distinct clinical entity. Attempts to differentiate them diagnostically from the early manifestations of "true" psychoses, especially schizophrenic illnesses, for which imprisonment has to varying degrees been identified as an important trigger, make reference to the narrowness of the scope of paranoid beliefs in prison psychoses, with a tendency for delusions to be limited to the immediate environment but not to relate directly to fellow prisoners (Nitsche and Wilmanns 1911). Another differentiating factor which has been identified as crucial is the termination of prison psychotic phenomena with the interruption or ending of imprisonment, even though there may be "remnants" in the form of "now affect-free remains of delusions" or querulant or hypochondriac character traits (Birnbaum 1931).

4.3.3 Schizophrenic Illnesses

In comparison with other prisoners, prisoners with schizophrenic illnesses have a greater tendency to develop difficulties in obeying prison rules and in sustaining work roles within the institution. Because they more often behave in aggressive and violent ways, they are more often put under lock and key (Morgan et al. 1993). Occasionally, beliefs about being poisoned may be the basis for a refusal to eat, which is misinterpreted as a hunger strike. Even in countries where routine medical examinations are required, the presence of schizophrenic illnesses is not infrequently overlooked, especially when negative symptoms dominate the clinical picture (Anderson et al. 1996).

4.3.4 Depressive Illnesses

Depressive illnesses are sometimes missed in prison medical practice (Herrman et al. 1991). When a nonconformist prisoner presents with physical complaints without organic correlates, in clinical practice in the criminal justice system a diagnosis is too often made of simulation or of a psychosomatic disorder, without due consideration being given in differential diagnosis to the possibility of an affective illness. A depressive illness may also underlie apparently disruptive conduct in prison, e.g. when prisoners vandalise their cells (Maden et al. 1994).

Simulation of psychiatric disorder, e.g. through feigning of psychotic symptoms, is not frequent and can constitute a coping strategy aimed at improving conditions of imprisonment or obtaining prescribed medications for the prisoner's own use or for dealing.

4.4
Vexatious Complainers

Prisoners who are vexatious complainers generally begin with prolific writing activities, which are extremely time-consuming for the staff responsible for dealing with them. These activities may relate to judgements or decisions of the court felt to be unjust or to real or imagined neglect or discrimination during imprisonment.

Half of the complaints which need to be dealt with in a penal institution generally originate from just a few prisoners. For them this behaviour may not only be a kind of "occupational therapy", embarked on as a way of coping with loneliness and emptiness, but may also result from a subjectively experienced "cul-de-sac" in the course of imprisonment. The complaints often stop if an acceptable shared perspective and plans for the sentence can be arrived at with the prisoner before he or she becomes entirely alienated from the system.

5
Suicide and Suicide Attempts in Prison

5.1
Suicide Rate

In many places, suicide is the leading cause of death in prison (Tomasevski 1992). Published suicide rates range from 43 to 465 per 100,000 prisoners per annum (Liebling 1992). Based on the results of international suicide research, there is a consensus that the suicide rate in penal institutions is several times higher than for the general population (e.g. Cooke and Michie 1996; DuRand et al. 1995; Frühwald 1996). An explanation which has been advanced for this is that the most commonly used method – hanging, which is used by 85% of those committing suicide (e.g. Rieger 1971; Thole 1976) – is associated with a more limited chance of being saved than the methods frequently used in suicide attempts outside prison. Here it is also important to take into account the fact that the prison population is not a representative sample of the general population, but has overrepresented within it groups who are at especially high risk of suicide, above all those with addictions (Cooke and Michie 1996). More precise comparative studies, taking account of these selection factors, have not yet been published. There is a greater suicide risk during remand on custody than among sentenced prisoners, especially at the beginning of imprisonment. For example, the proportion of all suicides in prison which are recorded as having occurred during the first month of imprisonment has been reported as 42% (Thole 1976), 46% (Bogue and Power 1995) and even 73% (DuRand et al. 1995). This may be explained by the condition described as "imprisonment shock", i.e. by imprisonment as a stress factor (Harding and Zimmermann 1989), which varies greatly from individual to individual. Among those with dependence disorders, withdrawal symptoms have been identified as important stressors (Bogue and Power 1995). This period is also the one during which uncertainty about the future is greatest (Backett 1987). The population in prison on remand is characterised by a higher turnover, in particular a higher rate of new admissions than among sentenced prisoners, and there is also evidence for a higher prevalence of psychiatric disorders among those on remand (Bogue and Power 1995).

5.2
Risk Among Those Who Have Committed Homicide and Sexual Offences

A higher risk of suicide has been described among those guilty of homicide and of sexual offences (e.g. Thole 1976; Dooley 1990). This has been explained in terms of the long sentences anticipated by this population, the psychodynamic model of "internalisation of aggression" (e.g. Bogue and Power 1995) and the helplessness and hopelessness which may characterise long sentences. As yet it is an open question how far the process of stabilisation which has been described in longitudinal studies (e.g. Konrad 1994) among those serving long sentences might also be a sign of damage by imprisonment, reflecting sinking vitality and diminishing abilities to cope with life outside prison.

5.3
Risk Factors

Risk factors for suicidal behaviour, for which detection could be made more effective by the use of semi-structured interview instruments (Malone et al. 1995), are the following (e.g. Dooley 1990; Felthous 1994; Hurley 1989; Ivanoff et al. 1996):

- The presence of psychiatric disorders (above all schizophrenia, affective disorder, dependence syndrome), which has sometimes already led to inpatient psychiatric treatment
- Direct or indirect suicide threats
- Concrete suicidal ideation, in which exploratory discussion reveals thoughts about how in practice to prepare for and carry out the act

- Preparatory actions (e.g. setting aside a naked razor blade, preparatory cuts on the arm, collecting tablets)
- Previous suicide attempts
- Experience of loss, e.g. divorce
- Limited ability to cope with loss, frustrations, aggressions
- Accused of an aggressive or sexual offence

5.4 Characteristics of Suicides

On the basis of the current status of research knowledge, the predictive value of risk factors relating to history, such as the "broken home situation" (Rieger 1971) or suicide in the family or environment (Thole 1976), or of risk factors specific to imprisonment, such as confinement under particularly secure conditions, overcrowding and staff shortage (Felthous 1994), has not yet been adequately demonstrated. Here, the great variations between institutions in the characteristics of those who commit suicide must be taken into account; thus, in U.S. prisons, those involved are more often young, sober individuals accused of serious offences, compared with older intoxicated people with less serious offences in police custody (DuRand et al. 1995).

5.5 Suicide Prevention

Successful prediction of a behaviour such as suicide in prison, which is uncommon overall, remains difficult, above all because of lack of specificity of risk factors. The construction of a suicide profile, even when regional differences in the prison population are taken into account, is only of limited usefulness (Lester and Danto 1992). Beyond this, suicide prevention needs to involve caution in uncontrolled distribution of medication to those at risk of suicide, which creates the danger of collection and ingestion of a lethal dose.

5.6 Suicide Attempts

Not only the suicide rate, but also the rate of suicide attempts, among which the most frequently observed form is wrist cutting (Liebling 1992), is higher in prisons than in the general population (Ivanoff et al. 1996). In a sample of Canadian prisoners, previous suicide attempts were identified among 22.8%, compared with 7.1% of a comparable sample of the general population (Bland et al. 1990). In terms of history, prisoners who attempt suicide generally differ only slightly from comparison group probands, with more frequent experience of family violence and sexual abuse as the main area where a difference can be identified. A more important factor is the current prison situation, e.g. difficulties with fellow prisoners, lack of work or persisting sleep disturbances (Liebling 1995). A clear and reliable differentiation between "serious" and "not serious" suicide attempts cannot be made from either observation of manipulative behaviour or extent of planning and preparation or from the lethality of the methods used (Haycock 1989).

6 Deliberate Self-Harm

6.1 Sex Differences

Deliberate self-harm can be defined as a self-inflicted, direct physical injury which is not intended to be life-threatening (Herpertz and Saß 1994). The major self-harming behaviours reported in prison are injuries by cutting, self-poisoning and inflicting burns. The prevalence of patients manifesting such behaviour has been reported as 6.5% among male prisoners (Toch 1975), rising to 24% among prisoners with antisocial personality disorders (Virkkunen 1976). A history of self-harming behaviours can be obtained from 32% of female compared with 17% of male sentenced prisoners in England and Wales (Maden et al. 1994); this difference mainly reflects different frequencies of self-poisoning with tablets. Among female prisoners, it is rare for self-injurious behaviours to be reported as having taken place exclusively in the context of imprisonment; here, the previous history generally reveals self-harm outside the penal institution, pointing to the importance of individual vulnerability in explaining this phenomenon. During the period of imprisonment, the prevalence rate for both sexes is around 5%. In an Australian sample of female prisoners (Hurley and Dunne 1991), there was a history of at least one incident of self-harm among 22% of those interviewed, with self-harm having occurred exclusively before imprisonment in 11%.

6.2 Possible Explanations

Self-injurious behaviour is rarely a manifestation of a psychotic illness. In a study of U.S. prisoners, half of

the self-harmers gave a goal-directed motivation for their behaviour (Franklin 1988). Stressful experiences leading up to self-harm often involve real or imagined rejections or, less frequently, experiences of loss or of failure (Herpertz and Saß 1994). In borderline personality disorders, where acts of self-harm are common, such acts may serve as self-stimulation in the context of an experience of pain which is possibly abnormal, overcoming states of emptiness and boredom and interrupting unpleasant moods. They may also be a weapon, used to express anger against significant others (Frühwald 1996). Biological explanations emphasise the status of the opioid system, the dopaminergic and/or the serotinergic systems as substantial factors within a complex structure of causality involving life history, situational stresses and personality characteristics (Herpertz and Saß 1994).

6.3 Motives

Ideas from learning theory interpret deliberate self-harm as a dysfunctional form of problem solving (e.g. aimed at provoking attention or maintaining a relationship in which care is received) which can become established through positive and/or negative reinforcement. Self-harmers who report goal-directed motives more often give a history of such behaviour than those with suicidal intent (Franklin 1988). The psychodynamic approach of ego psychology emphasises the co-existing auto-aggressive, self-punishing tendencies and relieving, protecting and ego-stabilising functions (for a review, see Lester and Danto 1992).

In the range of motivations of those who harm themselves, there is a wide overlap with the intentions indicated by individuals who present having swallowed foreign bodies such as razor blades, batteries or cutlery; the spectrum extends from responses to command hallucinations experienced by people with schizophrenia to suicidal tendencies or goal-directed action aimed at achieving a transfer to a hospital situated outside the penal institution. Multiple episodes of swallowing foreign bodies are not uncommon (Karp et al. 1991). Compared with a control group, metal swallowers have a raised prevalence of schizophrenic illnesses, personality disorders and dependence disorders (Tacke et al. 1975).

One of the considerations which needs to be taken into account in prison practice is the secondary illness gain through the diagnostic and therapeutic measures resulting from presenting with symptoms. Benefits which prisoners may aim to achieve include alteration of prison conditions, sometimes extending to transfer outside the restrictive conditions of prison and even achieving an opportunity for escape.

7 Special Situations in Custody

7.1 Solitary Confinement

Solitary confinement in custodial institutions in most countries is a legally regulated exception to normal practice, which may, for example, be arranged in response to certain requirements regarding security. While isolation under experimental conditions over a 4-day period did not provoke any substantial psychopathological changes (Walters et al. 1969), it seems that with substantially longer periods, free-floating anxieties, increased reactivity to external stimuli and distortions of perception, disruptions of concentration, derealisation, pseudo-hallucinations and hallucinations in various sensory modalities (Rasch 1976; Grassian 1983) may manifest themselves. The picture will be shaped by a complex interaction between variables related to the individual (e.g. presence of a personality disorder, extent of positive symptoms among people with schizophrenia) and the type of isolation (above all degree of sensory deprivation, size of room). States of agitation and acts of self-harm may be observed as a result of this, and it can be assumed that the suicide risk is increased (Bernheim 1994). The development of an acute confusional state followed by partial amnesia and delusions of persecution, often accompanied by hallucinations of an anxious character, has also be described. The symptoms generally fade a few hours after the end of isolation (Grassian 1983). On the other hand, among schizophrenics who are susceptible to acute symptoms, a relapse of positive symptoms appears to occur (for a review, see Grassian and Friedman 1986).

7.2 Life Sentences

In very recent times, initiatives for abolishing lifelong sentences have re-ignited the debate about the possible damaging psychological consequences of longer sentences. It has been suggested that prisoners may become incompetent at living, may lose their will to carry on with life, decline physically and mentally and become deadened and that the likelihood of damaging consequences from imprisonment grows with increasing length of imprisonment. In the first decades of this century, it was suspected that there was a connection between lifelong imprisonment and psychiatric illnesses. However, the model of prison-related damage as having a uniform course with defined phases was soon abandoned.

7.2.1 Psychological Consequences

Taylor (1961) postulated a general psychopathological syndrome among long-term prisoners, characterised by apathy, flattening of affect and loss of initiative; a corresponding set of interpretations and causal pathways was proposed. Long-term imprisonment was seen as associated with monotony, loss of autonomy and loss of contact with the outside world (Walker 1983). Further psychopathological changes identified were early senility among older people, personality disorders, depression, states of nervous exhaustion and psychosomatic illnesses.

7.2.2 Moderating Variables

Variables proposed as moderating such prison damage have on the one hand been the prisoner's own characteristics, such as personality traits, age and state of health, but also attitude to the crime and the prison sentence, and on the other hand the nature of the prison environment, i.e. the atmosphere of the institution, basic living conditions (Walker 1983) and also remaining scope for action and behavioural characteristics of the prison officers.

7.2.3 Results of Empirical Studies

Discussions of possible connections between psychological disorders and longer prison sentences based on subjective understanding and personal experience have led to contradictory conclusions reported in the literature. In a study of self-reported damage from imprisonment, around half the subjects, individuals who had been released but were originally sentenced to life imprisonment, reported negative effects due to the prison sentence served (Peper and Kramer 1978).

The empirical studies of an English research group (Banister et al. 1973; Heskin et al. 1973) of a cross-sectional sample of 175 prisoners who had served different periods in prison did not detect any reduction of intellectual abilities. In their psychological study of personality, they found an increase in (predominantly intropunitive) hostility, a tendency to introversion and a reduction in self-esteem with increasing time in prison. In a follow-up study around 19 months later (Bolton et al. 1976), there were no indications of a general deterioration in psychological state, but rather an increase was observed in verbal intelligence, as well as a decrease in hostility, linked to an increase in emotional maturity. In a longitudinal study of 23 Austrian prisoners (Lapornik et al. 1996), no significant changes in intellectual capability were found, but there were indications of a reduction in concentration and attention capacities in the *Syndrom-Kurztest* (SKT; Short Syndrome Test), although without motivational effects being taken into consideration.

The results of the English research group were in essence confirmed by a cross-sectional study in Berlin (Rasch 1981). Rasch (1981) also pointed out that the subjective impression of a maturing process could not be evaluated, as the scores obtained for ego strength, emotional lability and immaturity for the group as a whole lay beyond the norm. He concluded by suggesting that enduring damage cannot be assumed necessarily to be the sequel to longer periods of imprisonment.

In a systematic longitudinal study, Zamble (1982) established that during the course of imprisonment, after an average of 7.1 years, the long-term prisoners he examined were increasingly occupied with work or other structured activities and their frequency of contact with "outside" had largely been maintained. As measured by psychometric testing, the level of emotional disturbance in the group as a whole fell, as did stress-related medical problems and the frequency of disciplinary measures. The conclusion drawn from this was that the likelihood of damaging consequences from imprisonment was low. Zamble (1982) suggests that the change in living environment at the beginning of the sentence causes substantial psychological complaints, but that the constancy of the prison atmosphere then leads to slow and gradual improvement.

7.2.4 Increasing Psychological Adaptability

There was also evidence of increasing psychological stabilisation in a more recent empirical study with a mean follow-up time of 9.5 years (Konrad 1994), based on the longitudinal observation of qualitative phenomena (the development of psychiatric diagnoses and physical illnesses) and results from psychological testing, e.g. results on the *Freiburger Persönlichkeitsinventar* (FPI; Freiburg Personality Inventory) and the 16-PF (*Persönlichkeitsfaktorentest*; Personality Factors Test). However, this should not be immediately equated with an absence of damage from prolonged imprisonment; it primarily simply indicates an increasing ability among the subjects to adapt psychologically to prison conditions as the years pass, internalising the reality of the prison environment.

The empirical findings could also be interpreted as reflecting the prisoner's constantly declining inner resistance against the situation of being imprisoned, with a corresponding reduction in psychological and physical indicators of stress. Thus it has been observed

that reactive aggressiveness as measured in psychological tests – which within the normal range functions as an indicator of intact capacity for self-assertion – moves significantly away from the normal range with increasing length of imprisonment. Subjects' tendency to give in increases, which can also be interpreted as an indicator of resignation. Such an interpretation suggests that parallels might be drawn with long-term hospitalisations, in which, with increasing periods of comprehensive care within a "total" institution, competence in living and the capacity to cope adequately with life outside the institution diminishes more and more. Understood in this way, damage due to imprisonment cannot therefore be evaluated using indicators which are measured exclusively during imprisonment.

8 References

Abram KM (1990) The problem of co-occurring disorders among jail detainees. Law Hum Behav 14: 333–345

Adams K (1986) the disciplinary experiences of mentally disordered inmates. Crim Just Behav 13: 297–316

Anderson HS, Sestoft D, Lillebaek T, Gabrielsen G, Kramp P (1996) Prevalence of ICD-10 psychiatric morbidity in random samples of prisoners on remand. Int J Law Psychiatry 19: 61–74

Arboleda-Flórez J (1994) An epidemiological study of mental illness in a remanded population and the relationship between mental condition and criminality. Doctoral dissertation, University of Calgary, Calgary

Backett SA (1987) Suicide in scottish prisons. Br J Psychiatry 151: 218–221

Banister PA, Smith FV, Heskin KJ, Bolton N (1973) Psychological correlates of long-term imprisonment. I. Cognitive variables. Br J Criminol 13: 312–322

Bellino TT (1973) The Ganser syndrome: a contemporary forensic problem. Int J Offend Ther Comp Criminol 17: 136–137

Bernheim JC (1994) Suicides and prison conditions. In: Liebling A, Ward T (eds) Deaths in custody: international perspectives. Bourne, Bournemouth, pp 91–108

Binswanger R (1979) Probleme der Gefängnispsychiatrie. Nervenarzt 50: 360–365

Birnbaum K (1931) Kriminalpsychopathologie und Psychobiologische Verbrecherkunde. Springer, Berlin Heidelberg New York

Bland RC, Newman SC, Dyck RJ, Orn H (1990) Prevalence of psychiatric disorders and suicide attempts in a prison population. Can J Psychiatry 35: 407–413

Bogue J, Power K (1995) Suicide in Scottish prisons, 1976–93. J Forensic Psychiatry 6: 527–540

Bolton N, Smith FV, Heskin KJ, Banister PA (1976) Psychological correlates of long-term imprisonment. IV. A longitudinal analysis. Br J Criminol 16: 38–47

Brooke D, Taylor C, Gunn J, Maden A (1996) Point prevalence of mental disorder in unconvicted male prisoners in England and Wales. BMJ 313: 1524–1527

Chiles JA, Cleve EV, Jemelka RP, Trupin EW (1990) Substance abuse and psychiatric disorders in prison inmates. Hosp Community Psychiatry 41: 1132–1134

Cooke DJ, Michie C (1996) Suicide in Scottish prisons: a methodological note. Leg Crim Psychol 1: 287–293

Council of Europe (1987) European prison rules. Council of Europe, Publications Section, Strasbourg

Dell S, Robertson G, James K, Grounds A (1993) Remands and psychiatric assessments in holloway prison. I. The psychotic population. Br J Psychiatry 163: 634–640

Dooley E (1990) Prison suicide in England and Wales, 1972–87. Br J Psychiatry 156: 40–45

Duff P (1997) Diversion from prosecution into psychiatric care. Who controls the gates? Br J Criminol 37: 15–34

DuRand CJ, Burtka GJ, Federman EJ, Haycox JA, Smith JW (1995) A quarter century of suicide in a major urban jail: implications for community psychiatry. Am J Psychiatry 152: 1077–1080

Edwards AC, Morgan DW, Faulkner LR (1994) Prison inmates with a history of inpatient psychiatric treatment. Hosp Community Psychiatry 45: 172–174

Felthous AR (1994) Preventing jailhouse suicides. Bull Am Acad Psychiatry Law 22: 477–488

Franklin RK (1988) Deliberate self-harm. Crim Just Behav 15: 210–218

Frühwald S (1996) Kriminalität und Suizidalität. ZfStrVo 45: 218–224

Grassian S (1983) Psychopathological effects of solitary confinement. Am J Psychiatry 140: 1450–1454

Grassian S, Friedman N (1986) Effects of sensory deprivation in psychiatric seclusion and solitary confinement. Int J Law Psychiatry 8: 49–65

Gunn J, Maden A, Swinton M (1991) Treatment needs of prisoners with psychiatric disorders. BMJ 303: 338–341

Harding T, Zimmermann E (1989) Psychiatric symptoms, cognitive stress and vulnerability factors. A study in a remand prison. Br J Psychiatry 155: 36–43

Haycock J (1989) Manipulation and suicide attempts in jails and prisons. Psychiatr Q 60: 85–98

Herpertz S, Saß H (1994) Offene Selbstbeschädigung. Nervenarzt 65: 296–306

Herrman H, McGorry P, Mills J, Singh B (1991) Hidden severe psychiatric morbidity in sentenced prisoners: an Australian study. Am J Psychiatry 148: 236–239

Heskin KJ, Smith FV, Banister PA, Bolton N (1973) Psychological correlates of long-term imprisonment. II. Personality variables. Br J Criminol 13: 323–330

Hodgins S (1995) Assessing mental disorder in the criminal justice system: feasibility versus clinical accuracy. Int J Law Psychiatry 18: 15–28

Hodgins S, Côté G (1990) Prevalence of mental disorders among penitentiary inmates in Quebec. Can Ment Health 38: 1–4

Hurley W (1989) Suicides by prisoners. Med J Aust 151: 188–190

Hurley W, Dunne MP (1991) Psychological distress and psychiatric morbidity in women prisoners. Aust New Zeal J Psychiatry 25: 461–470

Ivanoff A, Jang SJ, Smyth NJ (1996) Clinical risk factors associated with parasuicide in prison. Int J Offend Ther Comp Criminol 40: 135–146

Jemelka R, Trupin E, Chiles JA (1989) The mentally ill in prisons: a review. Hosp Community Psychiatry 40: 481–491

Jordan BK, Schlenger WE, Fairbank, Caddell JM (1996) Preva-

lence of psychiatric disorders among incarcerated women. II. Convicted felons entering prison. Arch Gen Psychiatry 53: 513–519

Karp JG, Whitman L, Convit A (1991) Intentional ingestion of foreign objects by male prison inmates. Hosp Community Psychiatry 42: 533–535

Knigge F (1932) Über psychische Störungen bei Strafgefangenen. Arch Psychiatr Nervenkr 96: 127–148

Konrad N (1994) Psychische Störung und lange Freiheitsstrafe. In: Jung H, Müller-Dietz H (eds) Langer Freiheitsentzug – wie lange noch? Plädoyer für eine antizyklische Kriminalpolitik. Forum, Bonn, pp 125–141

Konrad N (1995) Der sogenannte Schulenstreit – Beurteilungsmodelle in der Forensischen Psychiatrie. Psychiatrie, Bonn

Konrad N (1997) Psychiatrie im Justizvollzug. Recht Psychiatr 15: 51–59

Langelüddeke A, Bresser PH (1976) Gerichtliche Psychiatrie. De Gruyter, Berlin

Lapornik R, Lehofer M, Moser M et al (1996) Long-term imprisonment leads to cognitive impairment. Forensic Sci Int 82: 121–127

Lester D, Danto BL (1992) Suicide behind bars. Charles, Philadelphia

Liebling A (1992) Suicides in prisons. Routledge, London

Liebling A (1995) Vulnerability and prison suicide. Br J Criminol 35: 173–187

Maden A, Swinton M, Gunn J (1994) Psychiatric disorder in women serving a prison sentence. Br J Psychiatry 164: 44–54

Malone KM, Szanto K, Corbitt EM, Mann JJ (1995) Clinical assessment versus research methods in the assessment of suicidal behavior. Am J Psychiatry 152: 1601–1607

Mechler A (1981) Psychiatrie des Strafvollzugs. Fischer, Stuttgart

Morgan DW, Edwards AC, Faulkner LR (1993) The adaptation to prison by individuals with schizophrenia. Bull Am Acad Psychiatry Law 21: 427–433

Nitsche P, Wilmanns K (1911) Die Geschichte der Haftpsychosen. Z Ges Neurol Psychiatr 3: 353–382, 497–524

Peper B, Kramer H (1978) Problemschwerpunkte bei der Wiedereingliederung von begnadigten Lebenslänglichen. Bewährungshilfe 25: 1–13

Porporino FJ, Motiuk LL (1995) The prison careers of mentally disordered offenders. Int J Law Psychiatry 18: 29–44

Rasch W (1976) Die Gestaltung der Haftbedingungen für politisch motivierte Täter in der Bundesrepublik Deutschland. Monatsschr Kriminol Strafrechtsreform 59: 61–69

Rasch W (1981) The effects of indeterminate detention. Int J Law Psychiatry 4: 417–431

Rieger W (1971) Suicide attempts in a federal prison. Arch Gen Psychiatry 24: 532–535

Robertson G, Dell S, James K, Grounds A (1994) Psychotic men remanded in custody to Brixton prison. Br Med J 164: 55–61

Roesch R (1996) Mental health interventions in pretrial jails. In: Davies G, Lloyd-Bostock S, McMurran M, Wilson C (eds) Psychology, law and criminal justice. De Gruyter, Berlin, pp 520–531

Roesch R, Ogloff JRP, Eaves D (1995) Mental health research in the criminal justice system: the need for common approaches and international perspectives. Int J Law Psychiatry 18: 1–14

Schleusener J (1976) Psychische Veränderungen als Reaktion auf die Haftsituation. Z Strafvollzug Straffälligenhilfe 25: 19–23

Sieverts R (1979) Haftpsychologie. In: Sieverts, R, Schneider HJ (eds) Handwörterbuch der Kriminologie. De Gruyter, Berlin, pp 445–455

Smith CE (1987) Prison psychiatry and professional responsibility. J Forensic Sci 32: 717–724

Steadman HJ, Morris SM, Dennis DL (1995) The diversion of mentally ill persons from jails to community-based services: a profile of programs. Am J Public Health 85: 1630–1635

Tacke B, Hanisch A, Knaack M, Rode I (1975) Untersuchung psychiatrischer und psychologischer Faktoren, welche für Selbstbeschädigungen (das sog. Metallschlucken) von Häftlingen in Strafanstalten bestimmend sind. Westdeutscher Verlag, Opladen

Taylor AJW (1961) Social isolation and imprisonment. Psychiatry 24: 373–376

Teplin LA (1990) Detecting disorder: the treatment of mental illness among jail detainees. J Consult Clin Psychol 58: 233–236

Teplin LA (1994) Psychiatric and substance abuse disorders among male urban jail detainees. Am J Public Health 84: 290–293

Teplin LA, Abram KM, McClelland GM (1996) Prevalence of psychiatric disorders among incarcerated women. I. Pretrial jail detainees. Arch Gen Psychiatry 53: 505–512

Thole E (1976) Suicid im Gefängnis. Z Strafvollzug Straffälligenhilfe 25: 110–114

Toch H (1975) Men in crisis. Aldine, Chicago

Tomasevski K (1992) Prison health. International standards and national practices in Europe. Helsinki Institute for Crime Prevention and Control, affiliated with the United Nations, Helsinki

Virkkunen M (1976) Self-mutilation in antisocial personality (disorder). Acta Psychiatr Scand 54: 347–352

Walker N (1983) Side-effects of incarceration. Br J Criminol 23: 61–71

Walters RH, Callagan JE, Newman AF (1969) Effect of solitary confinement on prisoners. Am J Psychiatry 119: 771–773

Wilmanns K (1924) Die Abhängigkeit der Haftpsychosen vom Zeitgeist. Monatsschr Kriminol Strafrechtsreform 15: 308–333

Zamble E (1982) Behavior and adaptation in long-term prison inmates. Crim Just Behav 19: 409–425

CHAPTER

22

M.M. Fichter

Psychiatry and the Homeless

Translator: E. Taub

1 Introduction

Homelessness is context dependent; it changes over time and differs in different cultural areas. The diversity of terms used for it over the years itself demonstrates this fact. In pre-industrial Germany, the homeless were called *Vagierer*, *Kammesierer*, and *Hippenbuben* (all pejorative terms for migrant beggars and confidence-men); in the nineteenth century, they were known as *die verarmten Korrigenden* ("impoverished persons requiring correction") and were sent off to workhouses; in the early twentieth century, they were *Landstreicher und Vagabunden* ("tramps and vagabonds," Wilmanns 1906) or, in psychiatric circles, *die unsteten Psychopathen* ("vagrant psychopaths"); under Nazi rule, they became *die arbeitsscheuen Nichtsesshaften* ("work-avoiding persons without fixed residence"); and after the Second World War they became "the homeless medically ill" in English-speaking countries and *die Wohnungslosen* ("the dwelling-less") or *die Obdachlosen* ("the shelter-less") in the German-speaking countries. These changing expressions partly reflect a change in the population designated, and partly a change in society's attitude toward it.

In the early years of the twentieth century, German psychiatric research on the homeless achieved international prominence through the work of Karl Wilmanns, above all through his monograph *Zur Psychopathologie des Landstreichers* ("The Psychopathology of Tramps"), which appeared in 1906. Working in the tradition of Kraepelin, Wilmanns showed that a considerable percentage of the "tramps and vagabonds" then confined in workhouses and prisons were suffering from dementia praecox. Consequently, he advocated the humane accommodation and psychiatric treatment of mentally ill "tramps," publicly supported preventive education of children and adolescents at risk, and provisions to combat unemployment and alcoholism. Kurt Schneider's concept of psychopathy was also applied to research on the homeless (Schneider 1934).

In Nazi Germany, starting with the *Bettlerrazzia* ("Raid on Beggars") of September 1933, homeless persons with "pathological inferiorities" were confined to workhouses and concentration camps. Several thousand perished in concentration camps. Later applications for reparations were turned down on the ground that Sec. 1 of the the German Federal Compensation Law (*Bundesentschädigungsgesetz*, BEG) provided for reparations only for injustices committed out of motives of racial, political, or religious persecution.

Rising unemployment since the 1970s, increasing housing costs, and the policy of "deinstitutionalization" (which resulted in a considerable reduction in the number of beds for psychiatric patients) exacerbated the problem of homelessness in most Western industrial countries. After a latency period, there also ensued an intensification of research on the homeless, particularly in English-speaking countries.

In North America and Northern Europe in the 1980s and 1990s, social measures were introduced that removed homeless people from the streets by making living space available to them in shelters and boarding houses. These measures were effective on the whole, although the psychiatric and other medical care received by these formerly homeless persons is still largely inadequate. Legal measures were taken in the United States in the 1980s (McKinney Homeless Assistance Act). A further discussion of the historical development of homelessness is contained in Greifenhagen and Fichter (1996).

2 Epidemiology

In the last 20 years, a number of methodologically sound studies, involving representative population samples, were performed in North America, and later in Germany, to determine the prevalence of mental illness in the homeless. These studies required that the population of "homeless persons" be rigorously defined and that a representative sample of it be taken. Special research methods were needed, such as those applied successfully by Koegel et al. (1988) in Los Angeles and by Fichter et al. (1996) in Munich. Moreover, the development of operational psychiatric diagnosis (DSM-III, DSM-III-R, DSM-IV, ICD-10) and of standardized and structured interviews for the reliable assessment of psychiatric manifestations and the provision of diagnoses further increased the accuracy of psychiatric and epidemiologic studies of the homeless.

The instantaneous and lifetime prevalence of mental illness, including alcohol and drug abuse/dependency, among homeless populations in the 1980s and 1990s is now well documented (Jones et al. 1991; Fichter et al. 1996). Several American studies showed that the lifetime prevalence of severe mental illness (schizophrenia or typical affective illnesses) among the homeless was 15%–30%, and of alcohol and drug abuse/dependency 60%–70% (Fischer et al. 1986; Koegel et al. 1988; Robertson 1992).

Many epidemiological studies of mental illness in the homeless have shown that the homeless make far

less use of psychiatric and other medical services than the general population. That this group is medically underserved is even more evident if we consider that the prevalence of mental illness is much higher in the homeless than in the general population. One of the conclusions of the UCLA Homeless Health Study (part of the Rand Corporation's Course of Homelessness project carried out in the United States) was that homeless persons gave other matters that are important in everyday life and survival higher priority than their health: solving the recurrent daily problems of finding food, clothing, and shelter required so much effort that they had little opportunity to look after their health (Gelberg et al. 1997).

Research on homelessness and mental illness in the Federal Republic of Germany was reviewed by Rössler et al. (1994) and by Greifenhagen and Fichter (1996). Fichter et al. (1996, 1997) reported the results of the first epidemiological study of a representative sample of homeless men in Munich, the First Munich Homeless Project. In June and December 1988, in the first stage of this study on the prevalence of mental illnesses, 456 homeless men were briefly interviewed to provide a "preliminary sampling" in order to determine the population to be studied. A second stage resulted in the definitive selection of a sample of 172 homeless men, of whom 146 (85%) were studied with the aid of a standardized psychiatric interview based on the DSM-III criteria (Diagnostic Interview Schedule; DIS). These interviews took place in 1989/1990.

The homeless men were 26–60 years old, with an average age of 43 years. A large number of them either had never been married (55%) or were divorced (38%). Their level of education was below average (13% had not finished secondary school). They had first become homeless at an average age of 33.7 years (standard deviation, 10.4 years); the duration of homelessness, according to their responses in interviews, was 9 years on average (standard deviation, 8.1 years). Recurrent homelessness was not uncommon, and most of these homeless men had no regular work. A total of 89.7% had an annual income of less than 20,000 marks (approximately $12,500), and 67.8% had an annual income of less than 10,000 marks (approximately $6,250).

These homeless men were found to be relatively sedentary; 79% had lived exclusively in Munich during the 12 months before the interview, 12.3% in Munich and one other city, and the rest in three or more cities. Their preferred sleeping place in the 4 weeks before the interview was outdoors in 70.5%, in a homeless shelter or boarding house in 19.9%, with acquaintances or relatives in 18.5%, and in unfinished or demolished buildings in 14% (multiple listing was allowed). In the time since the study was performed, these numbers have likely shifted to a large extent in the direction of residential homes, shelters, and boarding houses.

The homeless men themselves stated that they were homeless because of financial problems (65.1%), unemployment or loss of employment (49.3%), divorce (8.2%), or other family problems (11.0%). Mental problems were said to be a reason for homelessness by 17.1%, alcohol problems by 37.0%, drug problems by 2.7%, and general medical problems by 12.3% (listing of multiple reasons was allowed).

Life on the street has its dangers: 40.4% of the homeless men surveyed had been victims of physical violence leading to injury on at least one occasion; 33.6% had been robbed, 28.8% had been victims of theft, and 9.6% had been victims of sexual harassment or sexual abuse. The homeless men interviewed had relatively frequent contact with the police: 78.1% had been arrested at least once, usually for smaller offenses such as riding the subway without a ticket or petty theft, while 18.5% of the homeless men had been convicted of robbery, battery, or another crime. Karl Wilmanns already noted at the beginning of the twentieth century that mentally ill homeless persons were relatively frequently arrested and imprisoned, and it was only late in the course of illness that they entered psychiatric treatment.

Table 1 contains a summary of the more important findings obtained in Munich in 1989 (Fichter et al. 1996). A total of 138 out of 146 homeless men studied (94.5%) had an axis I diagnosis of a mental illness according to DIS/DSM-III. The most common diagnostic categories were substance abuse, affective diseases, anxiety disorders, and schizophrenia. The prevalence rates were much higher than in reference samples of the general population. There was also a relatively high psychiatric multimorbidity; 53.4% had two or more lifetime psychiatric diagnoses. Only 27.6% of the homeless men had ever been treated in a psychiatric inpatient unit, 11% had been in an alcoholic rehabilitation clinic, and 2.8% in a drug rehabilitation clinic.

The same group that had conducted the First Munich Homeless Project described above carried out a further, comprehensive public health study in Bavaria, in which 262 homeless men were studied at the time of an initial interview, and again 3 years later, with respect to the prevalence of mental (and serious physical) illness. In the same period of time, the effect of an intervention (assignment of fixed, permanent housing) on the further course of homelessness and health was studied in a separate group of subjects. This Public Health Homeless Project arrived at essentially the same conclusions as the First Munich Homeless Project (Fichter and Quadflieg 1999). Reviews on the

Table 1. Lifetime prevalence of mental illness according to the DIS/DSM-III criteria among 146 homeless men in Munich in 1989 (after Fichter et al. 1996)

	Homeless men in Munich (%)	Representative population sample in Los Angeles (%)[a]
Schizophrenia	12.4	0.5
Affective illness	41.8	8.8
Anxiety disorder	22.6	8.9
Cognitive impairment	8.9	NR
Alcohol abuse/ dependency	91.1	23.9
Drug abuse/ dependency	17.8	13.9
Any DIS/DSM-III axis I diagnosis	94.5	NR
No psychiatric diagnosis	5.5	NR
Only one DIS/DSM-III axis I diagnosis	41.1	NR
Two DIS/DSM-III axis I diagnoses	29.4	NR
Three DIS/DSM-III axis I diagnoses	18.5	NR
Four or more DIS/ DSM-III axis I diagnoses	5.5	NR

DIS, Diagnostic Interview Schedule; NR, not reported.
[a]Epidemiological Catchment Area (ECA) Study of 3055 men from Koegel et al. (1988).

prevalence of mental illness in the homeless may be found in Robertson and Greenblatt (1992), Schutt and Garrett (1992), Scott (1993), Rössler et al. (1994), Kuhlman (1994) and Bhugra (1996).

As for the frequency of mental illness in homeless women, there is a difference between women caring for one or more children (prevalence less than among homeless men) and women not caring for children (in this study, prevalence higher than among homeless men). The prevalence figures among homeless women vary from study to study because of differences in the sample populations (Merves 1992; Zima et al. 1996).

3 Possible Causative Factors and Course

A number of potential causative factors for the generation and perpetuation of homelessness have been discussed in the literature. Table 2 classifies these possible causative factors according to the level of social organization at which they are observed.

On the broad *cultural level,* discrimination against minorities and the basic social attitude toward marginal groups may play a role. On the *institutional level,* unemployment, an inadequate supply of housing, psychiatric deinstitutionalization without the provision of other measures by the community, and inadequate psychiatric and other medical services for the homeless may all contribute further. The services offered may be poorly integrated, not suitable to the needs of the recipients, or poorly accessible (*organizational level*). Inadequate social support, e.g. from family and friends, may exacerbate the problem (*group level*).

Finally, the *individual* may exhibit traits that elevate the risk of homelessness, such as the presence of mental illness, childhood traumas (Herman et al. 1997), deracination through emigration (Silove et al. 1997), unemployment (Fergusson et al. 1997), or

Table 2. Possible causative factors for homelessness (modified from Morse 1992)

Level of observation or causative factor	Can produce or maintain homelessness (example)
Cultural level	Discrimination against minorities, indifference to marginal groups
Institutional level	Unemployment, inadequate supply of housing, curtailment of financial support, policy of psychiatric deinstitutionalization with deficient establishment of corresponding measures in the community, lack of psychiatric and other medical services for the homeless
Local community	Unilateral local political measures
Organizational level	Inadequate organization and coordination of services offered, poor accessibility, mismatched to needs
Group	Inadequate social support to the individual
Individual	Illnesses, handicaps, personal preferences, inadequate adaptation to new (e.g. occupational) requirements

alcoholism (Koegel and Burnam 1988; Fichter et al. 1997; Welte and Barnes 1992; Podschus and Dufue 1995). Bremner et al. (1996) found a relationship between a low initial intelligence quotient and later homelessness; there was an even closer relationship between a decline in the intelligence quotient over time and homelessness, which may be explained by the presence of other risk factors (mental illness, schizophrenia, or substance abuse).

Very little is known about the course of homelessness and the course of mental illnesses in the homeless. Studies of these questions can also reveal factors that perpetuate illness (Koegel and Burnam 1991; Craig et al. 1996).

4 Intervention and Prevention

The conceptual scheme of causative factors for homelessness on several levels of observation, as illustrated in Table 2, suggests a similar multilevel approach to the classification of interventional measures. On each level of observation, there may be a corresponding intervention. Factors on the cultural, national, or institutional level, such as discrimination against minorities, a high rate of unemployment, or an inadequate supply of housing, cannot be effectively counteracted on the individual level. The same holds for the policy of psychiatric deinstitutionalization.

In view of the fact that, in the 1980s and 1990s, a very large part of the homeless in the United States and Northern Europe either were suffering from mental illness or had a history of mental illness, often with lasting sequelae, we as psychiatrists are obliged to do our best to improve the lot of the homeless mentally ill. This can, and should, be done simultaneously on each of the levels listed in Table 2. We can do the following (Levine and Rog 1990; Katz et al. 1992; Breakey 1997):

1. Provide help to individuals
2. Contribute to better coordination of the existing provisions for the homeless and their better adaptation to current needs
3. On the institutional level, influence the development of collateral provisions, and the establishment of provisions with a sufficiently low threshold to reach the affected persons

So far, there has been much talk and little action in this area, and scientific evaluations are sparse. There have been a few empirical evaluations in recent years. Lipton et al. (1988) studied 49 severely mentally ill homeless persons in the emergency room of a hospital in New York. The subjects were assigned either to a control group ("routine discharge planning") or to an intervention group receiving "individualized case management, coordination of public assistance or social security benefits, medication monitoring, money management, meals, activity therapy, and, when appropriate, referrals to psychosocial and rehabilitation programs" (Lipton et al. 1988, p. 41). Over the ensuing year, the patients in the intervention group spent fewer nights in the hospital and reported greater satisfaction with their life situation.

Building further on these results, Morse et al. (1992) compared several different conditions of treatment ("traditional outpatient mental health treatment vs. daytime drop-in centers offering food, recreational facilities and referrals vs. a continuous treatment team that provided therapeutic services to help clients to cope with personal problems, link with psychiatric and other services and learn community living skills"). Patients in the "continuous treatment" group were homeless for fewer days, made more use of the opportunities offered to them, and were more satisfied with what they received.

Cauce et al. (1994) applied a similar study protocol in a group of 300 adolescents with comparable results. Studies on the reduction of alcohol or drug dependency were reviewed by Mercier et al. (1992). The results were rather disappointing with regard to abstinence, but significant effects were found with regard to social adaptation, arrests, convictions, stability of living situation, and access to medical services.

Susser et al. (1997) studied 69 men who were transferred from a homeless shelter to an apartment in the community or city. Half of the subjects were randomized to standard care and the other half to an intervention group with additional care for 9 months ("critical time intervention"). The results indicated that additional care lowers the recidivism rate for homelessness. Toro et al. (1997) evaluated the course of 200 homeless persons randomized either to a control group or to a group receiving "intensive case management intervention." Follow-up at 6, 12, and 18 months revealed positive effects on the severity of psychopathology, harmful life events, and the quality of the home environment.

5 Overview

Many efforts have been made in the past decade to take the homeless off the streets and put a roof over their heads. The mentally ill homeless, however, cannot be made well by a roof over their head alone. The mentally ill homeless (and former homeless), in whatever country they may be, need help with their personal affairs (e.g. social contact, financial, occupa-

tional) and, above all, competent psychiatric treatment. Some of the illnesses affecting the homeless are inherently accompanied by limited insight into the illness and limited motivation to undergo treatment (alcohol and drug dependency, schizophrenia, manic phases). It is thus necessary to organize and coordinate the provision of care for the homeless so that it may also reach those who do not seek help for themselves, especially in situations where bureaucratic or other impediments make it even less accessible.

Low-threshold care provisions must be brought actively to the affected individuals ("outreach"). Because the services available today are highly fragmented, what is needed is care of the affected individuals in a manner corresponding to their needs, initiatives by individual organizations and facilities, accompanying evaluation, and continuous improvement and adaptation to the changing needs of the homeless population.

6 References

*Bhugra D (ed) (1996) Homelessness and mental health. Cambridge University Press, Cambridge

Breakey WR (1997) Editorial: It's time for the public health community to declare war on homelessness. Am J Public Health 87(2): 153–155

Bremner AJ, Duke PJ, Nelson HE, Pantelis C, Barnes TR (1996) Cognitive function and duration of rooflessness in entrants to a hostel for homeless men. Br J Psychiatry 169: 434–439

Cauce AM, Morgan CJ, Wagner V et al (1994) Effectiveness of intensive case management for homeless adolescents: results of a 3-month follow-up. J Emotion Behav Disord 2: 219–227

Craig TKJ, Hodson S, Woodward S, Richardson S (1996) Off to a bad start. A longitudinal study of homeless young people in London. (Final Report to the Mental Health Foundation, London)

Fergusson DM, Horwood LJ, Lynskey MT (1997) The effects of unemployment on psychiatric illness during young adulthood. Psychol Med 27: 371–381

*Fichter MM, Quadflieg N (1999) Alcoholism in homeless men in the mid-nineties: results from the Bavarian public health study on homelessness. Eur Arch Psychiatr Clin Neurosci 249: 34–44

Fichter MM, Koniarczyk M, Greifenhagen A, Koegel P, Quadflieg N, Wittchen HU, Wölz J (1996) Mental illness in a representative sample of homeless men in Munich, Germany. Eur Arch Psychiatry Clin Neurosci 246: 185–196

Fichter MM, Quadflieg N, Greifenhagen A, Koniarczyk M, Wölz J (1997) Alcoholism among homeless men in Munich, Germany. Eur Psychiatry 12: 64–74

*Fischer P, Shapiro S, Breakey WR, Anthony JC, Kramer M (1986) Mental health and social characteristics of the homeless: a survey of Baltimore shelter users. Am J Public Health 76: 519–524

Gelberg L, Gallagher TC, Andersen RM, Koegel P (1997) Competing priorities as a barrier to medical care among homeless adults in Los Angeles. Am J Public Health 87(2): 217–220

*Greifenhagen A, Fichter MM (1996) Psychiatrische Obdachlosenforschung. Von der "Psychopathologie des Landstreichers" zu den "Homeless Mentally Ill". Nervenarzt 67: 905–910

Herman DB, Susser ES, Struening EL, Link LB (1997) Adverse childhood experiences: are they risk factors for adult homelessness? Am J Public Health 87: 249–255

Jones JM, Levine IS, Rosenberg A (eds) (1991) Homelessness. Am Psychol 46 (special issue): 1109–1111

Katz S, Nardacci D, Sabatini A (1992) Intensive treatment of the homeless mentally ill. American Psychiatric Press, Washington DC

*Koegel P, Burnam A (1988) Alcoholism among homeless adults in the inner city of Los Angeles. Arch Gen Psychiatry 45: 1011–1018

Koegel P, Burnam A (1991) The Course of Homelessness Study: aims and design. Paper presented at the 119th annual meeting of the American Public Health Association, November 1991, Atlanta

Koegel P, Burnam MA, Farr RK (1988) The prevalence of specific psychiatric disorders among homeless individuals in the inner city of Los Angeles. Arch Gen Psychiatry 45: 1085–1092

Kuhlman TL (1994) Psychology on the streets: mental health practice with homeless persons. Wiley, Chichester

Levine IS, Rog DJ (1990) Mental health services for homeless mentally ill persons: federal initiatives and current service trends. Am Psychol 45: 963–968

Lipton FR, Nutt S, Sabatini A (1988) Housing the homeless mentally ill: a longitudinal study of a treatment approach. Hosp Community Psychiatry 39: 40–45

Mercier C, Fournier L, Pelandeau N (1992) Program evaluation of services for the homeless: challenges and strategies. Eval Progr Plan 15: 416–417

Merves ES (1992) Homeless women. Beyond the bag lady myth. In: Robertson MJ, Greenblatt M (eds) A national perspective. Plenum, New York, pp 229–244

Morse GA (1992) Causes of homelessness. In: Robertson MJ, Greenblatt M (eds) A national perspective. Plenum, New York, pp 3–17

Morse GA, Calsyn RJ, Allen G, Tempelhoff B, Smith R (1992) Experimental comparison of the effects of three treatment programs for homeless mentally ill people. Hosp Community Psychiatry 43: 1005–1010

Podschus J, Dufue P (1995) Alkoholabhängigkeit unter wohnungslosen Männern in Berlin. Sucht 41(5): 348–354

Robertson MJ (1992) The prevalence of mental disorder among homeless people. In: Jahiel R (ed) Homelessness: a prevention-oriented approach. John Hopkins University Press, Baltimore, pp 57–86

Robertson MJ, Greenblatt M (1992) Homelessness. A national perspective. Plenum, New York

Rössler W, Salize HJ, Biechele U (1994) Psychisch kranke Wohnsitzlose – Die vergessene Minderheit. Psychiatr Prax 21: 173–178

Schneider K (1934) Die psychopathischen Persönlichkeiten, 3rd edn. Deuticke, Leipzig

Schutt RK, Garrett GR (1992) Responding to the homeless. Policy and practice. Plenum, London

Scott J (1993) Homelessness and mental illness. Br J Psychiatry 162: 314–324

Silove D, Sinnerbrink I, Field A, Manicavasagar V, Steel Z (1997) Anxiety, depression and PTSD an asylum-seekers: associations with pre-migration trauma and post-migration stressors. Br J Psychiatry 170: 351–357

*Susser E, Valencia E, Conover S, Felix A, Tsai WY, Wyatt RJ (1997) Preventing recurrent homelessness among mentally ill men: a "critical time" intervention after discharge from a shelter. Am J Public Health 87: 256–262

Toro PA, Passero Rabideau JM, Bellavia CW, Daeschler CV, Wall DD, Thomas DM, Smith SJ (1997) Evaluating an intervention for homeless persons: results of a field experiment. J Consult Clin Psychol 65: 476–484

Welte JW, Barnes GM (1992) Drinking among homeless and marginally housed adults in New York State. J Stud Alcohol 53(4): 303–315

*Wilmanns K (1906) Zur Psychopathologie des Landstreichers. Eine klinische Studie. Barth, Leipzig

Zima BT, Wells KB, Benjamin B, Duan N (1996) Mental health problems among homeless mothers. Arch Gen Psychiatry 53: 332–338

CHAPTER

23

A.J. Holland

Mental Retardation: A Psychiatric Perspective

My thanks go to Robbie Patterson for considerable administrative and secretarial support and to colleagues in the community teams for adults with learning disability, Lifespan NHS Trust, and colleagues at the University of Cambridge for help and advice.

1 Introduction

1.1 Historical Perspectives

This past century has seen dramatic changes in almost every aspect of the lives of people with mental retardation. Among the most striking are the attitudes of society and the changing philosophy away from segregation towards integration. Equally as striking have been those advances in fields such as cytogenetics, molecular genetics and teratology that have enabled some of the causes of mental retardation to be identified. These two examples, taken from the diverse fields of the social and biological, vividly illustrate how significant this change has been. In the early 1900s, the eugenics movement was a powerful force and genetic theories were used to justify what are now seen as harsh and inhumane attitudes. Equally, biological knowledge was in its infancy. Inborn errors of metabolism were only just beginning to be investigated, and the human chromosome complement was a long way off being described. In the later part of this century, increasing understanding of the process of early human development at both the biological and psychological levels led to studies of whether mental retardation was a result of "delayed" and/or "deviant" development (Hodapp et al. 1990). Research in the behavioural sciences has centred on trying to understand more about the characteristics and aetiology of specific disabilities such as autism and the factors which contribute to the occurrence of behavioural and psychiatric disorders. These have included epidemiological studies such as those of Rutter et al. (1970) and Birch et al. (1970), which demonstrated the high rates of such disorders in children and adults with mental retardation. Since then, an extensive and primarily psychological literature on causation of behaviour disorders has also developed. The challenge now is to attempt to integrate developmental and biobehavioural models.

1.2 Terminology

The use of terms such as "mental retardation" cannot fully convey the complexity of abilities and disabilities present in any given individual but, as has been seen historically, can instead easily become a term of abuse and can be used to discriminate against the person concerned. The issues are well illustrated by the fact that the terminology varies across countries and changes over time. In its classification of mental disorders (ICD-10), the World Health Organization (1992) uses the term "mental retardation" as it is still used in the United States (American Psychiatric Association 1994). In the United Kingdom, terms such as "idiot", "imbecile" and "feeble-minded" were used in legislation in the early 1900s, followed by "mental sub-normality", "mental handicap" and now "learning disability". In Australia, and increasingly in the United States, the term "developmental disability" is used. Labels are necessary to enable resources to be directed to those people with special needs, but they should not offend and they should reliably inform. In this chapter, the term "mental retardation" has been chosen primarily because it is still the most widely used. Because labels in this area all too easily become pejorative, it is essential that they are used with respect and with the aim of helping the person concerned rather than in a manner which results in segregation and denial of rights.

1.3 Human Development

Understanding how specific abilities and disabilities arise and identifying the nature of the interventions that are needed requires an appropriate conceptual framework. In the case of mental retardation, this particularly requires a developmental perspective and recognition of how the importance of biological and environmental factors changes over time (Clarke and Clarke 1984). Normal human development, from conception to adult life, follows a path which is largely predictable. Prenatal development is characterised by the systematic differentiation of organ systems and parts of the body, resulting in a viable fetus at the time of birth. In childhood, growth and physical development continues and the reaching of developmental milestones such as smiling, sitting, standing, walking and early language development becomes apparent. More advanced and uniquely human abilities are also acquired, such as literacy and numeracy, the ability to engage in symbolic play and understanding and communicating complex ideas and feelings through an increasingly sophisticated language. This process of cognitive development starts prenatally (for a review, see Karmiloff-Smith 1995) and is dependent on the regulation of the expression of particular genes in specific organ systems over time. The modulation of gene expression by intra- and extracellular environmental factors occurs within the context of a social environment which meets both the nutritional and emotional needs of the individual (for an historical perspective on "psychobiological development", see Gottlieb 1996).

There is variation in the precise timing of these developmental milestones because of inherent differ-

ences between individuals and their environments. At birth or in the neonatal period, a child may be identified as having a syndrome commonly associated with mental retardation and/or the presence of developmental delay, with significant intellectual impairment and learning disabilities becoming apparent in early life (see Table 1 for examples and below for further discussion). In this context, the role of paediatricians and clinical geneticists is to determine the likely cause of the developmental delay, advise whether specific treatments are indicated and, if requested by the family, to give advice as to the likely outcome of any future pregnancies. At this early stage, establishing the cause of the developmental delay is also important. It may help the parents adjust to having a child who may have a specific genetic or environmentally determined disorder, and it may help in determining whether other problems, e.g. sensory or physical impairments, are likely to occur in the future.

There are also rare situations where the presence of developmental delay and subsequent mental retardation is not immediately apparent, but early identification of the presence of a specific disorder and the starting of treatment are important. This is particularly so for congenital hypothyroidism and the rare inborn error of metabolism phenylketonuria. The former is treated with thyroid replacement and the latter with the use, until adult life, of a diet low in the amino acid phenylalanine. Screening for both these disorders shortly after birth is routine in many countries, as prompt treatment prevents the development of what would otherwise be significant intellectual impairment and mental retardation.

2 Definitions

2.1 Diagnostic Criteria

At each stage in the progression of development, the extent to which the child or adult has attained those abilities expected for that chronological age can be assessed and the presence or absence of any impairments or disabilities judged. When the process of development falls significantly below what is expected in terms of intellectual ability and the acquisition of particular skills, the term "mental retardation" may apply. This term therefore implies the presence of a disorder of development which has affected the ability to acquire social and living skills (adaptive behaviours) in association with significant impairment of intelli-

Table 1. Examples of syndromes commonly associated with varying degrees of mental retardation either due to a gene mutation or the presence of a chromosome abnormality

Syndrome	Genotype	Phenotype
Angelman syndrome	Deletion chromosome 15q11–13, maternal chromosome	Abnormal gait, characteristic facial appearance, excessive laughing
Down syndrome	Trisomy 21	Characteristic facial appearance, flattened occiput, short stature, shortened and curved little finger
Fragile X syndrome	CGG repeats FMR-1, gene Xq27.3	Long face, prominent ears, large testes post-puberty, speech abnormalities, gaze avoidance, females can be mildly affected
Lesch-Nyhan syndrome	HPRT gene mutation, Xq26–27.3	Severe self-injurious behaviour, choreoathetosis, motor delay, dysarthric speech
Prader-Willi syndrome	Deletion chromosome 15q11–13, paternal chromosome	Neonatal hypotonia, delayed sexual development, short stature, characteristic facial appearance, over-eating leading to obesity
Rett syndrome	Possibly X-linked dominant	Females only affected, regression in development 1–2 years of age, repetitive movements hands and body, abnormal breathing
Tuberous sclerosis	Mutations chromosome 9q34 and 16p13	White and shagreen skin patches, facial rash (adenoma sebaceum), hamartomas, renal lesions
Williams syndrome	Deletion elastin gene chromosome 7q11.23	Characteristic facial appearance, aortic and/or pulmonary heart valve stenoses, hypercalcaemia

There are several hundred syndromes resulting in significant mental retardation, and readers are advised to consult genetic and paediatric texts for a complete description of the phenotype of the above syndromes and for a full list of syndromes. These have been chosen either because they are relatively common or because they illustrate a specific genetic/chromosomal abnormality.

gence. The latter is usually taken to be a score on an established intelligence quotient (IQ) test greater than two standard deviations below the mean. For those tests with means of 100 and standard deviations of 15 points, this is a score of less than 70. A summary of the DSM-IV criteria for mental retardation is given in Appendix A. As can be seen, the presence of a significant intellectual impairment in itself is insufficient to determine the presence or absence of mental retardation, and additional evidence of a functional disability is required. Although the benefits and pitfalls in the measurement of intellectual capacity (IQ) have been debated, it does serve as an objective, standardised and reliable measure in assessing an individual's ability to perform cognitive tasks which, in turn, are likely to relate to their ability to solve everyday problems. The IQ therefore remains an important guide to the severity of mental retardation. Measurement of "adaptive behaviour" and general functioning has been more problematical, but the use of assessment instruments that are well standardised has been a considerable advance. Examples of these include the revised Vineland (Sparrow et al. 1984) and the revised Adaptive Behaviour Scales (Nihira et al. 1993) of the American Association for Mental Retardation (AAMR). These informant-based schedules are divided into different domains (e.g. socialisation), giving an overall comprehensive picture of an individual's strengths and weaknesses.

The term "mental retardation" refers to a highly heterogeneous group of people who have in common evidence of developmental delay that affects their ability to learn. However, individuals differ widely in the extent of their disabilities, the presence or absence of other impairments and disabilities (e.g. physical and sensory) and in the cause and exact characteristics of their impairments and disabilities. Like the rest of the general population, they also vary in personality and life experience. No one term can adequately describe this group of people with such varied and complex needs, and for this reason a more hierarchical definition has been adopted by the World Health Organization.

2.2 Impairments, Disabilities and Handicaps

The World Health Organization (1980) has proposed a classification scheme which helps to overcome some of these difficulties and, most importantly, also helps to guide appropriate intervention. The definitions of the above terms, and how they link together, are given in Appendix B and Fig. 1. In this scheme, mental retardation can be understood on a number of different levels. At the level of "impairment" (intellectual impairment), the organ system involved is primarily the brain. It is impairment in this organ system for genetic, chromosomal or environmental reasons (e.g. as listed in Table 1) which has primarily affected the acquisition of developmentally determined skills and the process of learning. The "disability" (learning disability) which follows is that of the ability to learn and acquire those skills which come with development and are necessary for independent life. The extent of the disability may vary, e.g. from profound delay in the acquisition of language, motor and self-help skills on the one hand to more subtle difficulties in the acquisition of skills such as numeracy and literacy, this only becoming apparent when the child starts school. As is discussed later in the chapter, the level of disability may relate to the degree of brain impairment but may be influenced by the extent of interventions and additional support available. The intervention at this level is one of special education and additional help with physical and other disabilities. In the United Kingdom, this involves the process of "statementing", which identifies the special educational needs of the child and entitles them to nursery education from the age of 3 and full-time education from the age of 5 until they are 19. The "handicap" (mental handicap) experienced by the person is a measure of the extent to which the person is disadvantaged as a result of their disabilities. Importantly, this is not something which is necessarily fixed by the level of the disability, but is seen as an interaction between the disabilities on the one hand and the extent of support and intervention on the other. For example, a person who has not acquired the ability to appreciate danger and therefore cannot be out by him-/herself will have a very restricted quality of life unless family and care staff are available to go with the person to shops, on buses etc. The level of handicap can therefore vary significantly depending on the level of support and attitude of carers supporting the person concerned. The better the level of support available to the individual, the less the likely disadvantage and handicap experienced by the person him- or herself.

This system of conceptualising mental retardation helps to overcome the problems of stigma which are

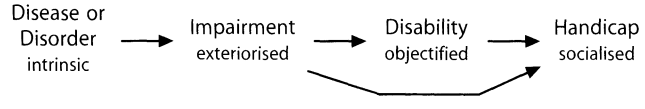


Fig. 1. Link between impairments, disabilities and handicaps

associated with more commonly used labels and provides the mechanism whereby the needs of any particular person and his or her family can be considered. These include tasks as varied as the identification of the extent and the cause of the impairment as well as intervention at the educational and social levels.

3 Epidemiology

3.1 Prevalence Rates

The above definitions help to illustrate how difficult it is to obtain accurate estimates of the number of people considered to have mental retardation. Firstly, definitions vary and there is no obvious dividing line between what is mental retardation and what is not. Secondly, prevalence rates are highly dependent on whether a true population sample is identified or whether an "administrative sample" (i.e. those known to mental retardation services) are used. Thirdly, rates are very age dependent in that those with profound mental retardation have an increased mortality rate and many with mild mental retardation leave services once they finish school. Finally, there are likely to be cultural differences depending on the educational system and expectations in any particular society or socio-economic group. Figures should therefore be considered with caution.

The World Health Organization's system of impairments, disabilities and handicaps used to conceptualise the complex disabilities referred to as mental retardation can be refined and operationalised so that, for example, for epidemiological purposes more accurate estimates of the extent of such disabilities in any community can be obtained. The appropriate use of IQ scores can be used to obtain some measure of the proportion of the population who have a significant intellectual impairment and are therefore at risk for learning disability and for being disadvantaged or "handicapped" as a result. Intellectual performance measured by such tests has been shown to have an almost normal distribution in populations with a skew to the left due to the presence of biologically determined disorders affecting brain development. Approximately 2%–2.5% of a population fall greater than two standard deviations below the mean of 100 (IQ < 70). Of those with IQs below 70, approximately 80% have mild mental retardation (more accurately, mild intellectual impairment), 12% moderate, 7% severe and less than 1% profound.

Although there were variations in early prevalence studies due to the use of different definitions, there is general agreement from a number of studies that the prevalence rate for severe mental retardation is 0.3%–0.4% (for a review, see Fryers 1991). These estimates are obtained primarily from administrative samples, but given the level of need of the individuals in the studies, it is likely that the majority had been identified. The estimates can therefore be seen as reliable. When it comes to aetiology, the distinction between those with IQs between 50 and 70 (mild) and those with IQs less than 50 (moderate to severe) becomes important (see also the epidemiological studies by Rutter et al. 1970; Birch et al. 1970).

3.2 Learning Difficulties

The percentage of children defined as having learning difficulties within a population is greater than those with an intellectual impairment. In the United Kingdom, the Warnock Committee estimated that about 20% of children are likely to have learning difficulties at some point requiring special educational help. The reasons for their special educational needs can be as diverse as severe mental retardation, the presence of a chronic physically disabling condition, emotional and behavioural problems and specific learning difficulties. Given these diverse reasons, the severity and extent of "learning disabilities" are not easy to define and may well depend on a variety of factors including, for example, the school system, the extent of social disadvantage and definitions of specific and generalised learning disabilities. While some children will have been identified prior to starting school as having profound developmental delay affecting their intellectual and learning abilities, others will only be identified as having "learning difficulties" after starting school. Similarly, for some the severity of the learning disability is such that it will have a profound effect on their future lives, whereas with others it may not seriously affect future prospects.

3.3 Lifespan Differences

There are also changes over the lifespan from conception, through childhood, adult life and old age. At conception, many chromosome disorders which would otherwise have resulted in a child with severe developmental delay and profound mental retardation result in a spontaneous miscarriage. At birth, it has been estimated that approximately 1% of children have a chromosome anomaly which, depending on what form it takes, may result in the presence of mental retarda-

tion (Jacobs 1990). However, developments in prenatal screening programmes for Down syndrome and inherited disorders allowing parents, depending on national laws, to decide whether or not to have a termination of pregnancy may affect these rates. Children with profound mental retardation, some with progressive metabolic disorders, have a reduced life expectancy, and in others there may be high rates of congenital disorders resulting in an increased mortality rate. At the milder end of the spectrum, the presence of intellectual impairment and learning difficulties may not be suspected or confirmed until the child starts school and is found to be behind other children in their learning ability. Some of course will have specific learning difficulties, such as dyslexia, rather than more general learning difficulties. Rates may remain relatively stable during statutory school years, but as children with past special needs leave full-time education they may no longer be considered in need of special education or other specialist services and may therefore no longer appear in prevalence figures. Examples of how age-specific prevalence rates of severe mental retardation vary are given by Fryers (1991). In childhood, they are estimated at between 3 to 5 per thousand of the population, dropping to 2.5 or less after 45 years of age.

4 Aetiology

The factors which contribute to the ultimate level of an individual's disability and their relationship to each other are likely to be complex. The starting point is the World Health Organization's system of classification. In the case of some individuals, the impact of, for example, an inherited disorder on brain development may have been so profound that it has a very powerful effect on the ultimate level of that person's disability. In other cases, the level of impairment may have a less profound effect and educational and social factors and life opportunities may be the significant influence.

4.1 Sub-cultural and Organic Division

The same arguments that apply to studies of the epidemiology of mental retardation can also apply to studies of aetiology. For example, the proportion of people considered to have a genetic, chromosomal or environmental cause of mental retardation will critically depend on how the base rate of people with mental retardation is defined. If it is those people known to services, then there will be a bias towards those with biologically determined and more severe mental retardation who are identified at birth or shortly afterwards, whereas a true population sample ascertained on the basis of IQ or special educational needs will include a much more diverse population.

Historically, there has been a division between those with an obvious biological reason for their mental retardation and those without such an obvious cause, generally resulting in milder disabilities. These groups have been referred to as having organic and sub-cultural mental retardation, respectively. In the case of the former, the mental retardation is due to a single major effect. It is this group which predominately accounts for the deviation of the normal distribution of IQ to the left, whereas the latter (sub-cultural group) has been considered to be the extreme left side of the normal curve and part of a continuous variable, the aetiological factors being familial (both polygenic and environmental). Table 2 summarises the main differences between these two groups (see also Scott 1994).

However, this distinction between mild/familial and severe/biological is less robust than previously supposed for the following reasons. Firstly, new biologically determined causes of mild mental retardation (e.g. female carriers of the fragile X syndrome) have been described. Secondly, it is recognised that people with specific syndromes (e.g. Down syndrome) usually have a distribution of abilities and some will have only mild intellectual impairments. Thirdly, even in clearly genetically or chromosomally determined disorders,

Table 2. Differences between biologically determined and sub-cultural mental retardation

Biological	Sub-cultural
Moderate/severe impairment	Mild or borderline impairment
Significant impairments in adaptive functioning	Minor or no impairment in adaptive functioning
Frequent evidence of CNS abnormalities	Evidence of CNS abnormalities uncommon
Family members usually of normal intelligence	Intellectual abilities mildly impaired in family members
Dysmorphic characteristics common	Dysmorphic characteristics unusual
Other impairments and disabilities common	Other impairments and disabilities unusual
Equal distribution across families of different socio-economic status	More common in families of lower socio-economic status
Social adversity unusual	Social adversity more common

there are other factors which have an influence on the ultimate cognitive profile of that person and there is not necessarily a direct relationship between, for example, the genetic mutation on the one hand and the level of disability on the other. Furthermore, such relatively crude methods fail to take into account the distinction between the levels of impairment, disabilities and handicaps which, as has already been argued, may be influenced by different factors at any one time as well as over time.

4.2 Genetic and Chromosomal Causes of Impairment

Biologically determined aetiological factors which have a significant effect on brain development and thus on intellectual development are clearly important. These biologically determined causes of impaired brain development can be divided into the following:

1. Those due to the loss or addition of whole or parts of specific chromosomes (autosomes and sex chromosomes) or the rearrangement of chromosomes
2. Those due to the effects of a single gene mutation which either has a direct effect on brain development or indirectly affects brain development through the accumulation of metabolites
3. Those due to the effects of pre-, peri- or postnatal trauma

The disorders of chromosome numbers (mainly Down syndrome) and other chromosomal anomalies account for approximately 30% of all causes of significant mental retardation. Over 500 single gene disorders associated with mental retardation have been described, but as a group they only account for a few per cent of cases (Wahlstrom 1990). Although a large number of syndromes associated with mental retardation have been described, there remains a significant number in which the genetic or chromosomal basis for the syndrome has not been identified or in which neither a cause or syndromal cluster of signs has been established. However, increasing numbers of single gene disorders, different mutations in the same gene, different genetic mechanisms and previously unrecognised sub-microscopic chromosomal deletions are now being described which are associated with mental retardation (e.g. see Flint et al. 1995).

It is beyond the scope of this chapter to cover in detail all the different causes of mental retardation. A few particularly important syndromes are discussed below and some listed in Table 1, either because they are numerically important or because they illustrate psychiatric or genetic points. For further details about chromosomal disorders, see Bolton and Holland (1994). The role of genetic factors in the causation of mental retardation has been reviewed by Simonoff et al. (1996).

4.2.1 Down Syndrome

The single most important disorder almost invariably associated with mental retardation is Down syndrome. In approximately 95% of patients, this is due to the inheritance of three rather than two copies of chromosome 21 (trisomy 21) in all somatic cells as a result of chromosomal non-dysjunction during meiosis. In the other few per cent, it is either due to the presence of an unbalanced chromosomal translocation or to the presence of mosaicism. In the former, there is a rearrangement of two pairs of chromosomes including 21, and in the latter only a proportion of the cells have trisomy 21 due to the occurrence of non-dysjunction of chromosome 21 in a somatic cell rather than a gamete. Maternal age is well established as the most significant factor contributing to the likelihood of non-dysjunction occurring in the female gamete: The probability of having a child with Down syndrome increases from approximately 1:2000 in early adult life to 1:45 when maternal age at conception is in the 40's (for a review on Down syndrome, see Berg 1993). From a psychiatric perspective, the relationship between Down syndrome and the high risk for Alzheimer disease has become increasingly important because of the considerably improved life expectancy experienced by people with Down syndrome. It is clear that increased cerebral amyloid disposition occurs from early in life, and by their 30s the majority of people with Down syndrome will have Alzheimer-like plaques and tangles in their brains. Rates of Alzheimer disease are, however, less than the neuropathological findings would predict. Although studies vary, Alzheimer disease affects a few per cent of people with Down syndrome in their 30s, increasing to approximately 40% in their 50s. Although the reasons for this early onset of Alzheimer disease in people with Down syndrome is not known for certain, the gene for the amyloid precursor protein (APP) has been shown to be located on chromosome 21 and a reasonable hypothesis is that excessive amyloid production is an important factor. The differential diagnosis of apparent cognitive and functional decline in later life includes not only Alzheimer disease but also hypothyroidism, the effects of sensory impairments and depression (for a review, see Holland and Oliver 1995; see also Sect. 6.3).

4.2.2 Prader-Willi and Angelman Syndromes

As the techniques available for the study of chromosomes have improved and with the development of

new chromosome banding techniques followed by the ability to mark specific DNA sequences using fluorescent in situ hybridisation (FISH), an increasing number of chromosomal deletions have been identified. In the case of two syndromes (Prader-Willi and Angelman syndromes), this has led to the discovery of the importance of gender-specific genomic imprinting in the causation of different phenotypes. In the case of both syndromes, the majority of affected individuals have a deletion on chromosome 15 (15q11-13). In the case of Prader-Willi syndrome, this is on the chromosome 15 of paternal origin and, in Angelman syndrome, of maternal origin. Particular genes on the other chromosome 15, at the same site as the deletion, are differentially imprinted according to the gender of the parent of origin. Thus both copies of different genes in the two syndromes are lost through deletion and because they are normally switched off by being imprinted. The effects of gender-specific genomic imprinting also explains the cause of those cases of Prader-Willi and Angelman syndrome not associated with a chromosome 15 deletion. In the majority of these cases, both copies of chromosome 15 are from one parent and hence both copies of specific genes (depending on the gender of the parent of origin) are imprinted and therefore not expressed. In the case of Prader-Willi syndrome, the problems of over-eating and severe obesity together with an apparently increased propensity to severe temper tantrums has been of psychiatric interest (see Sect. 6.3).

4.2.3 Fragile X Syndrome

More recently, the fragile X syndrome has been recognised as an important cause of mental retardation since the characteristic fragile site on the X chromosome (Xq27.3) was described by Lubs in 1969 after culturing the chromosomes of an affected individual in a medium low in folic acid. The genetic basis of this X-linked disorder is the presence of a triplet repeat sequence (CGG) in the FMR-1 gene. The normal repeat sequence is less than 50, but pre-symptomatic carriers have between 50 and 200 repeat sequences and mildly affected female carriers and affected males have 200 or more. Males are more severely affected and have moderate to severe mental retardation, a characteristic litanic and cluttered form of speech, gaze avoidance and other social impairments, minor dysmorphic facial characteristics (e.g. large ears, long narrow face) and, post-pubertally, large testes. Increased rates of autism have been reported, but this is disputed and these findings may be due to the gaze avoidance and other social impairments present. Approximately 1 in 2500 males in Caucasian populations are affected (for a review, see De Vries et al. 1994).

4.3 Environmental Causes of Impairment

The occurrence of environmentally determined causes of mental retardation may well depend on economic and socio-cultural factors as well as health policies. In the developing world, serious perinatal problems, severe malnutrition and chronic infections are likely to be more significant factors than chromosomal disorders as causes of impaired early development and sub-optimal functioning in later life. In the developed world, it is estimated that approximately 20%–30% of significant mental retardation may be caused by environmentally determined factors, the majority of which are due to perinatal injury, particularly in very low birth weight babies. Improvements in neonatal care are resulting in greater survival, and particularly for those with birth weights below 1500 g there is an increasing risk of future physical and/or intellectual impairment as a consequence of periventricular haemorrhage (for details, see Reynolds 1996). Congenital viral infections such as cytomegalovirus and toxoplasmosis remain important potential causes. Rubella, as a cause of severe fetal abnormalities, has been largely eliminated in many countries through immunisation campaigns. Excessive maternal alcohol abuse leading to the fetal alcohol syndrome (see Porter et al. 1984) and, more recently, maternal human immunodeficiency virus (HIV) infection can both lead to babies with significant impairment of brain development and therefore of intellectual ability. While specific environmentally determined biological factors may give rise to significant mental retardation, others may have a less obvious effect, reducing the potential of the people concerned. Examples include poor maternal and child nutrition and the effects of lead. From a worldwide perspective, these examples are likely to be of great importance.

4.4 Polygenic Effects

The effects of these organically determined factors on intellectual development is usually such that developmental delay and the other features of mental retardation are apparent in early life and place limitations on ultimate levels of attainment. In the case of those with mild levels of intellectual impairment and with no obvious organically determined disorder, the picture is more complex. Studies of the influences on intellectual ability (IQ) have demonstrated that, within those populations studied, genetic influences are significant and are likely to be polygenic. However, the somewhat artificial dichotomy between "nature" and "nurture"

has increasingly been seen to be unhelpful. Interactive and more complex models of the relative roles and relationships between what is inherited and what is in the environment are being developed (for a review, see Simonoff et al. 1996). In this review, the role of shared and non-shared environmental influences and the numerous complexities which have to be considered at both the genetic and environmental level are discussed. Plomin and Daniels (1987) have also extensively reviewed this field, specifically with regards to personality and intelligence, and have concluded that individual (non-shared) environmental effects are more significant than shared environmental effects. For example, individuals may respond to similar environmental influences differently and age and gender may modify the response.

As genetic and biological mechanisms are identified and more complex models of understanding are developed, then some key questions about intervention can begin to be addressed. These include the following issues:

1. How to identify the mechanism whereby a specific single gene or chromosome disorder gives rise to impaired brain development and intellectual impairment and what contributes to individual variation within syndromes. In the case of phenylketonuria understanding the genetic defect and its consequences has led to a preventative strategy, but the same may be true for disorders such as Down syndrome.
2. Whether early intervention modifies the future severity of any disability and the extent to which this is the case regardless of the cause or severity of the impairment. This is a complex issue. Past research has suggested only limited effects (Zigler and Berman 1983). However, more complex models that take into account both biological and environmental factors may show differences.
3. How the extent of a given impairment, disability and resultant disadvantage can be minimised. As has already briefly been discussed, this not only includes the role of intervention in childhood but also the availability of resources and support to those with mental retardation in later life.

In psychiatric practice, it is necessary to try to integrate knowledge from this whole range of scientific study, including the causes of and nature of the mental retardation, the additional mental health or physical problems, together with dynamic and family factors that may contribute to the level and extent of any person's disability as well as to his or her general well-being and quality of life. Particularly for those with mild mental retardation, cycles of disadvantage can be seen which include the effects of poor parenting, a failure to engage in education, poor nutrition, lack of opportunities and the presence of abuse and a neglectful environment.

5 Changing Needs

The definitions of mental retardation now used and particularly the refinement of classification, such as that of the World Health Organization, provide the basis whereby the general and more specific needs of both children and adults with mental retardation can be considered. The needs of people with mental retardation and the needs of their families and other carers change over time. At birth and in early childhood, the focus may be on the identification of the cause of the developmental delay and the treatment of other congenital problems which may be present. In developing countries, the prevention of malnutrition and infections would be the priority. Later, the need is to provide help to the child and their families and to maximise the development of skills through specialist educational support. In adult life, the availability of appropriate support networks and supported employment opportunities may be the priority. At each of these stages, it is necessary to be able to characterise the extent and nature of the person's abilities and disability, and in this context it is important to be clear about definitions and to use any labels which might be applied in a positive manner.

Traditionally, distinctions have been drawn between what are predominately social care needs and what are health needs. Although this can be useful, this distinction can be somewhat artificial. Both physical and mental health can be influenced by the quality of the social care environment and, similarly, the state of a person's mental health and the extent to which he or she presents with problematic behaviour can influence the quality of the social care environment. High rates of problem behaviours reported in institutional settings are not simply consequent upon the quality of the environment, but also relate to the fact that it was the presence of problem behaviour in the first place which resulted in admission to an institutional setting.

5.1 Special Needs

The general and mental health needs of people with mental retardation require special consideration for a number of reasons. Firstly, as described below, there are much higher rates of associated physical and sensory impairments, psychiatric and behavioural dif-

ficulties and other illnesses. Secondly, particularly for those with more severe disabilities, there may be the presence of many health problems and therefore the need for often complex and multiple interventions. Thirdly, given the limited intellectual and language development, which is a central feature of mental retardation, the recognition and diagnosis of specific illnesses can be problematic. A person may not be able to describe or easily localise pain or report other symptoms; thus physical illnesses go unnoticed until they become severe. Finally, while children have the support of their parents in seeking medical help, if unwell, adults with mental retardation may not be able to take the initiative in seeking help and may find accessing primary and secondary health services very difficult. Similarly, routine health screening procedures such as breast examination and cervical smears in women may not happen. In these circumstances, carers have the responsibility to be aware of possible health problems and to help the person access health services as and when required. Community surveys investigating the physical health of people with mental retardation have reported significant rates of undetected illness.

5.2 General Health Needs

The starting point in considering the general health needs of children and adults with mental retardation is that they have the same needs as those without such disabilities but in addition may have additional needs for the reasons given above. Many of the biologically determined causes of mental retardation are also associated with the presence of additional impairments and resultant disabilities. These include high rates of physical and sensory impairments and disabilities as well as additional illnesses, such as epilepsy. In some cases, very specific risks are associated with particular causes of mental retardation. For example, people with Down syndrome are at risk for congenital heart disease or, in later life, sensory impairments, thyroid disorders and Alzheimer disease with age-specific prevalence rates for these disorders increasing from a few per cent in those in middle age to up to 50% in those aged 50 years and over (see Berg et al. 1993).

For the same reasons as given earlier, accurate estimates of the prevalence of these additional impairments is problematic. Not only are the studies primarily on administrative samples, but they also include different age-groups and different service settings. Prevalence rates of physical impairments and disabilities, sensory impairments and epilepsy all increase with the degree of mental retardation. In the case of epilepsy, prevalence rates rise to 50% in those with profound disabilities (for a review, see Bird 1997) and may also include different types of seizures which are difficult to control, resulting in the use of more than one anti-convulsant.

5.3 Mental Health Needs

There is convincing evidence that people with mental retardation have higher rates of mental health problems than the general population. However, such epidemiological studies face similar difficulties as have been described earlier. These include the problems associated with the definition of "mental retardation" and the difficulty of diagnosis of specific mental disorders in those with impaired language development. Inevitably, diagnostic criteria have to be modified. However, if they are too loose, they are no longer valid, and if too strict, they underestimate the true prevalence of mental disorder in this population.

There is no readily accepted definition of "mental health" other than the rather unsatisfactory one of "the absence of mental disorder". However, what is meant by the term "mental disorder" has over many years become more refined through the development of diagnostic criteria such as defined in the International Classification of Mental and Behavioural Disorders (ICD-10; World Health Organization 1992) and the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV; American Psychiatric Association 1994). The range of mental disorders as defined in the above classification systems are found affecting people with mental retardation as well. As with the general population, appropriate diagnosis is essential, as it is the diagnosis and associated formulation that primarily determines treatment. As is well recognised, the range of mental disorders includes those due primarily to identified organic factors (e.g. delirium and dementia), the functional psychoses (e.g. schizophrenia), affective disorders (e.g. depression, manic depressive psychosis), neurotic disorders (e.g. obsessive compulsive disorders, phobias) and personality, behavioural and developmental disorders. In some cases, the diagnostic process is less helpful in informing treatment and is in danger of having a negative influence on outcome (e.g. in the case of personality or behavioural disorders). The diagnosis of a "behavioural disorder" assumes the exclusion of an underlying organic disorder or mental illness which might be directly causing the behaviour, and clearly in these circumstances a more sophisticated psychological analysis is required.

The Isle of Wight study of children (Rutter et al. 1970), the Aberdeen studies (Birch et al. 1970) and a number of other studies have reported population prevalence rates of psychiatric disorder of between 25% and 40%. This includes the full range of disorders.

The differences across studies reflect differences in methodology, specifically the diagnostic criteria used and the nature of the population studied. Prevalence rates of psychiatric disorder in general increase with the severity of the intellectual impairment from 20% in those with mild impairment to 45% in those with severe intellectual impairments. Similarly, rates of "pure" behaviour problems increase with the severity of impairment from a few per cent to between 20% and 30%, respectively (Kushlick and Blunden 1974; Richardson and Koller 1985).

The development of better diagnostic instruments such as the PAS-ADD schedules (Moss et al. 1996) now enable more accurate studies to be undertaken. Early studies using the instrument in a population of people with mental retardation over the age of 50 found rates of psychiatric disorder of 12% and of dementia of 12% (Moss et al. 1993; Patel et al. 1993). Cooper (1997) developed her own semi-structured informant interview and in a study of older people with mental retardation in a defined area compared prevalence rates of psychiatric disorder with that in a control group of people with mental retardation identified from the local register. Rates of psychiatric disorder were generally higher in the older group (68.7% vs. 47.9%). In both cases, behaviour disorders made up the single most prevalent group. Rates of past affective disorder were significant in the older and control groups (22.4% and 11%, respectively). The difference in these "lifetime prevalence" rates may simply reflect the difference in age in the two groups. The prevalence rate of dementia in the older group reached 21.6%. This study paid close attention to issues of diagnosis using ICD-10 criteria with the addition of a category of behaviour disorder. However, the population studied was an administrative sample and therefore may have had an over-representation of people with mental retardation and additional problems, as they are more likely to be known to services.

The recognition that a person with mental retardation may also be suffering from a severe psychiatric illness is particularly important because of the treatment implications. Prevalence rates of schizophrenia, in particular, have been reported to be three times greater in people with mild intellectual impairments compared to the general population (lifetime prevalence of 0.03 vs. 0.01). Initially, early-onset schizophrenia was considered to be a cause of significant mental retardation and the term "pfropfschizophrenia" was used. The presentation of schizophrenia in people with mild intellectual impairments is not significantly different from those without such an impairment.

Consideration of the mental health needs of people with mental retardation therefore should include creating the right social care environment to minimise the risk of abuse and to maximise opportunities, ensuring that mental health problems are detected and that appropriate assessment is available, leading to the necessary treatment and management strategies. These are outlined in more detail below.

6
Assessment and Treatment of Psychiatric and Behaviour Disorders

In the field of mental retardation, there are a profusion of different terms (e.g. psychiatric disorder, challenging behaviour, behaviour disorder) and a variety of perspectives from different disciplines. Most of these terms, such as "challenging behaviour", do not imply any understanding as to aetiology but are simply shorthand descriptive terms. The use of the word "behaviour" in such terms does not mean the approach is necessarily a behavioural one, just as calling something a "psychiatric disorder" does not mean that a psychiatric approach is necessarily the correct one or the only one needed.

People with more severe mental retardation have abnormalities in brain development resulting in impairments in skills and language development and increasing their vulnerability to psychiatric and emotional difficulties. For people with mild mental retardation, impaired brain development may be less severe but factors such as social disadvantage may help to account for the presence of psychiatric and emotional disorders. Furthermore, marked differences, particularly in the level of language development, give rise to considerable variations in the way in which individuals with mental retardation are able to describe their thoughts and feelings and therefore the way in which emotional distress and/or major psychiatric illness presents. Given the extreme diversity of people with mental retardation, it follows that the factors that should be considered with regards to mental health will be very varied across the range of disability.

6.1
Challenging Behaviour and Psychiatric Disorder

The high rates of these problems, together with the impact they can have on the lives of children and adults with mental retardation and their families, means that they account for many of the referrals to specialist community teams. The task of services is to undertake appropriate assessments in order to try and understand the reasons for the index problem and how it can most effectively be treated and managed.

Given the heterogeneous nature of this group of people defined as having mental retardation, it is not possible to make general statements about the presentation of psychiatric disorder and its treatment which are readily applicable to all those with mental retardation. The extent of impairment and, in particular, the level of language development will be crucial in determining how disorders such as depression might present. A psychiatric perspective is but one approach to understanding problem behaviour and is primarily a diagnostic one. It is also clear that psychological theories about learning and how behaviour can be shaped and reinforced over time are particularly important when it comes to understanding the reasons for problematic behaviour and when developing appropriate management strategies. The presence of problem behaviour therefore necessitates that multidisciplinary assessments take place, the facts contributing to the behaviour be identified and appropriate interventions initiated.

6.2 Psychiatric Assessment

The role of the psychiatrist when asked to see a person with mental retardation is essentially twofold. Firstly, through a detailed history taking (often from an informant, particularly if the person has limited language), mental state and physical examination together with necessary investigations to establish whether an additional mental disorder (over and above mental retardation) or physical illness is present, which might account for the index problem. The key issue is whether there has been a change in the person's behaviour, general well-being or intellectual or functional abilities. If so, this requires an explanation. The differential diagnosis will depend upon the clinical features but may include the development of a physical illness, the onset of a psychiatric illness such as depression, the impact of some life change or the development of some organic psychiatric disorder such as dementia. The diagnosis will depend upon the nature of the change, associated features (e.g. features of depression, the presence or absence of abnormal mental experiences, cognitive changes) in the context of a comprehensive picture over time of the person's life and any significant events. Additional information maybe required from a carer, specifically about whether a person's mental state appears to have changed, the person's apparent mood, their sleep and appetite pattern and whether he or she has lost interest in things or may be hallucinating. Secondly, a psychiatric assessment will also include a detailed developmental and family history, the purpose of which is to establish the cause, nature and extent of mental retardation and to address the questions as to whether the nature and cause of the mental retardation might contribute to or account for the index problem. For example, observed behaviour difficulties in a person with a developmental history diagnostic of autism may significantly help in the understanding of his or her behaviour. Changes in routine and excessive demands may be particularly difficult for people who are autistic and result in behaviour problems. Treatment in this case may involve helping carers understand the nature of autism and encouraging them to develop appropriate management strategies and, as far as is possible, a routine and predictable environment.

6.3 Behavioural Phenotypes

Knowing the cause of the mental retardation may also be important, as particular genetic or chromosomally determined disorders may be associated with specific psychiatric disorders or problem behaviours. For example, the behaviour of older people with Down syndrome may change because they develop hypothyroidism or Alzheimer's disease. Other examples of close links between specific syndromes and particular psychiatric or behaviour problems include the link between Prader-Willi syndrome and excessive eating leading to severe obesity, Lesch-Nyhan syndrome and self-injurious behaviour and Williams syndrome and a specific pattern of language development. These associations have been referred to as the "behavioural phenotype" of the specific genetic or chromosomally determined disorder (see O'Brien and Yule 1995). These associations are important, partly because they alert people to the potential for specific difficulties but also because of possible treatment implications.

6.4 Treatment

Changes in behaviour, general well-being and/or social and living skills in people with mental retardation should always lead to a more detailed assessment to address the question of the cause or causes of this apparent change. This invariably requires different approaches from different disciplines. All too easily it is assumed that such behaviour is simply due to the person's mental retardation and that nothing can be done about it. Treatment is based on both diagnosis and a holistic picture of the person and their circumstances. The range of treatments commonly used in psychiatry may well be applicable, taking into account how the person's impairments and disability may make a difference. In addition, particular attention

should also be paid to the environment within which the person lives. For example, it has become increasingly clear that significant amounts of physical and sexual abuse go unreported and undetected. The quality of the environment and the nature of the support available certainly influences the likelihood of problem behaviour or depression.

Given the fact that epilepsy is particularly common in people with mental retardation, the possibility should be considered that pre-, peri- or post-ictal states as well as the fit itself may contribute to problem behaviour. More effective treatment of the fits should then resolve the behaviour difficulties. The excessive use of and/or the use of inappropriate anti-convulsants (e.g. phenobarbitone) can contribute to further impairment of cognitive function or an altered mood state. A detailed clinical assessment and investigation of temporal relationships between seizures and behaviour, together with appropriate investigations (e.g. EEG, brain scan) should enable any link to be established (for a review, see Bird 1997).

Psychiatric illness is likely to require treatment with the appropriate neuroleptic, anti-depressant or mood-stabilising medication. However, there is a long and dubious history of using tranquillising medication to "control behaviour". The use of medication should be firmly guided by diagnosis and be used in the knowledge that it has been shown to be useful in treating that particular disorder (e.g. increasing aggression in the context of a depressive illness can be effectively treated with anti-depressant medication, as can a psychotic illness with neuroleptic medication). Medication acting on the central nervous system should be used with caution, as the response to such medication in those with a pre-existing disorder of brain development may be atypical. The risk of side effects from neuroleptic medication may be increased. As a general rule, doses of such medication should be low and increased with care together with regular monitoring of outcome.

In addition to cognitive and behavioural psychological interventions, interpretative psychotherapy and family therapy has been used, e.g. following bereavement or when family dynamics are thought to be playing a crucial part in the maintenance of the index problem. Given the additional problems that such therapies might encounter, because of the presence of impaired language development, music and art have been used as ways of helping people with limited or very limited verbal skills explore their emotions. It is clear that people with mental retardation can suffer major psychological traumas as a result of abuse or major life events. Music and art therapies may well provide ways of helping the individual resolve these issues or recovering from post-traumatic stress. Furthermore, the ability to make ones needs known and to understand and be able to communicate with others is a skill we take for granted, but when it is impaired, improving communication or developing new methods of communication (e.g. Makaton signing) may be a key factors in improving behaviour, enhancing skills and improving the person's quality of life. Speech and language therapy assessment and intervention are very important for this reason.

6.5
Functional Approaches

In addition to a diagnostic formulation, it is also important not only to consider what might be the causes of the particular change in behaviour or mental state, but to ask whether factors in the environment or in the type of care might be maintaining the problem. This can be particularly relevant where a behaviour appears to have a "function".

Psychological research and clinical practice has increasingly focused on the role that particular behaviours might have in enabling a person to have some control over their environment. Behaviour (e.g. aggression or self-injury) is seen to have a function (e.g. removing a particular demand or receiving social interaction). The task of psychological assessment is to identify the patterns of reinforcement and the environmental settings within which such behaviour is most likely to occur. This understanding then leads to management and/or environmental changes of a positive kind which in turn result in changes in the pattern of behaviour. The use of functional analysis and appropriate models of behavioural modification are now common place and can be highly effective (for a review, see Emerson 1995).

7
Social Care Needs and Normalisation

The shift away from institutional care towards a more focused, needs-led approach together with the changes brought about by the concepts of normalisation, first developed in Scandinavia (Bank-Mikkelsen 1969; Nirje 1973) and later in the United States with the concept of "social role valourization" (Wolfsenberger 1983), have been striking. These changes have led to a strong emphasis being placed on the rights of people with mental retardation to be integrated into mainstream society, to have the opportunities to establish and maintain friendships, the right to choose in matters of everyday life, the chance to develop competence through meaningful activity and the right to be

respected (O'Brien 1987). In the United Kingdom, the Government consultative paper *Better Services for the Mentally Handicapped* (Department of Health and Social Security 1971) was the starting point for major strategic change in the provision of care for people with mental retardation. These changes came about following major scandals concerning the quality of care and observed abuse in specific long-stay hospitals. Not only was there no theoretical basis for such institutions, other than those put forward by the eugenics movement, but the environments created did not foster development and easily became places of abuse. The process of closing large institutions and developing smaller, community-based homes continues to this day. Although there remains some debate about the role of what have come to be called "care villages", the basic principles of community care are accepted and not challenged.

As with health care needs, children and adults with mental retardation have similar basic social care needs as the rest of the population. These include a caring and supportive environment during childhood with appropriate emotional and physical support and opportunities to learn and develop skills necessary for independence in later life. In childhood, the focus has been on the development of special education either within mainstream schools or separately in special needs schools. The need for education is likely to continue into adult life and to include college courses providing further social and living skills training and literacy and numeracy teaching.

7.1 Needs Assessment

In the United Kingdom, there is now a statutory responsibility upon education authorities to provide special education and upon local social services departments to undertake needs-led assessments for all those with special needs. Such assessments should take place at key transitions in life (e.g. leaving full-time education and moving to adult services). The focus of this approach is to recognise the diversity of need both in individuals over time and across the group of people with mental retardation as a whole. While educational support during childhood has been excellent in many countries, the provision of day services for adults with mental retardation has been varied. A significant change has been the increasing development of supported employment schemes, whereby people with mental retardation can be helped to work in a variety of settings through the use of "job coaches". As with residential services, a variety of support is likely to be needed, the purpose of which is to provide meaningful occupation, the opportunity to develop further living and social skills and to have an acceptable quality of life.

7.2 Family and Paid Carers

With the closure of institutions and with the focus very much on "care in the community", the needs of both family carers (usually mothers) and of paid carers has received increasing attention. The stress involved in caring for a person with a significant mental retardation is well recognised. Levels of carer stress predominately relate to the presence or absence of additional behaviour problems rather than to the extent of disability. A good informal and formal support network, the availability of respite care and sharing the caring role seem to be key factors in minimising stress and the risk of neglect and/or abuse which can result. Recently introduced legislation in the United Kingdom now require social service agencies to specifically consider the needs of family carers.

7.3 Support Groups

The extent of support available to people with mental retardation and their carers varies considerably within and between countries. In addition to statutory health and social service agencies, there are an increasing number of specific syndrome support groups. These are a major source of information and advice. In many countries, organisations exists for even the rarest of syndromes or there are umbrella organisations which cover a range of syndromes.

7.4 Ethical and Legal Considerations

People with mental retardation may have special needs; they may be vulnerable to abuse and exploitation and may have considerable difficulties in making their needs known and being able to take significant decisions which affect their lives. For this reason, some of the potential legal and ethical issues which may arise regarding both social and health related issues should be considered.

In the case of children, parents or guardians have the responsibility for the care of those in their family whether they have a disability or not. The legal framework of most countries clearly places responsibility for the care of children and their protection from exploitation or abuse with the family. Parents have the legal right to take decisions on behalf of the child, where

possible taking his or her views into consideration. In adult life, the situation can be more complex, both ethically and in law. Tension exists between the right of an adult to self-determination and the need for the care and protection of people who may not be able to fully care for them. For example, in the United Kingdom, the absolute right of an adult to make the decision to accept or reject medical advice has been established through case law, providing that he or she is not incompetent to make that decision due to the presence of mental disability (e.g. due to mental illness, dementia, head injury). As far as medical treatment is concerned, for those adults who are considered incapable of making a decision, a doctor has a duty, under Common Law, to act in the best interest of the person concerned.

More generally, decisions are made everyday on behalf of people with significant mental retardation. Families and paid carers may involve him or her in that decision to the extent possible. However, under particular circumstances there is clearly a duty of care and there has to be the right balance between autonomy and risk. Examples of this include the decision as to whether a person with mental retardation can go out by themselves. Making such a decision would include an assessment of the skills of the person concerned and whether further help would allow the person to develop greater independence skills, an assessment of how any risk might be reduced and the views of the person him- or herself.

The over-paternalistic approaches which were commonplace are in danger of being replaced by an excessive wish to allow people with disability to make "choices" even when it is evident that they may not have the ability to make them. In some countries (e.g. Australia), guardianship laws have been enacted to provide a mechanism whereby a person has the right to make decisions on behalf of someone who lacks the capacity to make that decision for themselves. While such a legal framework has its benefits, it is also open to abuse and can become too restrictive. There is no easy solution to this issue. This tension has to be recognised and made explicit, and there is a need to balance the right to self-determination against the need to protect potentially vulnerable people.

8 Conclusions

This chapter has considered certain key themes. In summary, these are as follows:

- The history of the care of and attitudes towards people with mental retardation is not an honourable one. There have, however, been significant changes, and it is essential that people with mental retardation are considered with respect; when this label is applied, it must be for the benefit of those concerned.
- The term "mental retardation" refers to a highly heterogeneous group of people who have in common a history of early developmental delay, together with a significant intellectual impairment and impairment in social functioning, but differ in the extent and nature of their disabilities.
- The needs of both children and adults with mental retardation are often multiple and complex and change across the life span. Different disciplines and services will be involved, and the identification of needs can be aided by considering separately the reasons for and levels of impairments, disabilities and handicaps.
- People with mental retardation have high rates of behaviour and psychiatric disorder compared to the general population. Impaired language development can make the diagnosis of an additional mental disorder more difficult, but if changes in behaviour, mental state or well-being are observed or there is evidence of a loss of abilities, an assessment is essential. Multi-disciplinary assessments are often indicated. The role of the psychiatrist is to provide a psychiatric perspective to an understanding of the index problem. This includes the identification of additional medical or psychiatric disorders. Identifying the exact cause of the mental retardation can also be important.
- Detailed assessments provide the guide for intervention and specific treatment. Treatment and management strategies may include the full range of psychiatric, psychological and other psychotherapeutic interventions. Physical illness can become manifest as a change in behaviour and, particularly in those with limited language, should always be considered.
- People with mental retardation may have particular difficulties in making their needs known and because of their impairments and disabilities may be vulnerable to abuse and exploitation. Particularly in the case of adults, there can be a difficult balance between the rights of individuals to determine their own lives and the need for care and protection from risk.

Appendix A: Summary of the Diagnostic Criteria for Mental Retardation (DSM-IV)

A. Significant sub-average general intelligence
B. Significant limitations in adaptive functioning in at least two of the following:

- Communication
- Self-care
- Home living
- Social/interpersonal skills
- Use of community resources
- Self-direction
- Functional academic skills
- Work
- Leisure
- Health
- Safety

C. Onset before 18 years of age

Note

Significant sub-average intellectual functioning is defined as an IQ of about 70 or below. The choice of testing instrument should take into account the individual's socio-economic background, native language and other associated handicaps.

Adaptive functioning refers to how effectively individuals cope with common life demands and how well they meet the standards of personal independence expected of someone in their particular age-group, socio-cultural background and community setting. Adaptive behaviour may be influenced by individual and/or environmental factors, including the presence or absence of additional mental or physical disorders. Information on adaptive behaviour should be gathered from one or more independent sources.

The degree of severity of mental retardation may be specified on the basis of intellectual impairment taking into account other aspects of functioning.

Mental retardation may be classified as follows:

- Mild mental retardation:
  IQ level of 50–55 to approximately 70
- Moderate mental retardation:
  IQ level of 35–40 to 50–55
- Severe mental retardation:
  IQ level of 20–25 to 35–40
- Profound mental retardation:
  IQ level below 20 or 25

Appendix B: Impairments, Disabilities and Handicaps

Impairment

- Any loss or abnormality of psychological, physiological or anatomical structure or function
- Represents deviation from some norm in the individual's biomedical status
- Characterised by losses or abnormalities that may be temporary or permanent
- Includes the existence or occurrence of an anomaly, defect or loss in a limb, organ, tissue or other structure of the body or a defect in a functional system or mechanism of the body, including the systems of mental functioning
- Not contingent upon aetiology

Disability

- Any restriction or lack (resulting from impairment) of ability to perform an activity in the manner or within the range considered normal for a human being
- Concerned with compound or integrated activities expected of the person or of the body as a whole, such as represented by tasks, skills and behaviours
- Excesses or deficiencies of customarily expected activities and behaviour, which may be temporary or permanent, reversible or irreversible and progressive or regressive
- Process through which a functional limitation expresses itself as a reality in everyday life

Handicap

- A disadvantage for a given individual, resulting from an impairment or disability that limits or prevents the fulfilment of a role that is normal for that individual
- Places some value upon this departure from a structural, functional or performance norm by the individual or his or her peers in the context of their culture
- Relative to other people and represents discordance between the individual's performance or status and the expectations of his or her social/cultural group
- A social phenomenon, representing the social and environmental consequences for the individual stemming from their impairment and disability

9 References

American Psychiatric Association (1994) Diagnostic and statistical manual of mental disorders, 4th edn (DSM-IV). American Psychiatric Association, Washington DC

Bank-Mikkelsen NE (1969) A metropolitan area in Denmark: Copenhagen. In: Kreugel R, Wolfsenberger W (eds) Changing patterns in residential services for the mentally retarded. President's Committee on Mental Retardation, Washington DC

Berg JM (1993) Down Syndrome. In: Berg JM, Karlinsky H, Holland AJ (eds) Alzheimer disease, Down syndrome and

their relationship. Oxford Medical Publications, Oxford, pp 19–36

Berg JM, Karlinsky H, Holland AJ (eds) (1993) Alzheimer disease, Down syndrome and their relationship. Oxford Medical Publications, Oxford

Bird J (1997) Epilepsy and learning disabilities. In: Russell O (ed) The psychiatry of learning disabilities. Royal College of Psychiatrists seminars series. Gaskell, London

Birch HG, Richardson SA, Baird D, Horobin G, Illsley R (1970) Mental subnormality in the community: a clinical and epidemiological study. Williams and Wilkins, Baltimore

Bolton P, Holland AJ (1994) Chromosome abnormalities. In: Rutter M, Taylor EA, Hersov L (eds) Child and adolescent psychiatry: modern approaches. Blackwell, Oxford

Clarke AM, Clarke ADB (1984) Constancy and change in the growth of human characteristics. The 1st Jack Tizard Memorial Lecture. J Child Psychol Psychiatry 25: 191–210

Cooper SA (1997) Epidemiology of psychiatric disorders in elderly compared with younger adults with learning disabilities. Br J Psychiatry 170: 375–380

Department of Health and Social Security (1971) Better services for the mentally handicapped. DHSS, London

De Vries LBA, Halley DJJ, Oostra BA, Niermeijer MF (1994) The fragile-X syndrome: a growing gene causing familial intellectual disability. J Intellect Disab Res 38: 1–8

Emerson E (1995) Challenging behaviour: analysis and intervention in people with learning disabilities. Cambridge University Press, Cambridge

Flint J, Wilkie AOM, Buckle VJ, Winter RM, Holland AJ, McDermid HE (1995) Subtelomeric chromosomal deletions explain a significant fraction of idiopathic mental retardation. Nature Genet 9: 132–139

Fryers T (1991) Public health approaches to mental retardation: handicap due to intellectual impairment. In: Holland WW, Dettels R, Knox G (eds) Oxford text book of public health, 2nd edn. Oxford University Press, Oxford, pp 485–508

Gottlieb G (1996) A systems view of psychobiological development. In: Magnusson D (ed) The lifespan development of individuals. Cambridge University Press, Cambridge

Hodapp RM, Burack JA, Zigler E (eds) (1990) Issues in the developmental approach to mental retardation. Cambridge University Press, Cambridge

Holland AJ, Oliver C (1995) Down's syndrome and the links with Alzheimer's disease. J Neurol Neurosurg Psychiatry 59: 111–114

Jacobs PA (1990) Review: the role of chromosome abnormalities in reproductive failure. Reprod Nutr Dev [Suppl 1]: 63s–74s

Karmiloff-Smith A (1995) Annotation: the extraordinary cognitive journey from foetus through infancy. J Child Psychol Psychiatry 36(8): 1293–1313

Kushlick A, Blunden R (1974) The epidemiology of mental subnormality. In: Clarke AM, Clarke ADB (eds) Mental deficiency: the changing outlook, 3rd edn. Methuen, London, pp 31–81

Moss SC, Patel P, Prossser H, Goldberg D, Simpson N, Rowe S, Lucchino R (1993) Psychiatric morbidity in older people with moderate and severe learning disability. 1. Development and reliability of the patient interview (PAS-ADD). Br J Psychiatry 163: 471–480

Moss SC, Prosser H, Ibbotson B, Goldberg D (1996) Respondent and informant accounts of psychiatric symptoms in a sample of patients with learning disability. J Intellect Disab Res 40(5): 457–465

Nihira K, Leland H, Lambert N (1993) Adaptive behavior scale – residential and community, 2nd edn. Pro-Ed, Austin

Nirje B (1973) The normalisation principle: implications and comments. In: Gunzberg AC (ed) Advances in the care of the mentally handicapped. Bailliere Tindall, London

O'Brien J (1987) A guide to personal futures planning. In: Bellamy GT, Wilcox B (eds) A comprehensive guide to the activities catalogue: an alternative curriculum for youth and adults with severe disabilities. Brookes, Baltimore

O'Brien G, Yule W (eds) (1995) Behavioural phenotypes. MacKeith, London

Patel P, Goldberg D, Moss S (1993) Psychiatric morbidity in older people with moderate and severe learning disability. II. The prevalence study. Br J Psychiatry 163: 481–491

Plomin R, Daniels D (1987) Why are children in the same family so different from one another? Behav Brain Sci 10: 1–60

Porter R, O'Connor M, Whelan J (eds) (1984) Mechanisms of alcohol damage in utero. Ciba Found Symp 105: 1–296

Reynolds O (1996) Causes and outcome of perinatal brain injury. In: Magnusson D (ed) The lifespan development of individuals. Cambridge University Press

Richardson SA, Koller H (1985) Epidemiology. In: Clarke AM, Clarke ADB, Berg JM (eds) Mental deficiency: the changing outlook. Methuen, London, pp 356–400

Rutter M, Graham P, Yule W (1970) A neuropsychiatric study in childhood. Spastics International Medical Publication, London

Scott S (1994) Mental retardation. In: Rutter M, Taylor EA, Hersov L (eds) Child and adolescent psychiatry: modern approaches. Blackwell, Oxford, pp 616–646

Simonoff E, Bolton P, Rutter M (1996) Mental retardation: genetic findings, clinical implications and research agenda. J Child Psychol Psychiatry 37(3): 259–280

Sparrow SS, Balla DA, Cicchetti DV (1984) Vineland adaptive behavior scale (interview edition – survey version). American Guidance Service, Circle Pines

Wahlstrom J (1990) Gene map of mental retardation. J Ment Defic Res 34: 11–27

Wolfsenberger W (1983) 'Social role valorization': a proposed new term for the principle of normalization. Ment Retard 21: 234–239

World Health Organization (1980) International classification of impairments, disabilities and handicaps. World Health Organization, Geneva

World Health Organization (1992) ICD-10: international statistical classification of diseases and related health problems, 10th revision. World Health Organization, Geneva

Zigler E, Berman W (1983) Discerning the future of early childhood intervention. Am Psychologist 38: 894–906

Part 2
Psychiatry and Somatic Disorders

CHAPTER

1

H. Förstl, A. Jablensky

Organic Origin of Mental Disorders: An Introduction

If we had complete understanding in any one of the three areas of mental illness, namely, pathological anatomy, etiology or symptomatology, then not only from each of them could an integrated and thoroughgoing classification of psychoses be found, but each of them would also – a requirement that is the foundation of our scientific research in general – substantially coincide with the other two. Illnesses having the same causes would also always present the same symptoms and have the same autopsy results. From this basic idea it follows that the clinical grouping of mental disorders will be based *at the same time* on all three aids to classification, and they must still fall into line with experiences gained from the course, the outcome, and, indeed, the treatment.

Kraepelin (1899)

Translator: S. Henn

1 Introduction

One hundred years after Kraepelin's statement, the hope for an exhaustive knowledge of the structural, molecular biological, and clinical features of mental disorders has not yet been fulfilled. It appears doubtful whether sufficient knowledge of the individual levels of observation can ever be obtained and whether a satisfactory congruence between morphology, biochemistry, and symptomatology can actually be achieved. The scientific realism of Kraepelin, i.e., the assumption of a real and concrete substratum of terminologies and theories, independent of the observer, still does not appear to be supported firmly enough in psychiatry. The classification of mental illnesses according to ICD-10 and DSM-IV is still based mainly on clinical symptomatology (APA 1994; WHO 1991). Results of additional technical investigations are referred to first in order to prove or exclude "organic" factors and, thereby, to make a clinically suspected diagnosis more or less probable. Earlier difficulties and achievements in diagnosing organically caused mental disorders seem today only a little more understandable in view of the way in which it is taken for granted that modern investigative techniques can give information about some central nervous system (CNS) functional disorders. The results so gathered, however, are then additionally used for diagnostic arrangement into classification systems, whose hypothetical character becomes particularly clear in the revision phases and is even more obvious in psychiatry than in other medical specialities. In this chapter, several lines of development will be outlined which have led to our currently used nosology and understanding of organic causes of mental disorders.

2 From Classical Antiquity to the Nineteenth Century

2.1 Early Clinical Observations

In classical antiquity, mental disorders were regarded as direct consequences and, thereby, as signs of somatic diseases. As a result, there was no separation between organic and functional disorders and no differentiation of independent mental illnesses. In the Hippocratic corpus, there is talk for the first time of "phrenitis," a condition of confusion and sensory excitement. Celsus (25 B.C. to 50 A.D.) drew up an extensive psychiatric nosology, which, in addition to mania, melancholy, and hysteria, also cited lethargy (with fever, somnolence, and unfavorable prognosis) as well as phrenitis. Phrenitis was characterized by its acute appearance and fever. The recommended acute treatment was directed at the fever. Transition to a condition of a continuous cognitive deficit, a "dementia continua," was possible. As causes of phrenitis, Aretaeus (50–130) cited fever, alcohol, and other intoxications, e.g., with the anticholinergic scopolamine. Galen (129–199) localized the starting point of such disorders in the brain.

Until the time of the Enlightenment, variations on the classical concepts remained in use, although the meanings of the concepts were subjected to numerous changes. Kendall (1978) compared the early Hippocratic school with its empirical-biographical base to the theoretical base of Platonic universal teachings, which attempted to discover reality in universally valid ideas. Stimulated by Sydenham's suggestion to introduce a system into medicine after the model of Linnaeus' *Systema Naturae*, the theoretical-platonic base gained the upper hand until the eighteenth century. The lack of clinical experience on the part of many authors was a recognizable deficiency of their theoretically striking extensions of psychiatric classification systems.

Organically caused disorders were assigned to different forms of mania, melancholy, or madness according to their symptomatology. William Cullen (1777) differentiated the group of "neuroses" or nervous diseases as disorders of perception and movement without fever and without focal signs. Within neuroses, he differentiated "comata" (apoplexy and paralysis), "adynamiae" (syncopes, hypochondria), "spasmi" (e.g., convulsions, chorea, hysteria, hydrophobia), and "vesaniae" as disorders of intellectual achievements. In addition to congenital, acquired, and senile forms of "amentia," mania and melancholy were also included. Alexander Crichton's (1798) fine differentiation of amentia into six subforms has more characteristics of an elaborated psychopathological system than a nosological system (exhaustion, memory disorders, perception disorders, association disorders, and impaired abilities in judgment and execution).

2.2 Clinical-Anatomical Model

With few exceptions, nosology was based into the nineteenth century on cross-sectional observation of clinical disorders. In the seventeenth century, Thomas Willis (1672) had attempted to connect brain changes to mental disorders, but with little success. The dictum "mental diseases are brain diseases," which has frequently been ascribed to Griesinger (1845), comes from him. Giovanni Battista Morgagni (1761) proved

that many of the brain changes in mental illness which were described earlier represented chance findings which had nothing to do with clinical symptomatology. His studies on the pathological anatomy of internal diseases acquired great importance. Similar insights into neuropsychiatry still took time. The "organology" or "phrenology" of Franz Josef Gall and his followers proved to be both a stimulus and a strain for the serious investigation of clinical-neuropathological connections. They claimed to be able to read many talents and character features from the skull surface (cranioscopic psychodiagnosis), using these methods, however, in a more restrained fashion in assessing mental disorders (Spurzheim 1833). Only the description of the speech center by Broca (1861) and Wernicke (1874) created some confidence that certain clinical deficits could be connected with defined and recognizable cerebral changes. New knowledge in psychopathology could not keep pace with the swift and steady growth of knowledge in the pathoanatomical bases of neuropsychological deficits, i.e., aphasias, apraxias, and agnosias (Liepman 1905; Pick 1898; Wilbrand 1892). Following the description of important clinical syndromes at the beginning of the century and then a stagnant phase, the clinical-anatomical investigative approach likewise led to important successes in organic psychoses.

Thomas Sutton (1813) noticed that a subgroup of phrenitis patients, primarily alcoholic patients, showed special clinical features: onset of illness without high fever but with visual hallucinations, awkward movements (floccilegium), and occasional stereotypic executions of occupational actions (occupational delirium). In contrast to other forms of phrenitis, no improvement occurred with bloodletting, but only following administration of opium. He called this condition "delirium tremens." Pearson (1813) and Armstrong (1824) described similar conditions at almost the same time(an older description with the name "Kardiakos" comes from the Talmud; Hankoff 1972).

It took more than 50 years before Gayet (1875) and Wernicke (1881) – again almost at the same time – described characteristic anatomical consequences of alcohol abuse. Two of the three patients in whom Wernicke proved a superior hemorrhagic polioencephalitis were alcoholics and were admitted in delirium tremens. In a series of studies, Korsakow published additional clinical findings on the consequences of alcoholism. Among other things he described in them was the combination of anmestic syndrome and polyneuropathy (1887).

Progressive paralysis was a central theme of medical research in the nineteenth century, with significance reaching far beyond the study of syphilis. Bayle (1822) and Calmeil (1826) proved connections between progressive cognitive deficits, paralysis, and meningeal inflammation. This discovery is viewed today as the first description of a particular neuropsychiatric illness. Newly developed technology in the second half of the century made possible for the first time the implementation of reproducible histological and bacteriological investigations. The transmission of syphilis, its histological characteristics, and the demonstration of spirochetes in effluvia, of antibodies in cerebrospinal fluid, and, finally, of spirochetes in the brain together built up a chain of evidence which slowly led to greater precision in the concept of disease and also contributed to the description of some other, more important illnesses in modern psychiatry.

In 1894, Binswanger still expressed the view that the paralytic process of the illness "is undisputedly the consequence of a functional over-exertion of the CNS and, primarily, the cerebral cortex." In contrast to this, Alzheimer presumed "that in 70% of cases a connection with syphilis is certain or probable" (cited by Kraepelin 1899). Until discovery of the spirochetes and development of the Wassermann test, Kraepelin himself diagnosed one third of his patients as having a progressive paralysis, after which the rate sank to 5% (Hunter 1973).

The study of progressive paralysis led to a greater precision in the concept of dementia, albeit with a one-sided emphasis on "cognitive paradigms" (Berrios 1990). The term "dementia" or "amentia" was, as mentioned, in use since classical antiquity to denote a group of mental illnesses with disorders of intellectual abilities, but also of perception, affect, and intentionality. It included acute and chronic, primary and secondary, and congenital and senile forms. Cognitive deficits, above all of memory, were increasingly emphasized as the main feature of dementia. In addition, around 1900, the concept usually implied secondary appearance and irreversibility. Thus the number of subsumed illnesses was reduced and it became easier, within this less heterogeneous group, to track down anatomical and etiological bases of individual forms of dementia.

The effort to obtain a differential diagnosis of progressive paralysis contributed considerably to the characterization of other dementias with degenerative and vascular causes. That strokes could lead to grave cognitive deficits was known (Durand-Fardel 1843), and the term "apoplectic dementia" had already been created (Ball and Chambard 1881). Marce (1863) had established that a large part of senile dementia was not caused by progressive paralysis, but by vascular changes in the brain. Binswanger (1894) described clinical and macropathological features of a subcortical vascular encephalopathy.

Alzheimer (1895) differentiated in several studies a series of distinct vascular brain diseases with characteristic clinical features, among others a multi-infarct syndrome.

In earlier reviews and in his *Habilitationsschrift*[1] on progressive paralysis, Alzheimer mentioned forms of dementia with gland- or plaque-like interstitial depositions in the cerebral cortex. These had been described by Redlich (1898), among others, as miliary sclerosis of the cerebral cortex. The description of an individual patient (Alzheimer 1906) who became ill in late middle-age with a dementia having a particularly severe course and in whose cortex, in addition to plaques, a large number of previously unknown, intraneuronal fibrillary structures were found made Alzheimer's disease history, as the *casus primus*, the disease of the century, something which is probably not only owed to the scientific achievement of Alzheimer, but also to the influence of Kraepelin, who coined the eponym. Within 5 years, reports of more than ten patients with this form of dementia were published (Fuller 1912). Fischer (1907) had also demonstrated great numbers of plaques in senile dementia, and Simchowicz (1911) proposed quantitative neuropathological criteria with which to differentiate otherwise normal aging changes. Since then, the term Alzheimer's disease has also been broadened to include senile dementias with corresponding clinical and pathological features. In the same period, histological substrates of focal brain atrophy (Alzheimer 1911) and of tremor paralysis in Parkinson's disease (Lewy 1912) were described in Alzheimer's laboratory.

Alzheimer's large project to write a "pathological anatomy of mental diseases" remained unfinished. The *Anatomy of Psychoses*, edited by Spielmeyer (1930), contained extensive chapters on the pathology of neurological diseases, such as epilepsy, encephalitis, extrapyramidal motor illnesses, and consequences of intoxication and of CNS arteriosclerosis and degenerative dementias. The section on dementia praecox is short; depressive illnesses were not treated. Earlier attempts to present the whole of psychiatry as a science of diseases of the forebrain (Meynert 1884) or as an extension of the study of aphasia (Wernicke 1906) did not gain acceptance and were dismissed as "brain mythology."

3 Psychiatric Taxonomy in the Twentieth Century

3.1 Differentiation of Organically Caused Mental Disorders

As clinical observations and early successes in explaining their neuropathological bases and, sometimes, their etiology increased, new nosological concepts developed whose contours are also recognizable in present ICD-10 and DSM-IV classifications. Against the confusing background of seemingly innumerable putative causes (from infectious brain diseases and degenerative processes to nutritional-toxic influences and psychological trauma), it became possible to map clinical illnesses with some heuristic power at the turn of the century. As is also the case for many other models, psychiatric constructs developed in antithetical pairs (e.g., exogenous vs. endogenous, process vs. reaction, psychosis vs. neurosis). The influential and threefold division of mental disorders proposed by Kraepelin represents a form of these dichotomies. He compared the "organic" mental disorders, i.e., those associated with external causes and brain diseases, to functional psychoses (dementia praecox and manic depression) and to nonpsychotic mental disorders (neuroses and psychopathy), respectively. It quickly became clear that this simple schema did not permit an unambiguous and reliable classification of all mental disorders.

Contrary to the view expressed by Kraepelin, a specific psychopathological syndrome could not be attributed to each noxious substance, and it turned out that various systemic and brain diseases cause only quite a limited repertory of mental disorders. Whereas knowledge of neuropsychology could be further expanded and differentiated – not least by traumatizing war experiences – (e.g., Gelb and Goldstein 1920; Poppelreuther 1923), it constantly became clearer that unambiguous and integrated brain changes could not be understood using conventional postmortem neuropathological methods for a large number of mental disorders. This is the case both for progressive illnesses and for a series of clinical disorders with less pronounced deficits. A large number of clinical-phenomenological classifications were proposed for these mental reactions, but only a few are still in use today.

For Bonhoeffer (1908, 1910), dimming of consciousness was the main symptom of "acute exogenous type of reaction," i.e., those "types of mental damage" restricted to the large number of possible physical disorders. He differentiated the following: delirium; hallucinations with less pronounced loss of awareness and paranoid delusions; sudden motor and affective excitement with disorientation and failures in judgment (epileptiform type but without epileptic episodes); symptomatic stupor (from 1910: semi-conscious state) with slowing down, loss of awareness, loss of interest, impeded comprehension and reaction, and mild euphoria; finally, amentia with fluctuating consciousness disorders, incoherence, flight of ideas, mild ability to be diverted, and hallucinations. Only delirium and hallucinations still have comprehensible clinical significance today. It may be assumed that the others were possibly phenomena typical of the time.

[1]Translator's note: a postdoctoral thesis required to qualify as a university professor.

E. Bleuler (1916) differentiated between two types of "psychosyndrome" according to the damage observed, acute or chronic: the "diffuse brain psychosyndrome," largely synonymous with amnestic syndrome (Korsakow 1887; Moll 1915) and consisting of a loss of cognitive performance possibly accompanied by movement and perception disorders, and the "localized brain psychosyndrome," in which after focal or multifocal brain lesions, in spite of various locations of damage, a similar clinical picture appears with changes in drive and mood. Later, M. Bleuler (1954) added an "endocrine psychosyndrome," one that – with a variety of endocrine causes – can cause changes in social behavior, sleep, sexuality, movement, and sensitivity to warmth and cold, hunger and thirst.

The description by von Baeyer (1947) of "organic personality disorders" with lack of drive, cumbersomeness, affective disorder, over-familiar behavior, weakness, and increase or change in preexistent personality features is similar to E. Bleuler's definition of localized brain psychosyndrome.

A growing number of clinical observations showed that organic brain disorders could also be manifested similarly to schizophrenic illnesses or personality disorders. In a comprehensive literature review of 782 original articles, Davison and Bagley (1969) proved many systemic and brain illnesses to be the cause of a schizophreniform psychosis. On the basis of their symptomatology, organic and endogenous schizophrenia could not be reliably differentiated. It was clear that, in general, neither a noxious substance nor a lesion leads inevitably to a particular mental disorder, nor, on the contrary, does a particular psychopathological picture refer to a certain organic cause.

Kurt Schneider (1948) treated this diagnostic uncertainty with very restrictive criteria for the diagnosis of physically based psychoses: important physical finding, unambiguous connection, no indication of an alternative cause (e.g., through a familial genetic burden), favorable influence after improvement of the organic illness.

Lipowski (1975) advocated a broader concept, differentiating three kinds of mental disorders resulting from a systemic or brain illness:

- Disorders of organic origin in the narrower sense, with a variety of psychopathological disorders which are a direct consequence of diffuse or focal CNS damage or of a metabolic disorder
- Reactive disorders, i.e., psychoses, neuroses, personality or behavior disorders, which are a faulty adaptation to strains arising from mental illness and their psychological and social consequences
- Deviant behaviors with self-destructive refusal of compliance, denial of illness, or excessive dependency

At the same time, he emphasized the significance of cognitive disorders for the proof of organic causes, although a uniform and severe impairment of all cognitive abilities can only be found with extensive CNS changes. Disorders of organic origin can be classified in three ways by observing cognitive deficits: (1) according to the extent (global vs. selective, e.g., dementia and delirium vs. other disorders), (2) according to manifestation (severe vs. discrete, e.g., dementia and delirium vs. functionally evident disorders), and (3) according to the course (chronic vs. transient, e.g., dementia vs. delirium). Lipowski emphasized that the development of organically caused mental disorders is not only dependent upon the severity of the damage, but also upon dispositional and environmental factors. Kraepelin (1920) also came to share this view toward the end of his scientific career, when he moved away from his influential nosology and also from the concept of specificity of noxious substances.

3.2 Is the Concept of Organic Mental Illness Obsolete?

Attempts to dispense perhaps completely with the handy term "organic" – as proposed in DSM-IV – can prove difficult to enforce. The organic versus functional dichotomy continues to be useful in clinical practice and understanding, although it can scarcely be supported on theoretical grounds. Both of the currently most important classification systems, ICD-10 and DSM-IV, are similar in many important respects, but treat the problem of organicity differently. In DSM-IV (APA 1994), the concept of organic mental illness was abandoned, based on the consideration that this term implies the lack of a biological basis for nonorganic mental disorders. Those illnesses which were previously called organic now appear in sections on delirium, dementia and other cognitive disorders, psychological with medical illnesses, and drug-induced illnesses. The multiaxial structure of DSM fundamentally requires that every systemic or brain disease in which a causal relationship to a psychiatric symptomatology is assumed is recorded as comorbidity (axis III). This empirical base has an advantage, since it allows a theoretically unencumbered, unbiased collection of clinical data. Whether this is practical and leads to the accumulation of usable information remains to be shown.

In contrast to this, a more conservative approach was chosen in ICD-10 (WHO 1991), in which the classes of organic mental disorders were retained, although their boundaries were extended to include two groups of disorders, namely, a syndrome in which necessary and determining features are deficits

of cognitive functions or consciousness (i.e., the Bonhoeffer-Bleuler types of reaction and organic psychosyndrome), and a syndrome in which hallucinations, delusion, and disorders of affect, personality, and behavior are uppermost. In addition, ICD-10 differentiates between primarily organic psychosyndromes on the basis of specific brain diseases and secondary, symptomatic mental disorders caused by extracerebral illnesses.

The principle difference between DSM-IV and ICD-10 is concerned less with diagnosis and classification of diseases exhibiting cognitive deficits and disorders in consciousness than with the assessment of the gray zone of endogenous forms of mental and affective syndromes which are associated with functional disorders of the brain having variable genesis and severity. A schizophreniform illness with a somatic disease (e.g., epilepsy) according to ICD-10 could be coded in DSM-IV in several different ways: as schizophrenia on axis I (meeting the symptom and time criteria) with epilepsy (comorbidity on axis III), as a schizophreniform illness (if the diagnostic criteria of schizophrenia are not fulfilled) or as a psychotic illness with a somatic disorder, whereby the epilepsy would again then be encoded on axis III.

Research into "nonorganic" disorders in the last few decades has produced a series of organic correlates. In the case of schizophrenia, these results extend from ventricular enlargement and reduced hippocampal volumes to disordered neurons in the prefrontal cortex. In the case of delusional psychoses at higher ages, frequent neurodegenerative changes were found. In affective, anxiety, and compulsive disorders, as well as in personality disorders and many others, the significance of concrete organic factors becoming increasingly clear. It can be assumed that this research will lead to a fundamental contribution toward an integrated understanding of the interplay of psychological stressors and neurobiological factors.

With the increased sensitivity of neuroradiological and neurophysiological methods of investigation, the boundaries between "organic" and "functional" mental disorders are becoming blurred. At the same time, a revolution with a series of genetically determined diseases is indicated, which will shake conventional clinical nosologies and, at least in some areas, will make them obsolete, as diagnosis and intervention become possible even before symptoms are manifest. Similarly, the collection and classification of clinical and morphological data for these diseases will perhaps lose importance, as is presently the case for Kraepelin's symptomatology and pathology of neurosyphilis, which lost importance after the elucidation of its etiology.

4 References

Alzheimer A (1895) Die arteriosklerotische Atrophie des Gehirns. Allg Z Psychiatr 51: 809–811

Alzheimer A (1906) Über eine eigenartige Erkrankung der Gehirnrinde. Allg Z Psychiatr 64: 146–148

*Alzheimer A (1911) Über eigenartige Krankheitsfälle des späteren Alters. Z Ges Neurol Psychiatr 4: 356–885

APA (1994) Diagnostic and statistical manual of mental disorders, 4th edn (DSM-IV). American Psychiatric Association, Washington, DC

Armstrong J (1824) Practical illustrations of typhus fever. Collins and Haunay, New York

Ball B, Chambard E (1881) Demence apoplectique. In: Dechambre A, Lereboullet L (eds) Dictionnaire encyclopédique des sciences médicales. Masson, Paris

Bayle ALJ (1822) Recherches sur l'arachnite chronique. Thesis no. 247, Paris

Berrios G (1990) Alzheimer's disease: a conceptual history. Int J Geriatr Psychiatry 5: 355–365

*Binswanger O (1894) Die Abgrenzung der progressiven Paralyse I bis III. Berl Klin Wochenschr 49: 1103–1105, 1137–1139, 1180–1186

Bleuler E (1916) Lehrbuch der Psychiatrie, 1st edn. Springer, Berlin

Bleuler M (1954) Endokrinologische Psychiatrie. Thieme, Stuttgart

Bonhoeffer K (1908) Zur Frage der Klassifikation der symptomatischen Psychosen. Berl Klin Wochenschr 45: 2257–2260

Bonhoeffer K (1910) Die symptomatischen Psychosen im Gefolge von akuten Infektionen und inneren Erkrankungen. Deuticke, Leipzig

Broca P (1861) Nouvelle observation d'aphémie produite par une lésion de la moité postérieur des deuxième et troisième circonvolutions frontales. Bull Soc Anat Paris 36: 398–407

Calmeil LF (1826) De la paralysie considerée chez les aliénés. Baillière, Paris

Crichton A (1798) An inquiry into the nature and origin of mental derangement. Cadell and Davies, London

Cullen W (1777) First lines of the practice of physic. Elliott, Edinburgh

Davison K, Bagley CR (1969) Schizophrenia-like psychoses associated with organic disorders of the central nervous system: a review of the literature. Br J Psychiatry (special publication no 4): 114–178

Durand-Fardel M (1843) Traité de ramollissement du cerveau. Baillière, Paris

*Fischer O (1907) Miliare Nekrosen mit drusigen Wucherungen der Neurofibrillen, eine regelmäßige Veränderung der Hirnrinde bei seniler Demenz. Monatsschr Psychiatr Neurol 24: 361–372

Fuller S (1912) Alzheimer's disease (senium praecox): the report of a case and review of published ones. J Nerv Merit Dis 39: 440–455, 536–557

Gayet CJA (1875) Affection encéphalitique (encéphalite diffuse probable) localisée aux étages des pédoncles cérébraux et aux conches optiques. Arch Physiol Norm Pathol 2: 341–351

Gelb A, Goldstein K (1920) Psychologische Analysen hirnpathologischer Fälle. Barth, Leipzig

*Griesinger W (1845) Die Pathologie und Therapie der psychischen Krankheiten für Aerzte und Studierende. Krabbe, Stuttgart

Hankoff LD (1972) Ancient description of organic brain syndrome: the "Kardiakos" of the Talmud. Am J Psychiatr 129: 147–150

Hunter R (1973) Psychiatry and neurology – psychosyndrome or brain disease. Proc R Soc Med 66: 359–364

Kendell RE (1978) Die Diagnose in der Psychiatrie. Enke, Stuttgart

Korsakow SS (1887) Ob alkoholnom paralichie. Kushnereff, Moscow

*Kraepelin E (1899) Psychiatrie. Ein Lehrbuch für Studierende und Aerzte, vol 2, 6th edn. Barth, Leipzig

Kraepelin E (1920) Die Erscheinungsformen des Irreseins. Z Ges Neurol Psychiatr 62: 1–29

Lewy F (1912) Paralysis agitans. I. Pathologische Anatomie. In: Lewandowsky M (ed) Handbuch der Neurologie, vol 3. Springer, Berlin, pp 920–933

Liepman H (1905) Ueber Störungen des Handelns bei Gehirnkranken. Karger, Berlin

Lipowski ZJ (1975) Organic brain syndromes. In: Benson DF, Blumer D (eds) Psychiatric aspects of neurologic disease. Grune and Stratton, New York

Marce LV (1863) Recherches cliniques et anatomo-pathologiques sur la démence sénile et sur les différences qui la séparent de la paralyse générale. Gazette Med Paris 34: 433–435, 467–469, 497–502, 631–632, 761–764, 797–798, 831–833, 855–858

*Meynert T (1884) Psychiatrie. Klinik der Erkrankungen des Vorderhirns begründet auf dessen Bau, Leistungen und Ernährung. Braumüller, Vienna

Moll JM (1915) The "amnestic" or "Korsakov's" syndrome with alcoholic aetiology: an analysis of thirty cases. J Ment Sci 61: 424

Morgagni GB (1761) De sedibus et causis morborum per anatomen indagatis. Remondiniana, Venice

Pearson SB (1813) Observations on brain fever. Edinburgh Med Surg J 9: 326–332

*Pick A (1898) Beiträge zur Pathologie und pathologischen Anatomie des Centralnervensystems mit Bemerkungen zur normalen Anatomie desselben. Berlin, Karger

Poppelreuther W (1923) Zur Psychopathologie und Pathologie der optischen Wahrnehmung. Z Ges Neurol Psychiatr 83: 86–152

Redlich E (1898) Ueber miliare Sklerose der Hirnrinde bei seniler Atrophie. Jahrb Psychiatr Neurol 17: 208–216

Schneider K (1948) Klinische Psychopathologie, 2nd edn. Thieme, Stuttgart

Simchowicz T (1911) Histologische Studien über die senile Demenz. In: Nissl F, Alzheimer A (eds) Histologische und histopathologische Arbeiten über die Großhirnrinde mit besonderer Berücksichtigung der pathologischen Anatomie der Geisteskranken. Fischer, Jena, pp 267–443

Spielmeyer W (1930) Die Anatomie der Psychosen. In: Bumke O (ed) Handbuch der Geisteskrankheiten, part VII, vol XI. Springer, Berlin

Spurzheim JG (1833) Observations on the deranged manifestations of the mind or insanity. Marsh, Capen and Lyon, Boston

Sutton T (1813) Tracts on delirium tremens, on peritonitis and on some other inflammatory affections. Underwood, London

von Baeyer W (1947) Zur Pathocharakterologie der organischen Persönlichkeitsveränderungen. Nervenarzt 178: 21–28

Wernicke C (1874) Der aphasische Symptomenkomplex. Eine psychologische Studie auf anatomischer Basis. Cohn and Weigert, Breslau

Wernicke C (1881) Lehrbuch der Gehirnkrankheiten, vol II. Fischer, Kassel

*Wernicke C (1906) Grundriß der Psychiatrie in klinischen Vorlesungen, 2nd edn. Thieme, Leipzig

WHO (1991) Tenth revision of the international classification of diseases, chapter V (ICD-10). WHO, Geneva

Wilbrand H (1892) Ein Fall von Seelenblindheit und Hemianopsie mit Sectionsbefund. Dtsch Z Nervenheilkd 2: 361–87

Willis T (1672) De anima brutorum. Davis, London

CHAPTER 2

H. Lauter, A. Kurz

Clinical Assessment of the Dementias

Translator: E. Taub

1 Definition

The term "dementia" denotes an etiologically nonspecific and semiologically heterogeneous constellation of psychopathological manifestations. Dementia is primarily characterized by an acquired, persistent or progressive impairment of memory and cognition, occurring in the absence of a disturbance of consciousness and resulting in a limitation of the patient's ability to carry out his or her daily activities. The cognitive changes are associated with changes of affect control and social behavior.

This syndrome may be produced by many neurological and systemic diseases that affect different parts of the brain and have varying temporal courses. Just as the etiology of dementia is extremely heterogeneous, so, too, are its clinical manifestations. Cerebral damage primarily affecting frontal areas is typically associated with personality changes, loss of initiative, and paucity of language. Damage of the temporal and parietal lobes, in contrast, primarily produces disturbances of memory and orientation.

The recognition of the presence of dementia and the establishment of its underlying cause are fundamentally different components of the diagnostic process. Many organic brain disorders are indeed associated with an impairment of memory and other cognitive functions, but have the clinical manifestations not of a dementia syndrome, but rather of an "organic personality change" (as in some frontal lobe processes) or of a "mild cognitive dysfunction" (as in many cerebrovascular diseases; see Chap. 8, this volume, Part 2). In other cases, when the patient's intellectual baseline is high, a decline in cognitive ability may lead the clinician to suspect a dementing illness even though the diagnostic criteria for a dementia syndrome or other organic psychosyndrome are not (yet) fulfilled. These cases may be designated as "subdiagnostic" dementia (Helmchen et al. 1996). On the other hand, there are rare cases of patients with an unusually low intellectual baseline who may cross the diagnostic threshold for the dementia syndrome merely because of an additional, age-related cognitive decline, even though no brain illness is present.

2 Diagnostic Criteria

The diagnostic criteria for dementia have undergone several changes since the last edition of this textbook. With the introduction of DSM-III, operational diagnosis became standard for all psychiatric diseases, including the dementing diseases. DSM-III was rightly criticized, however, for inadequately defining the aspects of cognitive impairment that are most typical of dementia, and for failing to provide a satisfactory solution to the problem of making a categorical decision about the presence or absence of disease manifestations on the basis of dimensional functional impairments (Jorm and Henderson 1985). These deficiencies have been largely or completely remedied in the current psychiatric classification systems, ICD-10 and DSM-IV, both of which specify the diagnostically relevant features of dementia and indicate which diagnostic studies are appropriate. The global character of cognitive impairment is no longer emphasized, but the mere demonstration of an impairment of memory does not suffice for the diagnosis of dementia in either system. The patient's cognitive deficits must lead to a decline in his or her level of performance and must not be due to a disturbance of consciousness or to delirium.

Despite these common features, the two classification systems differ in several important ways. In ICD-10 (Appendix A), the typically chronic or progressive nature of the dementia syndrome is stressed, and the clinical manifestations are required to be of at least 6 months' duration, so that other, reversible conditions will not be misdiagnosed as dementia. This inclusion of prognostic data in the definition of dementia accords with the European tradition.

DSM-IV (Appendix B), in contrast, contains no such restriction; the course of dementia may be progressive, remitting, or stationary, depending on its underlying cause. The two systems also differ markedly in the type and number of required and supplementary defining features and in the algorithms specifying how these must be combined. For example, ICD-10 attaches far less diagnostic importance to aphasia, apraxia, and agnosia than DSM-IV. On the other hand, ICD-10 requires impairment not only of memory, but also of executive function as an obligatory prerequisite for the diagnosis of a dementia syndrome, and it further requires an impairment of emotional control, initiative, or social behavior.

The distinction between dementia and milder degrees of cognitive impairment not associated with comparably widespread psychopathological manifestations is drawn in both systems by the use of a threshold level of psychosocial functioning. The definition of dementia is thus made to depend on its effects on the patient's social relationships and ability to cope with the activities of daily living. According to ICD-10, mild dementia may be diagnosed as soon as memory impairment or other cognitive deficits impede the patient's activities of daily living, so that complicated everyday tasks and leisure activities can no longer be performed. According to DSM-IV, however,

dementia may be diagnosed only if each of the demonstrated cognitive deficits significantly limits the patient's social or occupational activities and causes a clear deterioration of his or her performance.

With regard to their clinical utility, both classification systems, despite their differences, serve adequately to facilitate the recognition of the dementia syndrome and to increase the reliability of its diagnosis. Unfortunately, however, it is still difficult to determine in many cases whether a mild cognitive impairment meets the rather vaguely defined threshold criterion for the diagnosis of dementia, especially because the social consequences of dementia in the individual case largely depend on the patient's life circumstances at the time, as well as on his or her level of education and ability to compensate for the deficits.

Moreover, the features said to be characteristic of dementia are not equally so for all dementing diseases and all stages in the progression of dementia. Memory impairment is mild or lacking in several types of cerebrovascular disease and lobar atrophy. The occurrence of episodes of delirium, said to rule out the presence of a dementia syndrome, is actually characteristic of Lewy body dementia. It thus makes sense that DSM-IV not only rejects the conventional distinction between organic and functional mental illness, but also avoids providing a unified, disease-independent definition of dementia. DSM-IV lists the psychopathological features constituting dementia primarily in the context of etiologically specified disease processes. This decision to avoid defining unified, overarching organic psychosyndromes can be understood as an attempt to establish a closer correspondence between individual organic brain diseases and their characteristic psychiatric manifestations and physical findings. This tendency has the further consequence that progressive diseases of the central nervous system leading to dementia can often be diagnosed at an early stage, when the psychopathological and social defining features of the dementia syndrome are still absent. The diagnosis of a dementing disease is thus no longer coupled to the ascertainment of a disease-independent dementia syndrome. This trend is likely to continue as diagnostic tests become ever more refined and to result in the increasing availability of early therapeutic and even preventive interventions.

3 Differential Diagnosis of the Dementia Syndrome

Unlike dementia, which usually takes a chronic progressive course, *delirium* is an organically caused brain disorder of abrupt onset and limited duration, characterized by clouding of consciousness, reduced ability to maintain, focus, and shift attention, marked fluctuation of clinical findings, and cognitive and perceptual impairment, as well as severe impairments of psychomotor function and of the sleep–wake cycle.

The commoner causes of delirium are acute somatic illnesses and disorders, drug intoxications, and withdrawal phenomena. Delirium generally lasts no more than a few days or weeks. Delirium may be superimposed on preexisting dementia. In such cases, its cause must be sought in concurrent illnesses or substance-related problems.

The *amnestic syndrome* consists of an isolated impairment of memory and is thus generally easily distinguished from dementia, in which several areas of cognitive functioning are simultaneously impaired. Detailed neuropsychological study of amnestic patients may disclose impairment of visuospatial processing or other subtle cognitive deficits that would otherwise escape notice. Loss of motivation and behavioral abnormalities, normally seen in frontal brain damage, are found in the Wernicke-Korsakoff syndrome.

Dementia is characterized by the loss of previously present cognitive abilities. If an adequate history is not available, dementia may be difficult to distinguish from *mental retardation* (either congenital or acquired early in life), from *inadequate education*, or from severe *personality disorders*, especially when the latter result in lifelong impairment of social adaptation.

Elderly individuals may have purely *age-related cognitive changes*. Cognitive ability declines with advancing age to a highly variable extent (Reischies 1997). The areas of performance that are most affected are speed, memory, and visuospatial ability (Crook et al. 1986). The cognitive changes of normal aging differ from those seen in dementia in that they are milder, progress much more slowly, and rarely affect everyday activities (Helmchen et al. 1997). Nonetheless, it may be quite difficult to distinguish age-related cognitive changes from dementing illness in very old patients of low primary intelligence.

Dementia must also be distinguished from organic psychosyndromes that are clearly too severe to represent a purely age-related cognitive decline, but nevertheless not severe enough to meet the diagnostic criteria for dementia, delirium, or amnesia – the latter primarily because they do not affect everyday life significantly enough to rise above the psychosocial threshold criterion for dementia. Such cognitive impairments come under the headings of *mild cognitive disorder* (ICD-10), *cognitive disorder, not otherwise specified* (DSM-IV), or *mild neurocognitive disorder* (DSM-IV Research Criteria). A prerequisite for the provision of any of these diagnoses is the objective demonstration of a neurological, medical, or substance-induced disease state underlying the

psychopathological signs and symptoms. A mild cognitive abnormality lacking any identifiable cause cannot be assigned a diagnosis under either of the current classification systems. Cases of this type are common and occupy a still inadequately defined intermediate zone between normal cognitive aging and dementia. Various terms have been proposed for such cases (see Chap. 10, this volume).

It is currently considered certain that some of these cases represent a stage of Alzheimer's disease preceding the onset of overt dementia, and that mild cognitive disorder will progress to dementia within 1 year in approximately 15% of patients (Petersen 1995; Grundman et al. 1996; Petersen et al. 1997). An intensive research effort is currently being devoted to the problem of identifying neurodegenerative diseases in the pre-dementia stage with the aid of markers in the blood or cerebrospinal fluid or of sensitive imaging techniques (Kurz et al. 1998). Such patients are an important target group for future therapeutic techniques intended to prevent dementia.

Several longitudinal studies have shown that dementia is misdiagnosed as depression in more than one quarter of all patients and, conversely, that depression is often misdiagnosed as dementia. Misdiagnoses of the latter type are due to the frequent occurrence of memory impairment, lack of concentration, and cognitive difficulties in depression. This combination of cognitive and affective disturbances has been designated as "pseudodementia" (Kiloh 1961) or, alternatively, as the "dementia syndrome of depression" (Folstein et al. 1978). In most such patients, however, the cognitive impairment is not severe enough to merit the diagnosis of dementia, and its pattern is often atypical for the dementia syndrome. In particular, patients with "depressive pseudodementia" do not have aphasia, agnosia, or apraxia, disturbances of orientation, or impairment of their everyday activities because of their cognitive disturbance (Des Rosiers et al. 1995). Further features distinguishing such patients from those with organic cognitive disturbances include the previous occurrence of depressive episodes, a family history of affective disorders, and a discordance between the subjective perception of severe intellectual deficits and the relatively mild objective findings (Stoppe and Staedt 1994).

A close and consistent correlation between the degree of cognitive impairment and the severity of depression cannot be demonstrated. Moreover, even when there is marked clinical improvement with respect to the affective manifestations, cognitive performance typically recovers only in part (Alexopoulos et al. 1993). The degree of reversibility of cognitive disturbances in depression has likely been overestimated. When deficits of memory and intellect persist, a combination of dementia and depression should be suspected.

4 Clinical Assessment

A comprehensive mental status examination and physical examination must be performed to confirm or exclude the diagnosis of dementia in every patient in whom it is suspected. The examination begins with careful questioning of the patient and of a reliable informant. A history obtained from an informant other than the patient is the most important set of data for the diagnosis of dementia. The history includes information on when, how, and in what circumstances the illness first became manifest, its subsequent course, its effect on the ability to carry out activities of daily living, behavioral changes, previous illnesses, medication use, alcohol consumption, and the possible occurrence of similar illnesses in the patient's family. The history also includes information concerning the patient's primary personality, level of education, life history, and social situation. After the history has been taken, a detailed psychiatric examination is performed, which must include a test of cognitive functions.

History-taking and psychiatric examination serve the purpose of translating the grammar of the operationally defined classification criteria into a series of interview questions, through which information is obtained about the defining features of the dementia syndrome and of other disorders that must be differentiated from it, and which take account of the patient's intellectual capacity. If the examination yields sufficient evidence favoring a diagnosis of dementia, the severity of dementia should be gauged with a test of cognitive performance, such as the Mini-Mental Status Test (MMST). If the diagnosis is difficult to confirm, the use of additional standardized assessment techniques or neuropsychological testing for cognitive deficits should be considered. In some cases, a diagnosis of dementia cannot be made in a single examination and requires observation of the patient's behavior and clinical course over a longer period; this is especially true when a reliable informant is lacking, when there is concurrent delirium, and when subtle personality changes need to be evaluated.

The remaining steps of the evaluation are intended to establish the cause of the dementia syndrome. First, a detailed physical examination is performed, with special attention to the neurological examination. A series of baseline laboratory tests provides an overview of the functional state of the major organ systems and serves to identify possible causes of dementia as well as other, potentially treatable somatic illnesses. Beyond

these basic diagnostic techniques, further, possibly invasive procedures may be indicated if the initial investigation leads to the suspicion of particular underlying diagnoses. Further laboratory testing is sometimes necessary. A lumbar puncture is indicated in atypical, rapidly progressive forms of dementia or when an inflammatory process affecting the central nervous system is suspected.

The elements of the psychiatric examination can be summarized as follows:

1. History-taking from the patient
2. Questioning of a close relative
3. Determination of family history
4. Determination of psychiatric findings
5. Clinical dementia test to determine the severity of cognitive impairment
6. Physical examination, with particular attention to the neurological examination
7. Baseline laboratory tests: blood profile, erythrocyte sedimentation rate, urinalysis, electrolytes, hepatic enzymes, protein electrophoresis, thyroid-stimulating hormone (TSH), vitamin B_{12} and folic acid, *Treponema pallidum* hemagglutination assay (TPHA)
8. Electroencephalography (EEG)
9. Computed tomography (CT)/magnetic resonance imaging (MRI)

The evaluation of the patient does not end with the confirmation of the presence of dementia and the determination of its etiology. The *nomothetic diagnosis*, by which the patient's abnormal state is abstracted and categorized as conforming to a particular element of a disease classification system, must be complemented by *idiographic diagnosis*, i.e., the consideration of a large number of factors that are of diagnostic, prognostic, and therapeutic importance to the individual patient. One such factor is the possible presence of concurrent noncognitive disorders (Burns et al. 1990), which are often more responsive to symptomatic and palliative treatment than the cognitive deficits and the limitation of everyday activities. The assessment also includes the identification of retained abilities that may aid in rehabilitation and the detection of secondary factors, such as physical illnesses and handicaps or sensory deficits, as well as evaluation of the patient's present and future needs for social and nursing support or for assistance with legal and financial matters.

In many cases, a position must be taken regarding the patient's competence to provide informed consent to diagnostic and therapeutic measures. The ethical principles and legal criteria that apply in this area, and their practical implementation in methods of psychiatric assessment, have long been a subject of intense discussion in the biomedical community in all Western countries (Helmchen and Lauter 1995; Koch et al. 1996).

The various aspects of idiographic assessment can be briefly summarized as follows:

1. Psychiatric aspects:
   a) Particular manifestations of psychopathology (e.g., depression, hallucinations, delusions, sleep disturbance)
   b) Behavioral abnormalities (e.g., disorientation, danger of running away)
   c) Degree of independence and need for nursing care
2. Somatic aspects:
   a) Other somatic illnesses
   b) Sensory deficits (visual or auditory impairment)
   c) Motor deficits (gait impairment, danger of falling)
3. Social aspects:
   a) Living situation (alone, with family, in institution)
   b) Presence of other person(s) with a relationship to the patient
   c) Quality of relationship
   d) Burden on other person(s)
4. Personal aspects:
   a) Life history
   b) Personal values
   c) Awareness of illness and suffering
   d) Retained abilities
5. Legal aspects:
   a) Competence to give informed consent
   b) Need for legal guardianship
6. Prognostic aspects:
   Stage of progression of the dementing process

The goals of this type of assessment are the education and counseling of the patient and his or her family, the provision of social assistance, and the initiation of measures for treatment and rehabilitation.

5 Subjective Aspects

Statements made by demented patients reveal that they are aware of their cognitive impairment not only at the inception of the dementing process, but also in its advanced stages. Many aspects of these patients' behavior are not direct consequences of the underlying changes in the brain, but can rather be understood as natural reactions to, and attempts to cope with, the perceived loss of functional ability (Kurz et al. 1988; Haupt and Kurz 1990).

Family members are less disturbed by the patient's cognitive impairment itself than by the resulting difficulty of communication, progressive loss of

initiative, and changes in personality and behavior, including anxiety, aggression, sleep disturbances, agitation, and delusions. Particularly disturbing to family members is their often protracted uncertainty over whether the observed changes in the patient are merely aspects of his or her personality or expressions of an underlying disease. The patient's relatives are often poorly informed about the nature, manifestations, and course of dementing illnesses. They suffer from the increasing limitation of their own life possibilities and from often continuous tension, as well as role conflicts and guilt feelings. These stresses, in turn, may lead to the psychoreactive generation or exacerbation of health problems (Morris 1988). Awareness of the subjective aspects of dementia on the part of the physician is an indispensable element of the clinical evaluation and is a prerequisite for sympathetic interaction with patients and their families and for the alleviation of their suffering.

6 Standardized Techniques of Assessment

Psychiatric history-taking and examination are generally carried out in the form of a free interview. The free interview has the advantage that an experienced clinician can adjust the questions asked to the particular cognitive abilities of the patient under examination, or of his or her accompanying family members, and thus maintain a relaxed, conversational atmosphere that helps the evaluation attain its goal. Nonetheless, the data obtainable by this technique necessarily suffer, to some extent, from a certain informational and observational inadequacy and cannot be expressed in quantitative form. The use of standardized instruments of assessment can counteract these deficiencies. Such instruments, including clinical psychometric tests, structured interviews, and specific neuropsychological tests for the assessment of cognitive disorders, are discussed in the following section.

6.1 Psychometric Tests

Some of the typical features of dementia, such as memory impairment, are also found in normal elderly individuals and are distinguished as pathological only by their severity. Diagnostic assessment, to be useful, must therefore not just categorize these deficits, but also rate their severity. The degree of cognitive and psychosocial impairment is assessed by means of brief tests or rating scales that measure the severity of performance impairments, behavioral disturbances, and handicaps. They incorporate cutoff values that demarcate the different grades of severity of the deficits under study or serve as a threshold beyond which the diagnosis of dementia is probable. Such instruments can thus contribute to the evaluation of dementia. It should be remembered, however, that they only inform us about the behavior or cognitive performance of a subject at the time of examination and do not indicate what factors may be responsible for the deficits and abnormalities they reveal. Poor performance on a cognitive test may be due to dementia, but may also be the product of many other factors, including advanced age, poor attention, lack of motivation, depression, or physical illness. Clinical diagnosis, therefore, does not always accord with the results of psychometric tests (O'Connor et al. 1991).

From the methodological viewpoint, psychometric tests may be classified as objective tests of performance, self-assessment tests, and tests involving assessment by other individuals; they may, alternatively, be classified according to the source of the data (e.g., patient, examiner, family member). In Table 1, tests are classified according to the specific features that they test.

Several instruments are designed exclusively for the quantification of cognitive impairment. Other psychometric tests deal with the behavioral abnormalities and depressive symptoms of demented patients. Some tests serve to assess competence in the activities of daily living and provide a measure of retained independence and of the degree of need for social or nursing assistance. Some standardized instruments include batteries of performance tests and rating scales and are used for the simultaneous assessment of cognitive and noncognitive disturbances. There are yet other rating scales for the severity and stage of progression of individual dementing illnesses. A last group of instruments is used to assess changes in the patient's condition, particularly in the evaluation of therapeutic approaches.

6.2 Structured Interviews

Free questioning and assessment on the basis of the examiner's experience are not sufficiently reliable or valid for use in the evaluation of mild cognitive disturbances, particularly in field studies on large population samples, which generally employ multiple examiners.

Semistandardized, structured interview techniques are preferable for use in such situations. The questions have a prescribed form and structure, and the encoding and processing of the responses are standardized,

Table 1. Commonly used psychometric tests in dementia

Feature tested	Test	Reference
Cognitive impairment	Mini-Mental Status Test (MMST)	Folstein et al. (1975)
	Syndrom-Kurz-Test (SKT)	Erzigkeit (1983)
	Brief Cognitive Rating Scale (BCRS)	Reisberg et al. (1983)
	Mattis Dementia Rating Scale (DRS)	Mattis (1976)
	Hierarchic Dementia Scale (HDS)	
	Clock test	Ploenes et al. (1994)
Behavioral abnormalities	Behavioral Pathology in Alzheimer's Disease (BEHAVE-AD)	Reisberg et al. (1987)
	Cohen-Mansfield Agitation Inventory (CMAI)	Cohen-Mansfield and Billig (1986)
	Behavior Rating Scale for Dementia of the Consortium to Establish a Registry for AD (CERAD-BRSD)	Tariot et al. (1995)
	Neuropsychiatric Inventory (NPI)	Cummings (1996)
Depressive symptoms	Geriatric Depression Scale (GDS)	Yesavage et al. (1983)
	Dementia Mood Assessment Scale (DoMAS)	Sunderland et al. (1988)
Competence and functional abilities	Instrumental Activities of Daily Living (IADL) Scale	Lawton et al. (1969)
	Functional Assessment Staging (FAST)	Reisberg (1988)
	Progressive Deterioration Scale (PDS)	DeJong et al. (1989)
	Interview for Deterioration in Daily Living Activities in Dementia (IDDD)	Teunisse (1995)
	Disability Assessment for Dementia (DAD)	Gauthier et al. (1993)
Combined assessment of cognitive and noncognitive functions	Nürnberger Alters-Inventar (NAI)	Oswald and Fleischmann (1995)
	Alzheimer's Disease Assessment Scale (ADAS)	Mohs and Cohen (1988)
	Clinical Assessment Geriatric Scale (SCAG)	Shader et al. (1974)
Severity and stage of progression of dementia	Global Deterioration Scale (GDS)	Reisberg et al. (1982)
	Clinical Dementia Rating (CDR)	Hughes et al. (1982)
Changes in condition over the course of treatment	Clinician's Global Impression of Change (CGIC)	Guy (1976)
	Clinician's Interview-Based Impression of Change (CIBIC, CIBIC plus)	Schneider et al. (1997)
	Alzheimer's Disease Cooperative Study Clinician's Global Impression of Change (ADCS-CGIC)	Schneider and Olin (1997)

but the examiner still has considerable freedom to adapt the line of questioning to the situation of the particular individual being interviewed.

Some of these instruments have proved their value primarily in the evaluation of dementing and depressive illnesses in advanced age; these include the Initial Subject Protocol (ISP; Berg et al. 1982), the Cambridge Mental Disorders of the Elderly Examination (CAMDEX; Roth et al. 1986), the Geriatric Mental State Schedule (GMS; Copeland et al. 1976, 1986, 1988), the Comprehensive Assessment and Referral Evaluation (CARE; Gurland et al. 1977), the Canberra Interview for the Elderly (CIE; Henderson 1992), and, in German, the *Strukturierte Interview zur Diagnose von Demenzen vom Alzheimer Typ, der Multiinfarkt- (oder vaskulären) Demenz und Demenzen anderer Ätiologie* (SIDAM: Zaudig et al. 1991; Zaudig 1996). Zaudig (1995) has provided a detailed overview of these instruments.

The instruments named here also take account of information supplied by the patient's family, obtained in part through interviews with additional informants, and contain a number of internationally used rating scales. Diagnostic evaluation can thus proceed according to a preset algorithm, which has been tailored to the psychiatric classification systems currently in use.

6.3 Neuropsychological Tests

Neuropsychological tests are useful primarily for the early recognition of organic brain illnesses and the detailed study of their course. They are also used for the assessment of specific impairments of cognitive function, for the purpose either of providing a guide

for rehabilitation, or of precisely localizing the patient's clinical findings to structural or functional brain damage at a specific site. There are neuropsychological tests of different aspects of memory, linguistic and visuospatial ability, and cognitive executive and control functions, including certain aspects of attention, thinking, and problem-solving behavior. The test results can be compared with normal values for the patient's age, sex, and level of education. A simple and easily applied technique for neuropsychological assessment was developed by the Consortium to Establish a Registry for Alzheimer's Disease (CERAD) (Welsh et al. 1994). The CERAD battery tests the patient's performance on verbal and nonverbal learning and memory tasks, naming, verbal fluency, and visual construction. It includes the MMST.

7 Screening Tests

Many dementing illnesses are not diagnosed until late in their course. Mild dementia is often overlooked, not only by physicians, but even by family members, especially those of very old patients (Ross et al. 1997). The question thus arises of whether elderly persons at especially high risk for dementia should undergo screening, as is done for other medical problems in this age-group. An ideal screening test would detect probable cases rapidly and efficiently, but would not necessarily have to provide a definitive diagnosis. The currently available tests are low in specificity, so that screening would have to be followed by a detailed psychiatric examination in many cases.

The Seven-Minute Screen (Solomon et al. 1998) is a new screening technique of this type that might find widespread application in general practice. It consists of four simple tests that may be conducted by auxiliary medical personnel, require no specialized knowledge for their diagnostic interpretation, and take a total of less than 8 min to apply on average. This screening procedure is very highly specific and sensitive for the differentiation of cognitively intact elderly persons from patients with mild Alzheimer's disease. It remains to be seen how well this procedure will perform under the conditions of general medical practice. The question of whether screening for dementia is ethically justified or medically defensible, in view of the possibly unnecessary worry it would cause elderly patients (Cooper and Bickel 1984; Henderson 1994), can now be answered more positively than a few years ago, in view of the better therapeutic options now available.

8 Comorbidity

An increased frequency of depressive disorders has been found repeatedly in cohorts of demented patients (Lammi et al. 1989; Henderson 1990; Reifler et al. 1982).

Field studies on community samples, however, have not demonstrated any correlation between categorically diagnosed dementia and depression. With the aid of dimensional evaluative techniques, it was revealed that depressive symptoms are more frequent with increasing cognitive impairment, but severe performance impairments due to dementia are somewhat negatively correlated with the presence of depressive symptoms (Helmchen et al. 1993). It must be borne in mind that severely demented subjects often cannot be included in field studies, and their responses often cannot be evaluated. Informant interviews, too, have only limited ability to establish the presence of depressive symptoms in demented patients, because relatives generally rate the severity of depression more highly than the patients themselves, particularly when dementia has reached an advanced stage (Karrer 1995). The validity of a diagnosis of depression in the presence of dementing disease is further called into question by the fact that several characteristic features of depression, including loss of initiative, slowing, social withdrawal, sleep disturbance, and vegetative symptoms, may also be the result of cognitive impairment, and their inclusion in a diagnostic scheme for depression may artificially exaggerate the correlation between dementia and depression.

Demented patients are more likely than individuals of the same age who are not mentally ill to suffer from multiple organic illnesses and physical handicaps. Somatic comorbidity contributes to the generation of depressive symptoms, some of which may be of purely somatic origin. The physician must bear this diagnostic ambiguity in mind when deciding whether treatment with antidepressants is indicated.

9 Ethical Problems in Research

The necessary improvements in the area of early diagnosis will only come about through a steady increase in scientific knowledge. This, in turn, depends on the voluntary participation of patients in epidemiological and clinical research studies that will yield a better understanding of the causes of dementia and of

potential early recognition of disease-specific traits. Severely demented patients, however, are not competent to give their informed consent to such studies. According to German law, individuals not competent to give informed consent may be included in a research project only when they are expected to derive an immediate benefit from it. This may be the case for therapeutic trials, but not for epidemiologic, pathogenetic, or diagnostic studies. It may be asked whether nontherapeutic research on incompetent patients ought to be legally permitted in some cases, e.g., in exceptionally promising scientific projects involving no particular hardship or risk to the patient, when the scientific question cannot be answered by studying competent patients, and when provisions are made for the patient's protection. This problem has recently come under discussion in Germany (Helmchen and Lauter 1995) and was extensively dealt with in the European Council's hearings on the Human Rights Convention in Biomedicine; it has also received much critical attention in the public at large (see Chap. 18, Vol. 1, Part 2).

Appendix A. Diagnostic Criteria for Dementia According to ICD-10

1. A decline in memory, which is most evident in the learning of new information. The decline should be objectively verified by obtaining a reliable history from an informant, supplemented, if possible, by neuropsychological tests or quantified cognitive assessments.
2. A decline in other cognitive abilities characterized by deterioration of judgment and thinking, such as planning and organizing, and in the general processing of information. Evidence for this should be obtained when possible from interviewing an informant, supplemented, if possible, by neuropsychological tests or quantified objective assessments. Deterioration from a previously higher level of performance should be established.
3. The disturbances described in 1 or 2, above, must be severe enough to adversely affect the patient's activities of daily living.
4. Absence of clouding of consciousness.
5. A decline in emotional control or motivation, or a change in social behavior, manifest as at least one of the following:
   a) emotional lability,
   b) irritability,
   c) apathy,
   d) coarsening of social behavior.

All five criteria must be fulfilled. For a confident clinical diagnosis, the abnormalities listed in 1 and 2 should have been present for at least 6 months; if the period since the manifest onset is shorter, the diagnosis can only be tentative.

Appendix B. Common Symptom Presentation in Disorders Listed in the "Dementia" Section of DSM-IV

A. Development of multiple cognitive deficits, which are manifest as

1. memory impairment,
2. at least one of the following four disturbances:
   a) aphasia (deterioration of language function)
   b) apraxia (impaired ability to execute motor activities despite intact motor abilities, sensory function, and comprehension of the required task)
   c) agnosia (failure to recognize or identify objects despite intact sensory function)
   d) disturbances in executive functioning (impaired ability to think abstractly and to plan, initiate, sequence, monitor, and stop complex behavior).

B. The items in both Criterion A1 and Criterion A2 must be severe enough to cause significant impairment in social or occupational functioning and must represent a decline from a previous level of functioning.

C. Dementia is not diagnosed if these symptoms occur exclusively during the course of a delirium. However, a delirium may be superimposed on a preexisting dementia, in which case both diagnoses should be given.

10 References

*Alexopoulos GS, Meyers BS, Young RC, Mattis S et al (1993) The course of geriatric depression with "reversible dementia": a controlled study. Am J Psychiatry 150: 1693–1699

Berg L, Hughes CP, Coben LA, Danziger WL et al (1982) Mild senile dementia of Alzheimer type: research diagnostic criteria, recruitment, and description of a study population. J Neurol Neurosurg Psychiatry 45: 962–968

**Burns A, Jacoby R, Levy R (1990) Psychiatric phenomena in Alzheimer's disease I–IV. Br J Psychiatry 157: 72–94

*Coen RF, Swanwick GRJ, O'Boyle CA, Coakley D (1997) Behaviour disturbance and other predictors of carer burden in Alzheimer's disease. Int J Geriatr Psychiatr 12: 331–336

Cohen-Mansfield J, Billig N (1986) Agitated behaviors in the elderly I. A conceptual review. J Am Geriatr Soc 34: 711–721

Cole MG, Dastoor DP (1983) The Hierarchic Dementia Scale. J Clin Exp Gerontol 5: 219–234

*Cooper B, Bickel H (1984) Population screening and the early detection of dementing disorders in old age: a review. Psychol Med 14: 81–95

Copeland JR, Kelleher MJ, Kellett JM, Gourlay AJ et al (1976) A semi-structured clinical interview for the assessment of diagnostics and mental state in the elderly: the Geriatric Mental State Schedule. I. Development and reliability. Psychol Med 6: 439–449

Copeland JRM, Dewey ME, Griffiths-Jones HM (1986) A computerized psychiatric diagnostic system and case nomenclature for elderly subjects: GMS and AGECAT. Psychol Med 16: 89–99

Copeland JRM, Dewey ME, Henderson AS (1988) The Geriatric Mental Status (GMS) used in the community: replication studies of the computerized diagnoses AGECAT. Psychol Med 18: 219–232

Crook TH, Bartus RT, Ferris SH, Whitehouse P et al (1986) Age-associated memory impairment: proposed diagnostic criteria and measures of clinical change – Report of a National Institute of Mental Health Workgroup. Dev Neuropsychol 2(4): 261–276

Cummings J (1996) The Neuropsychiatric Inventory: assessing psychopathology in dementia patients. Neurology 48[Suppl 6]: S10–S16

De Jong R, Ostersund OW, Roy GW (1989) Measurement of quality-of-life changes in patients with Alzheimer's disease. Clin Ther 11: 545–554

Des Rosiers G, Hodges JR, Berrios G (1995) The neuropsychological differentiation of patients with very mild Alzheimer's disease and/or major depression. J Am Geriatr Soc 43: 1256–1263

Folstein MF, McHugh PR (1978) Dementia syndrome of depression. In: Katzman R, Terry RD, Bick LK (eds) Alzheimer's disease, senile dementia, and related disorders. Raven, New York, pp 87–96

*Folstein MF, Folstein SE, McHugh PR (1975) "Mini-Mental State": a practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res 12: 189–198

Gauthier L, Gauthier S, McIntyre M, Wood-Dauphinee S (1993) Assessment of functioning and ADL. Abstracts of the Sixth Congress of the International Psychogeriatric Association 9

*Grundman M, Petersen RC, Morris JC, Ferris S et al (1996) Rate of dementia of the Alzheimer type (DAT) in subjects with mild cognitive impairment. Neurology 46[Suppl]: A403

Gurland BJ, Kuriansky J, Sharpe L (1977) The Comprehensive Assessment and Referral Evaluation – CARE – : rationale, development, and reliability. Int J Aging Hum Development 8: 9–42

Guy W (1976) Clinical Global Impressions (CGI) ECDEU Assessment Manual for Psychopharmacology. US Department of Health and Human Services, Public Health Service, Alcohol Drug Abuse and Mental Health Administration, NIMH Psychopharmacology Research Branch, Rockville, pp 218–222

Haupt M, Kurz A (1990) Alzheimersche Krankheit: Erleben, Empfinden und Reaktionsformen des Kranken. Z Gerontol 23: 211–213

Helmchen H, Lauter H (1995) Dürfen Ärzte mit Demenzkranken forschen? Stuttgart, Thieme

Helmchen H, Linden M (1993) The differentiation between depression and dementia in the very old. Ageing Society 13: 589–617

**Helmchen H, Linden M, Wernicke T (1996) Psychiatrische Morbidität bei Höchstbetagten. Ergebnisse aus der Berliner Altersstudie. Nervenarzt 67: 739–750

Henderson AS (1990) Co-occurrence of affective and cognitive symptoms: the epidemiological evidence. Dementia 1: 119–123

Henderson AS (1992) The Canberra Interview for the Elderly: a new field instrument for the diagnosis of dementia and depression by ICD-10 and DSM-III-R. Acta Psychiatr Scand 85: 105–113

Henderson AS (1994) Early detection. In: Copeland JRM, Abou-Saleh MT, Blazer DG (eds) Principles and practice of geriatric psychiatry. Wiley, Chichester, pp 267–271

Hughes CP, Berg L, Danziger WL, Coben LA et al (1982) A new clinical scale for the staging of dementia. Br J Psychiatry 140: 566–572

Jorm AF, Henderson AS (1985) Possible improvements to the diagnostic criteria in DSM III. Br J Psychiatr 140: 394–399

Karrer S (1995) Einflußfaktoren auf die Beurteilung von Depressivität bei Patienten mit Alzheimerscher Krankheit. Thesis, Technische Universität, Munich

Kiloh LG (1961) Pseudo-dementia. Acta Psychiatr Scand 37: 336–351

Koch HG, Reiter-Theil S, Helmchen H (eds) (1996) Informed consent in psychiatry. Nomos, Baden-Baden

Kurz A, Feldmann R, Lauter H (1988) Leben mit der Demenz. Fundam Psychiatry 2: 3–7

Kurz A, Riemenschneider M, Buch K, Willoch F et al (1998) Tau protein in cerebrospinal fluid is significantly increased at the earliest clinical stage of Alzheimer disease. Alzheimer Dis Assoc Disord 12: 372–377

Lammi UK, Kivelä SL, Nissinen A (1989) Mental disability among elderly men in Finland: prevalence, predictors and correlates. Acta Psychiatr Scand 80: 459–468

Lawton MP, Brody EM (1969) Assessment of older people: self-maintaining and instrumental activities of daily living. Gerontologist 9: 176–186

Mattis S (1976) Dementia Rating Scale. In: Bellak R, Karasu B (eds) Geriatric psychiatry. Grune and Stratton, New York, pp 108–121

Mohs RC, Cohen L (1988) Alzheimer's Disease Assessment Scale (ADAS). Psychopharmacol Bull 24: 627–628

*Morris RG (1988) Factors affecting the emotional wellbeing of the caregivers of dementia sufferers. Br J Psychiatry 153: 147–156

*O'Connor DW, Politt PA, Hyde JB (1991) Clinical issues relating to the diagnosis of mild dementia in a British community survey. Arch Neurol 48: 530–543

Oswald WD, Fleischmann UM (1995) Nürnberger Alters-Inventar. Hogrefe, Göttingen

Petersen RC (1995) Normal aging, mild cognitive impairment, and early Alzheimer's disease. Neurologist 1(6): 326–344

*Petersen RC, Parisi JE, Hohnson KA, Waring SC et al (1997) Neuropathological findings in patients with a mild cognitive impairment. Neurology 48[Suppl]: A102

Ploenes C, Sharp S, Martin M (1994) Der Uhrentest: Das Zeichnen einer Uhr zur Erfassung kognitiver Störungen bei geriatrischen Patienten. Z Gerontol 27: 246–252

Reifler BV, Larson E, Hanley R (1992) Coexistence of cognitive impairment and depression in geriatric outpatients. Am J Psychiatry 139: 623–626

Reisberg B (1988) Functional Assessment Staging (FAST). Psychopharmacol Bull 24: 653–659

Reisberg B, Ferris SH, De Leon M, Crook T (1982) The Global Deterioration Scale for assessment of primary degenerative dementia. Am J Psychiatry 139: 1136–1139

Reisberg B, Schneck MK, Ferris SH, Schwartz GE et al (1983) The Brief Cognitive Rating Scale (BCRS): findings in primary degenerative dementia (PDD). Psychopharmacol Bull 19: 47–51

**Reisberg B, Borenstein J, Salob SP, Ferris SH et al (1987) Behavioral symptoms in Alzheimer's disease: phenomenology and treatment. J Clin Psychiatry 48(5)[Suppl]: 9–15

*Reischies FM (1997) Normales Altern und leichte Demenz: Auswirkungen normalen Alterns auf kognitive Leistungen und die Differenzierung von der leichten Demenz. In: Förstl H (ed) Lehrbuch der Gerontopsychiatrie. Enke, Stuttgart, pp 366–377

Ross GW, Abbott RD, Petrovitch H, Masaki KH et al (1997) Frequency and characteristics of silent dementia among elderly Japanese-American men. The Honolulu-Asia Aging Study. JAMA 277: 800–805

*Roth M, Tym E, Mountjoy CQ, Huppert FA et al (1986) CAMDEX. A standardized instrument for the diagnosis of mental disorder in the elderly with special reference to the early detection of dementia. Br J Psychiatry 149: 698–709

Schneider LS, Olin JT (1997) Clinical global impressions in Alzheimer's clinical trials. Int Psychogeriatr 8: 277–288

Schneider LS, Olin LT,b Doody RS et al (1997) Validity and reliability of the Alzheimer's Disease Study – Clinical Global Impression of Change (ADCS-CGIC). Alzheimer Dis Assoc Disord 2[Suppl 1]: 13–23

Shader RI, Harmatz JS, Salzman C (1974) A new scale for clinical assessment in geriatric populations: Sandoz Clinical Assessment – Geriatric (SCAG). J Am Geriatr Soc 22(3): 107–113

Solomon PR, Hirschoff A, Kelly B, Relin M et al (1998) A 7 minute neurocognitive screening battery highly sensitive to Alzheimer's disease. Arch Neurol 55: 349–355

Stoppe G, Staedt J (1994) Die frühe diagnostische Differenzierung primär dementer von primär depressiven Syndromen im Alter – ein Beitrag zur Pseudodemenzdiskussion. Fortschr Neurol Psychiatr 61: 172–182

Sunderland T, Alterman IS, Yount D, Hill JL et al (1988) A new scale for the assessment of depressed mood in demented patients. Am J Psychiatry 145: 955–959

Tariot PN, Mack JL, Patterson MB, Edland SE et al (1995) The Behavior Rating Scale For Dementia of the Consortium to Establish a Registry for Alzheimer's Disease. Am J Psychiatry 152: 1349–1357

Teunisse S (1995) Activities of daily living scales in dementia: their development and future. In: Levy R, Howard R (eds) Developments in dementia and functional disorders in the elderly. Wrightson Biomedical, pp 85–95

Welsh KA, Butters N, Mohs RC, Beekly D et al (1994) The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). V. A normative study of the neuropsychological battery. Neurology 44: 609–614

Yesavage JA, Brink TL, Rose TL (1983) Development and validation of a geriatric depression screening scale: a preliminary report. J Psychiatr Res 17:37

*Zaudig M (1995) Demenz und "leichte kognitive Beeinträchtigung" im Alter. Diagnostik, Früherkennung und Therapie. Huber, Bern

Zaudig M, Hiller W (1996) SIDAM Handbuch. Huber, Bern

Zaudig M, Mittelhammer J, Hiller W (1991) SIDAM-A structured interview for the diagnosis of dementia of the Alzheimer type, multiinfarct dementia and dementias of other etiology according to ICD-10 and DSM-III. Psychol Med 21: 225–236

CHAPTER

3

H. Bickel

Descriptive Epidemiology of Dementias

Translator: S. Henn

1 Introduction

The number of older individuals and their proportion they account for of the total population has risen steeply in industrial countries over the course of the twentieth century. In Germany, for example, the number of people over 65 increased fourfold to more than 12 million and is still increasing; at the same time, the number of those over 80 years of age increased tenfold (Statistisches Bundesamt 1985, 1994). In the wake of these demographic changes, the dementias have become a serious problem for the health care system.

Descriptive epidemiology endeavors to ascertain the distribution of dementias in the population and to provide information about time-related changes in frequency of the illness, locally varying distributions, and associations with personal characteristics such as age and gender. Routine statistics are of little value. Since the majority of patients do not seek help from specialists, the treatment statistics of psychiatric institutions or data from psychiatric case registers reflect only a portion of the morbidity present in the population. Statistics related to causes of death are also unproductive, since they likewise record only a fraction of the cases of illness. Epidemiology is therefore dependent upon field studies with random sampling of representative populations for a description of the actual spread of dementias.

Illness frequency in this study will be expressed in the form of variously defined measures of prevalence and incidence. Prevalence, which can be relatively easily estimated from cross-sectional studies, gives the portion of the population which suffers from dementia at a particular time, while incidence, which requires a more expansive and time-consuming longitudinal study, estimates the frequency of new illnesses within a particular period. The two values are relevant for different kinds of problems. Prevalence is above all important for the purposes of planning care, since it records the existence of the illness and can give clues to the existing need among the patients for medical and social services. Prevalence does not, however, allow compelling conclusions to be made regarding the distribution of risks for the illness, as prevalence depends both on the incidence and on the duration of illness, without it being possible to show in the cross-section studies which of these influencing factors is responsible for the prevalence difference. On the other hand, incidence, which is independent of duration of illness, can give information about etiologically important differences in risk between populations or within a single population.

2 Prevalence

Results relating to the prevalence of dementias in the elderly population are available from numerous field studies carried out in different parts of the world. Reported total rates for those over 65 years of age show considerable scatter, although the majority of studies show rates varying between 4% and 8%. Provided that the estimates refer to moderate and severe dementias, none of the newer studies exceeds a value of 8%, while the inclusion of milder dementias brings the values to 10% and more.

The comparability of single studies is greatly limited, however, by methodological differences, not least different age distributions in the populations studied. Some authors have therefore attempted to derive more stable estimates from meta-analyses of age-specific prevalence rates. The results are presented in Fig. 1 together with those from the most extensive European study to date (Ott et al. 1995).

Jorm et al. (1987) took into account all the studies which have been published between 1945 and 1985, while the so-called EURODEM Study (Hofman et al. 1991) was restricted to European studies with comparable sample definitions and case criteria. The analysis by Ritchie and Kildea (1995) was based on studies in which the criteria of DSM-III were applied. In spite of these differences in the selections made in the studies, the results show good agreement with one another. A sizable increase in prevalence is found with increasing age. The rates rise more than 20-fold from below 2% among those aged 65–69 years of age to more than 30% in those over 90.

Jorm et al. (1987) concluded from their statistical analysis that the prevalence rises exponentially, with a doubling of the rate in constant age intervals for each 5.1 years. This model of an exponential increase in the age range from 65 to 85 years is in accord with newer data and is congruent with results from both German (Häfner and Löffler 1991) and European (Hofman et al. 1991) studies. However, it is not appropriate to extrapolate to the oldest group, as above 90 years this model very quickly leads to excessive values and predicts a prevalence rate of 100% at 98.5 years of age.

It can be gathered from recently completed studies that prevalence increases markedly from the 85- to 89-year-old age-group to the over-90 group (Ebly et al. 1994; Wernicke and Reichies 1994; Fichter et al. 1995; Helmchen et al. 1996). However, whether a further increase occurs after 90 years or whether the rate approaches an asymptote is controversial. The limited data for the group over 95 leads us to suppose

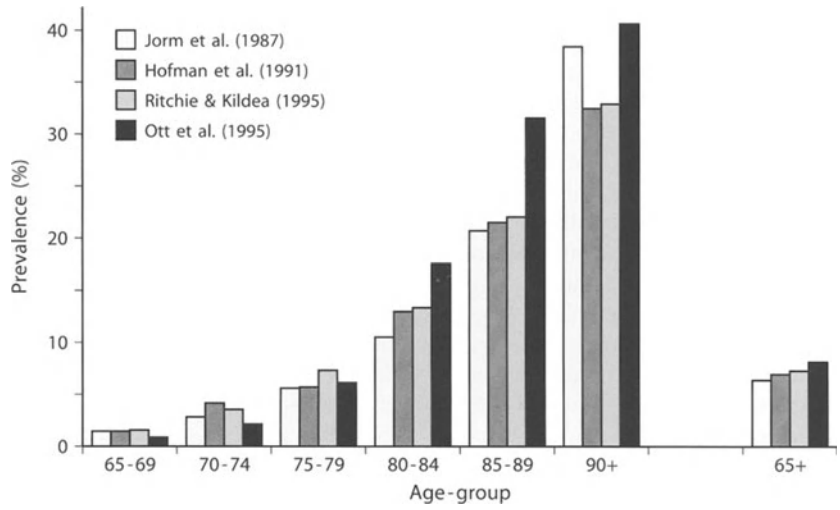


Fig. 1. Age-specific prevalence of dementia. *Asterisk*, standardized overall prevalence rates by age-group for those over 65 in the German population in 1992

a prevalence in the range of 28%–35% for moderate to severe dementias and between 40% to almost 60% for mild to severe dementias. Sobel et al. (1995) reached similar conclusions in the largest study to date on those over 100. In a country-wide study of all inhabitants of Finland who were 100 or older, they determined that 33% suffered from severe dementias and altogether 55.9% at most had mild to severe forms of the illness. These results are compatible with the conclusions from the meta-analysis by Ritchie and Kildea (1995), according to which prevalence grows exponentially over a wide age range, but the increase flattens out over 80 years and no additional increase worth mentioning can be observed over 95 years.

The information about gender-related differences in age-corrected prevalence rates is inconsistent. For the most part, no significant differences were found. However, in the EURODEM Study, men under 75 showed higher rates than women, and those above 75 showed lower rates.

If the age-specific rates are transferred to the present population in Germany (Statistisches Bundesamt 1994), a total prevalence of between 6.4% and 8.1% for those over 65 years is calculated. However, it has to be taken into account that the estimates include milder forms of dementias to a varying extent. The largest proportion of cases in absolute numbers can be expected with the given age configuration among the 80- to 89-year-olds (50%–60%), with proportions of between 25% and 35% among the 65- to 79-year-olds and from 12% to 16% among those over 90 years. In accordance with their overrepresentation in the elderly population, more than 70% of dementias should be found among women.

2.1 Presenile Dementias

Owing to the rarity of dementia among younger people, field studies are usually confined to the older population. For this reason, relatively little is known about the frequency of presenile illnesses. The EURODEM Study estimates prevalence in the 30- to 59-year-old age-group at 0.1%. Similar results are reported from studies based on case registers. The underestimation of the rates is probably low, since early cases of the illness, in contrast to the senile dementias, generally lead to the use of specialist care. Kokmen et al. (1989) report rates of 40–86 per 100,000 individuals in those aged between 45 and 59 years. Mölsä et al. (1982) found rates of 51 per 100,000 among those who are 45–54 years and rates of 144 per 100,000 among those 55–64 years old. According to these results, a prevalence of less than 0.1% to a maximum of 0.2% between the ages of 40 and 64 years can be concluded. People with dementia who are less than 65 years old should therefore comprise no more than 5% of the total number of the patients in western countries, where a high proportion of the population is older.

Newens et al. (1993), in a survey of results on the prevalence of presenile Alzheimer's dementia, report rates of between 18 and 47 cases per 100,000 individuals. In their own, very careful study in North England, they found rates of presenile forms of Alzheimer's dementia which continually rise with age, rising from 2.4 per 100,000 (45–49 years) to 11.8 (50–54 years), 35.6 (55–59 years), and 87.3 (60–64 years). The rates among men and women do not differ from one another.

2.2
Relative Frequency of Forms of Dementia

A rough classification of the illnesses is made in field studies separating an Alzheimer type of dementia, vascular dementia, and other forms of dementia. Little is known about the validity of this assignment, which is largely based on anamnestic information. Without additional technical diagnostic methods, which are of course planned in newer investigations but are often tolerated by only a minority of patients, the certainty of the diagnosis is probably very limited. Moreover, it is increasingly more likely that other brain illnesses, which can be hidden among Alzheimer-like dementias, such as frontal-lobe degeneration or dementia with Lewy bodies, are more frequent than previously assumed.

With these restrictions, Alzheimer's dementia is by far the most common form of dementia according to almost all field studies in Europe and North America. It accounts for between 50% and more than 80% of current cases. The large studies compiled from Canada (Canadian Study of Health and Aging Working Group 1994) and Rotterdam (Ott et al. 1995) conclude that 64% and 72%, respectively, of the illnesses originate from Alzheimer's dementia, followed by vascular dementia, accounting for 19% or 16%, respectively. The remainder of the illnesses, at most 5%–15%, in addition to cases of unknown origin, include a majority of specific forms of dementia, in which the majority are dementias connected to Parkinson's disease or to alcoholism.

To the extent that gender differences are taken into account, a higher portion of Alzheimer's dementia is found among women and a higher portion of vascular dementia among men, as the EURODEM Study showed (Fig. 2). The steep age-related increase in the prevalence rates appears to be caused primarily by Alzheimer dementia. Admittedly, vascular dementias and other forms of dementia are as frequent as Alzheimer dementia below 80 years of age. However, they increase to a lesser extent with increasing age. In contrast, the prevalence of Alzheimer's dementia rises very sharply from one age range to another, so that according to many studies it is responsible for nearly four fifths of dementia among the oldest elderly.

Results from Japan and China differ clearly from the pattern of distribution found in western countries. In these countries, which have a similarly high prevalence of dementias as Europe and North America, the vascular dementias are thought to be far more common than Alzheimer's disease. However, more recent Japanese studies indicate that the geographical difference is smaller than was suspected a few years ago (Ogura et al. 1995; Yoshitake et al. 1995). In addition, a study carried out in Shanghai justifies the assumption that the different results are based in part at least upon different diagnostic criteria (Zhang et al. 1990). Earlier reports, according to which Alzheimer's disease is not found in some African populations (Osuntokun et al. 1991), were unable to be confirmed by a study carried out with standardized methods in Ibadan, Nigeria, and Indianapolis (USA). Admittedly, a significantly higher prevalence rate of Alzheimer's dementia was found among the African-Americans than in the Nigerian sample, although at present it is unknown whether there is a different risk of illness in the two populations, which show a similar ethnic origin, or whether the prevalence differences are due to the different length of survival time of patients in Africa and in the United States (Hendrie et al. 1995).

3
Incidence

Although incidence has a far greater significance for the distribution of the risk of illness than does prevalence, which is dependent upon duration of illness, to date only a small number of prospective field studies have been carried out. According to the

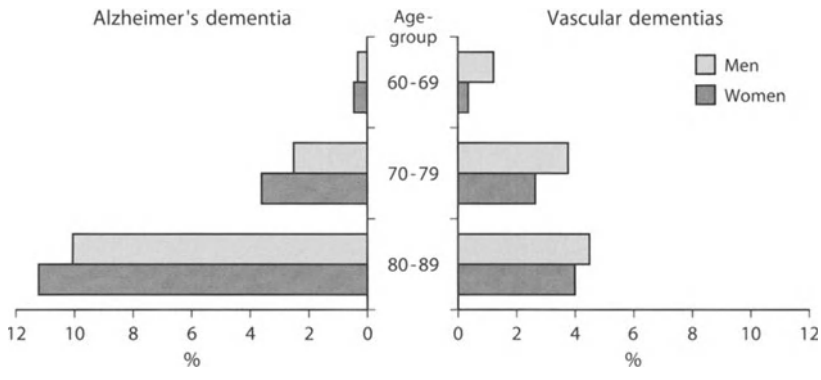


Fig. 2. Age- and gender-specific prevalence of Alzheimer's dementia and vascular dementias according to the results of the EURODEM Study (Rocca et al. 1991a,b)

Table 1. Age-specific incidence of dementia according to results of prospective field studies

Authors	Area of study	Size of random sample (*n*)	Yearly incidence rate (per 1000 individuals)	
			By age-group	Elderly population total
Hagnell et al. (1981)	Lundby, Sweden (1947–1957)	655	60–69: 5.1 70–79: 18.8 80+: 57.3	60+: 16.3
	Lundby, Sweden (1957–1972)	696	60–69: 2.9 70–79: 14.8 80+: 33.7	60+: 10.7
Aronson et al. (1991)	New York, USA	488	75–79: 13.0 80–84: 35.0 85+: 60.0	75+: 34.0
Copeland et al. (1992)	Liverpool, UK	1070	65–74: 3.8 75–84: 11.8 85+: 28.7	65+: 9.2
Bachman et al. (1993)	Framingham, USA	2117	65–69: 1.4 70–74: 5.3 75–79: 10.3 80–84: 16.2 85+: 23.6	No details given
Boothby et al. (1994)	Gospel Oak, London, UK	813	65–69: 8.0 70–74: 6.0 75–79: 37.0 80+: 39.0	65+: 20.0
Bickel and Cooper (1994)	Mannheim, Germany	489	65–69: 4.7 70–79: 12.2 80+: 39.6	65+: 15.4
Letenneur et al. (1994)	Gironde and Dordogne, France	2792	65–69: 2.2 70–74: 6.8 75–79: 17.1 80–84: 31.9 85–89: 42.9 90+: 73.8	65+: 16.3
Paykel et al. (1994)	Cambridge, UK	1195	75–79: 23.0 80–84: 46.0 85–89: 85.0 90+: 82.0	75+: 43.0
Yoshitake et al. (1995)	Hisayama, Japan	828	–	65+ Men: 19.3 Women: 20.9
Hebert et al. (1995)	Boston, USA	2313	65–69: 6.0 70–74: 10.0 75–79: 20.0 80–84: 33.0 85+: 84.0	No details given
Clarke et al. (1996)	Nottingham, UK	1042	65–69: 7.2 70–74: 13.2 75–79: 16.3 80–84: 34.6 85+: 21.7	65+: 15.8

results cited in Table 1, within 1 year, 1%–2% of people over 65 develop a dementia illness. Admittedly, no meta-analysis of age-specific incidence rates was carried out, but a steep, almost exponentially increased risk of new illness with increasing age is indicated. On average, the 12 investigations show that the number of new illnesses each year rises from 3 per 1000 individuals (65–69 years) to 7 per 1000 (70–74 years), 17 per 1000 (75–79 years), 33 per 1000 (80–84 years), up to about 50 per 1000 (>85 years).

Men and women are affected approximately equally; the gender-related differences in risk of illness are slight and inconsistent. As with existing cases, Alzheimer's dementia also prevails among the new illnesses, accounting for 42% in the Japanese study and between 55% and 75% in studies from western countries. The sizeable increase with age for incidence rates seems to be caused primarily by the steep rise in Alzheimer's dementia, which in the investigation by Letenneur et al. (1994), for example, was responsible for only one third of the first illnesses occurring in those aged 65–74 years, but for two thirds in those aged 75–84 years and for 96% of new cases in people over 85 years of age.

It is still a matter of speculation whether the increase in the incidence rates continues in the highest age-group, whether a plateau is reached, or even whether the rates reverse among the long-lived. This question is difficult to answer, since the estimated values show very broad confidence intervals due to the low numbers of people in the high age-groups. However, none of the studies support a reduction in the risk of disease at an age of more than 85–90 years. Rather, the indications are increasing that the incidence among the very elderly is very much higher than was previously assumed. While the studies showed in Table 1 report rates which deviate sizably from one another among the high age-groups, the investigations which have dealt with the very elderly show similarly high values. Thus Gussekloo et al. (1995), in the Netherlands, determined a 1-year incidence rate of 6.9% for those over 85 years of age; in Sweden, Aevarsson and Skoog (1996) established a rate of 9% for the same age-group; and Johansson and Zarit (1995) found rates of 9.9% and 7.7% in two surveys with 84- to 90-year-olds. Studies from Munich and Mannheim came to identical conclusions. According to varying diagnostic criteria, Fichter et al. (1996) estimated the yearly rate of new illness among those over 85 years of age at 7.2%–14.4%. Bickel (1996) calculated a rate of 7.5% for 85- to 89-year-olds and a rate of 11% for those over 90 years.

3.1 Presenile Dementias

In middle age, the risk of illness is slight, as expected for all forms of dementia. Mölsä et al. (1982) report a yearly rate of 10.2 new illnesses per 100,000 individuals in the age-group from 45 to 54 years and a rate of 27 per 100,000 in those aged 55–64 years. In the study by Kokmen et al. (1993), which looks at 5-year intervals from 1960 to 1984, the corresponding rates amount to an average of 10.7 and 46.9 first illnesses per 100,000 individuals.

The incidence of presenile Alzheimer's dementia is 6.4 or 2.6 cases per 100,000 in the 45- to 54-year-old age-group and 16.5 or 24.7 cases in the 55- to 64-year-old age-group in the two studies, respectively. Newens et al. (1993) report a rate of 3.4 per 100,000 between 40 and 60 years of age, with a continuous increase with age from 0 (40–44 years) to 0.9 (45–49 years), 4.9 (50–54 years), 8.1 (55–59 years), and 14.5 per 100,000 (60–64 years). Discontinuities among the age-specific incidence, which could be a clue to varying disease processes in presenile and senile forms of Alzheimer's dementia, were not observed in the present investigations.

3.2 Geographical and Time-Dependent Risk Differences

Geographical differences in the incidence of dementias can hardly be assessed in the absence of comparable data. Transnational studies with standardized methods, which could provide information about risk differences, have only recently been initiated. Without exception, presently available incidence data come from industrialized countries. They scatter just as widely within individual countries as between countries and provide no basis for the assumption that marked differences exist between European countries and countries with populations of predominantly European ancestry.

Convincing evidence for changes with regard to time is also lacking. Of course, the number of dementia patients has certainly increased in recent decades, but this increase can be explained by higher life expectancy and the sharp increase in the number of older people. The two single epidemiological studies which have been concerned with time-related differences in incidence come to opposite conclusions. In the Lundby study, in which a small region in southern Sweden was studied by the same research group over a period of 25 years, a lower, although not significantly decreased, incidence was found in the period from 1957 to 1972 than in the preceding period from 1947 to 1957 (Hagnell et al. 1981). If a downward trend did actually exist, it does not appear to have continued in subsequent years, for recent data from other studies do not point to an incidence that is still decreasing. On the contrary, Kokmen et al. (1993) in Rochester (USA) established a significant increase in the incidence rates in their register study spanning the years 1965–1984. However, this increase was only observed in those aged over 85 years, while the rates remained constant for the younger age-groups. Since the Rochester study is based upon cases which were known to the medical facilities, it remains doubtful whether there is a real increase or only an

apparent one caused by increasing use of medical assistance.

3.3 Life-Span Risk

A question that has received little attention despite its great relevance is that of the risk to the individual of becoming ill with a dementia over the course of his or her lifetime. This question has both a practical and a theoretical aspect. Practically significant is the question of the likelihood that a dementia will actually appear over the course of one's life, i.e., how high the individual risk is based on current life expectancies, or the size of the portion of the population which develops dementia. Theoretically significant is the question of the likelihood that a dementia will appear at a given age if death does not occur prematurely. In other words, will everyone become ill with dementia if he or she lives to becomes old enough, or is only a portion of the population predisposed?

Only a small amount of data is available relating to both questions. In a U.S. retrospective study of a representative random sample of deaths of people aged over 65, based upon information from surviving relatives, it was determined that 10%–15% of those who died aged between 65 and 72 years, 18%–25% of those who died aged between 75 and 84 years, and 32%–40% of those who were over 85 years of age at death had suffered severe cognitive impairments in their last year of life (Lentzner et al. 1992). A similarly designed study in Mannheim, which did not only ascertain cognitive disorders, but also used case criteria for dementia illnesses, came to comparable conclusions (Bickel 1996). According to this study, the number of dementia patients among the deceased rose from 4.5% in those who died between 65 and 69 years to 10% (70–74 years), 25% (75–79 years), 36% (80–84 years), and nearly 50% (>85 years). Altogether, about 30% of all deceased older individuals had suffered from a moderate to severe dementia during their lifetime. If age-specific prevalence rates and the higher rate of mortality of dementia patients are taken into consideration, this result appears plausible, indicating that nearly one third of those who now live beyond 65 years will become ill with dementia in the further course of aging.

Whereas the proportion of those affected among the deceased is in principle accessible to direct observation, the likelihood of becoming ill with dementia at a particular age can only be estimated indirectly from the age-specific incidence rates. The assumption must then be made that those who die before having reached the age of interest without having developed the illness would in the further course of aging have had the same risk as survivors from whom the incidence rates were derived. From this assumption, statements can be made about what portion of the population would become ill with dementia by a particular age if all were to survive until this age or until the onset of dementia, depending upon which occurs first.

The most sophisticated calculations of age-related risk of morbidity come from the Lundby study (Hagnell et al. 1981). According to the results, in the study period from 1947 to 1957, the cumulative risk for a moderate to severe dementia by age 89 amounted to 53.6% for men and 58.4% for women. In the study period from 1957 to 1972, the cumulative risk for both genders by age 89 was estimated to be 40%. Divided according to the two most important diagnostic groups, there turned out to be a risk for all grades of severity of Alzheimer's dementia of 25.5% for men and 31.9% for women and a risk for vascular dementia of 29.8% and 25.1%, respectively (Hagnell et al. 1992).

However, the Lundby data omit the very elderly. A retrospective study carried out in Mannheim (Bickel 1996) permits estimates for this age-group to be made, according to which the risk amounts to only 2% by age 70, rising to 6% by age 75, 12% by age 80, and 36% by age 85. At 89 years of age, the cumulative risk is in agreement with the Lundby study (50%), whereby the risk for Alzheimer's dementia is 28% and the risk for a vascular dementia is 22%. However, the probability of illness already reaches a value of 75% by 95 years of age and almost 90% by 100. It cannot be concluded from this steep increase that age-independent disease processes are responsible for the origin of dementias, but it seems a likely supposition that almost everyone will develop dementia if he or she reaches a very advanced age.

In contrast to this, the cumulative risk up to age 65 appears to be very low. If an estimation is based on the published incidence rates, the risk of becoming ill with presenile dementia should lie at clearly less than 1% for all forms taken together and at less than 0.3% for an Alzheimer's dementia.

4 Duration of Illness and Mortality

Epidemiological and clinical studies agree in showing that the remaining life expectancy of dementia patients is considerably reduced. Van Dijk et al. (1991) examined 90 publications on the mortality of patients with dementia and calculated an average 2-year death rate of 25% (range, 5%–35%) for outpatients, 50% (range, 35%–70%) following admission to a nursing home, and

60% (range, 40%–71%) following admission for treatment as a psychiatric inpatient. In a recent population-based longitudinal study carried out in Mannheim on prevalent cases averaging more than 80 years of age, the 2-year mortality rate was 25% for mild dementia, 45% for moderate dementia, and 55% for severe dementia (Cooper et al. 1996).

The total duration of illness is difficult to determine because of the usually insidious onset of dementia disorders. In some studies supported by information from relatives concerning time of appearance of the first symptoms, average survival times of 4.7–8.1 years for Alzheimer's dementia and 5.2–6.7 years for vascular dementias were found. The duration becomes shorter with increasing age at onset of illness, although all age-groups exhibit large differences among individuals in the remaining life expectancy. If the first symptoms appeared prior to 65 years of age, the duration of illness amounted to an average of 8–10 years. If they began between the ages of 65 and 80, the duration of illness was 5–8 years, and if they began at an age of more than 80, it was 3–5 years. Since at higher ages the general state of health is worse and associated illnesses accumulate, the decreasing survival rates permit no conclusion about a more rapid progression of dementia among the elderly.

If the length of survival is assessed in relation to other time frames more significant for care systems, such as the time of diagnosis or of an inpatient admission to a home or hospital, 6 years is found for illnesses beginning before 65 years of age and about 3 years for those developing later.

Compared with the mortality in the age-matched general population or, more precisely, in age-matched control groups of individuals without dementia, the mortality risk for patients with dementia is two- to fivefold higher. Although the survival time for illnesses arising early is longer than for those developing late, the discrepancy with the usual life expectancy is very high. This is the case above all for younger patients, since dementias developing in the very elderly reduce the remaining life expectancy to a lesser extent, as life expectancy is already short at this age. Nevertheless, a two- to threefold increased risk of mortality was found even with dementia patients over 85 (Aronson et al. 1991; Johannsson and Zarit 1995). In demented patients between 65 and 75, however, relative risks greater than fivefold were reported. The field study by Katzman et al. (1994) found a relative mortality risk of 5.4 for Alzheimer's dementia and 7.2 for other forms of dementia.

Factors associated with survival time include age, gender, and, in particular, severity of dementia. By controlling for other variables, Heyman et al. (1996), in an extensive sample of Alzheimer outpatients, found a 47% higher mortality risk with an increase in age of 5 years, twice the risk for men, and an increased risk of 36% for each additional level of severity when graded on a scale of four levels. The shorter duration of illness for men could explain the previously reported tendency toward higher prevalence rates among women. On the other hand, the decreased length of survival with increasing aging leads us to suppose that the increased prevalence seen in the elderly is not due to the accumulation of long-lasting illnesses, in particular, but that this increase in prevalence is actually lessened by decreasing survival time and that the figures reflect an even steeper increase of incidence.

The life expectancy of patients with dementia has risen in recent decades, but whether this will lead to increasing prevalence rates is still debatable, although there is some circumstantial evidence for such a trend. In several cohorts of hospital inpatients treated in the years 1957–1987, Wood et al. (1991) confirmed an increase in survival of 11.7 months for women and 7.4 months for men, an increase which could not be explained either by age differences or by differences in severity of impairments. Beard et al. (1994), on the basis of a population-based case register for the period between 1960 and 1984, found an 18% decline in the death rate for Alzheimer's patients at the end of each decade. The reasons for longer survival should more likely be sought in the improved treatment of concurrent illnesses rather than in changes in the natural course of the illness.

5
Dementias and Care Needs

5.1
Increasing Need for Care

During the course of dementias, the progressive amnestic and cognitive losses inevitably lead to a great need for help and care. In mild dementia, coping with difficult demands is no longer possible, but the affected individuals are not yet constantly dependent upon others. However, with moderate dementia, provided that memory loss, apraxia, loss of motivation, and disorientation have not yet led to complete dependency, supervision and instruction at least are necessary. Finally, in severe dementia, in which the patient is frequently bedridden, severe behavior disorders and bladder and stool incontinence appear, permanent help is necessary with even the simplest daily tasks. Cooper et al. (1992) confirmed this in a field study, in which 85.3% of moderately to severely demented patients over 65 and 16.7% of mildly demented patients were judged to need permanent care.

5.2 Comparison with Other Illnesses

The relative significance of dementias for the care needs of the elderly in comparison with other illnesses is difficult to estimate in view of the small number of representative studies that take dementias into account appropriately and in view of differences in definition and registration of the need for care. However, it is becoming increasingly clear that a large proportion of those needing care, especially those with the highest care needs, are dementia patients. Cooper et al. (1992), in a study of older general practice patients, established that 67.1% of those needing care suffered from severe dementia and a further 12.8% from mild dementia; only 20.1% of those needing a considerable amount of care were free of serious cognitive disorders. Harrison et al. (1990) reported a 63% prevalence of dementia in an extensive random sample of older people who were being cared for in inpatient institutions or outpatient services. When they differentiated this group according to the severity of dependence upon help from others, the portion of demented individuals rose from 44.5% at the lowest care level to more than 90% at the two highest care levels.

5.3 Dementia Patients in Old-Age and Nursing Homes

The significance of dementias for loss of independence and need for care in old age is supported by studies on residents in nursing homes. As can be seen in Fig. 3, a large proportion of the places in resident old-age facilities are occupied by dementia patients. In old-age homes, the prevalence rates fluctuate between 17% and 36%. In nursing homes, the proportion of demented patients rises to values between 51% and 72%. In addition, close correlations can be seen in all facilities between the individual amount of care required and the severity of the dementia. The greater the helplessness is and the more care services that have to be provided, the more frequently the patient is suffering from a severe grade of dementia (Nygaard et al. 1987). Thus, according to a study in Mannheim, the proportion of nursing home residents who are dementia patients, which at 65% is already high, increases to 80% among those residents requiring intensive care (Bickel 1996). Physical illnesses seem to play an ever smaller role in comparison with the mental handicaps as increased intensity of care is needed.

In many countries, half of all dementia patients are already being cared for in resident care facilities. Where large numbers of places are available, as in Scandinavian countries, the proportion increases to 55%–75% (Fratiglioni et al. 1994: Juva et al. 1993), while in other countries it is at most between 40% and 50% (Robertson et al. 1989; O'Connor et al. 1989). However, there is not much data about the situation in Germany. The ratios determined by Cooper (1984) at the beginning of the 1980s are frequently still assumed to be valid, according to which only 20.9% of dementia patients are cared for in resident old-age homes and nearly 80% are cared for by relatives. However, more recent results lead us to suppose that in Germany, too, about 40% of older people suffering from moderate to severe dementia are now cared for in nursing

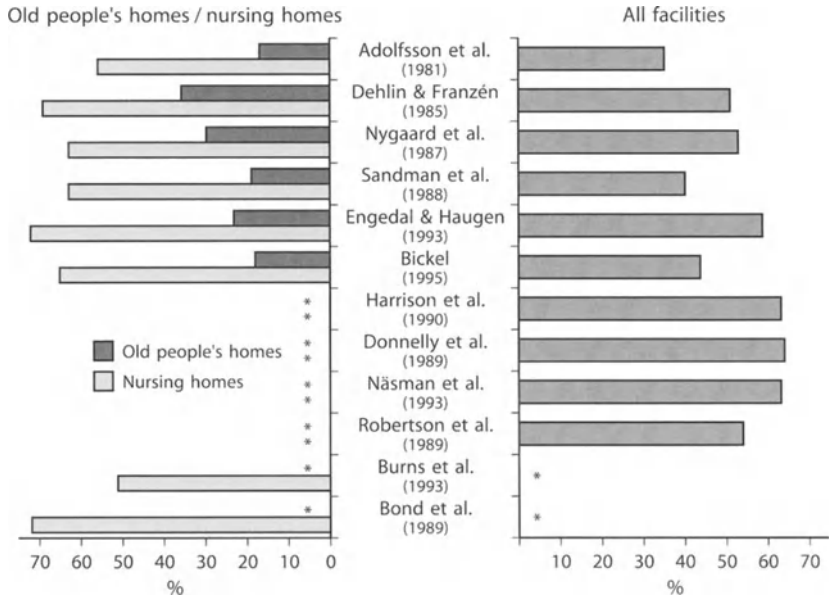


Fig. 3. Prevalence of dementias among residents of inpatient old-age facilities. *Asterisks*, no details given

homes and only 60% in their own home environment (Bickel 1995).

Dementia patients have obviously become the main reason for using nursing homes. In the United States, a dementia illness has been established in two thirds of new admissions (Rovner et al. 1990). In Mannheim, according to information from relatives, 50% of those entering nursing home are admitted because of dementia (Bickel 1995). Sooner or later, most patients are dependent upon residential long-term care. Each year, more than 20% of dementia patients living in private households enter nursing homes (Haupt and Kurz 1993). Longitudinal studies concur that between 65% (Bickel 1995) and 75% (Welch et al. 1992; Severson et al. 1994) move to nursing homes in the course of illness and remain there for an average of 2–3 years. Only one quarter to one third of patients can be cared for in private households until the end of their life. Relatives seem to be overtaxed in particular by home care if patients lose bladder and bowel control and behavioral problems increasingly appear (O'Donnell et al. 1992).

6 Development of Patient Numbers

The increase in chronic illnesses and handicaps combined with the constant growth in the aging population is confronting the health care system with mounting challenges. Since dementias, unlike other groups of illnesses, apparently play a decisive role in care needs in old age and placement requirements in residential care facilities, empirically based forecasts for changes in the numbers of patients are especially important for planning purposes. Jorm et al. (1988) have attempted, on the basis of demographic projections for 29 countries, to predict the development of numbers of dementia patients for the period between 1980 and 2025. They assume that the prevalence in all countries shows the same age dependency and that the age-specific rates remain constant over time. As their model calculations demonstrate, all countries have to prepare themselves for an increase in patient numbers in the wake of demographic changes. The expected increases will certainly be very different; Japan, for example, can count on an extreme increase of dementia patients of about 215%. However, the patient numbers will probably also double or triple in countries such as Australia, New Zealand, Canada, the United States, and some eastern European countries. In Germany and other western European countries, however, a greater proportion of the demographic aging process has already occurred. Future increase here among the elderly will, of course, bring with it more burdens that also need to be taken seriously. However, patient numbers will rise less steeply than in countries in which, at the beginning of the 1980s, a comparatively low proportion of the population was over 65. Thus, in what used to be West Germany, the estimated rise is about 40% by the year 2025. For the former East Germany, a slight decrease was predicted even far into the 1990s, only then to be followed by a continuous rise shortly before the turn of the millennium.

Häfner and Löffler (1991) were able to confirm the predictions for the former West German states on the basis of current data. Supported by a population prediction beginning with the year 1989, a growth of about 10% by the year 2000 can be expected according to their calculations. This is expected to rise to 30% by the year 2010 and to 40% by the year 2020. No changes worth mentioning should occur between 2020 and 2030, but the number of patients should again increase by the year 2040 and then be about 50% above the starting value in 1989.

According to available prevalence estimates (see Fig. 1), 780,000–990,000 dementia patients can be expected among the 12.2 million older people in Germany. This prediction, based on conservative assumptions, corresponds to an increase in patient numbers of about 400,000–500,000. A greater rise could occur with a further gain in life expectancy and a continuation of the tendency for a longer duration of illness. Furthermore, with the decreasing number of offspring, it can be expected that a growing proportion of patients will be dependent upon public care facilities. If lasting success is not achieved in the prevention and treatment of dementias, the foreseeable changes in the configuration of age-groups in Germany will also result in a constantly rising need for care for decades to come. A decrease in patient numbers cannot be anticipated until the year 2040 at the earliest, when the low birth-rate years reach the older age-groups.

7 References

Adolfsson R, Gottfries CG, Nyström L, Winblad B (1981) Prevalence of dementia disorders in institutionalized Swedish old people. The work load imposed by caring for these patients. Acta Psychiatr Scand 63: 225–244

Aevarsson O, Skoog I (1996) A population-based study on the incidence of dementia disorders between 85 and 88 years of age. J Am Geriatr Soc 44: 1455–1460

Aronson MK, Ooi WL, Geva DL, Masur D, Blau A, Frishman W (1991) Dementia. Age-dependent incidence, prevalence, and mortality in the old. Arch Intern Med 151: 989–992

Bachman DL, Wolf PA, Linn RT et al (1993) Incidence of dementia and probable Alzheimer's disease in a general population: the Framingham study. Neurology 43: 515–519

Beard CM, Kokmen E, O'Brien PC, Kurland LT (1994) Are patients with Alzheimer's disease surviving longer in recent years? Neurology 44: 1869–1871

Bickel H (1995) Demenzkranke in Alten- und Pflegeheimen: Gegenwärtige Situation und Entwicklungstendenzen. In: Forschungsinstitut der Friedrich-Ebert-Stiftung (ed) Medizinische und gesellschaftspolitische Herausforderung: Alzheimer-Krankheit. Der langsame Zerfall der Persönlichkeit. Friedrich-Ebert-Stiftung, Bonn, pp 49–68

*Bickel H (1996) Pflegebedürftigkeit im Alter. Ergebnisse einer populationsbezogenen retrospektiven Längsschnittstudie. Gesundheitswesen 58[Suppl 1]: 56–62

Bickel H, Cooper B (1994) Incidence and relative risk of dementia in an urban elderly population: findings of a prospective field study. Psychol Med 24: 179–192

Bond I, Atkinson A, Gregson BA (1989) The prevalence of psychiatric illness among continuing-care patients under the care of departments of geriatric medicine. Int J Geriatr Psychiatry 4: 227–233

Boothby H, Blizard R, Livingston G, Mann AH (1994) The Gospel Oak Study stage III: the incidence of dementia. Psychol Med 24: 89–95

Burns BJ, Wagner HR, Taube JE, Magaziner J, Permutt T, Landerman LR (1993) Mental health service use by the elderly in nursing homes. Am J Public Health 83: 331–337

*Canadian Study of Health and Aging Working Group (1994) Canadian Study of Health and Aging: study methods and prevalence of dementia. Can Med Assoc J 150: 899–913

Clarke D, Morgan K, Lilley J, Arie T, Jones R, Waite J, Prettyman P (1996) Dementia and "borderline dementia" in Britain: 8-year incidence and post-screening outcomes. Psychol Med 26: 829–835

Cooper B (1984) Home and away: the disposition of mentally ill old people in an urban population. Soc Psychiatry 19: 187–196

Cooper B, Bickel H, Schäufele M (1992) Demenzerkrankungen und leichtere kognitive Beeinträchtigungen bei älteren Patienten in der ärztlichen Allgemeinpraxis. Ergebnisse einer Querschnittuntersuchung. Nervenarzt 63: 551–560

Cooper B, Bickel H, Schäufele M (1996) Early development and progression of dementing illness in the elderly: a general-practice based study. Psychol Med 26: 411–419

Copeland JRM, Davidson IA, Dewey ME et al (1992) Alzheimer's disease, other dementias, depression and pseudodementia: prevalence, incidence and three-year outcome in Liverpool. Br J Psychiatry 161: 230–239

Dehlin O, Franzén M (1985) Prevalence of dementia syndromes in persons living in homes for the elderly and in nursing homes in southern Sweden. Scand J Prim Health Care 3: 215–222

Donnelly CM, Compton SA, Devaney N, Kirk S, McGuigan M (1989) The elderly in long-term care. 1. Prevalence of dementia and levels of dependency. Int J Geriatr Psychiatry 4: 299–304

Ebly EM, Parhad IM, Hogan DB, Fung TS (1994) Prevalence and types of dementia in the very old: results from the Canadian Study of Health and Aging. Neurology 44: 1593–1600

Engedal K, Haugen PK (1993) The prevalence of dementia in a sample of elderly Norwegians. Int J Geriatr Psychiatry 8: 565–570

Fichter MM, Meller I, Schröppel H, Steinkirchner R (1995) Dementia and cognitive impairment in the oldest old in the community. Prevalence and comorbidity. Br J Psychiatry 166: 621–629

Fichter MM, Schröppel H, Meller I (1996) Incidence of dementia in a Munich community sample of the oldest old. Eur Arch Psychiatry Clin Neurosci 246: 320–328

Fratiglioni L, Forsell Y, Torres HA, Winblad B (1994) Severity of dementia and institutionalization in the elderly: prevalence data from an urban area in Sweden. Neuroepidemiology 13: 79–88

Gussekloo J, Heeren TJ, Izaks GJ, Ligthart GJ, Rooijmans HGM (1995) A community based study of the incidence of dementia in subjects aged 85 years and over. J Neurol Neurosurg Psychiatry 59: 507–510

Hagnell O, Lanke J, Rorsman B, Öjesjö L (1981) Does the incidence of age psychosis decrease? A prospective, longitudinal study of a complete population investigated during the 25-year period 1947–1972: the Lundby study. Neuropsychobiology 7: 201–211

Hagnell O, Öjesjö L, Rorsman B (1992) Incidence of dementia in the Lundby study. Neuroepidemiology 11[Suppl 1]: 61–66

Harrison R, Savla N, Kafetz K (1990) Dementia, depression and physical disability in a London borough: a survey of elderly people in and out of residential care and implications for future developments. Age Ageing 19: 97–103

Häfner H, Löffler W (1991) Die Entwicklung der Anzahl von Altersdemenzkranken und Pflegebedürftigkeit in den kommenden 50 Jahren – eine demographische Projektion auf der Basis epidemiologischer Daten für die Bundesrepublik Deutschland (alte Bundesländer). Öffentl Gesundheitswesen 53: 681–686

Haupt M, Kurz A (1993) Predictors of nursing home placement in patients with Alzheimer's disease. Int J Geriatr Psychiatry 8: 741–746

Hebert LE, Scherr PA, Beckett LA et al (1995) Age-specific incidence of Alzheimer's disease in a community population. JAMA 273: 1354–1359

*Helmchen H, Linden M, Wernicke T (1996) Psychiatrische Morbidität bei Hochbetagten. Ergebnisse aus der Berliner Altersstudie. Nervenarzt 67: 739–750

*Hendrie HC, Osuntokun BO, Hall KS et al (1995) Prevalence of Alzheimer's disease and dementia in two communities: Nigerian Africans and African Americans. Am J Psychiatry 152: 1485–1492

Heyman A, Peterson B, Fillenbaum G, Pieper C (1996) The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). XIV. Demographic and clinical predictors of survival in patients with Alzheimer's disease. Neurology 46: 656–660

*Hofman A, Rocca WA, Brayne C et al (1991) The prevalence of dementia in Europe: a collaborative study of 1980–1990 findings. Int J Epidemiol 20: 736–748

Johansson B, Zarit SH (1995) Prevalence and incidence of dementia in the oldest old: a longitudinal study of a population-based sample of 84–90-year-olds in Sweden. Int J Geriatr Psychiatry 10: 359–366

Jorm AF, Korten AE, Henderson AS (1987) The prevalence of dementia: a quantitative integration of the literature. Acta Psychiatr Scand 76: 465–479

Jorm AF, Korten AE, Jacomb PA (1988) Projected increases in the number of dementia cases for 29 developed countries:

application of a new method for making projections. Acta Psychiatr Scand 78: 493–500

Juva K, Sulkava R, Erkinjuntti T, Valvanne J, Tilvis R (1993) Prevalence of dementia in the city of Helsinki. Acta Psychiatr Scand 87: 106–110

Katzman R, Hill LR, Yu ESH et al (1994) The malignancy of dementia. Predictors of mortality in clinically diagnosed dementia in a population survey of Shanghai, China. Arch Neurol 51: 1220–1225

Kokmen E, Beard MC, Offord KP, Kurland LT (1989) Prevalence of medically diagnosed dementia in a defined United States population: Rochester, Minnesota, January 1, 1975. Neurology 39: 773–776

Kokmen E, Beard CM, O'Brien PC, Offord KP, Kurland LT (1993) Is the incidence of dementing illness changing? A 25-year time trend study in Rochester, Minnesota (1960–1984). Neurology 43: 1887–1892

Lentzner HR, Pamuk ER, Rhodenhiser EP, Rothenberg R, Powell-Griner E (1992) The quality of life in the year before death. Am J Public Health 82: 1093–1098

Letenneur L, Commenges D, Dartigues JF, Barberger-Gateau P (1994) Incidence of dementia and Alzheimer's disease in elderly community residents of south-western France. Int J Epidemiology 23: 1256–1261

Mölsä PK, Marttila RJ, Rinne UK (1982) Epidemiology of dementia in a Finnish population. Acta Neurol Scand 541–552

Näsman B, Bucht G, Eriksson S, Sandman PO (1993) Behavioral symptoms in the institutionalized elderly – relationship to dementia. Int J Geriatr Psychiatry 8: 843–849

*Newens AJ, Forster DP, Kay DWK, Kirkup W, Bates D, Edwardson J (1993) Clinically diagnosed presenile dementia of the Alzheimer type in the Northern Health Region: ascertainment, prevalence, incidence and survival. Psychol Med 23: 631–644

Nygaard HA, Breivik K, Bakke K, Brudvik E, Moe TJ (1987) Dementia and work load evaluation of the elderly. Compr Gerontol 1: 65–68

O'Connor DW, Pollitt PA, Hyde JB et al (1989) The prevalence of dementia as measured by the Cambridge Mental Disorders of the Elderly Examination. Acta Psychiatr Scand 79: 190–198

O'Donnell BF, Drachman DA, Barnes HJ, Peterson KE, Swearer JM, Lew RA (1992) Incontinence and troublesome behaviors predict institutionalization in dementia. J Geriatr Psychiatry Neurol 5: 45–52

Ogura C, Nakamoto H, Uema T, Yamamoto K, Yonemori T, Yoshimura T (1995) Prevalence of senile dementia in Okinawa, Japan. Int J Epidemiology 24: 373–380

Osuntokun BO, Ogunniyi AO, Lekwauwa GU, Oyediran ABOO (1991) Epidemiology of age-related dementias in the Third World and aetiological clues of Alzheimer's disease. Trop Geogr Med 43: 345–351

*Ott A, Breteler MMB, van Harskamp F, Claus JJ, Van der Cammen T JM, Grobbee DE, Hofman A (1995) Prevalence of Alzheimer's disease and vascular dementia: association with education. The Rotterdam study. Br Med J 310: 970–973

Paykel ES, Brayne C, Huppert FA et al (1994) Incidence of dementia in a population older than 75 years in the United Kingdom. Arch Gen Psychiatry 51: 325–332

Ritchie K, Kildea D (1995) Is senile dementia "age-related" or "ageing-related"? – evidence from meta-analysis of dementia prevalence in the oldest old. Lancet 346: 931–934

Robertson D, Rockwood K, Stolee P (1989) The prevalence of cognitive impairment in an elderly Canadian population. Acta Psychiatr Scand 80: 303–309

Rocca WA, Hofman A, Brayne C et al (1991a) Frequency and distribution of Alzheimer's disease in Europe: a collaborative study of 1980–1990 prevalence findings. Ann Neurol 30: 381–390

Rocca WA, Hofman A, Brayne C et al (1991b) The prevalence of vascular dementia in Europe: facts and fragments from 1980–1990 studies. Ann Neurol 30: 817–824

Rovner BW, German PS, Broadhead J, Morriss RK, Brandt LJ, Blaustein J, Folstein MF (1990) The prevalence and management of dementia and other psychiatric disorders in nursing homes. Int Psychogeriatrics 2: 13–24

Sandman PO, Adolfsson R, Norberg A, Nyström L, Winblad B (1988) Long-term care of the elderly. A descriptive study of 3600 institutionalized patients in the county of Västerbotten, Sweden. Compr Gerontol 2: 120–133

Severson MA, Smith GE, Tangalos EG et al (1994) Patterns and predictors of institutionalization in community-based dementia patients. J Am Geriatr Soc 42: 181–185

Sobel E, Louhija J, Sulkava R et al (1995) Lack of association of apolipoprotein E allele e4 with late-onset Alzheimer's disease among Finnish centenarians. Neurology 45: 903–907

Statistisches Bundesamt (1985) Bevölkerung gestern, heute und morgen. Kohlhammer, Mainz

Statistisches Bundesamt (1994) Statistisches Jahrbuch 1994 für die Bundesrepublik Deutschland. Metzler Poeschel, Wiesbaden

van Dijk PTM, Dippel DWJ, Habbema JDF (1991) Survival of patients with dementia. J Am Geriatr Soc 39: 603–610

Welch HG, Walsh JS, Larson EB (1992) The cost of institutional care in Alzheimer's disease: nursing home and hospital use in a prospective cohort. J Am Geriatr Soc 40: 221–224

Wernicke TF, Reischies FM (1994) Prevalence of dementia in old age: clinical diagnoses in subjects aged 95 years and older. Neurology 44: 250–253

Wood E, Whitfield E, Christie A (1991) Changes in survival in demented hospital inpatients 1957–1987. Int J Geriatr Psychiatry 6: 523–528

**Yoshitake T, Kiyohara Y, Kato I et al (1995) Incidence and risk factors of vascular dementia and Alzheimer's disease in a defined elderly Japanese population: the Hisayama study. Neurology 45: 1161–1168

Zhang M, Katzman R, Salmon D et al (1990) The prevalence of dementia and Alzheimer's disease in Shanghai, China: impact of age, gender, and education. Ann Neurol 27: 428–437

CHAPTER

4

W.E. Müller, H. Förstl

Pharmacological and Nonpharmacological Approaches to the Treatment of Dementia

Translator: E. Taub

1 Introduction

All approaches to the treatment of dementia are intended primarily to improve cognitive deficits and the factors responsible for them, and secondarily to improve the accompanying "noncognitive" manifestations. Both of these types of disease manifestation are amenable to treatment by pharmacological and nonpharmacological means. In this chapter, we shall discuss the medical treatment of dementia and the social and psychological interventions that may be directed against the disease. The use of neuroleptics and antidepressants to treat disturbances accompanying dementia is discussed in Vol. 2, Part 1, Chap. 11.

2 Antidementive Agents

The class of nootropic agents includes a number of medications that have been reported in the international literature to have a beneficial effect on dementia. These agents are used to treat the primary clinical manifestation of dementia, i.e., the impairment of cognitive performance, and the resulting limitation of the patient's ability to carry out everyday activities (Table 1).

The term "nootropic" was coined by Giurgea (1982), who used it as a designation for cognition-enhancing agents (literally, substances directed at the mind, from the Greek *nous*, "mind"). The new term was inspired by the novel pharmacological properties of piracetam, the first nootropic agent. Unlike nonspecific stimulants, such as caffeine (Coper et al. 1987; Giurgea 1982), nootropic agents activate higher, integrative cerebral functions. Coper et al. (1987) suggested several years ago that the term nootropic be used for all substances that clinically improve disturbed adaptive performance in the context of age-related cognitive changes or in the dementia syndrome.

In the current extended definition, the term nootropic now refers to a class of substances that may differ from each other pharmacologically, but share common therapeutic properties. Thus centrally acting calcium antagonists such as nimodipine, glutamate antagonists such as memantine, and acetylcholinesterase inhibitors such as tacrine and donepezil are counted as neuroleptics, according to the current definition, because they are clinically effective in age-related cognitive disturbance or in the dementia syndrome. The term "antidementive agents" covers all of these agents as well and removes the ambiguity of classification brought about by conflicting older and newer uses of the term nootropic.

2.1 Mechanisms of Action

The primary neural mechanism of action of most of the currently available antidementive agents is unknown. The distinction previously drawn between primarily vasoactive and primarily metabolically active substances is no longer supported by more recent experimental data. The improvement of cerebral energy metabolism and (in some cases) of cerebral perfusion brought about by these agents is, in our opinion, not the source of their therapeutic efficacy, but rather a secondary effect and is thus a property of the entire group rather than of individual substances.

According to current concepts, most antidementive agents exert their beneficial effect by improving synaptic neurotransmission and thus improving the disturbed communicative ability of functionally impaired neurons (Müller 1988). Antidementive agents (with the exception of acetylcholinesterase inhibitors) probably do not exert their effect on the specific histopathological abnormalities found in neurodegenerative or vascular dementia, but rather cause a nonspecific improvement in the performance of retained normal structures. Recent studies have led to the conclusion that the normalization of membrane changes plays an important role in this process (Stoll et al. 1996; Müller et al. 1997). The presumed nonspecificity of the target of antidementive agents probably explains why they are just as effective in primary neurodegenerative as in vascular forms of dementia, as is now well documented for a few substances, including nicergoline. Acetylcholinesterase inhibitors, however, have been experimentally tested to date only in the treatment of Alzheimer's disease.

It is still too early for any definite conclusion concerning the extent to which an improvement of the rheological properties of the blood (e.g., by nicergoline, pentoxyphylline, and piracetam) contributes to the therapeutic effectiveness of these agents. Improvement of the rheological properties of the blood through changes in platelet aggregation and erythrocyte deformability probably does play a role in the effectiveness of some antidementive substances in the treatment of brain infarcts (e.g., piracetam; Herrschaft 1993; Orgogozo 1999).

There is evidence for well-defined targets for a few of these substances. Memantine acts as an antagonist at central *N*-methyl-D-aspartate (NMDA)-type glutamate receptors (Müller et al. 1995). Nimodipine is an antagonist of voltage-independent calcium channels (type L), similar to the peripherally active substances verapamil and nifedipine. The original hypothesis that nimodipine protects the central nervous system from excessive amounts of free intracellular calcium

Table 1. The most important antidementive agents currently available for the treatment of senile impairment of cerebral function and of mild to moderately severe dementia syndrome

Substance	Trade name	Approval
Antidementive agents with unexplained preclinical mechanisms of action		
Bencyclane	Fludilat	–
Dihydroergotoxin	Hydergin	+
Cyclandelate	Natil	–
Ginkgo biloba extract	Tebonin forte	+
Meclofenoxate	Helfergin	–
Naftidrofuryl	Dusodril	–
Nicergoline	Sermion	+
Pentoxifylline	Trental	–
Piracetam	Nootrop, Normabrain	+
Pyritinol	Encephabol	+
Xanthinol nicotinate	Complamin	–
Antidementives with defined mechanisms of action		
Donepezil (acetylcholinesterase inhibitor)	Aricept	+
Memantine (glutamate antagonist)	Akatinol	–
Nimodipine (calcium antagonist)	Nimotop	+
Rivastigmine (acetylcholinesterase inhibitor)	Exelon	+
Tacrine (acetylcholinesterase inhibitor)	Cognex	+

The agents marked with a "+" have received formal approval for this indication (Müller 1995) (donepezil, metrifonate, rivastigmine, and tacrine only for Alzheimer's disease). As individual evaluation is difficult, it is recommended that only approved agents be used. A number of older agents have not received formal approval, not because clinical data are lacking, but because the available data could no longer be considered in the approval process as a result of time constraints.

($[Ca^{2+}]_i$) seems to be an oversimplification in the light of recent studies (Müller et al. 1996). Nimodipine probably protects the aging central nervous system less from $[Ca^{2+}]_i$ overload than from $[Ca^{2+}]_i$ oversensitivity.

Tacrine (Davis et al. 1992; Davis and Powchilk 1995; Watkins et al. 1994), donepezil (Rogers et al. 1996), and rivastigmine (Corey-Bloom et al. 1998; Rösler et al. 1999) are inhibitors of the enzyme acetylcholinesterase, which catalyzes the breakdown of the neurotransmitter acetylcholine not only in the brain, but also in peripheral synapses. All of these agents slow the breakdown of acetylcholine and thus contribute to a functional compensation for the loss of cholinergic neurons in the basal nucleus of Meynert, which are involved in the control of cognitive function (Levy et al. 1997). Their therapeutic effect, however, is rather modest; the cholinergic neurons of the basal nucleus are indeed strongly affected in Alzheimer's disease, but many other neuronal and neurotransmitter systems also participate in the neurodegenerative process (Benzi and Moretti 1998).

2.2 Effects in Animal Models

The analysis of the mechanism of action of nootropic agents essentially relies on experiments in animals. Such experiments are used to study either the ability of antidementive medications to stabilize adaptive behaviors or their protective properties against central nervous system injury. Important biochemical and pharmacological properties of the antidementive agents have been documented, albeit not equally well for all agents. Properties concerning the acute improvement of performance (up to 3 months) include the following:

1. Rather nonspecific improvement of disturbed mechanisms of central nervous function
   - Transmitter release
   - Receptor density and functionality
   - Glucose utilization
   - Membrane fluidity
   - Protection against hypoxia
2. Specific substitution of the cholinergic deficit
   - Acetylcholinesterase inhibitors; muscarinic and nicotinic agonists

Properties concerning the slowing of the neurodegenerative process include the following:

1. General mechanisms of excitotoxic and programmed cell death
   - Protection against hypoxia, mitochondrial function
   - Protection against oxidative stress
   - Repair of membrane defects
   - Improvement of neuronal communication

2. Interference with specific pathological processes
   - τ-Hyperphosphorylation
   - Amyloid anabolism, catabolism, and neurotoxicity

The goal of experimentation in animals is to delineate biochemical changes in the central nervous system, effects on physiological regulatory systems, and changes in the animals' behavior. Because therapeutic effects can usually be seen only when there is a prior functional deficit, various experimental models of dementia have had to be devised. The models in use include the physiological changes of normal aging, focal or global brain injuries (hypoxia, ischemia, intoxication, focal structural lesion), and functional impairment through stress.

The models currently used to investigate the mechanism of action of these agents may be classified into five groups (Herrschaft 1992; Schindler 1989; Müller 1988):

1. Biochemical, histological, and neurohistochemical studies of the central nervous system
2. Locomotor and exploratory activity and emotional behavior
3. Coordination
4. Experiments on adaptive behavior, coordination, and learning under 10% hypoxia at normal atmospheric pressure
5. Other cognitive tasks

2.3 Clinical Pharmacological Aspects

The pharmacological property common to practically all antidementive agents is that they improve cognitive functions such as memory, learning, and concentration. This property becomes evident primarily in the context of impairment of these functions of the central nervous system, on either an experimental or a pathophysiological basis (e.g., in dementia). This principal effect of the nootropic agents can be demonstrated not only in animal experiments (Schindler 1989), but also in clinical pharmacological studies on patients with impaired cognitive function (Coper et al. 1987). Furthermore, many antidementive agents bring about a normalization of the slowing in the α-frequency range of the electroencephalogram (EEG) often seen in the elderly, in addition to other signs of increased vigilance (Coper et al. 1987; Maurer et al. 1993).

2.4 Therapeutic Efficacy

Commonly raised objections to the antidementive agents concern the inadequacy of many early experimental studies, the inconsistent demonstrations of efficacy of individual agents, and the frequently small difference between the effect of the drug and that of a placebo. More recent studies reveal that a blanket rejection of the therapeutic efficacy of these agents is no longer tenable (Anonymous 1997; Ihl and Kretschmar 1997; Müller 1995).

Most of the currently available antidementive agents were developed and clinically tested in an era when research methods in clinical gerontopsychiatry were still inadequate. This is one of the reasons why the effectiveness of several agents was inadequately documented and why the clinical experimental findings have often been inconsistent. Now that "impairment of cerebral function," the diagnosis usually used in the past, has been abandoned, the primary target population for the testing of antidementive agents consists largely of patients with primary degenerative dementia of various types, particularly Alzheimer's disease. Only after the efficacy of a drug has been demonstrated by studies on patients of this type are further tests performed to study possible therapeutic effects against other disease entities, such as vascular dementia. As has already been found in other diagnostic categories of psychiatric classification, the rough description of disease entities provided by ICD-9 does not allow diagnoses to be made in a manner adequate for scientific purposes. An operationalized diagnostic system for dementia, such as those of DSM-IV and ICD-10, is considered to be essential.

Target variables used in studies of the efficacy of antidementive agents should cover at least the following three areas:

- Psychopathological level: assessment of disease manifestations by the psychiatrist by means of the applicable clinical scales
- Use of objective psychological testing procedures by the psychologist
- Everyday behavior: assessment by family or nursing personnel

The more concordant the findings on these three levels, the better the evidence for efficacy. Instruments of assessment should be chosen not only on the basis of test-theoretic considerations; practicality must also be taken into account.

It is difficult to say whether the differences in efficacy of agent and placebo that have been demonstrated in studies of antidementive agents are clinically relevant, because there is no accepted operationalization of the concept of clinical relevance with respect to these substances. The first attempts at such an operationalization were made only recently (Kanowski et al. 1990; Herrmann and Kern 1987; Oswald and Oswald 1988).

Table 2. Expected outcome of treatment as a function of severity of cognitive impairment, according to the Global Deterioration Scale (GDS) (after Steinwachs 1996)

GDS stage	Clinical manifestations	Expected outcome of treatment with nootropic agents
3	Decline of occupational performance, difficulty reading longer passages, difficulty remembering new names	Improvement possible (a comprehensive interview is necessary for assessment)
4	Decline of ability to carry out complex tasks, difficulty functioning in strange places	Improvement of the clinical impression, of performance in psychometric tests, and in everyday behavior has been demonstrated for various nootropic agents
5	Patients cannot get by without assistance; they are temporally and spatially disoriented; difficulty in serial subtraction of 4 from 40	Clinically, only a slowing of progression can be expected; the available studies provide initial documentation of this (e.g., piracetam, tacrine, donepezil, Ginkgo biloba extract)
6	May forget spouse's name; difficulty counting backwards from 10; incontinence; needs help functioning in familiar surroundings	No study findings are available for this stage; suitable and recognized tests for assessment of patients in this stage are a topic of current research

The following defining criteria were developed (among others) and are essentially the same as the current American criteria (Small et al. 1997):

- 15%–25% difference between agent and placebo if treatment with the agent results in a therapeutic response in at least 50% of patients
- Statistical concordance of significant differences between the effects of agent and placebo on the different levels of assessment
- Cumulation of therapeutic effects on several assessment levels in individual subjects (i.e., the existence of a responder type) in the group of patients treated with the agent
- More pronounced slowing of the progression of the illness in the agent-treated group in longitudinal studies

Because validated, generally applicable instruments are not available for the assessment of patients with severe dementia, such patients cannot now be included in studies of the efficacy of antidementive agents. These studies, therefore, are typically performed only on patients with mild and moderately severe forms of dementia. These are probably also the patients in whom an effect is most likely to be demonstrated, because the disease process is not yet far advanced. Nonetheless, the methodological restriction remains unsatisfactory, and methods must be developed by which severe forms of dementia may also be included in efficacy testing, so that it can be judged whether an attempt at therapy in this patient group is reasonable. At present, a positive answer to this question can be derived only from a less than fully justifiable extrapolation from the results of studies on patients with mild and moderately severe dementia. In general, as far as the clinical relevance of the antidementive agents is concerned, the goals of treatment should not be set too high. Instead, it should be borne in mind that, in treating dementia, we are treating a chronic, progressive disease. Table 2 shows what may realistically be expected.

In particular, the important question of the extent to which the nootropic agents slow or halt the progressive neurodegenerative process cannot yet be definitively answered. Longitudinal studies (1 year or more) are available only for a few substances (donepezil, *Ginkgo biloba* extract, piracetam, tacrine); these provide preliminary evidence that the disease process may be slowed (Croisile et al. 1993; Knopman 1995; Le Bars et al. 1997; Rogers and Friedhoff 1998). This "slowing," however, must be viewed critically. As mentioned above, the development of the acetylcholinesterase inhibitors was based on the premise that they could provide a compensatory mechanism for a specific neurochemical deficit in patients with Alzheimer's disease that is relevant to their cognitive performance. In comparison with this seemingly promising theoretical concept, the therapeutic results have been rather modest. The reasons for this have been extensively discussed by Benzi and Moretti (1998).

The question remains open whether the newer acetylcholinesterase inhibitors are more effective antidementive agents than the older nootropic substances, such as piracetam and *Ginkgo biloba* extract. No reliable answer can be given at present, because no relevant comparative studies have been performed. Indirect comparisons can be made, however, because a large, placebo-controlled, double-blinded study of the efficacy of one of the older nootropic agents (*Ginkgo biloba* extract) in patients with Alzheimer's disease was recently published (Le Bars et al. 1997). The method of this study was similar to that of studies on the newer antidementive agents (Table 3).

The primary target criterion in these studies was a cognitive parameter, the Alzheimer's Disease

Table 3. Comparison of the efficacy of various antidementive agents in patients with Alzheimer's disease (after Müller 1999)

Agent (study)	Dose, duration	ΔD[a]	Responder rate
Tacrine (Knapp et al. 1994)	60 mg, 30 weeks	2.2	Placebo 25% Agent 40%
Donepezil (Rogers et al. 1998)	10 mg, 24 weeks	2.9	Placebo 27% Agent 54%
Rivastigmine (Corey-Bloom et al. 1998; Rösler et al. 1999)	12 mg, 26 weeks	3.8 1.6	Placebo 16% Agent 24%
Egb 761 (Le Bars et al. 1997; only Alzheimer's disease)	120 mg, 52 weeks	1.7	Placebo 13% Agent 29%
Nimodipine (Morich et al. 1996)	180 mg, 24 weeks	0.5	?

[a]Because the severity of dementia differed among studies and the duration of treatment was not the same, the agent-minus-placebo differences ΔD on the ADAS-cog (intend-to-treat analysis, ITT), a commonly used test of cognition, cannot be directly compared with one another. Responders were defined as patients whose ADAS-cog scores improved by 4 points or more.

Assessment Scale (ADAS)-cog subscale, which is primarily a measure of the patient's global cognitive impairment. When this parameter was used to study the effect of the newer acetylcholinesterase inhibitors, a difference of 2–3 points between the treated group and the placebo group was found after 6 months of treatment. In the recently published *Ginkgo biloba* study, a difference of 1.7 points was obtained after 12 months. The United States Food and Drug Administration (FDA) regards an improvement of 4 points on the ADAS-cog subscale as a definite success of treatment. In the recently published donepezil study (Rogers et al. 1998), an improvement of this magnitude was obtained after 6 months of treatment by 53.5% of the patients receiving donepezil and by 26.8% of those receiving placebo. The studies of tacrine yielded similar findings. In the *Ginkgo biloba* study (Le Bars et al. 1997), the responder rate after 12 months was 22% with *Ginkgo biloba* extract and 10% with placebo treatment. Thus, in studies of both older and newer nootropic agents, there were approximately twice as many responders, according to the FDA criterion, to treatment with the agent as there were to placebo treatment.

Thus, when the urgently needed direct comparative studies are performed, we should not expect them to show any spectacular difference in the efficacy of nootropic agents of the two types, though a small difference may in fact be present. These studies may help answer the considerably more important question of which patients are more likely to respond to acetylcholinesterase inhibitors and which are more likely to respond to one of the other antidementive agents.

Theoretical considerations imply that combination therapy may be reasonable, i.e., a fairly nonspecific background therapy with older nootropic agents, combined with a pharmacologically specific therapy with acetylcholinesterase inhibitors.

2.5 Undesired Effects

The antidementive agents are, generally speaking, very well tolerated, and undesired effects are rare (exceptions include hepatotoxicity of tacrine and side effects of donepezil, rivastigmine, and tacrine that result directly from acetylcholinesterase inhibition). Gastrointestinal disturbances have been observed with practically all agents, however, as have increased irritability and sleep disturbances. Sedating effects (fatigue) have also been reported in rare cases for a few agents. Agents with a direct hypotensive effect (codergocrine mesilate, nicergoline, nimodipine) may lead to disordered circulatory regulation (hypotension, lightheadedness).

2.6 Treatment Process

The process of treatment of vascular and neurodegenerative forms of dementia can be carried out as shown in Fig. 1. The treatment must be adequately monitored, primarily by means of standardized cognitive testing and behavioral assessment. The agent should not be changed before the 12th week of treatment.

2.7 New Pharmacological Treatment Approaches

In recent years, a number of direct muscarinic receptor agonists have been developed as an alternative to the acetylcholinesterase inhibitors. It is hoped that the stimulation of muscarinic receptors in patients with Alzheimer's disease will be a more specific form of

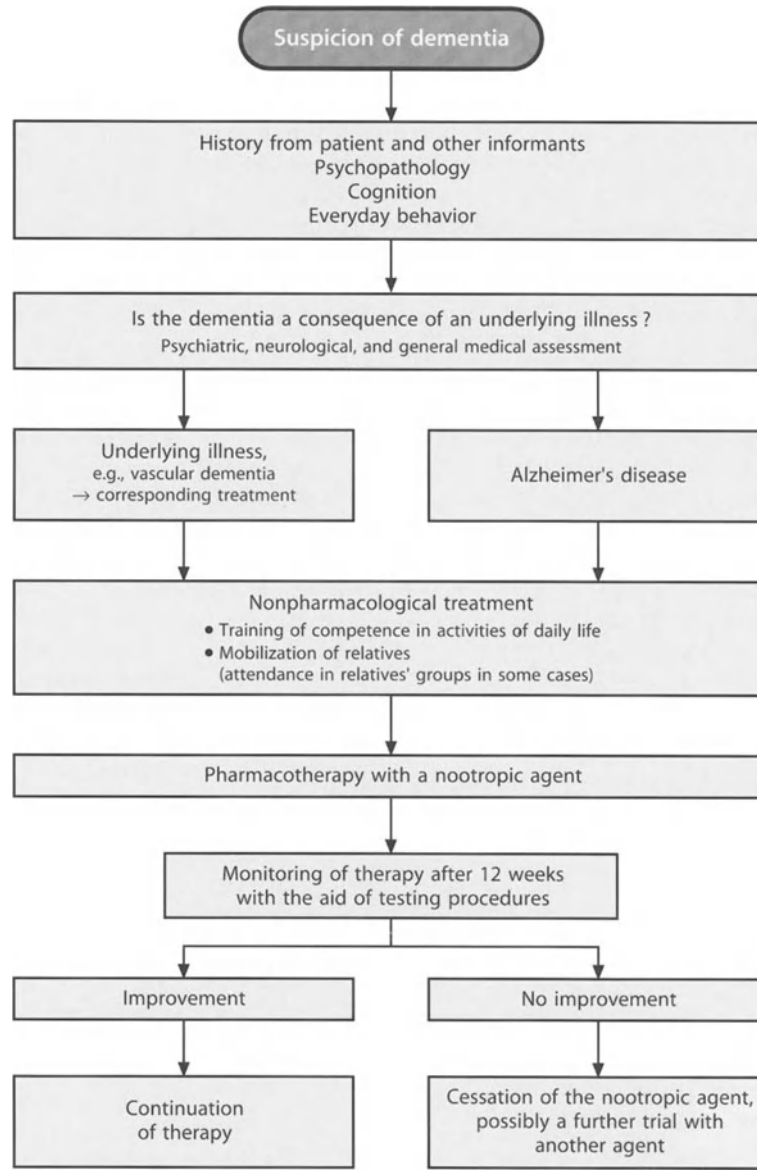


Fig. 1. Therapy for the dementias. (After Anonymous 1997)

treatment with fewer undesired effects. Initial clinical results are available from a controlled, double-blinded study of xanomeline, an M1- and M4-agonist (Bodick et al. 1997). Although a reliable comparison is difficult to make in this case, as in the other cases discussed above, it appears that the therapeutic effect of this agent is somewhat less than that of the acetylcholinesterase inhibitors. Furthermore, it is associated with an unexpectedly high frequency of undesired effects. Cautious interpretation of this study, as of previous studies, implies that cholinergic therapy of Alzheimer's disease with the current treatment strategies, as assessed by current diagnostic techniques, is of relatively limited usefulness. Proceeding from the assumption that β-amyloid is directly involved in the neurodegenerative process of Alzheimer's disease and that a sequence of further changes (including pathological membrane abnormalities, disturbances of intracellular calcium homeostasis, increased production of free radicals, secondary inflammatory processes, and, finally, cell death by exocytosis or apoptosis) is also involved, a number of agents are now undergoing clinical testing that might interfere with this cascade of different mechanisms (Heidrich et al. 1997). Among these substances are, for example, various antiphlogistic agents, the monoamine oxidase (MAO)-B inhibitor selegiline, the NMDA-antagonist memantine, the xanthine derivative propentophylline, and the free-radical scavenger ibendone. The clinical potential of these substances has not yet been definitively determined.

Thus, even when the second-generation drugs are considered, the therapeutic potential of the antidementive (nootropic) agents now available for the treatment of dementia remains modest. In the next few years, we must come to terms with this modest therapeutic potential and attempt to make optimal use of each of these agents. In the long term, we may expect, considering the major recent advances in the understanding of the pathogenesis of Alzheimer's disease, that new substances will be developed that will interfere more specifically with the mechanisms underlying neurodegeneration.

Many of the treatment strategies currently being pursued concern the anabolism, catabolism, and neurotoxicity of β-amyloid peptide, a substance of great importance in the pathophysiology of the disease. All of these developments are currently still in the preclinical phase. Even under the most optimistic projections, it will take several years before the first of these third-generation antidementive agents becomes available for clinical use.

3 Nonpharmacological Treatment Approaches

The pharmacotherapy of the dementias is fraught with side effects and high costs. The desired principal effects are often not obtained. Thus, nonpharmacological treatment approaches are more than merely auxiliary measures. Like symptomatically directed pharmacological treatment, these measures must be used in an individualized fashion based on the concerns of patients and others close to them. Nonetheless, there are standard situations and corresponding standard methods of treatment, even if their efficacy has not yet been adequately studied. Many reports of the use of these measures are of an anecdotal nature, and their efficacy has been assessed rather intuitively. The few systematic studies have been performed predominantly on institutionalized patients with Alzheimer's disease.

3.1 Psychotherapy, Cognitive Training, and Other Supportive Measures

When the diagnostic assessment of dementia is performed, careful attention must be paid to the patient's ability to understand and to his or her need for information. A desired explanation may not be withheld from the patient, nor may information concerning the diagnosed illness and its presumed prognosis be delivered in a perfunctory fashion. A mere act of informing, unaccompanied by an offer to help the patient come to terms with the illness, is not acceptable. Psychotherapeutic approaches, including attempts at cognitive restructuring, may yield a certain benefit in patients in the early phase of a degenerative brain illness who suffer from reactive depressive disturbances (Teri and Gallagher-Johnson 1991). Their efficacy, however, is temporally limited (if the course of illness indeed accords with the suspected diagnosis). Reactive depressive manifestations often improve as the cognitive deficits become more severe.

In mild dementia, cognitive training may lead to a certain improvement of memory and attention. A cognitive education program finely tuned to the specific deficits of a mildly affected patient may yield a small benefit (Commissaris et al. 1996). Nonetheless, a transfer of learned abilities leading to a lasting ability to cope with everyday demands cannot be demonstrated in patients with manifest dementia (Hofmann et al. 1996). Even the concrete training of practical functions (hygiene, dressing, kitchen work) has not yet been shown to be effective. Neuropsychological training methods found to be effective in the rehabilitation of head-injured and stroke patients yield only transient effects, at best, in patients with progressive brain diseases. These include the contextualization of information and visual associative learning (the "imagery method"), i.e., linguistic and visual double-coding of information.

The simplest mnemonic techniques, such as the use of a notebook, are also the most effective (Haupt 1997). On the other hand, so-called "brain jogging" in patients with manifest dementia is unfounded and deserves to be completely rejected. Even though suitable training methods may allow the recruitment of cognitive reserve abilities in normal aged persons, it is precisely these reserve abilities that are no longer present in demented patients. This is the central feature of the dementia syndrome (Zerfass et al. 1997). Frustration with such techniques is therefore inevitable and is much more disconcerting to the patients themselves than to their therapists and families. Reality orientation training (ROT) was developed for the treatment of patients with dementing illnesses and confusional states of other origins. Its aim is the improvement of orientation in space and time (Folsom 1986). The two forms of ROT are group ROT ("classroom ROT"), which involves a daily training program lasting half an hour with information about location, time, and situation, and 24-h ROT, which involves the continuous provision of orienting information. The ROT techniques are not fitted to the individual patient, and their effectiveness has not been satisfactorily demonstrated to date. It does seem that patients are more able to give verbal information about their

orientation after group ROT, and they are better able to locate themselves in the environment when they undergo 24-h ROT. This may be interpreted as a reinforcement and channeling of certain patterns of behavior and does not necessarily imply an improvement of cognitive functioning on a deeper level. The social contact, increased activity level, and improved mood associated with the performance of all of these techniques are important nonspecific factors that may make a major contribution to their effectiveness. It is worth noting that ROT has a positive effect on the caring persons' knowledge about the patients and their attitude to them.

Reminiscence therapy is intended to reinforce retained abilities through recollection of experiences and events that took place in the remote past. This approach thus emphasizes the patients' strengths and does not confront them with their weaknesses or subject them to an unnatural training setting. A prerequisite for this systematic occupation with one's own memories is a retained ability to communicate verbally. Aids to reminiscence used in treatment include pieces of music, photographs, newspaper clippings, and books (Norris 1986).

A variant of reminiscence therapy is self-preservation therapy (*Selbsterhaltungstherapie*), in which – with the participation of family members – the patient's personal identity is reinforced by the stimulation of autobiographical memories (Romero and Eder 1992). Validation therapy is based on a far-reaching recognition of the patient's "valid" point of view, which is supposed to be understood and accepted (Feil 1990).

3.2 Dealing with Patients and Structuring Their Environment

Because of their cognitive impairment, demented patients are scarcely able to learn and to adapt to their environment. These functions thus become largely the task of the caring partner or nursing staff. The care of the patient must be individually fitted to the patient's personality and current problems. Little can be said in general terms, and these few generalizations often sound trivial. For example, it is of little or no use to argue with or contradict patients, to lecture them, or to provoke a confrontation with them. The style of communication must be adapted to the patient's limited ability to understand and must thus be simple. Because patients can often receive nonverbal stimuli quite well even in advanced stages of disease, it is often helpful to combine verbal with gestural and mimetic signals. The most important factor is positive feedback, i.e., the acceptance and reinforcement of correct behaviors. Wrong responses can occasionally be steered back in the proper direction, or attention can be drawn away from them. When this no longer seems possible, a change of subject or a brief time-out may allow the conflict to be defused and forgotten.

The simplest behavioral-therapeutic techniques of positive and negative reinforcement prove effective as long as the patient's self-esteem is preserved by a basically positive attitude. It may be particularly important to discover the patient's retained abilities and needs (e.g., music, movement, social sense) and then rationally apply whatever competence is present to various tasks (e.g., setting the table, caring for pets, gardening).

Certain maladaptive and regressive modes of behavior require specific measures and should occasion a search for possible underlying factors. Patients often react aggressively when their intimate sphere is violated or when they feel threatened. Such reactions may also be fostered by delusions and hallucinations and may be treatable with neuroleptics, among other agents. Continuous shouting may be provoked by the pain of a treatable somatic illness or by an environment that provides the patient too little stimulation.

The patient's environment should be familiar and stimulating, but not alarming. Constancy of the environment is best maintained by people well known to the patient – close relatives, if possible – through the use of familiar clothing, pictures, and furniture, maintenance of the patient's dietary habits, and a regular program of daily activities. Holiday travels, transfers back and forth between the patient's home, the hospital, and the nursing home, renovation and overhaul of the living space, continuous change of personnel, and lack of people in a close relationship can all contribute to a deterioration of the patient's condition.

On the other hand, the patient's usual environment should not be too boring, but should include sources of stimulation. Sensory deprivation and social isolation should be avoided just as much as noise and commotion. Clear aids to orientation, pleasant colors and good lighting, visual and hearing aids, and pleasant music in line with the patient's taste are all beneficial. To be avoided is an uncontrolled exposure to stimuli that are easily misinterpreted, such as overhead announcements, frightening pictures, unsettling shadows, large mirrors, and constant bombardment with radio and television. Personal attention is irreplaceable, whether delivered in a fashion that places heavy demands on personnel – such as music therapy, art therapy, or ergotherapy – or simply by touching the patient. The strong need of many patients, even those with advanced dementia, for intimacy is often overlooked.

Patients should be allowed the freedom of movement they desire to the greatest possible extent. Limits should be set only when the safety of the patients

themselves, their partner, or those rendering care is threatened. It may suffice for a time simply to inform neighbors about the patient's condition, so that the patient can be taken home if required, or to make sure that bright clothing is worn, to lessen somewhat the danger of the patient's being struck by a passing vehicle. In later stages of illness, the possibility of injury must be dealt with even in the most narrowly confined environment.

3.3 Care of the Patient's Relatives

Relatives caring for the patient are subject to great mental stress, which – particularly in elderly women with elderly, demented husbands – leads to high morbidity (Rankin et al. 1992; Schultz et al. 1990). The attenuation of the caring partner's social network by the partner's own advanced age, as well as by the illness of the patient, is a major risk factor for mental illness in the caring partner, which, in turn, leads to the collapse of home care for the patient (Cooper et al. 1984; Gilhooly 1986; Zarit et al. 1986). The degree of mental stress is not closely related to the severity of dementia; rather, it is better accounted for by a number of concrete stress factors, such as the number and severity of nursing problems (e.g., incontinence), behavioral abnormalities (especially aggressive behavior), lack of peace and quiet at night, and lack of normal communication (Förstl and Geiger-Kabisch 1995; Hettiaratchy and Manthorpe 1993).

The possible forms of care of the patient's relatives are as different from each other as the relatives' expectations and needs. Care may take place during office visits, in extensive individual discussions with relatives, in a patient-oriented, long-term involvement, in a directed session when a patient is admitted to hospital, or in professionally conducted group therapy or a self-help group.

Jacques (1984) characterized three types of caring relatives with regard to their degree of engagement and their attitude:

- Relatives who gladly participate in care and require instruction and encouragement
- Relatives who feel a duty to sacrifice themselves to an unreasonable extent while caring for the patient
- Relatives who care for the patient against their own desires

There are significant differences between the attitudes of male and female caring relatives, as was shown in a study of groups of relatives of demented patients (Förstl and Geiger-Kabisch 1995). While wives and daughters mainly seek mutual emotional support and, by exchange with the group, allay their fears and learn to accept more easily what the future holds, men are more eager to obtain information about the causes of dementia and the possibilities of treating it and require concrete advice on how to deal with the patient.

Caring relatives' different degrees of engagement, and their different levels of "emotional" and "instrumental" interest, call for a large measure of flexibility in the kinds of help that are offered to them. The two goals of the care of the patients' relatives, i.e., support for the relatives themselves and improvement of their care for the patient, may be attained in various ways. An important feature of any such approach is a raising of the relatives' tolerance of the burdens placed on them by the patients' disturbing behavior; this will help lessen their frustration. These aims may be served not only by the provision of information, but also by acceptance into a group of patients' relatives who are undergoing similar experiences.

4 References

**Anonymous (1997) Empfehlungen zur Therapie der Demenz. AVP-Sonderheft Therapieempfehlungen 4: 2–12

**Benzi G, Moretti A (1998) Is there a rationale for the use of acetylcholinesterase inhibitors in the therapy of Alzheimer's disease. Eur J Pharmacol 346: 1–13

Bodick NC, Offen WW, Levey AL et al (1997) Effects of xanomelin, a selective muscarinic receptor agonist, on cognitive function and behavioral symptoms in Alzheimer disease. Arch Neruol 54: 465–473

Commissaris K, Verhey FRJ, Jolles J (1996) A controlled study into the effects of psychoeducation for patients with cognitive disturbances. J Neuropsychiatry 8: 429–435

Cooper B, Mahnkopf B, Bickel H (1984) Psychische Erkrankung und soziale Isolation bei älteren Heimbewohnern: eine Vergleichsstudie. Z Gerontol 17: 117–125

Coper H, Herrmann WM, Woite A (1987) Psychostimulantien, Analeptika, Nootropika. Dtsch Arztebl 84: 337–342

Corey-Bloom J, Anand R, Veach J (1998) A randomized trial evaluating the efficacy and safety of ENA 713 (rivastigmine tartrate), a new acetylcholinesterase inhibitor, in patients with mild to moderately severe Alzheimer's disease. Int J Geriatr Psychopharmacol 1: 55–65

Croisile B, Trillet M, Fondarai J, Laurent B, Mauguière F, Billardon M (1993) Long-term and high-dose piracetam treatment of Alzheimer's disease. Neurology 43:301–305

Davis KL, Powchilk P (1995) Tacrine. Lancet 345: 625–630

Davis KL, Thal L, Gamzu E et al (1992): A double-blind, placebo-controlled multicenter study of tacrine for Alzheimer's disease. N Engl J Med 327: 1253–1259

Feil N (1990) Validation. Ein neuer Weg zum Verständnis alter Menschen. Delle Karth, Vienna

Folsom J (1986) Reality orientation for the elderly mental patient. J Geriatr Psychiatr 1: 291–307

Förstl H, Geiger-Kabisch C (1995) "Alzheimer-Angehörigengruppe" – eine systematische Erhebung von Bedürf-

nissen und Erfahrungen pflegender Angehöriger. Psychiatr Prax 22: 68–71

Gilhooly MLM (1986) Senile dementia: factors associated with caregivers preference for institutional care. Br J Med Psychol 59: 165–171

Giurgea CE (1982) The nootropic concept and its prospective implications. Drug Dev Res 2: 441–446

Haupt M (1997) Psychotherapeutische Strategien bei kognitiven Störungen. In: Förstl H (ed) Lehrbuch der Gerontopsychiatrie. Enke, Stuttgart, pp 210–219

Heidrich A, Rösler M, Riederer P (1997) Pharmakotherapie der Alzheimer Demenz: Therapie kognitiver Symptome – neue Studienresultate. Fortschr Neurol Psychiatr 65: 108–121

Herrmann WM, Kern U (1987) Nootropika: Wirkungen und Wirksamkeit. Eine Überlegung am Beispiel einer Phase-III-Prüfung mit Piracetam. Nervenarzt 58: 358–364

Herrschaft H (1992) Klinische Bewertung der Wirksamkeit von Nootropika. In: Riederer P, Laux G, Pöldinger W (eds) Neuro-Psychopharmaka, vol 5. Springer, Berlin Heidelberg New York, pp 163–166

Herrschaft H (1993) Nootropika in der Behandlung des Hirninfarktes und bei der vaskulären Demenz. Psycho 19: 280–284

Hettiaratchy P, Manthorpe J (1993) A carer's group for families and patients with dementia. In: Jones G, Miesen BML (eds) Care-giving in dementia. Routledge, London, pp 419–434

Hofmann M, Hock C, Kühler A, Müller-Spahn F (1996) Interactive computer-based cognitive training in patients with Alzheimer's disease. J Psychiatr Res 30: 493–501

**Ihl R, Kretschmar C (1997) Zur Nootropikabewertung für die Praxis. Nervenarzt 68: 853–861

Jacques A (1984) Coping with the care of ambulant dementing older people: key issues for carers. In: Förstl H (ed) The slow death of intellect. Age Concern Scotland, Edinburgh

Kanowski S, Ladurner G, Maurer K et al (1990) Empfehlungen zur Evaluierung der Wirksamkeit von Nootropika. Z Gerontopsychol Gerontopsychiatr 3: 67–79

*Knapp MJ, Knopman DS, Solomon PR, Pendlebury WW, Davis CS, Gracon SI for the Tacrine Study Group (1994) A 30-week randomized controlled trial of high-dose tacrine in patients with Alzheimer's disease. JAMA 271: 985–991

Knopman D (1995) Tacrine in Alzheimer's disease: a promising first step. Neurologist 1: 86–94

*Le Bars PL, Katz MK, Berman N, Itil TM, Freedman AM, Schatzberg AF (1997) A placebo-controlled, double-blind, randomized trial of an extract of ginkgo biloba for dementia. JAMA 278: 1327–1332

Levy R, Förstl R, Müller WE (1997) Neurotransmitter-Substitution. In: Förstl H (ed) Lehrbuch der Gerontopsychiatrie. Enke, Stuttgart, pp 152–162

Maurer K, Ihl R, Frölich L (1993) Alzheimer. Grundlagen, Diagnostik, Therapie. Springer, Berlin Heidelberg New York

Morich FJ, Bieber F, Lewis JM et al (1996) Nimodipine in the treatment of probable Alzheimer's disease. Clin Drug Invest 11: 185–195

Müller WE (1988) Nootropika, die Therapie der Demenz zwischen Anspruch und Wirklichkeit. Münch Med Wochenschr 130: 575–579

Müller WE (1995) Therapie mit Nootropika, Möglichkeiten und Grenzen. Psycho 12: 742–751

Müller WE, Mutschler E, Riederer P (1995) Noncompetitive NMDA receptor antagonists with fast open-channel blocking kinetics and strong voltage-dependency as potential therapeutic agents for Alzheimer's dementia. Pharmacopsychiatry 28: 113–124

Müller WE, Eckert A, Hartmann H et al (1996) Zur Kalziumhypothese der Hirnalterung. Nervenarzt 67: 15–24

Müller WE, Koch S, Scheuer K et al (1997) Effects of piracetam on membrane fluidity in the aged mouse, rat and human brain. Biochem Pharmacol 53: 135–140

*Müller WE (1999) Nootropika ohne Azetylcholinesterase-Hemmer. Präklinische und klinische Bewertung. In: Förstl H (ed) Alzheimer-Demenz, Grundlagen, Klinik und Therapie. Springer, Berlin Heidelberg New York

Norris A (1986) Selbst-Erhaltungs-Therapie (SET): Konzept einer neuropsychologischen Therapie bei Alzheimerkranken. Z Gerontopsychol Gerontopsychiatr 5: 267–563

Orgogozo JM (1999) Piracetam in the treatment of acute stroke. Pharmacopsychiatry 32(Suppl 1): 25–32

Oswald WD, Oswald B (1988) Zur Replikation von Behandlungseffekten bei Patienten mit hirnorganischen Psychosyndromen im Multicenter-Modell als Indikation für klinische Wirksamkeit. Eine placebokontrollierte Doppelblind-Studie mit Pyritinol. Z Gerontopsychol Gerontopsychiatr 1: 223–241

Rankin ED, Haut MW, Keefover RW (1992) Clinical assessment of family caregivers in dementia. Gerontologist 32: 813–821

Rogers S, Friedhoff L (1998) Long-term efficacy and safety of donepezil in the treatment of Alzheimer's disease: an interim analysis of the results of a US multicentre open label extension study. Eur Neuropsychopharmacol 8: 67–75

*Rogers SL, Friedhoff LT (1996) The efficacy and safety of donepezil in patients with Alzheimers's disease: results of a US multicentre, randomized, double-blind, placebo-controlled trial. Dementia 7: 293–303

Rogers SL, Farlow MR, Doody RS, Mohs R, Friedhoff LT (1998) A 24-week, double-blind, placebo-controlled trial of donepezil in patients with Alzheimer's disease. Neurology 50: 135–145

Romero B, Eder G (1992) Selbst-Erhaltungs-Therapie-(SET-) Konzept einer neuropsychologischen Therapie bei Alzheimer-Kranken. Z Gerontopsychol Psychiatr 5: 267–282

*Rösler M, Anand R, Cicin-Sain A et al (1999) Efficacy and safety of rivatigmine in patients with Alzheimer's disease: international randomised controlled trial. Br Med J 318: 633–638

Schindler U (1989) Pre-clinical evaluation of cognition en hancing drugs. Prog Neuropsychopharmacol Biol Psychiatry 13: 99–115

Schultz R, Visintainer P, Williamson GM (1990) Psychiatric and physical morbidity effects of caregiving. J Gerontol Psychol Sci 45: 181–191

*Small WG, Rabins PV, Barry PP et al (1997) Diagnosis and treatment of Alzheimer disease and related disorders. JAMA 278: 1363–1371

Steinwachs KC (1996) Zum Therapieerwartungshorizont einer Nootropika-Behandlung bei primär degenerativer Demenz im Alter. Nervenheilkunde 15: 80–84

Stoll S, Scheuer K, Pohl O, Müller WE (1996) Ginkgo biloba Extract (Egb 761) independently improves changes in passive avoidance learning and brain membrane fluidity in the aging mouse. Pharmacopsychiatry 29: 144–149

*Teri L, Gallagher-Thompson D (1991) Cognitive-behavioural interventions for treatment of depression in Alzheimer patients. Gerontologist 31: 413–416

Watkins PB, Zimmermann HJ, Knapp MJ et al (1994) Hepatotoxic effects of tacrine administration in patients with Alzheimer's disease. JAMA 271: 992–998

Winblad B, Poritis N (1999) Memantine in severe dementia. Int J Geriartr Psychiatry 14: 135–146

Zarit S, Todd PA, Zarit M (1986) Subjective burden of husbands and wives as caregivers: a longitudinal study. Gerontologist 26: 260–266

Zerfass R, Daniel S, Förstl H (1997) Grundzüge des diagnostischen Vorgehens bei Demenzverdacht. In: Förstl H (ed) Lehrbuch der Gerontopsychiatrie. Enke, Stuttgart, pp 253–262

CHAPTER

5

A. Kurz, H. Lauter

Clinical Aspects of Alzheimer's Disease

Translator: E. Taub

1 Introduction

In the last 15 years, Alzheimer's disease (AD) has become a major focus of interest in medicine and neuroscience. Significant advances in our understanding of this disease have emerged from modern basic research. In the clinical field, too, progress has been made, albeit to a less spectacular extent. AD is now one of the most thoroughly clinically characterized diseases. Its wide variety of clinical manifestations and the variations its course may take are now well known from numerous cross-sectional and longitudinal studies. The symptomatology of AD has recently been extended by studies of its early clinical phases.

Further progress in the clinical area is important, because our ability to provide answers to many questions of epidemiology, etiology, pathogenesis, and treatment largely depends on the state of the art of clinical assessment. Furthermore, the results of clinical research are of great practical importance. Patients and their families want to be as well informed as possible about the nature and prognosis of the disease, so that they may decide wisely as they plan for the future and make the necessary arrangements for long-term care.

Currently available treatments for AD, as well as newly developed ones, can be properly tested for effectiveness only if the diagnosis is made in an early stage of the disease and changes of the patient's condition are accurately monitored as the disease progresses. To achieve these goals, our diagnostic conventions and instruments of assessment must be continuously refined, adapted to meet specific purposes, and updated in accordance with new scientific knowledge and medical technology. Alongside the clinical phenotype, biological markers of disease are now being increasingly used for early diagnosis. In this chapter, we shall discuss some of these rapidly developing trends.

2 Diagnostic Criteria

2.1 Three Sets of Criteria for Clinical Diagnosis

The operational criteria used for the clinical diagnosis of AD divide the diagnostic decision-making process into two separate steps. First, the presence of a dementia syndrome must be demonstrated. Second, cerebral or systemic illnesses other than AD that are possible causes of dementia must be excluded.

The two psychiatric classification systems ICD-10 (Appendix A) and DSM-IV (Appendix B) list the diagnostic criteria for the dementia syndrome (defined somewhat differently in each system), state that these are also the defining features of AD, and further require the absence of neurological, systemic, or substance-induced illnesses or disorders that might be responsible for the cognitive disturbances and behavioral changes. Both sets of criteria require an insidious disease onset and continuous progression of the cognitive disturbances. According to DSM-IV, the dementia syndrome should not be better explained by another psychiatric disease, such as depression or schizophrenia. Both classifications provide further codes that take the age at onset of disease, the presence of depression, delusions, hallucinations, or behavioral disturbances, the occurrence of atypical features, or the superimposition of dementia on delirium into account.

Alongside the defining criteria of ICD and DSM-IV, a set of consensus criteria developed by a working group under the aegis of the National Institute of Neurological and Communicative Disorders (NINCDS) and the Alzheimer's Disease and Related Disorders Association (ADRDA) (McKhann et al. 1984) has long proved valuable, particularly for research purposes (Appendix C).

This set of criteria enables the identification of "definite," "probable," and "possible" AD. A diagnosis of *probable AD* requires the exclusion of other neurodegenerative or systemic illnesses as causes of dementia. Test findings that are compatible with the diagnosis of probable AD are listed, as are others that speak against it. If the exclusion of other causes of dementia is incomplete because of inadequate information or the presence of accompanying illnesses, or if the clinical picture is atypical, a diagnosis of *possible AD* can be made. Patients having the clinical features of AD and, at the same time, a prior history of cerebrovascular disease, such as major or lacunar strokes or diffuse white matter disease, also fall into this category. This is the important group of patients with mixed forms of AD and dementia of cerebrovascular origin. These patients often fail to fulfill the rigorous diagnostic criteria for vascular dementia, which require both the demonstration of a relevant cerebrovascular disease and a causal relationship between cerebral ischemia and the appearance or progression of the cognitive disturbance (Román et al. 1993). A diagnosis of *definite AD* requires both the fulfillment of the clinical criteria for the diagnosis of probable AD and histopathological confirmation by brain biopsy or necropsy.

The three sets of diagnostic criteria all possess satisfactory interrater reliability (Kukull et al. 1990a; Lopez et al. 1990; Blacker et al. 1994). They are largely similar in content, but differ from one another in

several ways. For a diagnosis of AD, DSM-IV requires that the cognitive deficits have led to "significant" impairment of social or occupational function, while ICD-10 requires only a mild impairment of the patient's ability to function in everyday life. According to both the NINCDS-ADRDA and the ICD-10 criteria, focal neurological abnormalities appearing in later stages of the disease are compatible with AD, while DSM-IV makes no mention of focal neurological signs or symptoms in AD and lists them only as features of dementia of cerebrovascular origin. Increased muscle tone, gegenhalten, and other parkinsonian manifestations may, however, occur even in intermediate stages of progression of AD.

2.2 Histopathological Criteria

If the validity of clinical diagnosis is judged by its neuropathological confirmation, then it may be safely stated that clinical assessment of patients yields sufficiently reliable diagnoses, as long as the sets of diagnostic criteria mentioned above are applied. The validity of clinical diagnosis is very high in typical cases. Validity depends strongly, however, on which standard criteria are used in the neuropathological diagnosis of AD (Tierney et al. 1988). The demonstration of histopathological changes in the cerebral cortex by biopsy or necropsy cannot be considered a gold standard against which the accuracy of clinical diagnosis can be measured, because the diagnostically relevant morphological features of AD are not uniformly defined, and the neuropathologist thus has considerable latitude in the provision of a diagnosis. Recent sets of histopathological criteria – the National Institute of Aging (NIA) criteria (Khachaturian 1985) and the Consortium to Establish a Registry for Alzheimer's Disease (CERAD) criteria (Mirra et al. 1991) – use the semiquantitatively determined plaque density in several brain regions as the decisive factor in the morphological diagnosis of AD and its differentiation from the changes of normal aging. At the same time, however, account is taken of the age of the patient and the presence or absence of a clinical diagnosis of dementia. The inclusion of clinical information leads, of course, to an undesirable circularity of diagnostic logic when AD must be differentiated from the normal cerebral atrophy of advanced age.

Furthermore, the two neuropathological definitions of AD referred to above differ from each other in several important ways. Thus the NIA criteria require counting of all neocortical plaques, while the CERAD criteria require counting only of the neuritic plaques in the neocortex. The NIA criteria require the exclusion of other organic cerebral causes of dementia, such as infarcts or subdural hematomata, while the CERAD criteria allow the provision of a morphological diagnosis of AD even when another cerebral disease, e.g., a cerebrovascular process, is simultaneously present. Finally, the neuroanatomical diagnosis of AD according to the CERAD criteria relies exclusively on the counting of plaques and does not take the presence of neocortical neurofibrillary tangles into account, even though several studies (Bouras et al. 1994; Bierer et al. 1995; Nagy et al. 1995) have shown that the latter have a close relationship to the onset and degree of severity of clinical manifestations of AD and that they are an important element in the morphological differentiation of AD from the changes of normal aging.

These diagnostic difficulties are related to the fact that the morphological criteria for AD have not yet been adequately validated with reference to the brains of normal elderly individuals who had no clinical manifestations of dementia up to the time of their death. Available studies (Crystal et al. 1988; Katzman et al. 1988a; Snowdon 1997) do, however, reveal that widespread plaques and neurofibrillary changes can be demonstrated in the brains of at least a few individuals who were not demented during their lifetimes.

2.3 Accuracy of Clinical Diagnostic Criteria

In general, use of the clinical diagnostic criteria results in a correct diagnosis in approximately 75% of patients, with a sensitivity of approximately 90% and a specificity of approximately 75% (Khachaturian 1985; Boller et al. 1989; Kukull et al. 1990b). Recent clinicopathological comparisons have yielded still higher validity rates of 85% to over 90% (Galasko et al. 1994; Jellinger 1996). In interpreting such findings, it should not be forgotten that scientific studies are often performed on highly selected patient populations, from which poorly documented cases and cases with very mild or very early AD are automatically excluded, and in which diagnoses have often been made after long periods of observation by clinically experienced investigators.

In routine medical practice, and also in population-based research studies, the accuracy of diagnosis is probably considerably lower, because the patient histories and clinical findings are not as well documented and many patients are in early stages of the disease or have atypical manifestations. The relatively high concordance rate of the clinical and neuropathological findings should not, therefore, lead to a false sense of security in the diagnosis of AD in early stages.

2.4
Future Developments Toward Early Diagnosis

It may be expected that the clinical diagnostic guidelines will be modified in the future to take greater account of neurobiological factors. For the sake of diagnostic accuracy, it will probably be advisable to create a larger number of diagnostic categories than we now have with "possible" and "probable" AD and to make the defining features of the disease depend, in part, on the goal of the diagnostic assessment. Higher diagnostic accuracy is required for diagnostic and therapeutic research studies, or for problems of genetic counseling, than for early detection of the disease. It ought to be possible, under certain conditions, to diagnose AD before the cognitive disturbances and the restriction of the patient's everyday activities become severe enough to constitute a dementia syndrome, so that therapeutic or preventive measures may be introduced early and with the best prospect of success.

3
Clinical Manifestations

The clinical picture of AD has been described repeatedly since its classic statement by Alzheimer (1907, 1911) and, among others, Grünthal (1926), Sjögren et al. (1952), and Delay and Brion (1962); in the textbooks of Reisberg and colleagues (Reisberg 1983; Terry et al. 1993; Burns and Levy 1994; Gauthier 1996); and in the previous edition of this book (Lauter and Kurz 1989).

3.1
Cognitive and Noncognitive Disturbances

While the early authors, influenced by the then prevailing "cognitive paradigm" of the dementias, directed their attention largely to the cognitive disturbances of AD – impairment of memory and spatial orientation, word-finding difficulties, and other focal neuropsychological manifestations in which the patient's cognitive abilities seem to be "fading away" – interest has increased in recent years in other mental changes that are not related, or only indirectly related, to cognitive deficits (Burns et al. 1990; Förstl et al. 1992a,b, 1993; Absher and Cummings 1994; Folstein and Bylsma 1994; Kurz 1998). Among these are disturbances of initiative, of mood, of emotional control, of personality, and of the sleep–wake cycle and psychotic phenomena. Unlike the cognitive deficits and the limitation of everyday activities, these changes appear discontinuously. Among these changes, lessened initiative is the most likely to persist, and psychotic phenomena are the least likely (Devanand et al. 1997).

These noncognitive phenomena usually create much more stress for relatives and professional personnel than the cognitive impairment (Coen et al. 1997). They are also a common reason for the institutionalization of patients with AD in nursing homes. They are, however, more amenable to treatment than the cognitive deficits. The spectrum of available pharmacological treatments has recently been considerably widened and improved by the introduction of new neuroleptic and antidepressive agents that have fewer side effects than older preparations. Morphological changes in certain brain regions clearly play a major role in the appearance of noncognitive manifestations and behavioral changes; thus they may serve as a pathogenetic model for the generation of psychopathological disturbances in the psychoses, whose morphological substrate in the brain is less well defined (Förstl and Fischer 1994).

Changes of initiative are among the more common behavioral disturbances in AD. Typically, there is a diminution of initiative, along with a loss of interests, reduced readiness to communicate, and apathy, although restlessness and increased motor activity can also be seen, manifesting themselves as pacing, shouting, aggression, and socially disruptive personality changes. This category of noncognitive manifestations, in a broader sense, also includes changes of eating behavior, of the sleep–wake cycle, and of sexuality. These behavioral disorders generally become worse as the disease progresses.

3.2
Delusions

A total of 20%–30% of patients satisfying diagnostic criteria for AD have disturbances of thought content of various types. It is common for patients to think that they are being robbed or to suffer from jealousy; such ideas often serve as useful excuses for misperceptions or for having mislaid objects. Such delusions are usually more transient and less systematic than in schizophrenic or paranoid psychoses. They are not closely correlated with the severity of cognitive impairment. A relatively intact cerebral cortex seems to be a prerequisite for the elaboration of complex delusional ideas: computed tomography (CT) reveals that patients with such delusional disturbances have less marked ventricular dilatation than patients without paranoid manifestations (Burns et al. 1990).

3.3 Hallucinations

Hallucinations, usually visual, occur in 10%–20% of patients with AD. Perceptual disturbances of another kind, which may be designated illusory situational misperceptions, are considerably more common. These include, among other phenomena, the misidentification of near relatives (Capgras syndrome), a feeling of unfamiliarity with one's own usual living space (Landis und Cummings 1986), the nonrecognition of oneself in a mirror, or a confusion of televised images with directly experienced reality.

3.4 Depressive Manifestations

The simultaneous occurrence of AD and depressive disorders is of great practical and theoretical interest. The prevalence of depressive manifestations among AD patients is extraordinarily variable among published reports, in which it ranges from 0% to 86%. This variation is a product of differences in the origin and composition of patient collectives, in stages of progression of AD, in the length of time over which the patients were studied, in methods of assessment, and in the criteria used to determine whether a depressive disturbance was present. In the majority of recent studies containing adequate numbers of patients, the prevalence of depressive manifestations is on the order of magnitude of 15%–50% (Cummings et al. 1987; Rubin et al. 1988; Patterson et al. 1990; Förstl et al. 1992a,b, 1993). The prevalence of depressive disturbances in patients with AD is thus definitely higher than in age-matched nondemented individuals.

Many of the psychopathological features of depression overlap with those of dementia. The diagnosis of a depressive syndrome in AD is therefore difficult to make. Depressive illnesses such as major depression are rarer in patients with AD than depressive manifestations. Estimates of the prevalence of depressive manifestations based on information from patients' families are generally much higher than determinations based on information from the patients themselves (Mackenzie et al. 1989).

The relationship of depressive manifestations to the stages of AD is controversial. Most investigators report depressive manifestations to be more frequent in early disease stages, in patients whose cognitive abilities are still relatively well preserved (Reifler et al. 1982; Burns et al. 1990; Cooper et al. 1990). The Berlin Aging Study of the extremely old also revealed a higher mean severity of depressive manifestations, with increasing impairment of cognitive abilities, in mildly and moderately demented patients, while fewer depressive manifestations were seen in the severely demented (Helmchen and Linden 1993; Reischies et al. 1997).

The pathogenetic relationship of AD and depression is complex. The depressive manifestations might be regarded as the direct expression of the underlying neurobiological process. Evidence in favor of this point of view is provided by histopathological studies, which show more severe involvement of the aminergic neurons of the locus ceruleus, and of the dorsal raphe nucleus and the substantia nigra, in AD patients suffering from depression (Zubenko and Moossy 1988; Zweig et al. 1988), or else severe involvement of the noradrenergic locus ceruleus with relatively mild loss of cholinergic neurons in the basal nucleus of Meynert (Förstl et al. 1992a,b). On the other hand, depressed patients with AD also more frequently have a family history of affective disorders than nondepressed patients (Pearlson et al. 1990); this finding seems to indicate a specific genetic predisposition.

Although prior depressive illnesses (Jorm et al. 1991) – particularly when associated with reversible pseudodementia (Kral and Emery 1989; Alexopoulos et al. 1993) – elevate an individual's risk of developing AD, there is no evidence that the early phases of a depressive illness appear more frequently in AD patients with depressive manifestations than in patients whose AD is not associated with depressive features (Bolger et al. 1994). On the other hand, depressed AD patients probably differ from nondepressed AD patients, even before the appearance of dementia, in certain features of primary personality traits, such as greater fearfulness, depressivity, withdrawal, and sensitivity and lower self-esteem (Chatterjee et al. 1992).

The generation of depressive manifestations in patients with AD is perhaps a function not only of damage to the brain directly caused by the illness process, but also of an increased vulnerability of these patients to depression, which, in turn, represents a disposition of specific neuronal systems to become pathologically affected (Strauss 1995). Moreover, in the setting of this presumed biological vulnerability, the patients' awareness of their own cognitive deficits, of their reduced ability to cope with everyday life, and of the diagnosis of AD, with all that it implies, is a major stress factor that likely promotes the appearance of depressive manifestations.

3.5 Somatic Manifestations

Somatic signs and symptoms are not part of the typical clinical picture of AD in its early and intermediate

stages (for the stages of AD, see Sect. 4 below). Mild extrapyramidal signs, alterations of posture and gait, and disturbances of sphincter control do, however, appear in the stage of moderately severe dementia. In the stage of severe dementia, primitive reflexes, severe rigidity with gegenhalten, total incontinence, dysphagia, myoclonus, and grand mal seizures may occur.

4 Clinical Course

Recent neuropathological studies shed new light on the course of AD (Braak and Braak 1996). Because of these findings, it is now presumed that the neurodegenerative process leading to AD silently unfolds for many years, and perhaps decades, in the entorhinal cortex before it reaches the hippocampus, causing the appearance of the first clinical manifestations, and then goes on to affect the neocortex, thereby bringing the patient over the clinical threshold of dementia (Braak and Braak 1991; Gertz et al. 1996).

4.1 Clinical Silence, Predementia, and Dementia Stages

The course of AD may, therefore, be divided into three stages: a clinically silent stage, of as yet unknown length; a stage of so-called predementia, lasting at least 5 years (Reifler 1997), in which mild cognitive changes are noted; and a final stage of dementia, which lasts for a mean of 8–10 years (until the death of the patient). The latter stage is, in turn, subdivided into stages of mild, moderately severe, and severe dementia.

Unequivocal clinical deficits appear in the stage of predementia. They are produced by the involvement of the hippocampus in the pathogenetic process and are typified by mild cognitive impairment affecting the storage and retrieval of new information. Impairment of memory is usually noted. Tests of learning ability are, therefore, the most sensitive clinical instruments for the early detection of AD (Linn et al. 1995). At the same time, incipient thought disturbances and mild disturbances of attention may be present, and loss of initiative is also often noted (Linn et al. 1995). The ability to perform complex tasks is also sometimes impaired in the stage of predementia, even though this impairment does not lead to a restriction of the patient's usual everyday activities (Table 1).

The threshold of clinical dementia is reached when definite disturbances of executive functions, such as the ability to carry out complex activities, cope with professional duties, or organize the household (Barberger-Gateau et al. 1992; Oppenheim 1994), are added to the impairment of memory and when the deficits in these two cognitive areas limit the patient's ability to go about everyday activities.

Depending on the topographical spread of the neurodegenerative process, disturbances of language function, of object recognition, and of spatial ability may be present, sometimes to such an extent that they are among the patient's more prominent disease

Table 1. Cognitive symptoms over the course of Alzheimer's disease

Area of disturbance	Clinical stage of disease			
	Predementia	Mild dementia	Moderately severe dementia	Severe dementia
Storage of new information	+	++	+++	+++
Recall of memories ("remote memory")	–	(+)	++	+++
Thinking (drawing of logical conclusions, problem-solving, planning, organizing)	(+)	+	+++	+++
Spatial and temporal orientation	–	+	+++	+++
Orientation in a room	–	(+)	++	+++
Language (word-finding, naming, vocabulary, expression, reading, writing)	–	+/–	++	+++
Recognition of objects and persons	–	+/–	++	+++
Copying of figures, estimation of spatial relationships	–	+/–	++	+++
Manipulation of objects	–	+/–	++	+++

Degree of severity: –, absent; +/–, may or may not be present; (+), minimal; +, mild; ++, moderate; +++; severe.

Table 2. Impairment of activities of daily living over the course of Alzheimer's disease

Area of disturbance	Clinical stage of disease			
	Minor cognitive impairment	Mild dementia	Moderately severe dementia	Severe dementia
Occupation, financial matters	(+)	++	+++	+++
Household, shopping, hobbies	–	+	++	+++
Dressing, bathing, showering	–	–	+	+++
Eating, walking	–	–	–	++

Degree of severity: –, absent; (+), minimal; +, mild; ++, moderate; +++, severe.

manifestations. Disturbances of spatial orientation and word-finding difficulty are often noted very early in the course of the disease.

In the stage of moderately severe dementia, the cognitive disturbances become increasingly severe and, as a rule, spread to all areas of cognitive performance. The deterioration does not always proceed in linear fashion, however, and may be interrupted by long plateau periods (Haupt et al. 1993). As long as the patient can still mobilize sufficient reserve abilities and compensating mechanisms, the mental deterioration progresses relatively slowly. The progression accelerates as the dementia becomes worse.

As the patient's cognitive abilities become progressively more impaired, his or her ability to carry out the functions of everyday life also diminishes. At first, only the more demanding tasks are affected; later, however, even the simplest tasks involved in caring for oneself become impossible (Table 2).

The basic properties of the patient's personality, social conventions, and the ability to understand and express emotion are generally little affected by AD.

In the final stage of progression of the disease, all higher cognitive functions are extinguished and can hardly be differentiated from one another. Primitive motor routines, dysphagia, flexion contractures, incontinence, and persistent vegetative-apallic syndromes generally appear. Diminished resistance to somatic illness in combination with impairments of eating and swallowing promote the occurrence of pneumonia, which is the cause of death in approximately 70% of patients (Förstl et al. 1991). Severe damage of autonomic nuclear areas in the brain stem is certainly another possible direct or auxiliary cause of death.

4.2 Influence of Education and Occupation

The patient's level of education and occupational activity may exert an influence on the course of illness. More educated patients, and those whose occupations require the performance of complex tasks, high levels of exertion, or social communicative abilities, have more severe deficits of perfusion in the temporal and parietal lobes than those of less educated patients in occupations carrying less responsibility who are demented to a comparable extent (Stern et al. 1992, 1995a). They have a shorter duration of illness and a higher mortality (Stern et al. 1995b). The lifestyle of an individual thus clearly has an effect on the extent of his or her cognitive functional reserves. A higher level of education and the experience of lifelong coping with difficult professional tasks make some AD patients better equipped than others to compensate for the deficits that the illness produces. Consequently, these patients become overtly demented later in the course of disease and therefore seem to be less severely affected by comparable degrees of neuropathological abnormality, while the duration of illness is usually shorter.

4.3 Instruments for Monitoring Disease Course

Rating scales are often used to assess the progression of the disease. Three instruments have been commonly used. The Clinical Dementia Rating (CDR) Scale (Hughes et al. 1982) has five levels; an assessment of function in six specific areas is used to generate a score signifying absence of dementia (0) and questionable (0.5), mild (1), moderately severe (2), and severe (3) dementia. The CDR should be applied only on the basis of an extensive examination of the patient and an interview with a well-informed relative. The Global Deterioration Scale (GDS) (Reisberg et al. 1982), on the other hand, has seven levels. A score of 1 corresponds to a fully normal cognitive ability for the patient's age. Patients with a score of 2 complain of cognitive impairment, but none can be found on objective examination. A score of 3 corresponds to a mild cognitive impairment, such as an age-associated memory

disturbance. Scores of 4–7 correspond roughly to scores of 0.5–3 on the CDR. Lastly, the Functional Assessment Staging (FAS) scale (Reisberg 1988), which is used to assess the need for nursing care, makes finer distinctions in the later stages of the disease (the neglected second half of AD).

4.4 Duration and Mortality

Patients live a mean of 8 years from the appearance of the earliest manifestations of the disease; variation between individuals is considerable. The mortality of patients who become affected by the disease before age 65 is approximately 3.5 times higher than that of healthy individuals of the same age. The relative mortality is not as high in patients who develop the first manifestations of the disease in advanced age.

5 Heterogeneity

In general, the clinical manifestations are highly variable. In view of this heterogeneity, attempts have repeatedly been made to use the age of onset of the disease, psychopathological features, certain neurological manifestations, or histopathological characteristics to distinguish various subtypes of AD. Several authors (Bondareff et al. 1987; Blennow and Wallin 1992, 1994; Bondareff 1994; Wallin and Blennow 1996) speak of a classical type of AD, characterized by an early age of onset, quite severe parietotemporal manifestations, marked neuronal loss in the locus ceruleus, and a large number of neocortical plaques and neurofibrillary changes, and a second type, characterized by later onset, global cognitive deficits, confusional states, more severe leukoaraiosis and cerebrovascular changes. There is, however, a great deal of overlap between these two forms. There has been, as yet, no convincing demonstration that such subtypes are indeed distinguishable from one another and remain so over the course of the disease (Kurz et al. 1992; Förstl and Fischer 1994).

It cannot be stated with certainty whether the familial and sporadic forms of AD differ with respect to their clinical phenotype. Most of the studies performed on this question yielded a negative answer (Edwards et al. 1991; Duara et al. 1993). Loci on three different chromosomes are known to be responsible for autosomal dominant, familial AD in different families. Apparently, this genetic heterogeneity is reflected in an only modest clinical heterogeneity of familial AD (Rossor et al. 1996).

6 Neurobiological Markers

AD cannot be reliably distinguished from other cerebral diseases causing a dementia syndrome on the basis of clinical findings alone. A provisional clinical diagnosis of AD therefore requires the exclusion of other dementing diseases and can be definitively confirmed only by histopathological examination. This unsatisfactory situation has provoked a search for neurobiological features of the disease that might serve to differentiate AD from other causes of dementia. Candidates for such markers primarily include alterations of brain structure or function that can be detected with modern imaging techniques. Genetic and neurochemical markers are further possibilities.

6.1 Imaging Techniques

6.1.1 Structural Imaging

The two structural imaging techniques, computed tomography (CT) and magnetic resonance imaging (MRI), are used mainly to exclude cerebral diseases, such as mass lesions, inflammatory processes, infarcts, and disturbances of cerebrospinal fluid circulation, that may cause dementia but are not the result of a neurodegenerative pathogenetic process. Recently, however, a second mode of application of structural imaging has become increasingly important, namely, the detection of morphological changes in the brain that allow a differential diagnosis between AD and the cerebral involution accompanying normal aging, and also between AD and other causes of dementia. Planimetric and volumetric imaging of certain brain structures and the quantification of subcortical signal changes on T_2-weighted MRI have been used for these purposes.

A significant reduction in total brain volume and dilatation of the external and internal cerebrospinal fluid spaces are usually found in AD. There is, however, a significant range of overlap between the cortical atrophy and ventricular dilatation of AD and those of normal old age. The demonstration of generalized atrophy, therefore, is of little value in differential diagnosis.

Neuropathological findings support the conclusion that the structures of the mediobasal temporal lobe suffer quite extensive neuronal loss very early in the

course of the disease and thereby undergo a reduction of volume. These observations have been confirmed in vivo by MRI. Many studies have shown that the hippocampus is an average of 40% smaller in patients with intermediate-stage AD than in cognitively intact age-matched control subjects (Seab et al. 1988; Jack et al. 1992; De Leon et al. 1997) and that, even in mild AD, a 25% volume reduction is present (Killiany et al. 1993; Lehiricy et al. 1994; for further literature, see Hampel et al. 1997). The fact that the presence of hippocampal atrophy can be regarded as a very early diagnostic sign of the presence of AD was also confirmed in neuropathologically verified cases (Jobst et al. 1992, 1994). Nonetheless, all large studies revealed a significant range of overlap between the patient and control groups. Moreover, most studies to date have concentrated on the differentiation of AD patients from normal individuals of the same age. With regard to the clinical diagnosis of AD, however, it must borne in mind that other dementing illnesses may also lead to a significant volume reduction in the medial temporal lobe.

Medial temporal volume reduction apparently also allows a prediction of which individuals with mild memory impairment will likely develop dementia in the years to come. A longitudinal study was performed on 32 patients who had mild cognitive impairments, most of whom developed dementia over a 4-year period of observation (De Leon et al. 1996). Most of these cognitively impaired, elderly individuals already had a mild hippocampal volume reduction, demonstrable by MRI, on their initial imaging study, but, in contrast to patients with overt AD, the abnormality did not extend to the lateral portions of the temporal lobe or to other regions of the brain. This study provides no information as to whether hippocampal atrophy in cognitively normal individuals might also be associated with a greater than normal risk for the development of dementia.

Sequential study with CT or MRI may also help support a diagnosis of AD. The progression of brain atrophy from year to year is much more pronounced in AD patients than in cognitively unimpaired control subjects of the same age. This increased rate of progression of brain atrophy is particularly noticeable in the mediobasal portion of the temporal lobe. The reduction of hippocampal volume per year in AD patients is ten times as large as that in age-matched controls (Smith and Jobst 1996; A.D. Smith et al. 1996) and is closely related to the rate of cognitive deterioration.

A number of observations imply that longitudinal CT or MRI studies can also be used to predict the appearance of AD. In nondemented, mildly cognitively impaired individuals, an elevated rate of hippocampal atrophy is considered an early warning sign for the development of dementia (Fox et al. 1996; Smith and Jobst 1996; A.D. Smith et al. 1996). Repeated measurements reveal a nearly linear increase in the cerebrospinal fluid volume, even before the clinical manifestations become severe enough to cross the defining threshold of dementia.

A further attempt to obtain specific criteria for the diagnosis of AD from structural imaging techniques relies on the common white matter changes that are evident as hypodense zones in the periventricular white matter on CT images and are particularly well seen as areas of elevated signal intensity in T_2-weighted MRI sequences. The pathogenesis and clinical significance of this so-called leukoaraiosis is discussed in Chap. 8 (this volume, Part 2).

Moderately severe and severe leukoaraiotic changes are seen in 50%–80% of patients with vascular dementia and in 20% of age-matched normal control subjects older than 65. A possible connection to AD is evident from the studies of Brun and Englund (1986).

In 60% of neuropathologically confirmed cases of AD, these authors found white matter changes which they interpreted as incomplete infarcts secondary to fibrohyalinosis of the white matter vasculature and whose distribution was not correlated with the pattern of cortical degeneration.

Patients with presenile AD were, however, no different from normal control subjects in the extent of their leukoaraiotic changes. Only patients who developed AD in old age had significantly more areas of hyperintensity in the periventricular zone and deep white matter. These findings still require confirmation by comparative studies of patient cohorts with early and late disease onset. An unequivocal differentiation of AD from the normal physiological effects of aging on the brain and from vascular dementing processes is quite difficult because of the high degree of intergroup overlap, and also because the question of how best to quantify leukoaraiotic changes has not yet been satisfactorily resolved.

6.1.2 Functional Imaging

The functional imaging techniques include functional magnetic resonance imaging (fMRI), magnetic resonance spectroscopy (MRS), single photon emission computed tomography (SPECT), and positron emission tomography (PET). The latter two techniques are the ones most widely used at present. They allow the measurement of regional cerebral perfusion and metabolic rates, as well as the quantitative determination of the presynaptic uptake of neurotransmitters and their binding to various receptors. Even in mild AD, in which the presence of the typical defining features of dementia is questionable or these features are com-

pletely absent, a reduction of the metabolic rate for glucose can be demonstrated in several areas of the brain (G.S. Smith et al. 1992). The area most strongly involved is the temporoparietal association cortex, while the primary motor and sensory cortical fields are relatively spared.

The pattern of regional distribution of deficits of cerebral perfusion and metabolism in AD is different from that found in normal elderly individuals and in patients with other dementing illnesses (Herholz 1995) and can be found even in only mildly demented AD patients (Mielke and Heiss 1998). The extent of cognitive impairment is highly correlated with the metabolic rates of the most severely involved cortical association areas (G.S. Smith et al. 1992; Mielke et al. 1994). The diminution of neural activity that is reflected by the regional metabolic deficits is the result not only of neuron loss, but also of a functional impairment of residual neurons (G.S. Smith et al. 1992). Functional imaging techniques such as PET are also used to study the influence of therapeutic measures on cerebral perfusion and metabolism and the effects of drugs on certain neurotransmitter and receptor systems. Several brain-activating substances have been shown to produce a significant increase in cerebral metabolic rate (Heiss et al. 1988; Nordberg et al. 1998).

Cerebral perfusion and metabolism may also be studied during cognitive activation. Currently available data imply that measurements of metabolic rate during stimulation are a more sensitive test for neuronal dysfunction than measurements of glucose metabolism at rest (Pietrini et al. 1999). The pattern of activation of various brain areas by the performance of a given task is different in patients with AD than in normal control subjects. While performing a memory task, for example, the AD patients had a stronger activation of frontal areas, possibly reflecting compensatory recruitment of available neuronal resources (Woodard et al. 1998).

fMRI allows noninvasive measurement of regional cerebral blood flow by detection of the state of oxygenation of hemoglobin. Even in patients with mild AD, temporoparietal blood flow (compared to cerebellar blood flow) is diminished by approximately 20% (Harris et al. 1996). fMRI findings are highly concordant with PET findings. It may be inferred that these two techniques are of comparable usefulness in the diagnosis of dementing illnesses (Gonzalez et al. 1995).

MRS is a noninvasive method of quantitative measurement of biochemical parameters in the brain. In the study of neurodegenerative disorders, it is particularly helpful to use proton MRS (^{1}H-MRS) to measure the concentration of compounds that are markers for neuronal cell death (*N*-acetyl-aspartate, NAA) and glial proliferation (*myo*-inositol, MI) (Lazeyras et al. 1998). AD patients typically have a diminished level of NAA, and an elevated level of MI, in comparison to normal control subjects; these findings are independent of cerebral structural changes and are poorly correlated with the clinical severity of dementia (Heun et al. 1997; Schuff et al. 1998). These biochemical abnormalities are particularly marked in the temporal and parietal areas in AD and thus allow diagnostic differentiation from frontotemporal degenerative processes (Ernst et al. 1997). It is still unclear whether ^{1}H-MRS is a suitable method for the early diagnosis of AD.

The further development of imaging techniques will bring improvements both in hardware and in techniques of assessment. Higher resolution may enable the representation of cortical structures in the living brain all the way down to the cellular level. Interesting perspectives are also opened up by the development of radioligands that can be used in SPECT and PET for the imaging of cholinergic and serotonergic neurotransmitter systems. Combined structural and functional imaging, e.g., in the form of three-dimensional overlaid images, will certainly also gain in importance. Reliable measurement of changes in cerebral metabolism by PET is difficult to achieve in AD patients because of their concomitant loss of brain matter. The higher spatial resolving power of MRI can probably be used to distinguish genuine diminutions of brain activity from those that are merely secondary to volume reduction. Meanwhile, the interpretation of structural changes will also be facilitated by the concomitant use of functional imaging techniques.

6.2 Genetic Markers

Tests for the detection of mutations in the genes encoding amyloid precursor protein (APP) and presenilin-1 and presenilin-2, now known to be associated with autosomally inherited familial forms of AD, can now be performed in all laboratories of human genetics. The diagnosis of AD in affected family members can thereby be established with certainty. It can also be determined whether the healthy blood relatives of such patients carry the relevant mutation and are thus expected to develop the disease at some later time. Younger family members are often interested in such information so that they can decide to have, or not to have, children depending on the degree of their genetic load. It should be borne in mind, however, that the communication of a positive test result places a severe psychological burden on the affected person, and it may also have repercussions for other family members, who might thereby discover their own elevated risk of developing the disease. The performance of such a test

is ethically acceptable only after thorough counseling and adequate time for reflection and when provisions have been made for long-term psychotherapeutic care once the test result is delivered. Moreover, the genetic information thus obtained must, of course, be kept confidential (Lennox et al. 1994).

Most cases of AD are of multifactorial etiology, i.e., a combination of factors including brain aging, genetic factors, prior brain injury, and environmental influences is assumed to be responsible for the generation of the disease. The single genetic risk factor that has been identified with certainty to date is the ε4 allele of the gene for apolipoprotein E (ApoE) on chromosome 19 (see Chap. 7, this volume). This normal genetic variant is found three times as often in patients with AD as in normal age-matched controls (Kurz and Müller 1997). Approximately 50% of all patients do not have this allele, however, while it is indeed present in many healthy individuals. The ε4 allele is thus neither a necessary nor a sufficient condition for the development of AD. ApoE gene typing can, therefore, make only a limited contribution to the detection of AD.

In patients who already have unequivocal manifestations of dementia, however, the demonstration of the presence of an ε4 allele is of great value in differential diagnosis, increasing the likelihood that the affected individual has AD. Several studies of cases verified by necropsy (Roses 1996) have revealed that the overwhelming majority of demented patients carrying at least one ε4 allele satisfied the histopathological criteria for AD. Nonetheless, the specificity of this genetic marker is still less than 100%, as other dementing diseases are also associated with an elevated frequency of the ε4 allele. According to current concepts, ApoE gene typing cannot be used to replace any other diagnostic procedure that is normally required for the exclusion of causes of dementia other than AD (American College of Medical Genetics 1995).

In elderly individuals suffering from mild cognitive disturbances, the presence of an ε4 allele is a predictor for the progression of these deficits and the development of dementia (Coria et al. 1995; Petersen et al. 1995). Nonetheless, the present state of epidemiological knowledge does not permit reliable quantitative estimation of the risk of AD to which a person of a given age, with a given ApoE constellation, is subject, nor can the time of onset of the disease be predicted. ApoE gene typing is therefore not a suitable predictive test for asymptomatic individuals.

6.3 Neurochemical Markers

The search for neurochemical markers for AD proceeds from the conjecture that the formation of neurofibrillary tangles, the deposition of beta amyloid in brain tissue, and the loss of synapses that occur in AD might lead to detectable changes in the concentration of proteins participating in these processes in the cerebrospinal fluid or even in the blood. Many studies have been published recently showing that tau protein is elevated in the cerebrospinal fluid of patients with AD, as compared both with normal controls and with patients suffering from dementia of other causes (Galasko et al. 1996; Riemenschneider et al. 1997). The specificity of an elevated concentration of tau protein is only approximately 90%, however, as it may be found in patients with dementia of other causes or with other neurological illnesses (Schenk et al. 1996). There is controversy as to whether the elevation of tau protein concentration becomes even higher as the disease progresses and whether it is correlated with the severity of cognitive impairment. The concentration of tau protein is elevated in the cerebrospinal fluid of AD patients even in the very early clinical stage (Kurz et al. 1998) and in some patients with mild cognitive impairment. Elevated tau levels have also been found in healthy carriers of mutations on the APP or presenilin-1 genes.

Further studies are concerned with the various derivatives of APP. The ill members of a Swedish family with APP gene mutations were found to have significantly lower concentrations of soluble precursor protein in the cerebrospinal fluid (Almkvist et al. 1997). A lower level was also found in healthy carriers of the mutation in comparison to family members not affected by it. Several groups of investigators have recently shown that AD patients have a diminished concentration of insoluble beta amyloid protein in the cerebrospinal fluid (Galasko et al. 1998). The probable reason for this is the increased deposition of the protein in brain tissue, mainly in the form of plaques. Further neurochemical markers that may be of diagnostic significance include synapse-associated proteins and melanotransferrin.

7 Differential Diagnosis

AD must be distinguished from many other dementing processes. These may be grouped into four categories:

1. Dementias of cerebrovascular origin
2. States of dementia in degenerative illnesses of the basal ganglia and other focal cortical and subcortical processes
3. Hereditary metabolic disorders, most of which are evident in childhood, but some of which are not evident until adulthood

4. Potentially treatable dementias

The delimitation of these diverse dementing processes from one another is based on the acute or insidious nature of the onset of the illness, its temporal course, the presence or absence of other neurological manifestations, the cortical or subcortical character of cognitive impairment, and the nature and extent of personality changes, behavioral abnormalities, motor disturbances, and functional impairment in everyday life. Further information regarding the clinical manifestations, pathogenesis, and prognosis of these illnesses may be found in Chap. 9 (this volume, Part 2). The cerebral diseases most commonly misdiagnosed as AD are Parkinson's disease, cerebrovascular diseases, and frontal lobe degenerations (Klatka et al. 1996).

8 Combinations with Other Dementing Diseases

8.1 Cerebrovascular Diseases

AD and cerebrovascular diseases are the two most common causes of demented states in the elderly. Autopsy series of patients who had been demented during their lifetimes reveal the simultaneous presence of both cerebrovascular disease and AD in 10%–23% of patients (Katzman et al. 1988b; O'Brien 1988; Galasko et al. 1994; Jellinger 1996). Such cases are, therefore, just as common as the pure forms of dementia of cerebrovascular origin, which account for approximately 10% of patients.

The diagnosis of such a mixed form of dementia cannot be made reliably on clinical grounds alone. The ischemia score (Hachinski et al. 1975), like its various modifications, can indeed be used to distinguish AD from vascular dementia with sufficient accuracy; its sensitivity in the diagnosis of mixed neurodegenerative-vascular cases, however, is in the range of 17%–50% (Chui et al. 1992). The currently used systems of defining criteria for vascular dementia (Chui et al. 1992; Román et al. 1993) do not treat the definition of mixed vascular-neurodegenerative cases uniformly. The category of possible AD in the criteria of the NINCDS-ADRDA working group (McKhann et al. 1984) includes patients with the clinical features of AD in which marked cerebrovascular changes are also demonstrable, where the latter might constitute a separate contributing etiologic factor but are not considered the sole etiology of dementia. In view of the heterogeneity of these concepts of classification and the lack of validation for any of them, there are no currently available, generally applicable methods for the unequivocal diagnostic differentiation of AD and dementia of vascular origin from a mixture of these two types of illness.

There are a number of possible explanations of the manner in which vascular and neurodegenerative factors interact in the generation of these dementia syndromes of mixed etiology. Snowdon et al. (1997) studied the brains of 61 nuns (see also Sect. 9, below) in which the histopathological criteria for AD had been met. In nearly half of the subjects, one or more brain infarcts were also found. Dementia had been present during the lifetime of 57% of the subjects who had the typical changes of AD and no infarct, but in 75% of those who had AD changes combined with at least one large neocortical infarct, and in 93% of those who had AD changes and one or two lacunar infarcts in the basal ganglia, thalamus, or deep white matter. In a further group of patients who did not have the typical cerebral findings of AD, the frequency of infarcts was just as high, but there was no relationship between these infarcts and the development of clinical manifestations of dementia.

It may, therefore, be presumed that the occurrence of a few small infarcts in strategically located areas of the brain, superimposed on the histopathological changes of AD, places an excessive load on the functional capacity of the already damaged brain and leads to a considerable lowering of the clinical threshold for dementia. This may also explain the observation, made in recent case-control studies, that vascular risk factors also elevate the probability of a patient's developing clinically diagnosable AD (Hofman et al. 1997). The clinical picture of an AD patient with cerebrovascular changes is probably different from that of a patient with pure AD (for further details, see Chap. 8, this volume, Part 2).

8.2 Parkinson's Disease and Lewy Body Disease

Histopathological study of the brains of 650 patients who had been given the diagnosis of AD in various medical institutions was performed at a reference center for neurodegenerative diseases. Many of these patients were found to have had other causes of dementia. The disease most commonly misdiagnosed as AD was not vascular dementia, but Parkinson's disease, accounting for 6% of patients. A further 7.4% of the patients had a combination of AD and Parkinson's disease. The cognitive disturbances of Parkinson's disease usually correspond to a clinical syndrome characterized by bradyphrenia, memory impairment, difficulty with complex intellectual tasks, and changes in affectivity; this syndrome, when severe, resembles the picture of a "subcortical" dementia. These

disturbances are caused by a degeneration of neuronal systems in the brain stem that also affects their cortical projections. On the other hand, the great majority of patients who have both severe dementia and parkinsonian clinical manifestations have a combination of AD and Lewy body disease.

Lewy body disease is said to be present when the intracellular inclusion bodies discovered by Friedrich Lewy are found not only in subcortical nuclei (the substantia nigra and other brain stem centers), but also in the cerebral cortex. These changes have been repeatedly identified in recent years as a cause of dementia, and recent autopsy statistics document that they are among the more common causes of dementia in the elderly (Perry et al. 1990). The clinical manifestations of dementia in Lewy body disease are described in Chap. 9 (this volume, Part 2). A controversy exists as to whether the dementia of Lewy body disease should be considered an independent nosological category or a subtype of AD.

There is probably a spectrum of similar neuropsychiatric entities that has the following common features: (a) the presence of Lewy bodies in both the subcortical structures and the cerebral cortex and (b) a variable number of senile plaques, in the total, or near-total, or absence of neurofibrillary changes. The brain stem type, corresponding to idiopathic Parkinson's disease and presenting mainly with motor abnormalities, is at one end of this spectrum; here, the Lewy bodies are almost exclusively subcortical and are quite sparse in the limbic cortex and neocortex. At the other end of the spectrum, we find the cortical forms of Lewy body pathology, in which these inclusion bodies are extraordinarily profuse and widespread in the cerebral cortex, so that the presence of dementia is adequately explained whether or not plaques or neurofibrillary changes are also present. The motor manifestations are mild in these cases. The numerically most prominent Lewy body diseases, however, are those affecting children, which occupy the middle range of the Lewy body spectrum. Here, too, there are many cortical inclusion bodies, but they affect a lesser number of neurons. Plaques and neurofibrillary changes are also found, but in quantities and intensities that still remain under the threshold for clinical dementia. Cognitive deterioration is, therefore, presumably the combined effect of the AD-like changes and the presence of Lewy bodies.

9
Diagnosis in the Preclinical Stage

The current recognition of a protracted clinically silent phase of AD, followed by the stage of predementia, gives rise to the idea of new possibilities for early detection and even the hope that the dementia of AD might be preventable. We are immediately led to ask which of the currently available psychological and biological features of AD might be used as markers of the still latent disease process and might thus enable timely prediction of the later clinical onset of the disease.

9.1
Psychological Predictors

The search for predictors of AD begins with the search for abnormalities of behavior, personality, professional life, or cognitive performance that might be present before the onset of dementia; information about these abnormalities is obtained retrospectively from the patients and their families or from other, objective sources. Efforts of this type have, as a rule, not yielded any generally applicable conclusions.

One exception is provided by a study of Catholic nuns in the United States who were 75–95 years old at the time of the study (Snowdon et al. 1996). Use was made of autobiographical information that had been written down by the nuns when they had entered the order, at a mean age of 22 years. Higher levels of linguistic expression in these youthful writings were found to be correlated with better results on cognitive tests taken in old age. The 24 nuns who had died of neuropathologically confirmed AD belonged, without exception, to the group whose autobiographies had been characterized by a relative poverty of ideas. This interesting result does not, of course, allow us to determine whether the novices' inferior linguistic and cognitive abilities were a very early expression of subtle neuropathological changes or, alternatively, represented a normal cerebral variant that, for some unknown reason, provides a favorable substrate for the later development of Alzheimer's disease.

A second approach to the search for psychological predictors of disease consists of the performance of longitudinal studies on representative cohorts of originally healthy elderly individuals in order to determine which neuropsychological features may serve to distinguish those individuals who later develop AD from those who do not.

Predictors that emerged from such a study included poor performance with regard to learning ability, delayed reproduction of learned information, deficient use of semantic aids to memory, and disturbances in the areas of attention, verbal fluency, and nonverbal intelligence. These abnormalities appeared many years before the first clinical manifestations of dementia (Linn et al. 1995). Nonetheless, it must be

borne in mind in the interpretation of such results that samples of apparently healthy elderly individuals are inevitably "contaminated" by patients who already have minimal manifestations of the AD that will only later become overt. Lack of cognitive plasticity, i.e., an inability to improve one's level of cognitive performance through mental training, may also be regarded as a predictor of dementia in the elderly (Baltes and Raykov 1996). Although these findings have been validated by statistical group comparisons, they cannot be used reliably to predict dementia in individual cases.

Finally, it should be pointed out that individuals at especially high risk of developing AD were also included in studies that were intended to identify psychological risk factors for the disease. Individuals who complain of forgetfulness are not suitable for such studies. Complaints of forgetfulness are mainly associated with a depressive tendency, fearfulness, and neuroticism (Jorm et al. 1994), but are only poorly correlated with objectively demonstrable memory impairment and have almost no predictive value for the future development of AD (Flicker et al. 1993). In contrast, elderly subjects with objectively demonstrable memory disturbances, such as an age-associated memory disturbance or a mild cognitive impairment, very often go on to develop more severe cognitive impairment and then dementia (Flicker et al. 1991; Paykel et al. 1994; Grundman et al. 1996). Many such individuals probably already have AD in the stage of minimal clinical manifestations (Petersen et al. 1997).

9.2 Biological Predictors

Individuals at particular risk for the later development of AD include carriers of pathogenic mutations who have not yet reached the age of clinical onset. These carriers have been found to have abnormalities of verbal memory years before the clinical onset of the disease, which are later followed by abnormalities of visual attention, visual perception, and spatial analysis (Newman et al. 1994). Asymptomatic carriers of an APP or presenilin-1 mutation were found to have latent abnormalities of attention and memory (Amberla et al. 1996), as well as significant hippocampal volume reduction (Wahlund 1996), an elevated concentration of tau protein in the cerebrospinal fluid (Amberla et al. 1996), and a diminished concentration of soluble APP (Almkvist et al. 1997). All of these cognitive and neurobiological changes were absent in these individuals' healthy relatives who did not carry these mutations. It cannot be determined to what extent these results apply to families with other mutations or to cases of AD not caused by mutations of a single gene.

A preclinical diagnosis of AD cannot at present be made on the basis of psychological predictors alone. Progress in this area will probably come from various types of biological markers, including morphological predictors detectable by imaging techniques, genetic tests, and probably neurochemical findings as well. Studies in this area (Kennedy et al. 1995; Small et al. 1995; Reiman et al. 1996) have most often involved the combined application of PET and genetic and cognitive predictors. The cohorts of patients or probands studied were generally quite small; nevertheless, these studies represent a first step toward being able to distinguish mild, nonprogressive cognitive disturbances in the elderly from the first signs of a dementing illness. It is hoped that the preclinical diagnosis of AD will be enhanced by the combined use of neurobiological, genetic, and neuropsychological predictors.

In view of the wide variation in disease manifestations and the inadequate diagnostic accuracy of individual predictors, combinations of predictors will need to be used if the early diagnosis of AD is to be further improved. Such an improvement would be of great importance, as it would create an opportunity for the development of new strategies of therapeutic and preventive intervention that can be put into action in very early stages of the disease.

10 Institutional Care

Many patients with AD are cared for, not in institutions, but at home, by their families. The percentage of patients cared for at home varies across the industrial nations and is influenced by cultural traditions, governmental social and health policies, and the number of available institutional beds. Publications in German usually still refer to the findings reported more than 15 years ago by Cooper and Sosna (1983) to the effect that 20% of demented persons are cared for in institutions and 80% by their families. In the retrospective longitudinal study carried out by Bickel (1996), however, it was concluded that, if patients with very early and mild forms of dementia are excluded, 40% of demented patients are cared for in institutions. Indeed, most patients cared for at home at any given time will sooner or later, before they die, require transfer to an institution. Only about 35% patients live with their families until their death.

The likelihood of admission to an institution is increased by the same factors that generally determine the utilization of institutions and by the degree of severity of dementia, the extent of cognitive and functional impairment, and, above all, the presence

of noncognitive disturbances (Haupt and Kurz 1993). According to available data from Western countries (Bickel 1995), 17%–36% of the inhabitants of old people's homes and 51%–72% of the inhabitants of nursing homes are demented. The percentage of demented patients in residential homes, old people's homes, and nursing homes in Germany is currently 42.3%. Given that Germany has approximately 660,000 institutional beds, it can be calculated that there are approximately 280,000 demented patients in institutions, and their average length of stay is 2–3 years (Welch et al. 1992; Severson et al. 1994; Bickel 1996; Heyman et al. 1997). Even though these data are generally not broken down according to the cause of dementia (AD versus other causes), it can be safely assumed that most of these patients suffer from AD.

Because relatives are rarely able to provide the degree of care required by severely demented patients, even when assisted by outpatient services, and because psychiatric hospitals have now almost completely withdrawn from the long-term care of chronically ill elderly patients, nursing home care has become the standard mode of medical and nursing care for patients with AD who either can no longer be cared for by their families or have no families to care for them (see also Chap. 3, this volume, Part 2).

11 Cessation of Treatment

In the advanced stages of a dementing disease, when the patient is no longer capable of consenting to or refusing medical treatment, the question may arise whether life-prolonging treatment measures ought to be instituted or continued, e.g., in situations where the health and well-being of the patient are impaired by a second illness requiring treatment, or where his or her survival depends on the provision of nutrition and fluids by artificial means.

The decision-making process in such cases involves a consideration, among other questions, of whether the proposed medical intervention has a sufficient probability of success that it should be performed to lessen the patient's suffering or, alternatively, whether it would prolong the patient's life without lessening suffering and perhaps itself involve the infliction of additional suffering. A decision not to provide artificial nutrition or medical intervention may be well founded and ethically justified in some cases. Nonetheless, this must not become a routine matter. It cannot be assumed a priori that the quality of life of a patient in an advanced stage of dementia is necessarily negative, particularly because we know little about the experiential world of demented patients.

Despite the change of personality caused by the dementing illness, the patient's prior self is usually still perceptible in emotional contact. Thus the same ethical rules ought to apply to the question of cessation of treatment in these patients as in all other patients with severe physical illnesses who are no longer able to express their wishes. Moreover, the application of such principles is a demonstration of respect for the patient's self, which may still be perceived in communication even though it is overshadowed by illness, and counteracts the unfortunate human tendency to shut powerless, handicapped, or disagreeable individuals out of society (see also Chap. 12, Vol. 3). Society at large has a special responsibility to patients with AD and other dementing diseases. In an age when the costs of health care are increasing, these patients need and deserve our protection, now more than ever.

Appendix A. ICD-10 Clinical Diagnostic Guidelines for Dementia in AD

(a) Presence of dementia.
(b) Insidious onset with slow deterioration. While the onset usually seems difficult to pinpoint in time, realization by others that the defects exist may come suddenly. An apparent plateau may occur in the progression.
(c) Absence of clinical evidence, or findings from special investigations, to suggest that the mental state may be due to other systemic or brain disease which can induce dementia (e.g. hypothyroidism, hypercalcaemia, vitamin B12 deficiency, niacin deficiency, neurosyphilis, normal pressure hydrocephalus, or subdural haematoma).
(d) Absence of a sudden, apoplectic onset, or of neurological signs of focal damage such as hemiparesis, sensory loss, visual field defects, and incoordination occurring early in the illness (although these phenomena may be superimposed later).

Appendix B. DSM-IV Diagnostic Criteria for Dementia of the Alzheimer Type

A. The development of multiple cognitive deficits that manifest themselves as

1. an impairment of memory (impaired ability to learn new information or recall previously learned information) and

2. at least one of the following cognitive disturbances:
   a) aphasia (disturbance of speech)
   b) apraxia (impaired ability to carry out motor activities, despite intact motor function)
   c) agnosia (inability to recognize or identify objects, despite intact sensory function)
   d) disturbance of executive function (planning, organizing, sequencing, abstracting)

B. Each of the cognitive deficits in Criteria A1 and A2 causes significant impairment in social or professional functional areas and represents a clear deterioration from a previous level of function.
C. The course is characterized by insidious onset and continued cognitive decline.
D. The cognitive impairments in Criteria A1 and A2 cannot be attributed to:

1. other illnesses of the central nervous system that cause progressive deficits in memory and cognition (e.g., cerebrovascular disease, Parkinson's disease, Huntington's disease, subdural hematoma, normal pressure hydrocephalus, brain tumor)
2. systemic illnesses known to cause dementia (e.g., hypothyroidism, vitamin B12 or folate deficiency, niacin deficiency, hypercalcemia, neurosyphilis, HIV infection)
3. substance-induced illnesses.

E. Deficits are not exclusively present during delirium.
F. The disorder cannot be better explained by another disorder on Axis I (e.g., major depression, schizophrenia).

Appendix C. NINCDS-ADRDA Criteria for the Clinical Diagnosis of AD

(after McKhann et al. 1984)

I. The clinical criteria for the diagnosis of *probable* Alzheimer's disease include:

- dementia, demonstrated by clinical examination and documented by the Mini-Mental Status Test, the Blessed Dementia Scale, or a similar test, and confirmed by neuropsychological testing,
- deficits in 2 or more areas of cognition,
- progressive deterioration of memory and other cognitive functions,
- no disturbance of consciousness,
- onset between the ages of 40 and 90 years, usually over 65, and
- lack of systemic illnesses or other cerebral diseases that would alone suffice to explain the progressive deficits of memory and cognition.

II. The diagnosis of *probable* Alzheimer's disease is supported by:

- progressive deterioration of specific cognitive functions such as language (aphasia), motor performance (apraxia), and perception (agnosia)
- impaired everyday activities and altered behavioral pattern
- family history of the occurrence of similar illnesses, particularly when neuropathologically confirmed
- the following laboratory findings:
  - normal routine cerebrospinal fluid examination
  - normal findings or non-specific changes on EEG, such as increased slow-wave activity
  - evidence of brain atrophy by CT, with progression on repeated examination

III. After the exclusion of other causes of dementia, the following clinical features are compatible with the diagnosis of *probable* Alzheimer's disease:

- plateaus in the course of the illness,
- accompanying symptoms of depression, sleep disturbance, incontinence, delusions, illusory misperceptions, hallucinations, catastrophe reactions of verbal, emotional, or physical type, disturbances of sexual behavior, or weight loss,
- other abnormal neurological findings in a few patients, mainly in advanced stages of disease, including motor disturbances such as increased muscle tone, myoclonus, or gait disturbances,
- epileptic seizures in advanced stages of the illness, and
- normal CT findings for age.

IV. The following features make the diagnosis of *probable* Alzheimer's disease uncertain or unlikely:

- sudden "apoplectic" onset
- focal neurological findings such as hemiparesis, sensory loss, visual field defects, disturbances of coordination in the early stage
- convulsions or gait disturbances at the onset of the illness or very early in its course

V. The clinical diagnosis of *possible* Alzheimer's disease

- can be made when dementia is present and other neurological, psychiatric, or systemic illnesses are absent that would be sufficient to cause the dementia, and in the presence of variations with regard to onset, manifestations, and course,
- can be made if another systemic or cerebral illness is present that would be sufficient to cause dementia but cannot be regarded as the single cause of dementia in this case, and
- should be used in research studies when, in the absence of other identifiable causes, only a single, slowly progressive, severe cognitive deficit can be recognized.

VI. Criteria for the diagnosis of *definite* Alzheimer's disease are:

- the clinical criteria for probable Alzheimer's disease, and
- histopathological demonstration by biopsy or autopsy.

VII. The following findings may be suitable for the differentiation of subtypes of Alzheimer's disease, and are specified for research purposes:

- familial occurrence
- onset before age 65
- trisomy 21
- simultaneous presence of other relevant illnesses, such as Parkinson's disease

12 References

Absher JR, Cummings JL (1994) Cognitive and noncognitive aspects of dementia syndroms: an overview. In: Burns A, Levy R (eds) Dementia. Chapman and Hall, London, pp 59–76

*Alexopoulos GS, Meyers BS, Young RC, Mattis S et al (1993) The course of geriatric depression with "reversible dementia": a controlled study. Am J Psychiatry 150: 1693–1699

Almkvist O, Basun H, Wagner SL, Rowe BA et al (1997) Cerebrospinal fluid levels of alpha-secretase-cleaved soluble amyloid precursor protein mirror cognition in a Swedish family with Alzheimer disease and a gene mutation. Arch Neurol 54: 641–644

**Alzheimer A (1907) Über eine eigenartige Erkrankung der Hirnrinde. Allg Z Psychiatr 64: 146–148

*Alzheimer A (1911) Über eigenartige Krankheitsfälle des späteren Alters. Z Ges Neurol Psychiat 4: 356–385

Amberla K, Almkvist O, Basun H, Jensen M et al (1996) Tau levels in CSF are related to cognitive function in familial Alzheimer's disease. Neurobiol Aging 17[Suppl]: S166–S167

American College of Medical Genetics (1995) Consensus statement: statement on use of apolipoprotein E testing for Alzheimer disease. JAMA 274: 1627–1629

Baltes MM, Raykov T (1996) Prospective validity of cognitive plasticity in the diagnosis of mental status: a structural equation model. Neuropsychology 10: 549–556

Barberger-Gateau P, Commenges D, Gagnon M, Letenneuro L et al (1992) Instrumental activities of daily living as a screening tool for cognitive impairment and dementia in elderly community dwellers. J Am Geriatr Soc 40: 1129–1134

Bickel H (1995) Demenzkranke in Alten- und Pflegeheimen: Gegenwärtige Situation und Entwicklungstendenzen. In: Forschungsinstitut der Friedrich-Ebert-Stiftung (ed) Medizinische und gesellschaftspolitische Herausforderung: Alzheimer-Krankheit. Der langsame Zerfall der Persönlichkeit. Friedrich-Ebert-Stiftung, Bonn, pp 49–68

*Bickel H (1996) Pflegebedürftigkeit im Alter. Ergebnisse einer populationsbezogenen retrospektiven Längsschnittstudie. Gesundheitswesen 58: 56–62

Bierer LM, Hof PR, Purohit DP, Carlin L et al (1995) Neocortical neurofibrillary tangles correlate with dementia severity in Alzheimer's disease. Arch Neurol 52: 81–88

Blacker D, Albert MS, Basett SS, Go RCP et al (1994) Reliability and validity of NINCDS-ADRDA criteria for Alzheimer's disease. Arch Neurol 51: 1198–1204

Blennow K, Wallin A (1992) Clinical heterogeneity of probable Alzheimer's disease. J Geriatr Psychiatry Neurol 5: 106–113

Blennow K, Wallin A (1994) Heterogeneity in Alzheimer's disease: a European view. In: Burns A, Levy R (eds) Dementia. Chapman and Hall, London, pp 115–125

Bolger JT, Carpenter BD, Strauss ME (1994) Behavior and affect in Alzheimer's disease. Clin Geriatr Med 10(2): 315–337

Boller F, Lopez O L, Moossy J (1989) Diagnosis of dementia: clinicopathologic correlations. Neurology 39: 76–79

Bondareff W (1994) Subtypes of Alzheimer's disease. In: Burns A, Levy R (eds) Dementia. Chapman and Hall, London, pp 101–114

Bondareff W, Mountjoy CQ, Roth M, Rossor MN et al (1987) Age and histopathologic heterogeneity in Alzheimer's disease. Evidence of subtypes. Arch Gen Psychiatry 44: 412–417

Bouras C, Hof PR, Giannakopoulos P, Michel JP et al (1994) Regional distribution of neurofibrillary tangles and senile plaques in the cerebral cortex of elderly patients: a quantitative evaluation of a one-year autopsy population from a geriatric hospital. Cereb Cortex 4: 138–150

*Braak H, Braak E (1991) Neuropathological staging of Alzheimer-related changes. Acta Neuropathol 82: 239–259

Braak H, Braak E (1996) Evolution of the neuropathology of Alzheimer's disease. Acta Neurol Scand Suppl 165: 3–12

Brun A, Englund E (1986) A white matter disorder in dementia of the Alzheimer type: a pathoanatomical study. Ann Neurol 19: 253–262

*Burns A, Levy R (1994) Dementia. Chapman and Hall, London

**Burns A, Jacoby R, Levy R (1990) Psychiatric phenomena in Alzheimer's disease. Br J Psychiatry 157: 72–94

Chatterjee A, Strauss ME, Smyth KA, Whitehouse PJ (1992) Personality changes in Alzheimer's disease. Arch Neurol 49(5): 486–491

Chui HC, Victoroff JI, Margolin D, Jagust W et al (1992) Criteria for the diagnosis of ischemic vascular dementia proposed by the State of California Alzheimer's Disease Diagnostic and Treatment Centers. Neurology 42: 473–480

Coen RF, Swanwick GRJ, O'Boyle CA, Coakley D (1997) Behaviour disturbance and other predictors of carer burden in Alzheimer's disease. Int J Geriatr Psychiatry 12: 331–336

Cooper B, Sosna U (1983) Psychische Erkrankung in der Altenbevölkerung. Eine epidemiologische Feldstudie in Mannheim. Nervenarzt 54: 239–249

Cooper JK, Mungas D, Weiler PG (1990) Relation of cognitive status and abnormal behaviors in Alzheimer's disease. J Am Geriatr Soc 38: 867–870

Coria R, Rubio I, Bayón C, Cuadrado N et al (1995) Apolipoprotein E allelic variants predict dementia in elderly patients with memory impairment. Eur J Neurol 2: 191–193

Crystal H, Dickson DW, Fuld P (1988) Clinicopathological studies in dementia: nondemented subjects with pathologically confirmed Alzheimer's disease. Neurology 38: 1682–1687

Cummings JL, Miller B, Hill MA, Neshkes R (1987) Neuropsychiatric aspects of multi-infarct dementia and dementia of the Alzheimer type. Arch Neurol 44: 389–393

Delay J, Brion D (1962) Les démences tardives. Paris, Masson

De Leon MJ, Convit A, George AE, Golomb J et al (1996) In vivo structural studies of the hippocampus in normal aging and in incipient Alzheimer's disease. Ann NY Acad Sci 777: 1–13

De Leon MJ, George AE, Golomb J, Tarshish S et al (1997) Frequency of hippocampal formation atrophy in normal aging and Alzheimer's disease. Neurobiol Aging 18: 1–11

Devanand DP, Jacobs DM, Tang MX, Del-Castillo-Castandeda C et al (1997) The course of psychopathology in mild to moderate Alzheimer's disease. Arch Gen Psychiatry 54: 257–263

Duara R, Lopez-Alberola RF, Barker WW, Loewenstein DA et al (1993) A comparison of familial and sporadic Alzheimer's disease. Neurology 43: 1377–1384

Edwards JK, Larson EB, Hughes JP, Kukull WA (1991) Are there clinical and epidemiological differences between familial and non-familial Alzheimer's disease? J Am Geriatr Soc 39(5): 477–483

Ernst T, Chang L, Melchor R, Mehringer CM (1997) Frontotemporal dementia and early Alzheimer disease: differentiation with frontal lobe H-1 MR spectroscopy. Radiology 203: 829–836

Flicker C, Ferris SH, Reisberg B (1991) Mild cognitive impairment in the elderly: predictors of dementia. Neurology 41: 1006–1009

Flicker C, Ferris SH, Reisberg B (1993) A longitudinal study of cognitive function in elderly persons with subjective memory complaints. J Am Geriatr Soc 41: 1029–1032

Folstein MR, Bylsma FW (1994) Non-cognitive symptoms of Alzheimer's disease. In: Terry RD, Katzman R, Bick KL (eds) Alzheimer' disease. Raven, New York, pp 27–40

Förstl H, Fischer P (1994) Diagnostic confirmation, severity and subtypes of Alzheimer's disease. Eur Arch Psychiatr Neurol Sci 244: 252–260

Förstl H, Burns A, Luthert P, Cairns N (1991) Demenz und internistische Erkrankungen: die Häufigkeit innerer Krankheiten bei Alzheimer-Demenz, vaskulärer Demenz und anderen dementiellen Erkrankungen. Z Gerontol 24: 91–93

Förstl H, Burns A, Cairns N, Luthert P et al (1992a) Organische Grundlagen depressiver Symptome bei der Alzheimer-Demenz. Nervenarzt 63: 566–574

Förstl H, Burns A, Luthert P, Cairns N et al (1992b) Clinical and neuropathological correlates of depression in Alzheimer's disease. Psychol Med 22: 877–884

**Förstl H, Sattel H, Bahro M (1993) Alzheimer's disease: clinical features. Int Rev Psychiatry 5: 327–349

Fox NC, Warrington EK, Stevens JM, Rossor MN (1996) Atrophy of the hippocampal formation in early familial Alzheimer's disease. A longitudinal MRI study of at-risk members of a family with an amyloid precursor protein 717 Val-Gly mutation. Ann NY Acad Sci 777: 226–232

Galasko D, Hansen LA, Katzman R, Wiederholt W et al (1994) Clinical-neuropathological correlations in Alzheimer's disease and related dementias. Arch Neurol 51: 888–895

Galasko D, Motter R, Seubert P (1996) Interpreting cerebrospinal fluid tau levels in Alzheimer's disease. Neurobiol Aging 17(4S):S1

Galasko D, Chang L, Motter R, Clark CM et al (1998) High cerebrospinal fluid tau and low amyloid beta 42 levels in the clinical diagnosis of Alzheimer disease and relation to apolipoprotein E genotype. Arch Neurol 55: 937–945

*Gauthier S (1996) Clinical diagnosis and management of alzheimer's disease. Dunitz, London

Gertz HJ, Xuereb JH, Huppert FA, Brayne C et al (1996) The relationship between clinical dementia and neuropathological staging (Braak) in a very elderly community sample. Eur Arch Psychiatry Clin Neurosci 246: 132–136

Gonzalez RG, Fischman AJ, Guimaraes AR, Carr CA et al (1995) Functional MR in the evaluation of dementia: correlation of abnormal dynamic cerebral blood volume measurements with changes in cerebral metabolism on positron emission tomography with fludeoxyglucose F 18. Am J Neuroradiol 16: 1763–1770

Grundman M, Petersen RC, Morris JC, Ferris S et al (1996) Rate of dementia of the Alzheimer type (DAT) in subjects with mild cognitive impairment. Neurology 46 [Suppl]:A403

Grünthal E (1926) Über die Alzheimersche Krankheit. Eine histopathologisch-klinische Studie. Z Ges Neurol Psychiatr 101: 128–157

Hachinski VC, Iliff LD, Zilkha E, Du-Boulay GA et al (1975) Cerebral blood flow in dementia. Arch Neurol 32: 632–637

*Hampel H, Teipel SJ, Kötter HU, Horwitz B et al (1997) Strukturelle Magnetresonanztomographie in der Diagnose und Erforschung der Demenz vom Alzheimer-Typ. Nervenarzt 68: 365–378

Harris GJ, Lewis RF, Satlin A, English CD et al (1996) Dynamic susceptibility contrast MRI of regional cerebral blood volume in Alzheimer's disease. Am J Psychiatry 153: 721–724

Haupt M, Kurz A (1993) Predictors of nursing home placement in patients with Alzheimer's disease. Int J Geriatr Psychiatry 8: 741–746

Haupt M, Pollmann S, Kurz A (1993) Symptom progression in Alzheimer's disease: relation to onset age and familial aggregation. Results of a longitudinal study. Acta Neurol Scand 88: 349–353

Heiss WD, Hebold I, Klinkhammer P, Ziffling P et al (1988) Effect of piracetam on cerebral glucose metabolism in Alzheimer's disease as measured by positron emission tomography. J Cereb Blood Flow Metab 8: 613–617

*Helmchen H, Linden M (1993) The differentiation between depression and dementia in the very old. Ageing Soc 13: 589–617

Herholz K (1995) FDG PET and differential diagnosis of dementia. Alzheimer Dis Assoc Disord 9(1): 6–16

Heun R, Schlegel S, Graf-Morgenstern M, Tintera J et al (1997) Proton magnetic resonance spectroscopy in dementia of Alzheimer type. Int J Geriatr Psychiatry 12: 349–358

Heyman A, Peterson B, Fillenbaum G, Pieper C (1997) Predictors of time to institutionalization of patients with Alzheimer's disease: the CERAD experience, part XVII. Neurology 48: 1304–1309

Hofman A, Ott A, Breteler MMB, Bots ML et al (1997) Atherosclerosis, apolipoprotein E, and prevalence of dementia and Alzheimer's disease in the Rotterdam Study. Lancet 349: 151–154

Hughes CP, Berg L, Danziger WL, Coben LA et al (1982) A new clinical scale for the staging of dementia. Br J Psychiatry 140: 566–572

Jack CR, Petersen RC, O'Brien PC, Tangalos EG (1992) MR-based hippocampal volumetry in the diagnosis of Alzheimer's disease. Neurology 42: 183–188

Jellinger KA (1996) Diagnostic accuracy of Alzheimer's disease: a clinicopathological study. Acta Neuropathol 91: 219–220

Jobst KA, Smith AD, Barker CS, Wear A et al (1992) Association of atrophy of the medial temporal lobe with reduced blood flow in the posterior parietotemporal cortex in patients with a clinical and pathological diagnosis of Alzheimer's disease. J Neurol Neurosurg Psychiatry 55(3): 190–194

Jobst KA, Hindley NJ, King E, Smith AD (1994) The diagnosis of Alzheimer's disease: a question of image? J Clin Psychiatry 55(11) [Suppl]: 22–31

Jorm AF, Duijn CMV, Chandra V, Fratiglioni L et al (1991) Psychiatric history and related exposures as risk factors for

Alzheimer's disease: a collaborative re-analysis of case-control studies. Int J Epidemiol 20(2)[Suppl 2]: S43–S47

Jorm AF, Christensen H, Henderson AS, Korten AE et al (1994) Complaints of cognitive decline in the elderly: a comparison of reports by subjects and informants in a community survey. Psychol Med 24: 365–374

Katzman R, Lasker B, Bernstein N (1988a) Advances in the diagnosis of dementia: accuracy of diagnosis and consequences of misdiagnosis of disorders causing dementia. In: Terry RD (ed) Aging and the brain. Raven, New York, pp 17–62

Katzman R, Terry R, DeTeresa R, Brown T et al (1988b) Clinical, pathological, and neurochemical changes in dementia: a subgroup with preserved mental status and numerous neocortical plaques. Ann Neurol 23: 138–144

*Kennedy AM, Frackowiack RSJ, Newman SK (1995) Deficits in cerebral glucose metabolism demonstrated by positron emission tomography in individuals at risk of familial Alzheimer's disease. Neurosci Lett 186: 17–20

**Khachaturian ZW (1985) Diagnosis of Alzheimer's disease. Arch Neurol 42: 1097–1105

Killiany RJ, Moss MB, Albert MS, Sandor T et al (1993) Temporal lobe regions on magnetic resonance imaging identify patients with early Alzheimer's disease. Arch Neurol 50: 949–954

Klatka LA, Schiffer RB, Powers JM, Kazee AM (1996) Incorrect diagnosis of Alzheimer's disease. A clinicopathologic study. Arch Neurol 53: 35–42

Kral VA, Emery OB (1989) Long-term follow-up of depressive pseudodementia of the aged. Can J Psychiatry 34: 445–446

Kukull WA, Larson EB, Reifler BV, Lampe TH et al (1990a) Interrater reliability of Alzheimer's disease diagnosis. Neurology 40: 257–260

Kukull WA, Larson EB, Reifler BV, Lampe TH et al (1990b) The validity of 3 clinical diagnostic criteria for Alzheimer's disease. Neurology 40: 1364–1369

*Kurz A (1998) "BPSSD": Verhaltensstörungen bei Demenz. Nervenarzt 69: 269–273

Kurz A, Müller U (1997) Apolipoprotein E und Alzheimer-Krankheit. In: Rösler M, Retz W, Thome J (eds) Alzheimer-Krankheit. Abgrenzung normalen Alterns, Epidemiologie, Ätiologie, Pathogenese, Klinik, Behandlung, Ethik. Deutscher Studien-Verlag, Weinheim, pp 144–151

Kurz A, Haupt M, Pollmann S, Romero B (1992) Alzheimer's disease: is there evidence of phenomenological subtypes? Dementia 3: 320–327

Kurz A, Riemenschnieder M, Buch K, Willoch F et al (1998) Tau protein in cerebrospinal fluid is significantly increased at the earliest clinical stage of Alzheimer disease. Alzheimer Dis Assoc Disord 12: 372–377

Landis R, Cummings JL (1986) Loss of topographic familiarity. Arch Neurol 43: 132–136

Lauter H, Kurz A (1989) Demenzerkrankungen im mittleren und höheren Lebensalter. In: Kisker KP, Lauter H, Meyer JE, Strömgren E (eds) Psychiatrie der Gegenwart, vol 8, 3rd edn. Springer, Berlin Heidelberg New York, pp 135–200

Lazeyras F, Charles HC, Tupler LA, Erickson R et al (1998) Metabolic brain mapping in Alzheimer's disease using proton magnetic resonance spectroscopy. Psychiatry Res 82: 95–106

Lehiricy SM, Baulac M, Chivas J (1994) Amygdalo-hippocampal MR volume measurements in the early stages of Alzheimer disease. Am J Neuroradiol 15: 927–937

Lennox A, Karlinsky H, Meschino W, Buchanan JA et al (1994) Molecular genetic predictive testing for Alzheimer's disease: deliberations and preliminary recommendations. Alzheimer Dis Assoc Disord 8(2): 126–147

Linn RT, Wolf PA, Bachman DL, Knoefel JE et al (1995) The 'preclinical phase' of probable Alzheimer's disease. A 13-year prospective study of the Framingham cohort. Arch Neurol 52: 485–490

Lopez OL, Swihart AA, Becker JT, Reinmuth OM et al (1990) Reliability of NINCDS-ADRDA clinical criteria for the diagnosis of Alzheimer's disease. Neurology 40: 1517–1522

Mackenzie TB, Robiner WN, Knopman DS (1989) Differences between patient and family assessments of depression in Alzheimer's disease. Am J Psychiatry 14: 1174–1178

McKhann G, Folstein M, Katzman R, Price D et al (1984) Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA work group under the auspices of Department of health and Human Services Task Force on Alzheimer's Disease. Neurology 34: 939–944

Mielke R, Heiss WD (1998) Positron emission tomography for diagnosis of Alzheimer's disease and vascular dementia. J Neural Transm Suppl 53: 237–250

Mielke R, Herholz K, Grond M, Kessler J et al (1994) Clinical deterioration in probable Alzheimer's disease correlates with progressive metabolic impairment of association areas. Dementia 5: 36–41

Mirra SS, Heyman A, McKeel D, Sumi SM et al (1991) The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). II. Standardization of the neuropathologic assessment of Alzheimer's disease. Neurology 41: 479–486

Nagy Z, Esiri MM, Jobst KA, Morris JH et al (1995) Relative roles of plaques and tangles in the dementia of Alzheimer's disease: correlations using three sets of neuropathological criteria. Dementia 6: 21–31

*Newman SK, Warrington EK, Kennedy AM, Rossor MN (1994) The earliest cognitive change in a person with familial Alzheimer's disease: presymptomatic neuropsychological features in a pedigree with familial Alzheimer's disease confirmed at necropsy. J Neurol Neurosurg Psychiatry 57: 967–972

Nordberg A, Amberla K, Shigeta M, Lundqvist H et al (1998) Long-term tacrine treatment in three mild Alzheimer patients: effects on nicotinic receptors, cerebral blood flow, glucose metabolism, EEG, and cognitive abilities. Alzheimer Dis Assoc Disord 12: 228–237

O'Brien MD (1988) Vascular dementia is underdiagnosed. Arch Neurol 45: 797–798

Oppenheim G (1994) The earliest signs of Alzheimer's disease. J Geriatr Psychiatry Neurol 7: 116–120

Patterson MB, Schnell AH, Martin RJ, Mendez MF et al (1990) Assessment of behavioral and affective symptoms in Alzheimer's disease. J Geriatr Psychiatry Neurol 3: 21–30

Paykel ES, Brayne C, Huppert FA, Gill C et al (1994) Incidence of dementia in a population older than 75 years in the United Kingdom. Arch Gen Psychiatry 51: 325–332

Pearlson GD, Ross CA, Lohr WD, Rovner BW et al (1990) Association between family history of affective disorder and the depressive syndrome of Alzheimer's disease. Am J Psychiatry 147: 452–456

Perry RH, Irving D, Blessed G, Firbairn A et al (1990) Senile dementia of Lewy body type. A clinical and neuropathologically distinct form of Lewy body dementia in the elderly. J Neurol Sci 95: 119–139

Petersen RC, Parisi JE, Hohnson KA, Waring SC et al (1997) Neuropathological findings in patients with a mild cognitive impairment. Neurology 48[Suppl]: A102

Petersen RC, Smith GE, Ivnik RJ, Tangalos EG et al (1995) Apolipoprotein E status as a predictor of the development of Alzheimer's disease in memory-impaired individuals. JAMA 273: 1274–1278

Pietrini P, Furey MI, Alexander GE, Mentis MJ et al (1999) Association between brain functional failure and dementia severity in Alzheimer's disease: resting versus stimulation PET study. Am J Psychiatry 156: 470–473

Reifler BV (1997) Pre-dementia. J Am Geriatr Soc 45: 776–777

Reifler BV, Larson E, Hanley R (1982) Coexistence of cognitive impairment and depression in geriatric outpatients. Am J Psychiatry 139: 623–626

*Reiman EM, Caselli RJ, Yun LS, Chen K et al (1996) Preclinical evidence of Alzheimer's disease in persons homozygous for the ε4 allele for apolipoprotein E. N Engl J Med 334: 752–758

Reisberg B (1983) Alzheimer's disease. Free Press, New York

Reisberg B (1988) Functional assessment staging (FAST). Psychopharmacol Bull 24: 653–659

Reisberg B, Ferris SH, De Leon M, Crook T (1982) The Global Deterioration Scale for assessment of primary degenerative dementia. Am J Psychiatry 139: 1136–1139

*Reischies FM, Geiselmann B, Gessner R, Kanowski S et al (1997) Demenz bei Hochbetagten – Ergebnisse der Berliner Altersstudie. Nervenarzt 68: 719–729

Riemenschneider M, Buch K, Schmolke M, Kurz A et al (1997) Diagnosis of Alzheimer's disease with cerebrospinal fluid tau protein and aspartate aminotransferase. Lancet 350: 784

Román GC, Tatemichi TK, Erkinjuntti T, Cummings JL et al (1993) Vascular dementia: diagnostic criteria for research studies: report of the NINDS-AIREN International Workshop. Neurology 43: 250–260

*Roses A (1996) Apolipoprotein E in neurology. Curr Opin Neurol 9: 265–270

Rossor MN, Kennedy AM, Frackowiack RSJ (1996) Clinical and neuroimaging features of familial Alzheimer's disease. Ann NY Acad Sci 777: 49–56

Rubin EH, Zorumski CF, Burke WJ (1988) Overlapping symptoms of geriatric depression and Alzheimer-type dementia. Hosp Community Psychiatry 39: 1074–1079

Schenk D, Motter R, Kholodenko D, Lieberburg I et al (1996) A beta 42 and tau as markers of Alzheimer's disease. Neurobiol Aging 17(4S):S1

Schuff N, Amend DL, Meyerhoff DJ, Tanabe JL et al (1998) Alzheimer disease: quantitative H-1 MR spectroscopic imaging of frontoparietal brain. Radiology 207: 91–102

Seab JB, Jagust WJ, Wong TS (1988) Quantitative NMR measurements of hippocampal atrophiy in Alzheimer's disease. Magn Reson Med 8: 200–208

Severson MA, Smith GE, Tangalos EG, Petersen RC et al (1994) Patterns and predictors of institutionalization in community-based dementia patients. J Am Geriatr Soc 42(2): 181–185

Sjögren T, Sjögren H, Lindgren AGH (1952) Morbus Alzheimer and morbus Pick. A genetic, clinical, and patho-anatomical study. Acta Psychiatr Neurol Scand Suppl 82: 1–115

Small GW, Mazziotta JC, Collins MT, Baxter LR et al (1995) Apolipoprotein E type 4 allele and cerebral glucose metabolism in relatives at risk for familial Alzheimer disease. JAMA 273: 942–947

Smith AD, Jobst KA (1996) Use of structural imaging to study the progression of Alzheimer's disease. Br Med Bull 52: 575–586

*Smith AD, Jobst KA, Edmonds Z, Hindley NJ et al (1996) Neuroimaging and early Alzheimer's disease. Lancet 348: 829–830

Smith GS, De Leon MJ, George AE, Kluger A et al (1992) Topopraphy of cross-sectional and longitudinal glucose metabolic deficits in Alzheimer's disease. Pathophysiologic implications. Arch Neurol 49: 1142–1150

*Snowdon DA (1997) Aging and Alzheimer's disease: lessons from the Nun Study. Gerontologist 37: 150–156

Snowdon DA, Kemper SJ, Mortimer JA, Greiner LH et al (1996) Linguistic ability in early life and cognitive function and Alzheimer's disease in late life. Findings from the Nun Study. JAMA 275: 528–532

Snowdon DA, Greiner LH, Mortimer JA, Riley KP et al (1997) Brain infarction and the clinical expression of Alzheimer disease. The Nun Study. JAMA 277: 813–817

Stern Y, Alexander GE, Prohovnik I, Mayeux R (1992) Inverse relationship between education and parietotemporal perfusion deficit in Alzheimer's disease. Ann Neurol 32: 371–375

Stern Y, Alexander GE, Prohovnik I, Stricks L et al (1995a) Relationship between lifetime occupation and parietal flow: implications for a reserve against Alzheimer's disease pathology. Neurology 45: 55–60

Stern Y, Tang MX, Denaro J, Mayeux R (1995b) Increased risk of mortality in Alzheimer's disease patients with more advanced educational and occupational attainment. Ann Neurol 37: 590–595

Strauss ME (1995) Ontogeny of depression in Alzheimer's disease. In Bergener M, Brocklehurst JC, Finkel SI (eds) Aging, health and healing. Springer, Berlin Heidelberg New York, pp 441–456

Terry RD, Katzman R, Bick KL (1993) Alzheimer disease. Raven, New York

Tierney MC, Fischer RH, Lewis AJ, Zorzitto ML et al (1988) The NINCDS-ADRDA work group criteria for the clinical diagnosis of probable Alzheimer's disease: a clinicopathologic study of 57 cases. Neurology 138: 359–364

Wahlund LO (1996) Biological markers and diagnostic investigations in Alzheimer's disease. Acta Neurol Scand Suppl 165: 85–91

Wallin A, Blennow K (1996) Clinical subgroups of the Alzheimer syndrome. Acta Neurol Scand Suppl 165: 51–57

Welch HG, Walsh JS, Larson EB (1992) The cost of institutional care in Alzheimer's disease: nursing home and hospital use in a prospective cohort. J Am Geriatr Soc 40(3): 221–224

Woodard JL, Grafton ST, Votaw JR, Green RC et al (1998) Compensatory recruitment of neural resources during overt rehearsal of word lists in Alzheimer's disease. Neuropsychology 12: 491–501

Zubenko GS, Moossy J (1988) Major depression in primary dementia: clinical and neuropathologic correlates. Arch Neurol 45: 1182–1186

Zweig RM, Ross CA, Hedreen JC, Stelle C et al (1988) The neuropathology of aminergic nuclei in Alzheimer's disease. Ann Neurol 24: 233–242

CHAPTER

6

C. Hock, F. Müller-Spahn

Risk Factors for Alzheimer's Disease

Translator: S. Henn

1 Introduction

Since Alzheimer's dementia is an illness probably characterized by an early, preclinical, neuropathological stage which can last several years to perhaps decades, an exact knowledge of specific risk factors could contribute both to early diagnosis of Alzheimer's disease and to estimating an individual prognosis. Reduction of specific risk factors could then become an important component of a multifactorial therapy.

Relative risk (RR) and odds ratio (OR) are frequently used as parameters for the relevance of a potential risk factor. These parameters are calculated from case control studies, which, in addition to cross-sectional studies and prospective longitudinal studies (cohort studies), are one of the most common study designs in psychiatric epidemiology (see also Chap. 2, Vol. 1). In the case control method, investigators attempt to compile two (random) samples in a retrospective analysis, in which only one has a particular distinguishing feature of the illness and relevant variables of age, gender, and other definable variables are the same. Both groups are then investigated with regard to possible risk factors, and the RR or OR values are calculated. The formulas for calculating the RR and OR are explained in Chap. 2 (Vol. 1, Part 1). Values greater than 1 indicate an increased risk, and those less than 1 a decreased risk. In addition, the confidence interval is cited as a measure of scatter. Case control studies have therefore found increasing application, since statistically more reliable risk data can be collected with this method than with the simpler cross-sectional studies. At the same time, costs are lower than in longitudinal studies.

Frequently cited case control studies include, among others, work from the EURODEM Project (European Community Concerted Action Epidemiology and Prevention of Dementia, Department of Epidemiology and Biostatistics, Erasmus University Medical School, Rotterdam, The Netherlands; see Breteler et al. 1991; Graves et al. 1991; Jorm et al. 1991; Mortimer et al. 1991; Rocca et al. 1991; van Duijn et al. 1991, 1994) and work from the Canadian Study of Health and Aging (University of Ottawa, see Canadian Study of Health and Aging 1994). Frequently cited longitudinal studies include, among others, work from the Hisayama Study, a 7-year follow-up study of 828 nondemented inhabitants of the Japanese city of Hisayama (Yoshitake et al. 1995), and work from the current Rotterdam Study, in the course of which baseline investigations with respect to various neurological and internal medicine illnesses were carried out between 1990 and 1993 as a prospective applied project (see Hofman et al. 1997; Ott et al. 1995). In the course of this study of 7983 individuals over 55 years of age from the Ommoord district of Rotterdam, dementia screening tests were carried out in conjunction with recording numerous other parameters.

It should be emphasized that the term "risk factor" is to be understood exclusively in the sense of a statistical association between the risk of becoming ill with Alzheimer's disease and an identifiable factor, and that very different factors affecting the illness can give rise to this association.

2 Family Burden

2.1 Dementia

Most cases of Alzheimer's dementia do not follow an autosomal dominant inheritance pattern, but occur "sporadically." However, the indication of genetic factors and separation from sporadic occurrence for illnesses which begin late is methodologically very difficult. Mathematical models that help to calculate the influence of genetic factors in late-onset sporadic illnesses are currently being developed. Nevertheless, it has also been shown that genetic factors play a certain role in cases of Alzheimer's disease which do not follow an autosomal dominant inheritance pattern. In a meta-analysis of several case control studies, van Duijn et al. (1991) showed that the risk of illness more than tripled with the occurrence of a dementia illness in at least one first-degree relative (OR, 3.5; 95% confidence interval, 2.6–4.6). Hirst et al. (1994) made use of the so-called Kaplan-Meier method, which allows risk to be estimated by using survival curves, in order to show that the life-time risk of becoming ill with dementia, though clearly higher (23.4% $\pm$ 3.0%), does not reach 50% of first-degree relatives of Alzheimer's patients, which would be compatible with an autosomal dominant inheritance course. In the Canadian Study of Health and Aging, the OR for the risk factor of a family history of dementia was 2.62 (range, 1.53–4.51). The OR values increased with the number of affected family members. Thus the RR varied with the amount of familial illness. Several investigators (see Li et al. 1995) were able to show that the probability of an increased family frequency was higher with early onset of illness than with late onset. The cutoff scores, which indicated a difference in genetic risk, were, interestingly enough, not at 65 years, but at 55, 70, and 75 years. These data suggest that genetic factors are most pronounced with very early

onset of Alzheimer's dementia, but also that genetic factors can play a role in late-onset Alzheimer's disease.

Care in the interpretation of genetic information from family "loading numbers" seems to be advisable though, if only because family members can also be exposed in common to certain environmental noxious substances. The present genetic data suggest that, in addition to the single gene-determined familial type of Alzheimer's dementia with autosomal dominant inheritance course, there is also a polygenetically determined type with a weaker familial frequency and a third, purely sporadic type.

2.2 Parkinson's Disease

Idiopathic Parkinson's disease is associated with Alzheimer's dementia on various levels. Parkinson's disease and Alzheimer's dementia exhibit partially similar histopathological changes. In addition, common etiological hypotheses have been formulated concerning energy and oxidation metabolism, among other things. Several case control studies showed a higher number of Alzheimer patients with a family history of Parkinson's disease compared with those in control populations. A meta-analysis in the EURODEM project resulted in an RR of becoming ill with Alzheimer's disease of 2.4 (range, 1.0–5.8) for probands with a positive family history of Parkinson's disease (van Duijn et al. 1991). This result could not be confirmed in the Canadian Study of Health and Aging (OR, 0.86; range, 0.28–2.61). Generally accepted pathogenetic models for the association between Alzheimer's dementia and Parkinson's disease described in the EURODEM project are, however, not available.

2.3 Down Syndrome

Patients with Down syndrome carry a high risk of becoming ill with Alzheimer's disease, especially after the age of 40. This is probably because of the threefold gene dosage of the amyloid precursor protein (APP) gene, which contains the β-amyloid sequence and is located on chromosome 21. It is suspected that the increased APP gene dose leads to an increased production of APP, subsequently followed by an overloading of the nonamyloid breakdown pathway of APP (α-secretase pathway) combined with a higher amyloid production (via the β- and γ-secretase pathway). Mutations in the APP gene are, moreover, associated with familial forms of early-onset Alzheimer's disease. More than 11 studies of the association between Alzheimer's disease and familial history of Down syndrome are now available. Eight studies show an increased frequency of family history of Down syndrome in Alzheimer's patients, although only four of them were statistically significant; three studies showed no association. Based on the lower frequency of Down syndrome in the general population, data about the statistical lack of association must be considered with caution. An RR of 2.7 (range, 1.2–5.7) was found in the EURODEM project (van Duijn et al. 1991).

3 Maternal Age at Birth

Based on the above-mentioned association of Down syndrome with Alzheimer's disease, the maternal age at birth was investigated in view of the risk of becoming ill with Alzheimer's dementia. There are 13 studies on this subject. Five show a statistically significant increased risk for a maternal age of over 40 at the time of birth, while eight studies did not show this result. A meta-analysis of four case control studies in the EURODEM project showed a positive association with an RR of 1.7 (range, 1.0–2.9) for a maternal age of over 40 at birth. Surprisingly, an increased RR (1.5; range, 0.8–3.0) was likewise established for a lower age at birth (15–19 years). Thus both a higher and a lower age at birth may represent a risk factor for Alzheimer's dementia (Rocca et al. 1991). Up to now there have been no conclusive pathophysiological explanations for either of these results.

4 Traumatic Head Injury

An increased RR value of 1.82 (range, 1.20–2.67) for the factor of traumatic brain injury (defined as skull–brain trauma with subsequent loss of consciousness) was calculated after a meta-analysis of 11 case control studies in the course of the EURODEM project (Mortimer et al. 1991). Newer studies showed that skull fracture without loss of consciousness could also present a risk factor (Rasmussen et al. 1995; van Duijn et al. 1994). Neuropathological consequences of head injury can include an increased number of amyloid immunoreactive neurons in the entorhinal cortex and other cortical areas (McKenzie et al. 1994), among other things. These results suggest that a skull fracture, but above all repeated head traumas (e.g., in professional boxers), may promote the amyloid deposition

process. Interestingly, the existence of the ApoE4 allele thus seems to exert a potentiating effect on the traumatic head injury risk factor (Mayeux et al. 1995). This was interpreted as indicating synergistic effects between genetic and environmental risk factors. In an interactions analysis of genetic and environmental factors of Alzheimer's disease, the traumatic head injury risk factor was independent of that of familial history of dementia illness (van Duijn et al. 1994).

5 Depression

From a prospective study by Kral and Emery (1989), indications became apparent that large numbers of depressive symptoms could appear before the onset of dementia. The authors showed that 79% of 44 older patients with depressive symptoms, above all those with simultaneous existence of pronounced cognitive disorders, developed Alzheimer's disease in a subsequent period of 4–18 years. Jorm et al. (1991), in a meta-analysis of 12 case control studies in the course of the EURODEM project, were able to show a statistically significant association of a history of depression with the appearance of Alzheimer's disease (RR, 1.82; range, 1.16–2.86). This association with the history of depression remained significant both for the period under 10 years and for the period over 10 years before the onset of dementia. A separate analysis of patients with early and late onset of Alzheimer's disease showed that the significant positive association was actually valid only for the subgroup of patients with late onset (RR of 4.46 in Alzheimer's disease with late onset vs. RR of 0.71 in Alzheimer's disease with early onset). Antidepressive therapy showed no influence on the RR. It was not further subclassified here, e.g., with regard to the use of substances with or without an anticholinergic spectrum of effect. Devanand et al. (1996) confirmed these results in a longitudinal study of 478 probands from a community in North Manhattan, New York, and calculated an RR of 2.94 (range, 1.76–4.91) for depression, documented on the basis of the Hamilton Rating Scale for Depression. The interactions analysis by van Duijn et al. (1994) showed no differences between Alzheimer's patients with and without a family history of dementia with regard to the risk factor of history of depression. No influence of burdensome life events (e.g., loss of life partner, divorce, loss of a child) showed an association with the appearance of Alzheimer's disease. Regarding the association of depression in the history and the appearance of Alzheimer's disease, the early involvement of the serotonergic and catecholaminergic systems might play a pathogenetic role in the course of neurodegeneration in Alzheimer's disease.

6 Influence of Education and Career

In the course of the above-mentioned longitudinal Rotterdam study, a cross-sectional analysis of a random sample showed a significant increase in the prevalence of Alzheimer's disease in probands with lower levels of education (Ott et al. 1995). The level of education was ascertained with the help of a standard instrument which made it possible to divide education levels into seven groups. The four highest levels of education were grouped together in one category. Thus there were four categories of education: (1) noncompletion of secondary school (26% of the probands), (2) vocational training (20%), (3) technical colleges (15%), and (4) higher education and university (39%). Using education level 4 as a reference, the risk of becoming ill with Alzheimer's disease rose with education level 2 to an RR value of 2.0 (range, 1.3–4.1) and with education level 1 to an RR value of 3.2 (range, 2.2–4.6). In the latter case, the risk for other forms of depression, including vascular dementia, also rose. This result agrees with at least seven other published articles. A certain diagnostic bias might be present due to the effect of the level of education on performance in the psychometric tests used to screen the patients.

Snowdon et al. (1996), with the help of linguistic methods, analyzed handwritten biographies that were written by novices before they finally entered their order (The Nun Study). These biographical sketches by the 75- to 95-year-old nuns were written an average of 58 years previously. In this study, the biographies of 14 patients who had later developed a confirmed Alzheimer's disease were characterized by linguistic defects and distinctive features, such as poverty of ideas and a low level of grammatical complexity. In the nuns in whom cognitive limitations were determined at higher ages, too, the analysis of biographies written during an earlier period of their live already showed comparable defects. Linguistic disorders at an early age were judged to be predictors for cognitive disorders at higher ages.

Stern et al. (1995) have pointed out that the survival time following diagnosis of Alzheimer's disease in patients with a high level of education is lower than in patients having less education. This was interpreted to mean that the causative neurodegeneration in patients with high educational levels is already further advanced at the time of diagnosis. A further possibility which needs to be considered is that factors of poor life style and nutrition, as well as higher exposure to

toxins, might also play a role which cannot be separated from that of the level of education.

7 Vascular Risks

Vascular factors were associated at various levels with Alzheimer's disease. Pronounced degenerative changes of cerebral microvasculature exist in Alzheimer's disease, and significant amyloid deposits also are found in the vascular system outside the brain parenchyma. A diffuse reduction in the density of bone marrow stores is associated with disease of the small vessels, but also with progression of Alzheimer's disease.

In the course of the Rotterdam study, Hofman et al. (1997) investigated 282 patients with dementia and compared them with 1698 controls with regard to indications for atherosclerosis. With the aid of ultrasound tests of vessels in the neck and blood pressure measurements, a scale with degrees of severity from 0 to 3 (absence of to severe atherosclerosis) was drawn up. All indicators for atherosclerosis increased the risk of becoming ill with Alzheimer's disease (OR, 1.3–1.8). The existence of severe atherosclerosis increased the OR to 3.0 (range, 1.5–6.0); with the additional presence of an ApoE4 genotype, the OR rose to 3.9 (1.6–9.6). In comparison, the OR for vascular dementia in the latter group was 19.8 (range, 4.1–95.0). Thus these studies provided new statistical indications of a clear association of atherosclerosis with Alzheimer's disease and confirmed a very strong association with vascular dementia.

In the Hisayama study mentioned at the beginning, various vascular factors were investigated with regard to the risk of Alzheimer's disease (e.g., smoking, overweight, high blood pressure, increased blood fats differentiated after subfractionation, diabetes mellitus). Significantly increased RR values were shown for lower high-density lipoprotein (HDL) cholesterol levels as well as for diabetes mellitus, frequently associated with angiopathies (RR, 2.18; range, 0.97–4.90) (Yoshitake et al. 1995).

Some studies were in agreement in showing a significantly lower risk for smokers. Graves et al. (1991), in a meta-analysis of 11 case control studies, calculated an RR value of 0.78 (range, 0.62–0.98); the Hisayama studies described an RR of 0.73 (range, 0.77–1.80). This result could not, however, be confirmed in the Canadian Study of Health and Aging (OR, 1.17; range, 0.77–1.80) or in a British Medical Research Council study from 1994 (Prince et al. 1994). These different results have not yet been explained.

8 Thyroid Insufficiency

Hyperthyroid and hypothyroid factors and goiter and general thyroid illnesses were investigated in the course of the EURODEM project in a meta-analysis of eight case control studies. Hypothyroidism resulted in a significantly increased risk for Alzheimer's disease with an RR of 2.3 (range, 1.0–5.4) (Breteler et al. 1991). The minimal interval between the appearance of hypothyroidism and the onset of dementia was 6 years, with a mean of 20 years. Attempts at pathogenetic explanation focused on the influence of thyrotrophic hormones on the maturation of the central nervous system (CNS) and on neuronal cell growth. The EURODEM results are in agreement with two other studies, but in contradiction to another seven investigations. Thus the results remain unclear at present.

9 Exposure to Toxic Substances

Exposure to potentially neurotoxic substances has been repeatedly considered as a risk factor for Alzheimer's disease. Results have been shown to be statistically significant for adhesives (OR, 2.16; range, 1.25–3.70), pesticides (OR, 2.17; range, 1.18–3.99) (Canadian Study of Health and Aging 1994), and solvents (OR, 2.3; range, 1.1–4.7) (Kukull et al. 1995).

Aluminum has also been discussed as a risk factor for Alzheimer's disease. Increased aluminum has been found in the plasma of Alzheimer's patients and in the cortex (Basun et al. 1991). Aluminum was also able to be detected in the specific histopathological depositions of Alzheimer's disease, in the senile plaques, and in neurofibrillary bundles. The epidemiological studies carried out so far do not fulfill the standard international criteria with regard to diagnostic precision, definition of population, and statistical power. In 1989, Martyn et al. reported an association between presenile Alzheimer's disease and an aluminum content of over 110 μ/l in drinking water. However, a major criticism of this study was the lack of diagnostic precision: the diagnosis of Alzheimer's disease was made exclusively on the basis of a certificate of referral for computed tomography of the head. The currently accepted criteria given in ICD-10, DSM-IV, or NINCDS-ADRDA (McKhann et al. 1984) were not fulfilled. Moreover, it should be noted that drinking water contributes only about 10% of daily aluminum intake; other, more important sources are tea, antacids, or industrial dust, for example. An association between

Table 1. Risk factors for Alzheimer's disease

Factor	RR/OR value[a]	Reference
Familial burden with dementia	OR 3.5 (2.6–4.6)	van Duijn et al. (1991)
	OR 2.62 (1.53–4.51)	Canadian Study (1994)
Familial burden with Parkinson's disease	OR 2.4 (1.0–5.8)	van Duijn et al. (1991)
	OR 0.86 (0.28–2.61)	Canadian Study (1994)
Familial burden with Down syndrome	RR 2.7 (1.2–5.7)	van Duijn et al. (1991)
Maternal age at birth		
>40 years	RR 1.7 (1.0–2.9)	Rocca et al. (1991)
15–19 years	RR 1.5 (0.8–3.0)	Rocca et al. (1991)
Traumatic head injury	RR 1.82 (1.20–2.67)	Mortimer et al. (1991)
Depression	RR 1.82 (1.16–2.86)	Jorm et al. (1991)
	OR 0.87 (0.46–1.67)	Canadian Study (1994)
	RR 2.94 (1.76–4.91)	Devanand et al. (1996)
Influence of education and career	RR 2.0–3.2 (1.3–4.1, 2.2–4.6)	Ott et al. (1995)
	OR 4.0 (2.49–6.43)	Canadian Study (1994)
	RR 1.18 (0.61–2.27)	Yoshitake et al. (1995)
Vascular risks		
Serious atherosclerosis	OR 3.0 (1.5–6.0)	Hofman et al. (1997)
Diabetes	RR 2.18 (0.97–4.90)	Yoshitake et al. (1995)
Thyroid insufficiency	RR 2.3 (1.0–5.4)	Breteler et al. (1991)
Exposure to toxic substances		
Adhesives	OR 2.16 (1.25–3.70)	Canadian Study (1994)
Pesticides	OR 2.17 (1.18–3.99)	Canadian Study (1994)
Solvents	OR 2.3 (1.1–4.7)	Kukull et al. (1995)
Alcohol abuse	No increased RR	Graves et al. (1991)

RR, relative risk; OR, odds ratio.
[a]Confidence interval, 95%.

Alzheimer's disease and antacid consumption has not yet been substantiated (for a summary, see Nicolini et al. 1991).

10 Alcohol Abuse

In the course of the EURODEM analyses, Graves et al. (1991) documented weekly alcohol consumption and divided it into three categories of light (below 3.2 oz, approx. 96 g), medium (3.2–5.95 oz, approx. 96–185 g), and heavy consumption (more than 5.95 oz, approx. 185 g) of pure alcohol per week. In no category did alcohol consumption influence the risk of becoming ill with Alzheimer's disease. These results agree with those of the Canadian Study and the Hisayama Study.

11 Concluding Comments

The aforementioned risk factors are listed and summarized in Table 1. Pathogenetic models so far account for only part of these factors. In addition to the influence of risk factors (e.g., atherosclerosis, exposure to toxic substances, hypothyroidism), future therapeutic relevance may also lie in deciphering basic mechanisms relating to potential protective factors (e.g., smoking). Such knowledge might contribute to the development of causally orientated therapeutic strategies.

12 References

Basun H, Forssel LG, Wetterberg L, Winblad B (1991) Metals and trace elements in plasma and cerebrospinal fluid in normal aging and Alzheimer's disease. J Neural Transm 4: 231–258

Breteler MMB, van Duijn MC, Chandra V et al (1991) Medical history and the risk of Alzheimer's disease: a collaborative re-analysis of case-control studies. Int J Epidemiol 20(2)[Suppl 2]: S36–S42

**Canadian Study of Health and Aging (1994) The Canadian Study of Health and Aging: risk factors for Alzheimer's disease in Canada. Neurology 44: 2073–2080

*Devanand DP, Sano M, Tang MX et al (1996) Depressed mood and the incidence of Alzheimer's disease in the elderly living in the community. Arch Gen Psychiatry 53: 175–182

Graves AB, van Duijn CM, Chandra V et al (1991) Alcohol and tobacco consumption as risk factors for Alzheimer's disease. Int J Epidemiol 20[Suppl 2]: S48–S57

Hirst C, Yee IML, Sadovnick AD (1994) Familial risks for Alzheimer disease from a population-based series. Genet Epidemiol 11: 365–374

**Hofman A, Ott A, Breteler MMB et al (1997) Atherosclerosis, apolipoprotein E, and prevalence of dementia and Alzheimer's disease in the Rotterdam Study. Lancet 349(9046): 151–154

Jorm AF, van Duijn CM, Chandra V et al (1991) Psychiatric history and related exposures as risk factors for Alzheimer's disease: a collaborative re-analysis of case-control studies. Int J Epidemiol 20(2)[Suppl 2]: S43–S47

*Kral V, Emery O (1989) Long term follow-up of depressive pseudodementia. Can J Psychiatry 34: 445–447

Kukull WA, Larson EB, Bowen JD et al (1995) Solvent exposure as a risk factor for Alzheimer's disease. Am J Epidemiol 141: 1059–1071

*Li G, Silverman JM, Smith CJ et al (1995) Age at onset and familial risk in Alzheimer's disease. Am J Psychiatry 152: 424–430

Martyn CN, Osmond C, Edwardson JA, Barker DJP, Harris EC, Lacey RF (1989) Geographical relation between Alzheimer's disease and aluminium in drinking water. Lancet 1(8629): 59–62

Mayeux R, Ottmann R, Maestre G et al (1995) Synergistic effects of traumatic head injury and apolipoprotein-E4 in patients with Alzheimer's disease. Neurology 45: 555–557

McKenzie JE, Gentleman SM, Roberts GW, Graham DI, Royston MC (1994) Increased numbers of beta APP-immunoreactive neurones in the entorhinal cortex after head injury. Neuroreport 6(1): 161–164

McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM (1984) Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. Neurology 34(7): 939–944

Mortimer JA, van Duijn CM, Chandra V et al (1991) Head trauma as a risk factor for Alzheimer's disease: a collaborative re-analysis of case-control studies. Int J Epidemiol 20(2)[Suppl 2]: S28–S35

Nicolini M, Zatta PF, Corain B (eds) (1991) Aluminium in chemistry, biology and medicine. A series of advances, vol 1. Raven, New York

**Ott A, Breteler MMB, van Harskamp F, Claus JJ, van der Cammen TJM, Grobbee DE, Hofman A (1995) Prevalence of Alzheimer's disease and vascular dementia: association with education. Br Med J 310: 970–973

*Prince M, Cullen M, Mann A (1994) Risk factors for Alzheimer's disease and dementia: a case-control study based on the MRC Elderly Hypertension Trial. Neurology 44: 97–104

Rasmusson DX, Brandt J, Martin DB, Folstein MF (1995) Head injury as a risk factor in Alzheimer's disease. Brain Injury 9: 213–219

Rocca WA, van Duijn CM, Clayton D et al (1991) Maternal age and Alzheimer's disease: a collaborative re-analysis of case-control studies. Int J Epidemiol 20(2)[Suppl 2]: S21–S27

**Snowdon DA, Kemper SJ, Mortimer JA, Greiner LH, Wekstein DR, Markesbery WR (1996) Linguistic ability in early life and cognitive function and Alzheimer's disease in late life – findings from the Nun Study. JAMA 275(7): 528–532

*Stern Y, Tang MX, Denaro J, Mayeux R (1995) Increased risk of mortality in Alzheimer's disease patients with more advanced educational and occupational attainment. Ann Neurol 37: 590–595

van Duijn CM, Clayton D, Chandra V et al (1991) Familial aggregation of Alzheimer's disease and related disorders: a collaborative re-analysis of case-control studies. Int J Epidemiol 20(2)[Suppl 2]: S13–S20

**van Duijn CM, Clayton DG, Chandra V et al (1994) Interaction between genetic and environmental risk factors for Alzheimer's disease. Genet Epidemiol 11: 539–551

**Yoshitake T, Kiyohara Y, Kato I et al (1995) Incidence and risk factors of vascular dementia and Alzheimer's disease in a defined elderly Japanese population: the Hisayama Study. Neurology 45: 1161–1168

CHAPTER

7

R. Sandbrink, K. Beyreuther

Molecular Genetics and Molecular Biology of Alzheimer's Disease

Translator: E. Taub
We thank Prof. Dr. Christian Haass for critically reviewing the manuscript.

1 Introduction

Basic research on the molecular biology of Alzheimer's disease began with the study of the characteristic histopathological features of the disease, which revealed that β-amyloid (Aβ), amyloid precursor protein (APP), and tau protein all participate in the disease process at the molecular level. The molecular genetic characterization of the various genetic etiologies of the disease then led not only to confirmation that Aβ is important in pathogenesis, but also to the identification of further genes and gene products involved in the production of Alzheimer's disease – the presenilins and apolipoprotein E, whose ε4 allele is a risk factor for the disease.

Research today focuses on the identification of further genes conferring a risk of developing the disease and, above all, on the further characterization of the interactions between the various molecules that may participate in the disease process and on the description of their roles in it. It must be borne in mind that Alzheimer's disease is clearly an etiologically heterogeneous group of phenotypically very similar diseases or disease forms. This review article can only cover a relatively small and subjective selection of all relevant studies that have contributed to our current understanding of Alzheimer's disease on the molecular level; for further information, the reader is referred to the works mentioned in the text and to other review articles.

2 Histopathological Features as the Starting Point of Molecular Biological Research

According to the usually applied diagnostic criteria, such as those of the National Institute of Neurological and Communicative Disorders and the Alzheimer's Disease and Related Disorders Association (NINCDS-ADRDA) (McKhann et al. 1984), a definitive diagnosis of Alzheimer's disease can only be made *postmortem* through the additional demonstration of certain neurohistopathological features. These features were first associated with the clinical picture of dementia by Alois Alzheimer in 1906 and were subsequently the principal justification for the definition of Alzheimer's disease as an independent disease entity (Alzheimer 1907). As these histopathological characteristics were the starting point for molecular biological research in Alzheimer's disease, they will now be briefly summarized.

In several areas of the brains of patients with Alzheimer's disease, certain regular and characteristic depositions may be demonstrated with special staining methods, such as amyloid staining with Congo red or thioflavin or silver staining. These include the profuse extracellular amyloid deposits, whose principal component has been identified with molecular biological methods as Aβ (also called βA4-amyloid), a proteolytic breakdown product of APP (see below; Kang et al. 1987; for reviews, see in Müller-Hill and Beyreuther 1989; Selkoe 1994).

Aβ aggregates into fibrils of less than 10 nm in diameter, which are deposited in solid form and are histopathologically recognizable as amyloid plaques of up to 0.2 mm in diameter (Fig. 1). Aβ deposition in cerebral parenchyma may take the form of senile or neuritic amyloid plaques or of amorphous (diffuse) plaques without amyloid properties, depending on the form of aggregation of Aβ, and other factors. One of the most important factors is the length of Aβ. There are two forms of Aβ, which differ with respect to the presence or absence of extra amino acid residues at the carboxyl terminal: the more prevalent form by far, which has 39 or 40 amino acid residues (Aβ40), and another form with 42 or 43 residues (Aβ42/43), which aggregates considerably more readily and is, therefore, more amyloidogenic (for reviews, see Hardy 1997; Sandbrink and Beyreuther 1996; Sandbrink et al. 1996a; Selkoe 1997; Younkin 1995). This is discussed in detail in a later section.

In some patients with Alzheimer's disease, aggregated Aβ additionally appears as vascular amyloid in the cerebral and meningeal blood vessels; this condition is known as congophilic angiopathy or cerebral amyloid angiopathy (CAA). Interestingly, the amyloid of CAA is largely composed of the "short" Aβ40, while plaque amyloid is predominantly composed of the "long" Aβ42/43 (Iwatsubo et al. 1994). Even before Aβ was identified as their major component, it was presumed that plaques, which are demonstrably present in practically all patients with Alzheimer's disease, were closely related to the pathogenesis of the disease. This was because a histologically very similar amyloid pathology is found, at an early age of onset, in Down's syndrome (trisomy 21), in which, in addition to the typical oligophrenia, patients frequently develop an age-dependent dementia syndrome similar to Alzheimer's disease (Iwatsubo et al. 1995). This, too, will be discussed further below.

A further characteristic feature is the formation of so-called neurofibrillary tangles (NFT; see Fig. 1). These are flame-shaped agglomerations of abnormal filaments within neurons; when these cells die, the tangles may persist as extracellular objects ("ghost tangles"). The double helix-shaped, twisted filaments

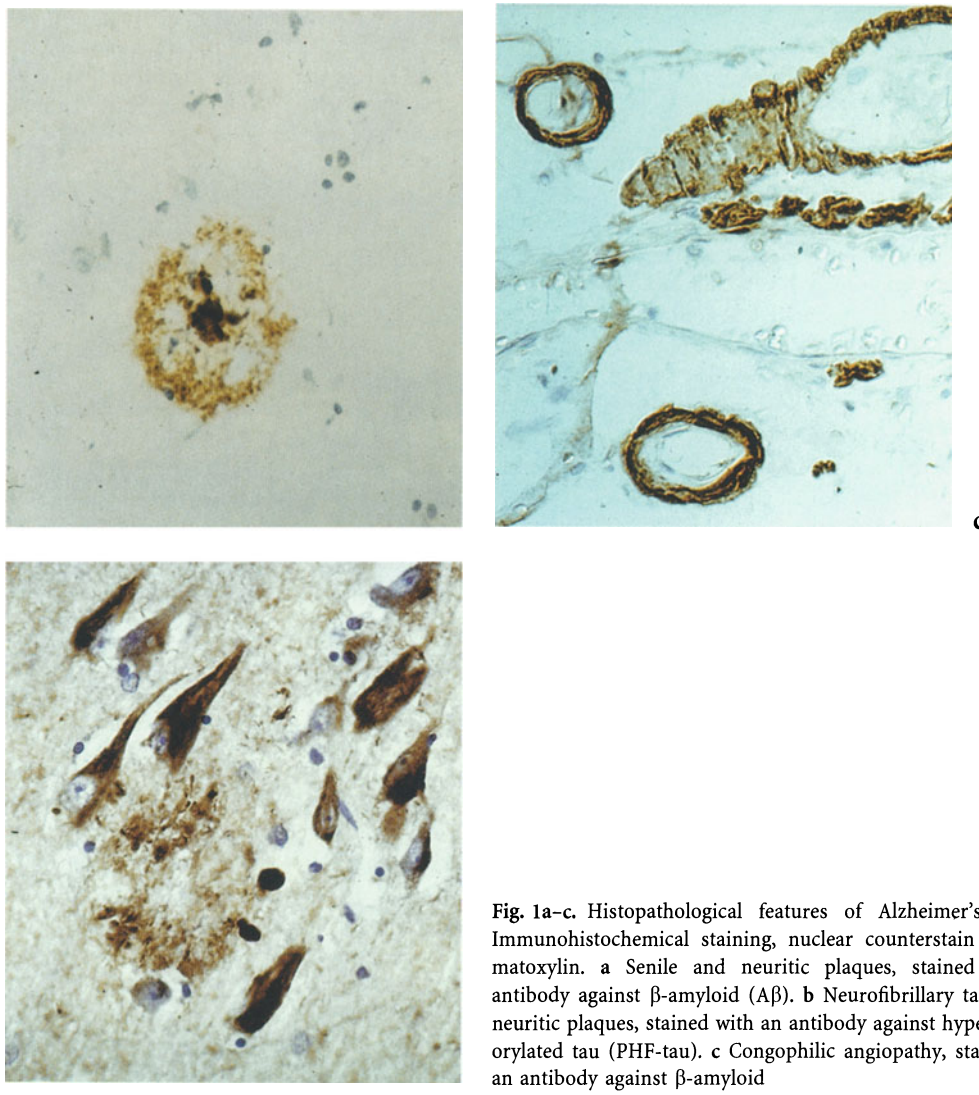


Fig. 1a–c. Histopathological features of Alzheimer's disease. Immunohistochemical staining, nuclear counterstain with hematoxylin. **a** Senile and neuritic plaques, stained with an antibody against β-amyloid (Aβ). **b** Neurofibrillary tangles and neuritic plaques, stained with an antibody against hyperphosphorylated tau (PHF-tau). **c** Congophilic angiopathy, stained with an antibody against β-amyloid

of the neurofibrillary tangles are called paired helical filaments (PHF) (Kidd 1963). These are aggregates with a diameter of 10–20 nm that are also found in a similar form in neuritic plaques and as threadlike structures in the neuropil (see Fig. 1).

These neurofibrillary tangles are primarily composed of pathologically altered tau protein, a neuronal microtubule-associated protein (Grundke Iqbal et al. 1986a,b; Kosik et al. 1986). The tau protein of paired helical filaments, unlike normal, soluble tau protein, is not bound to microtubules, but rather highly phosphorylated and partially fragmented (see below; for reviews, see Billingsley and Kincaid 1997; Goedert 1995; Trojanowski and Lee 1995). Neurofibrillary tangles are also found in other neurodegenerative diseases (for a review, see Feany and Dickson 1996). Furthermore, other varieties of tau protein deposition in neurons and glial cells are found in other diseases, including frontotemporal dementia with parkinsonism of chromosome-17 type (FTDP-17), which, as recently demonstrated, is caused by mutations in the tau protein gene (see below).

Both amyloid plaques and neurofibrillary tangles also appear in clinically healthy individuals at a rate that increases with age. This phenomenon may be interpreted as representing a "silent" preclinical phase of the disease; in other words, the first histopathological changes may take place several decades before the overt onset of the disease. With regard to amyloid deposition, the molecular genetic findings in hereditary (familial) Alzheimer's disease (discussed below) gave rise to the hypothesis of the Aβ cascade, according to which the formation and/or deposition of Aβ (in particular, Aβ42/43) sets in motion a complex series of pathological changes, ultimately leading to neurodegeneration, which produces the clinical picture of dementia (for reviews, see Hardy 1997; Sandbrink and Beyreuther 1996; Sandbrink et al.

1996a; Selkoe 1997; Younkin 1995). Before discussing these matters in detail, we will first provide a brief survey of the clinical genetic and molecular genetic findings in Alzheimer's disease.

3 Clinical Genetics and Etiological Heterogeneity

Alzheimer's disease accounts for at least half of all cases of dementia. In the English-speaking world, the disease is customarily designated as late-onset Alzheimer's disease (LOAD) when it becomes clinically manifest at an age of 65 years or older, and as early-onset Alzheimer's disease (EOAD) when it becomes manifest at an earlier age (in approximately 15%–20% of all patients with Alzheimer's disease). In the German-speaking countries, the same entities are often also called *senile Demenz vom Alzheimer Typ* (senile dementia of Alzheimer type, SDAT) and *präsenile Demenz* (presenile dementia), and the latter term has long been understood to refer to Alzheimer's disease in the proper sense. At present, the segregation of presenile cases into a separate disease entity on the basis of age alone is no longer viewed as justified, because the senile and presenile types have an essentially identical clinical course and identical histopathological features. An increasing number of authors now prefer to place the border between LOAD and EOAD at 60 years (EOAD patients would then account for 5%–10% of the total), as this is thought to be more in accordance with the current understanding of the various genetic etiologies of the disease.

Familial cases of Alzheimer's disease with autosomal dominant inheritance and complete penetrance were first described more than 60 years ago (Lowenberg and Waggoner 1934). In the prevailing current usage, as in this chapter, only these truly hereditary forms of Alzheimer's disease are referred to as familial Alzheimer's disease (FAD). For a large fraction of these FAD cases, the etiologically responsible, "deterministic" FAD gene can be identified. As discussed further below, the important genes are those encoding the presenilins (PS1 and PS2) and APP. The onset of disease is earlier in FAD than in Alzheimer's disease in general, but patients with FAD nevertheless account for only a small fraction (estimated at 5%–10%) of patients with Alzheimer's disease of onset before the age of 60 (Cruts et al. 1998; Sandbrink et al. 1996a). They thus presumably account for less than 1% of all cases of Alzheimer's disease.

In the great majority of patients with Alzheimer's disease, however, the disease is thought to be of multifactorial origin. Those cases often called "familial" in the purely descriptive sense, i.e. cases of patients with at least one first-degree relative with Alzheimer's disease, also belong in this category (for a review, see Sandbrink et al. 1996a). In such cases, which make up approximately 40%–50% of the total, a combination of genetic and other factors is thought to be etiologically responsible for the disease, just as in the so-called "sporadic" cases. As discussed below, the ε4 allele of the apolipoprotein E (APOE) gene has been identified as a genetic risk factor ("susceptibility gene").

A meta-analysis of several case-control studies yielded the conclusion that the first-degree relatives of a patient with Alzheimer's disease have a risk of contracting the disease that is approximately 3.5 times higher (confidence interval, 2.6–4.6) than that of control subjects (van Duijn et al. 1991). This finding was essentially confirmed by a more recent meta-analysis (Lautenschlager et al. 1996). In most of these studies, no statistically significant difference was found between EOAD and LOAD patients, which indicates that the patient cohorts studied included only a very small number of FAD patients. The clinical concept that the diagnosis of "probable Alzheimer's disease" is made by exclusion of other possible causes of dementia allows for the possibility that the disease may be heterogeneous (McKhann et al. 1984). As is now clear from the existence of different forms of Alzheimer's disease with different degrees of heritability, as well as from the existence of different molecular genetic findings, the disease is heterogeneous in etiology and might, indeed, be referred to as the group of "Alzheimer-type diseases," a term being used with increasing frequency. Despite many attempts, it has not been possible to date to provide a clear system of classification of subgroups of Alzheimer's dementia based on their clinical manifestations alone, although there may well be differences in the individual distribution of the cognitive and neurological deficits. Molecular biochemical methods might conceivably enable parameters to be defined which, in addition to the molecular genetic features described below, could be used to develop a reliable subclassification of Alzheimer's disease. This would be of great importance not only to scientific research, but also to treatment.

4 Molecular Genetics of Alzheimer's Disease

4.1 Amyloid Precursor Protein Gene

Three genes have been identified to date in which mutations can give rise to a familial form of Alzheimer's disease (FAD) (for reviews, see Hardy 1997; Sandbrink

et al. 1996a; Selkoe 1997). The earliest FAD gene to be identified was the APP gene on chromosome 21, i.e. the gene encoding the precursor protein of Aβ (Kang et al. 1987). APP is a glycoprotein with a single transmembrane domain that exists in eight different forms, depending on alternative splicing of three of a total of 18 exons (exons 7, 8, and 15) (Sandbrink et al. 1994a); these forms are named with reference to their length in amino acid residues (see below). The APP gene was recognized as a candidate FAD gene as soon as it was cloned in 1987. An FAD locus on chromosome 21 was demonstrated at approximately the same time (the AD1 locus).

There had been much earlier evidence for the existence on chromosome 21 (at least) of a genetic locus which, even in a nonmutated form, might be associated with the occurrence of pathological changes similar to those of Alzheimer's disease through a gene dosage effect: as already mentioned, such changes are regularly observed in Down's syndrome (trisomy 21). The first known mutation of the APP gene (at position 22 of the Aβ sequence) was described in 1990 (Levy et al. 1990; Van Broeckhoven et al. 1990). This APP-E693Q mutation was demonstrated in patients with hereditary cerebral hemorrhage with amyloidosis of the Dutch type (HCHWA-D), a rare disease in which recurrent cerebral hemorrhages occur and whose ultimate prognosis is poor; the hemorrhages occur because of massive Aβ deposition in the meningeal and cerebral microvasculature (see Fig. 1). Mutations of the APP gene are usually given geographical names; thus APP-E693Q is referred to as the Dutch mutation.

Shortly afterward, several point mutations at the valine site at position 46 of the Aβ sequence were demonstrated; these mutations are associated with an EOFAD phenotype (FAD with early onset of disease) (Chartier Harlin et al. 1991; Goate et al. 1991; Murrell et al. 1991). One of these mutations, the APP-V717I mutation (London), is the only APP mutation to date that has been found in a relatively large number of distinct FAD families. The average age of onset of Alzheimer's disease in patients with this V717I mutation is approximately the middle of the sixth decade of life.

A further APP mutation, the A692G mutation at position 21 of the Aβ sequence (Flemish mutation) (Hendriks et al. 1992), gives rise to the phenotype of Alzheimer's disease and/or massive CAA. The variable clinical expression of this mutation and its location one amino acid residue away from the site of the HCHWA-D mutation (E693Q) imply that HCHWA-D and FAD are phenotypic variants of a single disease (allelic phenotypes).

The Swedish FAD mutation, which was found in a large Swedish pedigree and is associated with an average age of onset of 55 years, consists of a K670N + M671L double mutation directly N-terminal to the Aβ sequence (Mullan et al. 1992). The recently identified mutation I716V, or Florida mutation, like

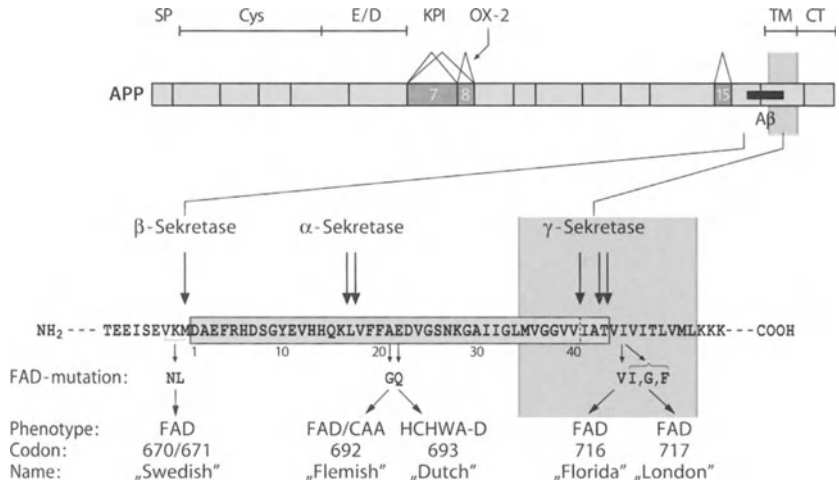


Fig. 2. Amyloid precursor protein (APP) and β-amyloid protein (Aβ). *Top*, exon and domain structure of APP; *bottom*, sequence in the area of the Aβ region with proteolytic cleavage sites (α-, β- and γ-secretase, *arrows*) and the six known familial Alzheimer's disease (FAD) mutations of the APP gene. APP domains: *SP*, signal peptide; *Cys*, cysteine-rich domain; *E/D*, domain with high percentage of negatively charged amino acid residues; *KPI*, Kunitz protease inhibitor domain; *OX-2*, domain resembling the lymphocyte receptor protein MRC OX-2; *TM*, transmembrane domain; *CT*, cytoplasmic domain. The different APP mutations bear geographic names. Phenotype: *FAD*, familial Alzheimer's disease; *CAA*, cerebral amyloid angiopathy (congophilic angiopathy); *HCHWA-D*, hereditary cerebral hemorrhage with amyloidosis – Dutch type

the London mutation, is located at the C-terminal end of the Aβ sequence (Eckman et al. 1997).

The age of onset of FAD resulting from one of these APP mutations varies, depending on the particular mutation, between approximately 40 and approximately 65 years. A typical feature of the FAD-APP mutations, although not of the FAD-presenilin mutations (see below), is an apparent influence of the APOE genotype (see below) on the age of onset of the disease (Houlden et al. 1998). No more than 20 families worldwide have been identified with such FAD-APP mutations; most of them have the V717I mutation (Fig. 2; Table 1; for reviews, see Hardy 1997; Sandbrink et al. 1996a).

As described in greater detail below, all FAD mutations of the APP gene lead to amino acid substitutions in the vicinity of one of the three secretase cleavage sites relevant to the pathogenesis of amyloidosis, which are designated α, β, and γ (see Fig. 2). These mutations lead to increased cleavage of APP at the corresponding sites, which, in turn, leads either to selectively increased formation of Aβ42/43, the longer form of Aβ, as in the London and Florida FAD-APP mutations (Eckman et al. 1997; Suzuki et al. 1994; for reviews, see Hardy 1997; Sandbrink et al. 1996a; Selkoe 1997), or to increased formation of all forms of Aβ, as in the Swedish FAD-APP mutation (Cai et al. 1993; Citron et al. 1992). Similarly, it is presumed that the pathological changes seen in Down's syndrome that resemble those of Alzheimer's disease come about through a gene dosage effect, in which the presence of an extra copy of the APP gene leads to overexpression of APP and thereby to increased production of Aβ.

4.2 Presenilins

Two further FAD genes were identified in 1996; their gene products are now known as presenilins (PS). The presenilin-1 (PS1) gene was the first of these to be found; with the aid of positional cloning, it was identified as the long-sought FAD gene on chromosome 14 (AD3 locus) (Sherrington et al. 1995). PS1 mutations are the most common known cause of FAD, but as far as can be determined from recent studies, they account for only about 20% of all cases of autosomal dominant FAD (Cruts et al. 1998). To date, at least 44 different PS1 mutations, located throughout

Table 1. Genes that play a role in Alzheimer's disease

Chromosome	Gene	Genetic defect or type of predisposition	Age of onset of disease (years)	Percentage of total cases of Alzheimer's disease	Molecular effect
FAD genes (autosomal dominant pattern of inheritance with, as a rule, complete penetrance)					
21	APP	Missense mutations: K670N+M671L ("Swedish"), A692G ("Flemish"), E693Q ("Dutch", phenotype: HCHWA-D), I716 V ("Florida"), V717I or G, F ("London")	Mid-50s (40–65)	<0.5% (≤20 families known, mainly V717I)	K670N+M671L and A692G: → Aβ elevated in general; I716 V and V717I (G, F): → Aβ42/43 elevated
14	PS1 (S182)	Missense mutations: at least 44 different kinds affecting 34 amino acid residues, including two splicing mutations (exons 4 and 9)	Mid-40s (24–78)	<1%–2%?, approx-20%–50% of all FAD cases	All mutations studied to date: → Aβ42/43 elevated
1	PS2 (STM2, E5–1)	Missense mutations: N141I (most cases), M239 V	Mid-50s (40–75)	<0.5%	N141I: → Aβ42/43 elevated
Susceptibility genes (risk genes)					
19	APOE	Polymorphism: ε4 allele associated with elevated risk of Alzheimer's disease (allele frequency in the normal population, 14%)	≥55	Approximately 30%–50%? (allele frequency in Alzheimer's disease)	Histopathology: → elevated β-amyloid plaque density

A role in the pathogenesis of Alzheimer's disease is certain for four genes: three familial Alzheimer's disease (FAD) genes and one susceptibility gene (APOE). Other proposed susceptibility genes whose association with the disease has not been confirmed are not listed in this table. The existence of further genes in both groups is assumed.

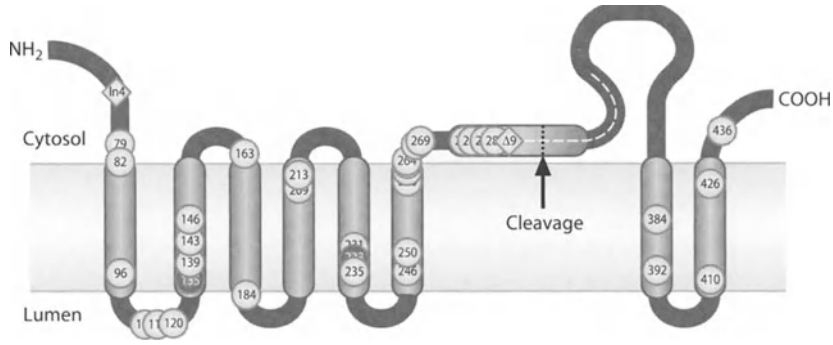


Fig. 3. Model of presenilin structure and position of the known familial Alzheimer's disease (FAD) mutations in the presenilins. *Open circles*, PS1 mutations; *cross-hatched circles*, PS2 mutations. The *Δ*exon-9 mutation (*Δ9*) in the splicing acceptor site in front of exon 9 of the PS1 gene leads to a frame-preserving deletion of exon 9 and to a missense mutation (*dotted*), while the mutation in the 5′ splicing donor region of intron 4 (*In4*) leads to shortened gene products. *Cleavage* denotes the site of proteolytic cleavage of PS1, which is thought to be similar for PS2; other processing pathways also exist

the coding region of the gene and affecting at least 34 different amino acid residues, have been identified in more than 72 families with FAD (and in two sporadic cases of Alzheimer's disease) (Fig. 3; see Table 1).

All but two of these mutations are missense mutations leading to the alteration of a single amino acid residue. One mutation is a nonsense mutation, resulting in the formation of an abnormally short gene product (Tysoe et al. 1998), and another mutation destroys the splice acceptor position before exon 9 and thereby causes a deletion without frameshift, involving a single amino acid substitution (see Fig. 3). Also worthy of mention is the recent demonstration that one particular amino acid substitution in the hydrophilic loop, E318G (exon 9), is not associated with an increased risk of developing Alzheimer's disease. This is apparently a rare polymorphism without pathological relevance (Mattila et al. 1998; Sandbrink et al. 1996b).

The average age of onset of Alzheimer's disease in patients with PS1 mutations ranges from 30 to 58 years, depending on the mutation (average, 45 years). The P117L mutation is associated with a particularly early mean age of onset (30 years); one patient even developed the disease at age 24 and died at 28 (Wisniewski et al. 1998). This mutation is located on exon 5, which encodes transmembrane domain II (see Fig. 3). Other mutations on exon 5 are also associated with a relatively early age of onset (average, approximately 40 years). Patients with the L235P mutation also tend to develop the disease at an early age (average, 32 years). All but one of these FAD-PS1 mutations (Rossor et al. 1996) have complete penetrance as far as can be determined at present. The great majority of PS1 mutations were found in only one FAD family each, so it may be safely assumed that further PS1-FAD mutations will be discovered in the future. The gene for presenilin-2 (PS2) on chromosome 1 (AD4 locus) was discovered only a short time after the discovery of that for PS1 because of the homology of these two substances (Levy-Lahad et al. 1995; Rogaev et al. 1995). Two PS2 mutations, both of the missense type, have been demonstrated in three families, at least one of which was very large. The better known of these mutations is the N141I mutation found in the so-called Volga German pedigree with an average age of onset of disease of 52 years (range, 40–85 years). This mutation is incompletely penetrant. The identification of the presenilin mutations led immediately to the question of whether these mutations, like FAD mutations in the APP gene, increase the production of Aβ42/43. That this is indeed so has been demonstrated for all FAD mutations in the presenilin genes, as for the FAD mutations in the APP gene, both in vivo (with plasma samples) and in vitro (with fibroblasts derived from FAD patients and in transfected cells) (Scheuner et al. 1996; reviews in Haass 1997; Hardy 1997; Sandbrink and Beyreuther 1996; Sandbrink et al. 1996a; Selkoe 1997). An increased amount of Aβ and, in particular, of the highly amyloidogenic protein Aβ42/43 is thus a common feature of all hitherto studied FAD mutations of the three currently known FAD genes.

4.3 Apolipoprotein E

The great majority of cases of Alzheimer's disease are of multifactorial etiology, with the participation of so-called risk or susceptibility genes. The APOE gene on chromosome 19 was identified as such a susceptibility gene in 1993 (Corder et al. 1993; Strittmatter et al. 1993a; for reviews, see Roses 1996; Sandbrink et al. 1996a), and this finding has subsequently been confirmed in a great many studies (AD2 locus). The APOE gene mainly appears as three different alleles, which are called ε2 (frequency in Europe, approximately 7%), ε3 (78%), and ε4 (15%). The corresponding gene products, designated as apoE, differ in two amino acid

residues, at positions 112 (C in E2 and E3, R in E4) and 158 (C in E2, R in E3 and E4).

The presence of the ε4 allele of APOE increases the risk of developing Alzheimer's disease, possibly in part because it lowers the average age of onset of disease. Specifically, a meta-analysis carried out several years ago of studies published up to that time revealed that the frequency of the ε4 allele of APOE in late-onset familial Alzheimer's disease is 48% (95% confidence interval, 45%–51%), and in the late-onset sporadic form 37% (confidence interval, 35%–39%), while it is somewhat lower in early-onset Alzheimer's disease (42% and 28% in familial and sporadic forms, respectively) (Van Gool et al. 1995).

It has been reported that, in LOAD families, the age of onset of Alzheimer's disease is lowered by approximately 59 years by each copy of the ε4 allele. In comparison to individuals without an APOE-ε4 allele, homozygous carriers of the APOE-ε4 gene have an approximately eightfold elevation of the risk of developing Alzheimer's disease at the age of 70 or older. Other studies yielded estimates of the relative risk of Alzheimer's disease (odds ratio) of 2.2–4.4 (in the presence of a single APOE-ε4 allele) and 5.1–17.9 (with two ε4 alleles) (for reviews, see NIA/AA 1996; Roses 1996; Sandbrink et al. 1996a).

A recent, comprehensive meta-analysis including 5930 patients and more than 8000 control subjects yielded a relative risk of developing Alzheimer's disease (odds ratio) of 14.9 (confidence interval, 10.8–20.6) in ε4/ε4 homozygotes of caucasian descent, 3.2 (2.3–3.8) in ε3/ε4 heterozygotes, and 2.6 (1.6–4.0) in ε2/ε4 heterozygotes, all of these in comparison to ε3/ε3 homozygotes (Farrer et al. 1997). Another recent study involving 310 LOAD families revealed a particularly strong effect of APOE-ε4 in the cohort between 61 and 65 years of age. Nonetheless, the influence of APOE-ε4 on the age of onset of disease, as determined by this study, is considerably less than that originally reported (Blacker et al. 1997).

Several studies provide evidence for a protective effect of the APOE-ε2 allele on the risk of developing Alzheimer's disease. The presence of APOE-ε2 not only diminishes the risk of developing the disease, but also raises the age of onset. According to the recent meta-analysis mentioned above, the relative risk in comparison to ε3/ε3 was 0.6 for both ε2/ε2 and ε2/ε3 (confidence intervals, 0.2–2.0 and 0.5–0.8, respectively). There is, however, controversy in the literature regarding the influence of the ε2 allele (for a review, see Sandbrink et al. 1996a).

According to a recent epidemiological study, carriers of the APOE-ε4 gene have an even higher risk of developing AD when arteriosclerotic changes are also present (Hofman et al. 1997). The association of the APOE-ε4 allele with arteriosclerotic changes and myocardial infarcts has been known for many years (see below); the novelty of this finding, therefore, is in the association of arteriosclerotic vascular disease and Alzheimer's disease. It must be pointed out explicitly that the APOE-ε4 "only" elevates the risk of developing the disease; more than half of all patients with Alzheimer's disease do not have even a single APOE-ε4 allele, and only some of the carriers of the ε4 gene are actually affected by the disease.

Like the FAD mutations, the various alleles of apolipoprotein E also seem to modulate the metabolism of Aβ. The principal piece of evidence for this is the multiply replicated finding of an association of the APOE-ε4 allele with the number and density of Aβ deposits in the brain (Rebeck et al. 1993; Schmechel et al. 1993). Biochemical studies suggest that the Alzheimer's disease-promoting effect is based not on altered production of Aβ, as in the case of the FAD mutations, but rather on increased aggregation or lessened degradation of Aβ (see below). Assuming that apoE has a "chaperone" function, it is conceivable that reduced binding of Aβ to apoE4 promotes the development of the disease, as has been suggested on the basis of in vitro studies (see below; Strittmatter et al. 1993b). In any case, several studies have shown that the overwhelming majority of patients with sporadic Alzheimer's disease do not have increased plasma concentrations of Aβ42/43 (e.g. Ida et al. 1996; Motter et al. 1995).

A recent study revealed that a certain variant of the APOE promoter, based on a polymorphism at position 491 of the APOE gene, seems to be associated with an increased risk of developing Alzheimer's disease, regardless of APOE genotype (ε2, ε3, or ε4) (Bullido et al. 1998). Because this variant is associated with increased expression of the APOE gene, this finding indicates a relationship between apoE formation and the risk of developing Alzheimer's disease. According to another study, a certain variant of another APOE promoter polymorphism is also associated with an increased risk of Alzheimer's disease (Lambert et al. 1998). The results of these studies accord with the observation that the development of amyloid pathology is significantly delayed in APOE knockout mice ($APOE^{-/-}$) that are simultaneously transgenic for the human V717I FAD mutation of the APP gene (see below; Games et al. 1995), as compared to transgenic control animals (Bales et al. 1997).

4.4 Other Possible Susceptibility Genes, Including Mitochondrial Mutations

Several genes other than APOE have been proposed in recent years as possible susceptibility genes, including

the A allele of the α_1-antichymotrypsin gene (ACT-A) and the five-repeat allele of the very low density lipoprotein (VLDL) receptor gene VLDL-R; the general validity of these proposed associations has not yet been conclusively proved. The HLA-A2 gene also seems to be associated with early onset of the disease (for a discussion, see Sandbrink et al. 1996a). Great interest was aroused by the observation that one of two alleles of a polymorphism in intron 8 of the PS1 gene seems to be associated with an increased risk of Alzheimer's disease (Wragg et al. 1996), but this finding could not be replicated in subsequent studies. More recently, an association of the K variant of the butyrylcholinesterase gene with an increased risk of LOAD was reported (Lehmann et al. 1997), as was a possible association between a particular allele of the bleomycin hydrolase gene and Alzheimer's disease (Montoya et al. 1998).

Very recent results confirming the presence of a long-conjectured relationship between Alzheimer's disease and certain mitochondrial mutations are particularly noteworthy (Davis et al. 1997). The abnormally low level of cerebral energy metabolism in Alzheimer's disease has been recognized for years (for a review, see Hoyer 1998). It has now been reported that patients with Alzheimer's disease have a significantly higher frequency of specific mutations of two mitochondrial genes, CO1 and CO2, that code for the catalytic center of cytochrome oxidase and, further, that these mutations occur simultaneously in most cases (Davis et al. 1997). More than 20% of mitochondrial genomes were in the mutated form in more than 60% of the patients with Alzheimer's disease who were studied, while such a high percentage of mutated mitochondrial genomes was found in only 20% of the control subjects. It was shown in vitro that these mutations cause dysfunction of the respiratory chain, one of whose effects is the increased production of free oxygen radicals (Davis et al. 1997). Such free radicals have long been suspected to participate in the pathogenesis of Alzheimer's disease, because they promote Aβ aggregation in vitro and, furthermore, in vitro experiments have revealed increased formation of APP as a result of an abnormality of cytochrome oxidase (see below).

Studies performed even more recently, however, indicate that the association reported by Davis and colleagues can be attributable to an artificial contamination of mitochondrial DNA preparations with nuclear DNA, which contains the sequences of certain CO1 and CO2 pseudogenes (Hirano et al. 1997; Wallace et al. 1997). The validity of the reported association of Alzheimer's disease with mitochondrial mutations has thus been put in doubt and requires further clarification.

5 β-Amyloid Cascade Hypothesis

The identification of FAD mutations in the APP gene was the first demonstration that a genetic alteration of the precursor protein of Aβ suffices to produce Alzheimer's disease (in its autosomal dominant hereditary form). In view of the facts that Aβ deposition is a typical feature of Alzheimer's disease, and that amyloid pathology is also regularly present in combination with dementia in Down's syndrome (trisomy 21, see above), the hypothesis of the so-called Aβ cascade was developed. According to this hypothesis, the formation and/or deposition of Aβ in Alzheimer's disease is thought to lead, by way of a complex pathological cascade, to neurodegeneration, and ultimately to the clinical picture of dementia (Hardy and Allsop 1991; for reviews, see Haass 1997; Hardy 1997; Sandbrink and Beyreuther 1996; Sandbrink et al. 1996a; Selkoe 1997).

As discussed above, molecular biological analysis of the various FAD mutations of the APP gene revealed that the common feature of these mutations is an increased production of the long form of APP, namely, Aβ42/43; its production is either selectively increased, as in the London and Florida mutations, presumably because of increased γ-secretase cleavage (Eckman et al. 1997; Suzuki et al. 1994), or else increased in tandem with increased production of all forms of Aβ, as occurs in the Swedish mutation because of increased β-secretase cleavage (Cai et al. 1993; Citron et al. 1992) and in the Flemish mutation because of decreased α-secretase cleavage (Haass et al. 1994). The Aβ cascade hypothesis was therefore modified to include the formation of Aβ42/43 as the decisive pathogenetic factor (Younkin 1995). The characterization of the presenilin mutations then showed that, to the extent that they have been studied, these mutations also lead to increased formation of Aβ42/43 both in vivo (plasma samples) and in vitro (in fibroblasts from FAD patients and in transfected cells) (Scheuner et al. 1996), as the Aβ cascade hypothesis would predict. Increased formation of Aβ and, in particular, of its strongly amyloidogenic form Aβ42/43 has thus been found in all hitherto studied FAD mutations of the three currently known FAD genes.

According to the Aβ cascade hypothesis, overexpression of Aβ42/43 in mice might also be expected to produce pathological changes similar to those of Alzheimer's disease. Corresponding transgenic animal models have, in fact, been established in the last few years, after many unsuccessful earlier attempts. Extensive amyloid deposition was demonstrated in the brains of a transgenic mouse line (the Athena exemplar

mouse) carrying an APP minigene (APP-cDNA, including abbreviated segments of introns 6, 7, and 8) with the V717F mutation under control of the PDGF promoter, in which there is an approximately fivefold overexpression of APP messenger RNA (APP-mRNA) [80% Kunitz protease inhibitor (KPI)-APP] (Games et al. 1995). It is not yet known with certainty whether these mice are behaviorally abnormal. Remarkably, they seem not to have neurofibrillary tangles, but they do have a certain degree of tau pathology in some of the plaques and in the neuropil (Games et al. 1995; Irizarry et al. 1997).

Extensive age-dependent amyloid deposition and a marked elevation of Aβ concentrations in the brain were also found in another transgenic mouse line (Tg2576), which expresses APP695sw (APP695 with the Swedish double FAD mutation) under control of the prion promoter (Hsiao et al. 1996). In these mice, too, neuritic plaques with reactive astrocytes and microglia seem to be present at 12 months (Frautschy et al. 1998), and they also develop memory deficits (Holcomb et al. 1998; Hsiao et al. 1996). Similar constructs in another murine strain (FVB) led to marked APP expression, severe neurological deficits, and death at the age of approximately 6 months in the absence of any demonstrable amyloid deposition (Hsiao et al. 1995). These findings suggest, and breeding experiments have since confirmed, that there are genetic factors in the murine genome that appear to modulate the formation and deposition of amyloid and the type of neurological manifestations produced.

A further FAD-APP transgenic mouse with severe amyloid pathology was recently described, the Sandoz mouse, which expresses APP751sw (APP751 with the Swedish double FAD mutation) under the control of a Thy1 promoter (Sturchlerpierrat et al. 1997). In all of these successful transgenic "murine models of Alzheimer's disease," the critical factor for the generation of amyloid pathology – aside from the use of a strain of mice with the right genetic background – is the adequate expression of the constructs, which is associated with extensive production of Aβ, in accordance with the amyloid cascade hypothesis.

The elevation of Aβ42/43 in the presenilin FAD mutations, which has been observed both in patients and in transfected cells (see above), was also demonstrated in corresponding transgenic mouse lines, which express PS1 with either the A246E mutation (Borchelt et al. 1996) or the M146L mutation (5.1 line; Duff et al. 1996). Both of these lines, however, lack amyloid pathology at any age studied to date, despite the elevation of Aβ42/43. A doubly transgenic mouse line obtained by crossing the 5.1 line with the Tg2576 line, which is transgenic for APP695sw (see above), was found to have a greater elevation of Aβ42/43 production than either of the singly transgenic lines. This was associated with a more rapid development of amyloid pathology than in the Tg2576 line and also with more pronounced behavioral abnormalities, although no neurofibrillary tangles were found at any age studied to date (Holcomb et al. 1998).

The same was demonstrated for the second mouse line named above; here, too, it was found that doubly transgenic mice expressing both FAD-PS1-cDNA with the A246E mutation and a chimeric murine/human FAD-APP construct with the Swedish mutation develop amyloid pathology significantly earlier than corresponding control animals that express either APP Swedish with wildtype PS1 or only the APP Swedish construct (Borchelt et al. 1997).

The Aβ cascade hypothesis receives further support from findings documenting the neurotoxicity of Aβ42/43, although it is unclear whether its important neurotoxic effects in vivo are actually of the same type (see below). In view of the molecular genetic findings, the Aβ cascade hypothesis is widely accepted today, at least for the familial form of Alzheimer's disease (FAD), even if it is perhaps a simplistic model of the pathogenesis of the disease. The literature does, however, contain discussion of other mechanisms of FAD pathogenesis, which will be dealt with further below.

Moreover, the validity of the Aβ cascade hypothesis in this simple form for the much more common, multifactorially based form of Alzheimer's disease is debated; the pathogenesis of these cases is presumed to be much more complex. It will, therefore, be necessary for us to discuss the molecular biology and cellular biology of APP and Aβ in detail in what follows and thereafter to describe the present state of understanding of further molecules that may be of importance in pathogenesis (presenilin, apolipoprotein E, tau etc.).

6
β-Amyloid and Amyloid Precursor Protein

6.1
Structure and Expression of Amyloid Precursor Protein

APP is a glycoprotein with a receptor-like structure consisting of a large N-terminal domain, a single transmembrane domain, and a short intracytoplasmic C-terminal segment (Dyrks et al. 1988; Kang et al. 1987; see Fig. 2). The Aβ sequence is 39–43 amino acids long; approximately one third of it lies in the transmembrane domain and the rest in the N-terminal ectodomain (see Fig. 2). The latter is characterized by

(proceeding from the N-terminal) a rather large region rich in cysteine lying behind the signal peptide and an acidic domain.

Alternative splicing of three of a total of 18 exons (7, 8, and 15) of the APP gene results in the production of APP in eight different isoforms (Sandbrink et al. 1994a), which are named in accordance with their length in amino acid residues. Neurons which show a particularly marked expression of APP predominantly produce APP695, which was the first APP isoform to be identified (Kang et al. 1987). APP695 was later shown not to contain the regions encoded by exon 7 (the KPI domain) and exon 8 (see Fig. 2). Non-neuronal cells, in contrast, predominantly produce APP isoforms that do possess the KPI domain (Kitaguchi et al. 1988; Ponte et al. 1988; Tanzi et al. 1988); the most important of these are the isoforms APP770 (with exons 7 and 8) and APP751 (with exon 7, but without exon 8), and also the two isoforms L-APP752 and L-APP733, which lack the region encoded by exon 15 (Sandbrink et al. 1994a).

APP isoforms without exon 15 are referred to as L-APP because they were first found in leukocytes and microglia (Konig et al. 1992). These L-APP isoforms can sometimes account for more than 50% of the total APP-mRNA in rat tissue, but they are not produced in neurons, as far as is known (Sandbrink et al. 1994a). Interestingly, it has been found in vitro that the L-APP isoforms are less amyloidogenic than APP695, which is produced mainly by neurons, possibly because of a difference in intracellular sorting (Hartmann et al. 1995). This effect appears to be independent of an L-APP-specific post-translational modification that is enabled by the formation of an ENE-GSG fusion sequence in the absence of the segment encoded by exon 15 (Pangalos et al. 1995a,b; Sandbrink et al. 1995; Thinakaran et al. 1995). This modification creates a recognition signal for a xylosyltransferase that can specifically modify L-APP by the addition of chondroitin sulfate moieties to proteoglycans; the L-APP thus produced is also designated appican (Pangalos et al. 1995a,b).

APP is ubiquitously expressed, with a strongly tissue-dependent distribution of its different isoforms (Sandbrink et al. 1994a). The APP gene thus possesses the characteristic features of a housekeeping gene, in accordance with the absence of a TATA box in the APP promoter (Salbaum et al. 1988). The promoter does contain several stress-responsive elements, including two binding sequences for the transcription factor NF-κB (Grilli et al. 1995; Salbaum et al. 1988). The latter probably form the basis for the induction of APP-mRNA expression, as has been observed in vivo in astrocytes and microglia, e.g. after ischemic lesions or axotomy (e.g. Banati et al. 1993, 1995; Siman et al. 1989).

6.2 Amyloid Precursor Protein Gene Family

APP is only one member of a family of genes displaying remarkable similarities in domain structure and also, to some extent, in alternative splicing. The APP gene family further includes the two APP-like proteins APLP1 and APLP2, which do not, however, contain the Aβ sequence; the different members of the APP gene family differ strongly from each other in this segment (the divergent domain), which lies N-terminal to the transmembrane region. APLP1 and APLP2 do, however, show close homology in their cysteine-rich domains and "acidic" domains as well as in their cytoplasmic portions. The APLP1 gene lacks exons corresponding to exons 7 and 8 of APP and therefore exists in only a single neuronally expressed isoform (Paliga et al. 1997; Wasco et al. 1992).

APLP2 resembles APP more closely than APLP1 does, not only in the extent of its homology to APP (Sprecher et al. 1993; Wasco et al. 1993), but also with respect to alternative splicing (Sandbrink et al. 1994b). The APLP2 gene possesses an alternatively spliced KPI exon (exon 7), although an exon corresponding to exon 8 is lacking. It is noteworthy that a portion of the divergent domain of APLP2 is encoded by an exon that is alternatively spliced in a tissue-dependent manner, like exon 15 of APP. Despite major differences in the sequence of this segment, an ENE-GSG fusion sequence appears in the L-APLP2 isoforms, as in the L-APP isoforms, that can serve as a recognition sequence for a xylosyltransferase (Sandbrink et al. 1995).

6.3 Amyloid Precursor Protein Processing: Secretory Amyloid Precursor Protein and β-Amyloid

APP undergoes numerous post-translational modifications in the endoplasmic reticulum and the Golgi apparatus, including sulfation and glycosylation in the ectodomain region N-terminal to the divergent region (Weidemann et al. 1989; see Fig. 2). Moreover, phosphorylation of APP has been described, not only in the N-terminal ectodomain (Hung and Selkoe 1994), but also in the cytoplasmic segment, in which there are several potential phosphorylation sites (Caporaso et al. 1992).

It was found quite early in the characterization of APP metabolism that APP is proteolytically cleaved N-terminal to the transmembrane domain; corresponding soluble derivatives of APP were demonstrated both in cell culture medium and in vivo (Weidemann et al. 1989). These derivatives are known

as secretory APP (APP_S or sAPP). Because sAPP does not contain the Aβ sequence (at least not completely), it is not a precursor of Aβ. It has been shown that sAPP is typically produced by proteolytic cleavage in the region of lysine 16 of the Aβ sequence, i.e. within the Aβ sequence (Esch et al. 1990).

This cleavage is catalyzed by a protease that has not yet been identified, which is provisionally designated α-secretase. The sAPP derivatives thus produced are correspondingly designated sAPPα. This α-secretase cleavage in the midst of the Aβ sequence prevents the proteolytic liberation of Aβ, which could then be deposited in aggregated form in the senile plaques; it is therefore referred to as the non-amyloidogenic metabolic pathway. The α-secretase cleavage apparently occurs during the transport of APP from the endoplasmic reticulum, via the Golgi apparatus and secretory vesicles, to the plasma membrane, i.e. in the course of the classical secretory metabolic pathway. α-Secretase may, in fact, be a group of several proteases that catalyze proteolysis at certain distances from the cell membrane in a manner largely independent of sequence (Maruyama et al. 1991). In any case, the Flemish FAD mutation referred to above is associated with diminished α-secretase cleavage (Haass et al. 1994).

There must, however, be a second, amyloidogenic metabolic pathway in addition to the nonamyloidogenic degradation of sAPP by α-secretase cleavage. Indeed, it was shown in 1992 that Aβ peptides are produced under physiologic conditions not only in vivo, but also in vitro (Haass et al. 1992a,b; Seubert et al. 1992; Shoji et al. 1992). The as yet unidentified one or more proteases that cleave Aβ in its N-terminal region are provisionally designated β-secretase, and the soluble APP derivatives thus formed are correspondingly designated sAPPβ. In a further step, the potentially amyloidogenic C-terminal derivatives can then be proteolytically cleaved in the C-terminal region, so that the intact Aβ molecule is released. Aβ molecules of different lengths are formed, varying between 39 and 43 amino acid residues long (see Fig. 2). The one or more proteases responsible for this step, the so-called γ-secretase(s), can also further cleave the C-terminal APP derivative formed by α-secretase cleavage, with resulting release of p3 peptide, whose molecular weight (MW) is approximately 3000 (for reviews, see Evin et al. 1994; Haass and Selkoe 1993; Selkoe 1994).

The soluble Aβ and p3 produced in this way were first demonstrated in cell culture media and in body fluids (Haass et al. 1992a,b; Seubert et al. 1992; Shoji et al. 1992). This Aβ turned out to be predominantly Aβ40, i.e. the shorter form of Aβ. The site of production of this secreted Aβ was then found to be a cellular compartment of low pH. It was inferred that APP is degraded by an endosomal-lysosomal pathway: APP molecules of the full length and potentially amyloidogenic C-terminal APP fragments produced by β-secretase cleavage, as well as no longer amyloidogenic APP fragments produced by α-secretase cleavage, are reinternalized from the cell surface and cleaved in an acidic intracellular compartment (by γ-secretase); the Aβ and p3 thus released are then newly secreted (Haass et al. 1992). The reinternalization of APP is probably mediated by an NPXY motif in the cytoplasmic segment of APP, which is also present in many cell surface receptors, e.g. the low density lipoprotein (LDL) receptor, and effects endocytotic uptake in clathrin-coated vesicles, followed by further processing in the endosomal-lysosomal metabolic pathway.

It was recently shown that large amounts of Aβ are present within cells as well, mainly in neurons (Tienari et al. 1997). The Aβ in question here is predominantly the "long" Aβ42/43 molecule, while Aβ40 is found in neurons in much smaller amounts than in non-neuronal cells. This intraneuronal Aβ42/43 in formed mainly in the endoplasmic reticulum, while intraneuronal Aβ40 is produced mainly in the Golgi apparatus and the trans-Golgi network (TGN) (Cook et al. 1997; Hartmann et al. 1997; Wild-Bode et al. 1997). Thus, in neurons (the cells that are damaged by Alzheimer's disease and that also have the most pronounced expression of APP in vivo), the highly amyloidogenic Aβ42/43 is produced in the same cellular compartment as that in which the presenilins are localized.

6.4 Regulation of Amyloid Precursor Protein Processing

A large number of studies document that the proteolytic processing of APP is regulated by numerous physiological and pathophysiological factors. For example, it has been shown that α-secretase cleavage is induced in PC12 cells by acetylcholine and after membrane depolarization (Nitsch et al. 1992). The increase in sAPPα production was more pronounced in NGF-differentiated PC12 cells, in accordance with the increased expression of both muscarinic receptors and APP itself in cells of neuronal phenotype (Haring et al. 1995). The production of sAPPα was similarly induced by electrical stimulation in cultured adult rat hippocampal slices, and maximal production was obtained under conditions considered optimal for the induction of long-term potentiation (LTP) (Nitsch et al. 1993).

The activation of metabotropic glutamate receptors also induces the production of sAPPα in hippocampal cultures (Lee et al. 1995). Induction of sAPPα release was similarly demonstrated for other receptor types

that are intracellularly coupled to the activation of the phospholipase C-dependent signal transduction pathway, i.e. the hydrolysis of phosphatidylinositol-4,5-diphosphate. In accordance with these findings, it had been previously been found that activators of protein kinase C, such as phorbol ester, have similar effects (Buxbaum et al. 1990). As shown by a recent study involving C6 glioma cells, the protein kinase C-dependent induction of sAPPα production and also basal sAPPα production can be inhibited by a high concentration of cyclic amino monophosphate (cAMP) (Efthimiopoulos et al. 1996).

These and other findings reveal that there are complex, phosphorylation-dependent regulatory mechanisms of APP processing. It appears, however, that these mechanisms do not depend upon phosphorylation of APP itself in the region of its cytoplasmic domain, because the effects described above also occur in the presence of a deletion of the cytoplasmic domain (Efthimiopoulos et al. 1996; Hung and Selkoe 1994) or of mutation of the phosphorylation sites that have been discussed (Efthimiopoulos et al. 1994; Jacobsen et al. 1994).

APP processing is also influenced by growth factors and cytokines. It was thus shown that nerve growth factor (NGF) and epidermal growth factor (EGF) induce sAPPα production in PC12 cells (Refolo et al. 1989). In general, factors promoting the production of sAPPα also appear to have neurotrophic effects, e.g. lengthening neuronal life span in vitro or promoting the growth of neurites. These effects are possibly due to the action of sAPPα itself (see below).

In many of these studies, it was found (insofar as the matter was studied) that the induction of sAPPα production was associated with a reduction of Aβ production, and vice versa. Thus the effect of protein kinase C activation on APP processing described above involved not only the induction of sAPPα production, but also reduced formation of Aβ (Hung et al. 1993; Jacobsen et al. 1994). Environmental influences, too, may act similarly; it was shown, for example, that glucose deprivation leads both to increased release of Aβ into the medium and to lessened formation of sAPPα (Gabuzda et al. 1994).

Conditions have also been described, however, under which both sAPPα formation and Aβ production are induced; this is the case when the intracellular calcium concentration is elevated by means of ionophors, for example (Buxbaum et al. 1994; Querfurth and Selkoe 1994). As intracellular calcium concentration can also be elevated by phospholipase C activation, this may provide a mechanism for protein kinase C-independent regulation of APP processing, through the appropriate receptors.

It must be emphasized that the studies published to date have all concerned the release of sAPPα and Aβ into medium. This fact alone should lead us to expect a more or less reciprocal relationship, because secreted Aβ seems to be formed predominantly by the endosomal–lysosomal pathway, which presumably requires endocytosis of the plasma membrane and can therefore no longer lead to Aβ production after α-secretase cleavage of APP. Such a reciprocal relationship need not exist, however, for intraneuronally formed Aβ42/43 and α-secretase. It does seem likely that conditions leading to a slowing of the passage of APP through the endoplasmic reticulum, for example, may be linked to increased production of Aβ42/43 and, simultaneously, to decreased formation of sAPPα (see Wild-Bode et al. 1997). For most of the experimental conditions referred to above, however, the corresponding studies have yet to be performed.

An inverse relationship of this sort between the formation of Aβ42/43 and of sAPPα does, in fact, appear to be present for the known FAD mutations of the APP gene; thus the Swedish FAD mutation also leads to decreased production of sAPPα (Cai et al. 1993; Citron et al. 1992), and similar findings were made for the London FAD mutation (Suzuki et al. 1994).

6.5 Amyloid Precursor Protein Function

The ubiquitous expression of APP indicates that this protein has a very basic function. It is quite conceivable that there is a spectrum of functional activity that depends on the APP isoform and post-translational modifications and, above all, on cell type. It may further be suspected that the transmembrane isoforms of APP and its secretory derivatives (sAPP) have distinct functions, and the possibility of a biological function for Aβ itself has also been discussed (for a review, see Mattson 1997).

Knockout mice whose APP genes have been inactivated ($APP^{-/-}$) have been used to show that lack of the APP gene in mice is associated with only mild neurological deficits at first, including reduced locomotion and gliosis (Zheng et al. 1995). At the age of approximately 7 months, there are cognitive deficits (revealed by the Morris water maze test) and a loss of synapses, and the average life expectancy is reduced (Zheng et al. 1995). Delayed neurite formation and shortened survival times have been described for in vitro cultured hippocampal neurons from these $APP^{-/-}$ mice, and even wild-type neurons in the vicinity of astrocytes from these $APP^{-/-}$ mice showed reduced axon growth and increased dendritic branching (Perez et al. 1997).

The relatively mildly affected phenotype of $APP^{-/-}$ mice probably results from a partial functional substi-

tution for APP by other members of the APP gene family, mainly APLP2, and vice versa (see above); thus APLP2 knockout mice ($APLP2^{-/-}$) also have a mildly affected phenotype, while approximately 80% of double knockout mice ($APP^{-/-}$ und $APLP2^{-/-}$) die in the first week of life, and the surviving 20% have marked neurological abnormalities (von Koch et al. 1997).

The actual function of APP, however, has not yet been precisely characterized. In general, a regulatory role of APP in cell–cell and cell–matrix interactions, in neurite growth, and, possibly, as a transient cell surface receptor is assumed. Such a function would also accord well with the ligand-binding properties of APP and the high degree of APP expression in stimulated, adhesive lymphocytes as well as in activated microglia. In the corresponding resting, nonadhesive cells, no APP expression is detectable.

These and other findings have led to the hypothesis that APP may function in the context of wound healing and repair processes. In neurons, which express APP constitutively in large amounts, APP has been found not only in perikarya but also in axons, dendrites, and synapses. APP is located both in the presynaptic membranes of the axons and in the postsynaptic membranes of the dendrites (Schubert et al. 1991). It appears that APP does not participate in the release of neurotransmitters or their binding to receptors, but, rather, plays a role in the maintenance, repair, and stabilization of synaptic contacts.

A receptor function for APP was suggested as soon as it was cloned, because of its structure as a glycoprotein with a single transmembrane domain (Kang et al. 1987). In accordance with such a function, APP is localized at the cell surface and reinternalized into the endosomal–lysosomal metabolic pathway in a manner mediated by the above-mentioned NPXY sequence; all of these features also characterize the LDL or EGF receptor, for example. Although a number of binding partners for APP have been suggested (as discussed below), a ligand internalized in this manner has not, as yet, been demonstrated.

If APP is to function as a cell surface receptor, coupling to an intracellular signal transduction pathway must also be postulated. It has been suggested that APP might activate the $G_{(0)}$ protein (Nishimoto et al. 1993). This observation in fact leads to an alternative model of Alzheimer's disease pathogenesis, in addition to the amyloid cascade hypothesis; there is evidence that FAD-APP (APP695 with the London mutation) effects a constitutive activation of the $G_{(0)}$ protein in neurons (Okamoto et al. 1996). This might, in turn, induce a predominantly negative transactivation of the cAMP-responsive element (CRE) (Ikezu et al. 1996) and/or a fragmentation of nucleosomal DNA mediated by the $G_{(0)}$ protein (Yamatsuji et al. 1996), as has been observed in apoptosis.

With respect to a possible receptor function of APP, it is also of interest that newly synthesized APP in neurons is transported first into the axons and then from the axons into the dendrites (Simons et al. 1995). The existence of this process, which is known as transcytosis, implies that APP might be able to bind to substances on the surface of axons and then transport them into the cell and to the dendritic surface. Substances that bind to APP and are possibly transported by it include very large molecules of the extracellular matrix, such as collagen and heparin sulfate proteoglycans, and also the two metal ions copper and zinc (Bush et al. 1993; Multhaup et al. 1996). The latter participate as cofactors in many important biological reactions, including the regulation of transcription, wound healing and repair processes, and the defense against reactive oxygen radicals. In the brain, disturbances of zinc and copper ion metabolism lead to severe impairments of neuronal function. It is of particular interest in this connection that it is precisely the Aβ domain of APP that is responsible for the (initial) axonal sorting (Tienari et al. 1996).

The neurotrophic activity of APP and its derivatives, already mentioned in the context of $APP^{-/-}$ knockout mice, has been documented many times. In PC12 cells, for example, both transmembrane APP and sAPP cause increased neurite growth (Milward et al. 1992). In another study, a corresponding effect of sAPP on neuroblastoma cells was traced to a particular region of the molecule (RERMS sequence, amino acids 328–332) (Jin et al. 1994). In the case of transmembrane APP, the promotion of neurite growth may be due to the already mentioned cell adhesion function of APP, which has been repeatedly demonstrated (see e.g. Schubert et al. 1989; Small et al. 1994). The binding of APP to molecules of the extracellular matrix has also been investigated and more precisely characterized in many studies (e.g. Kibbey et al. 1993; Multhaup 1994; Narindrasorasak et al. 1991). Interestingly, the biosynthesis of APP is also influenced through the extracellular matrix (Monning et al. 1995).

Modulation of the intracellular concentration of calcium has been discussed as a possible mechanism for the neurotrophic activity of sAPP, i.e. its promotion of neurite growth and synapse formation. The intracellular calcium concentration of hippocampal neurons in vitro was relatively reduced when sAPP was present, both in the resting condition and in the presence of glutamine-induced calcium inflow (Mattson et al. 1993b). It is suspected that sAPPα limits in this way the rise of intracellular calcium concentration that is responsible for the neurotoxic effect of Aβ (discussed above). In any case, sAPPα seems to protect neurons from Aβ toxicity and other oxidative injuries (Goodman and Mattson 1994). At

the same time, the intracellular calcium concentration also influences APP processing (see above), so complex feedback mechanisms are probably operative. The relationship of APP and its derivatives to calcium homeostasis, and other possible functions of APP, were recently discussed in an extensive review article (Mattson 1997).

In view of the fact that increased Aβ production is often associated with reduced sAPPα production, as discussed in Sect. 6.2, while sAPPα has a neuroprotective effect against a great many damaging processes, including Aβ toxicity (discussed below), many authors have concluded that a reduction of sAPPα itself figures importantly in the pathogenesis of Alzheimer's disease (for a review, see Mattson 1997). Reduced formation of sAPPα was seen in the FAD-APP mutations (see above), so such a mechanism is quite conceivably relevant to the pathogenesis of FAD as well.

The isoforms of APP that contain KPI domains are naturally thought to possess a corresponding function as protease inhibitors. This might be of importance in blood coagulation, for example, as factor IXa is inhibited by KPI-APP (APP770 and APP751) (Mahdi et al. 1995). The secretory derivatives of APP751 (sAPP751) are identical with the protease nexin II and are present as such in high concentration in platelet α-granules, which are secreted during platelet activation (Van Nostrand et al. 1991).

6.6 β-Amyloid Aggregation and Neurotoxicity

According to the amyloid cascade hypothesis, "short" Aβ40 and "long" Aβ42/43 differ considerably in their pathogenetic significance. The physicochemical properties of these two substances are such that Aβ42/43 has a markedly greater tendency toward aggregation. In general terms, it may be suspected that the driving force for the aggregation of Aβ protein into Aβ filaments, of which amyloid plaques are composed, lies in a change of conformation of regions of the Aβ protein that participate in polymerization. As an integral component of APP, this region of Aβ probably has an α-helical structure, while the amyloid filaments in aggregates take the form of a β-pleated sheet. These aggregates are insoluble, and the continued deposition of Aβ eventually leads to the formation of amyloid plaques, which consist of a multitude of such filaments, and are a pathological hallmark of Alzheimer's disease.

It is still not adequately known what factors in Alzheimer's disease provoke the aggregation of Aβ, which is clearly also formed under "normal" conditions, or how these factors may be reinforced. Relevant factors other than the above-mentioned production or impaired degradation of Aβ might include conformation-altering influences. The possible participation of radicals and pathological chaperones in such processes has also been discussed. It has been shown that reactive oxygen compounds can lead to the aggregation of synthetic Aβ protein in vitro (Dyrks et al. 1992). As aggregated Aβ also induces the formation of reactive oxygen radicals – as will be discussed – the possibility exists of a pathogenetically relevant positive feedback mechanism.

The most recent findings concerning the presence of mitochondrial mutations in patients with Alzheimer's disease, and the effect of antioxidants such as tocopherol (vitamin E) and estrogen, support the hypothesis that radicals play an important role in the pathogenesis of Alzheimer's disease. The disease-promoting effect of the E4 form of apolipoprotein E may also be due to a mechanism involving free radicals. A possible chaperone function has also been proposed for apolipoprotein E4 and other proteins, in which these proteins dispose of Aβ in the brain only in the reduced form. Their aggregation-promoting effect might then also be manifest in the setting of an elevated concentration of (free) Aβ, as will be discussed in the next section.

The observation that Aβ is produced in vivo and released by cells cultured in vitro into the culture medium, even under nonpathological conditions, led to the performance of many studies on the question of whether extracellular Aβ has a toxic effect on neurons (for reviews, see Iversen et al. 1995; Mattson 1997; Yankner 1996). In fact, such Aβ toxicity has been repeatedly found, but not in all studies. Later studies revealed that only the aggregated, fibrillary Aβ seems to be toxic and may even induce the formation of hyperphosphorylated tau (see below). This might provide an explanation for the stronger neurotoxic activity of the "long" Aβ42/43 in comparison to Aβ40, in accordance with the "simplest form" of the amyloid cascade hypothesis, which holds that it is the excessive quantities of aggregated Aβ42/43 in the extracellular space that cause toxic injury to neurons and, thereby, the neurodegeneration of Alzheimer's disease. The toxic effect may be due to an elevation of intraneuronal calcium concentration, which is attributed to the irreversible opening of non-*N*-methyl-D-aspartate (non-NMDA) calcium channels (Blanchard et al. 1997; for reviews, see Behl 1997; Mattson 1997; Mattson et al. 1993a).

In many studies, which are discussed in these reviews in greater detail, it was shown that aggregated Aβ induces oxidative stress in vitro, which appears to be associated primarily with an accumulation of hydrogen peroxide (H_2O_2) and membrane lipid oxidation (Behl et al. 1994; Harris et al. 1995). This oxidative stress and/or the possible activation of the RAGE receptor by Aβ (see below) might, in turn, result in the induction of the transcription factor NF-κB, for which a role in the

generation of neurodegenerative illnesses has long been hypothesized (Kaltschmidt et al. 1993) and whose transcription in vivo in fact appears to be activated in the neighborhood of amyloid plaques (Kaltschmidt et al. 1997). This factor might also release a pathogenetically relevant positive feedback mechanism by way of an elevation of APP expression, because corresponding elements are present in the APP promoter (Grilli et al. 1995). Increased neuronal transcriptional activity of NF-κB may have a neuroprotective effect, however, in that it counteracts oxidative stress and is, furthermore, anti-apoptotic (Barger and Mattson 1996; Lezoualc'h et al. 1998), so that the role of NF-κB in the pathogenesis of Alzheimer's disease currently remains unclear (for reviews, see Lezoualc'h and Behl 1998; O'Neill and Kaltschmidt 1997).

It is unclear whether such an extracellular toxic effect of Aβ, and particularly Aβ42/43, is actually relevant in vivo. In view of the micromolar concentrations used in most in vitro experiments on Aβ toxicity, it may be suspected that the local extracellular Aβ concentration in vivo is probably not high enough to have a toxic effect. Microglial cells, however, may function as mediators of Aβ toxicity in vivo, in that extracellular Aβ may stimulate them to release toxic radicals, etc. On the molecular level, specific binding of Aβ to the scavenger receptor of microglia (Elkhoury et al. 1996; Paresce et al. 1996) and/or to the receptor for advanced glycation end products (RAGE; Yan et al. 1996) has been proposed recently; binding of this type might lead, among other effects, to the above-mentioned activation of NF-κB and formation of reactive oxygen radicals. Perhaps this will explain the apparent protective effect of nonsteroidal anti-inflammatory drugs (NSAID) against Alzheimer's disease (Breitner 1996).

It is also worth noting that there appears to be no correlation between the extent of the increase of Aβ and the typical age of onset of illness in a particular FAD mutation, even though, in accordance with this model, an increase of extracellular Aβ42/43 has indeed been found, both in vitro and in vivo, in all FAD mutations. In view of the exclusively intracellular localization of the presenilins, intracellular Aβ may actually be the factor responsible for neurodegeneration; this would, of course, represent a deviation from the simple form of the amyloid cascade hypothesis. As described above, hippocampal neurons in vitro produce large quantities of Aβ42/43 intracellularly (Hartmann et al. 1997; Tienari et al. 1997). It may, therefore, be suspected that the corresponding intracellular compartments contain Aβ42/43 in much higher concentrations than the extracellular space. This Aβ42/43 could then fold into a β-pleated sheet structure within the neuron and interfere with its function. Under certain circumstances, however, it is precisely the still soluble low molecular weight aggregates (e.g. dimers, trimers) that are particularly active in this respect (for a discussion, see Sandbrink and Beyreuther 1996; Wild-Bode et al. 1997).

It should be pointed out once again in this connection that elevated plasma levels of Aβ42/43 are generally not found in patients with sporadic Alzheimer's disease (Ida et al. 1996). Younkin and colleagues found elevated Aβ42/43 levels in only about 10% of such patients studied (Scheuner et al. 1996). This finding, too, is consistent with an intracellular but not extracellular rise in Aβ42/43 concentration (Sandbrink and Beyreuther 1996).

6.7 A Physiological Function of β-Amyloid?

It remains unclear at present whether free Aβ protein has a physiological function or is "merely" a waste product of APP metabolism that takes on pathogenetic significance in high concentration. Furthermore, "short" Aβ40 and "long" Aβ42/43 might possibly have different functions.

A possible physiological function of Aβ has been sought in very many studies in which Aβ was given in far lower concentrations (pico- to nano-molar) than are necessary for its neurotoxic effect. The earliest studies of this type indicated a neurotrophic effect of (nonfibrillary) Aβ; it prolonged the survival of neurons in primary culture (Whitson et al. 1989; Yankner et al. 1990).

Nanomolar concentrations of Aβ are said to produce blockade of a certain potassium channel that also appears to be disturbed in Alzheimer's disease (Etcheberrigaray et al. 1994). In low concentrations, Aβ (in this case Aβ40) seems to promote the phosphorylation of various neuronal proteins at tyrosine residues (Luo et al. 1995) and thereby to lead to the activation of phosphatidylinositol-3-kinase (PI_3-kinase) (Luo et al. 1996). As tyrosine phosphorylation seems to play a decisive role in receptor regulation, e.g. with regard to growth factors and molecules of the extracellular matrix, Aβ in this form might possibly play a role in the modulation of neuronal plasticity and survival.

The induction of reactive oxygen molecules by Aβ may play a role in the regulation of vascular tone. Aβ has been shown to induce the production of superdioxide radical anions in endothelial cells in vitro, leading to inactivation of endothelium-derived relaxing factor (EDRF) and to lipid peroxidation, so that vasoconstriction or a diminution of vasodilatation occurs (Thomas et al. 1996). Aβ apparently also prevents the vasodilator effect of acetylcholine in this way, as has been demonstrated in coronary resistance vessels (Thomas et al. 1997).

A role for Aβ peptides as direct modulators of the neurotransmitter function of acetylcholine has also been proposed (for a review, see Auld et al. 1998). In

very low concentrations, and thus apparently independently of any detectable neurotoxic effect, Aβ seems to inhibit various cholinergic neurotransmitter functions, including the release of acetylcholine (Kar et al. 1996) and the reuptake of choline, as well as the regulation of cholinesterase activity (Pedersen et al. 1996). It is proposed that Aβ may contribute in this way to the often discussed vulnerability of certain cholinergic neuron populations, the loss of whose function is said to be related to the (limited) therapeutic effectiveness of acetylcholinesterase inhibitors in Alzheimer's disease.

7
Presenilins

7.1
Structure and Expression

The two presenilin genes, PS1 and PS2, are remarkably similar in structure: both consist of 13 exons with practically identical exon–intron borders, of which the first three (exons 1a, 1b, and 2 – the numbering is of historical origin) contain no coding (translated) sequences (for a review, see Hutton and Hardy 1997). Alternative splicing of both genes has been reported, but occurs differently in each: in the PS1 gene, for example, a VRSQ motif (codons 26–29), containing a potential phosphorylation site for protein kinase C and casein kinase II is alternatively spliced in a manner largely independent of the tissue in which it is found (Clark et al. 1995). Several other alternatively spliced transcripts have also been described (see Hutton et al. 1996).

The sequences of the presenilins are 67% identical at the protein level, and both are presumed to contain eight hydrophobic transmembrane domains and one membrane-associated domain. These are connected to one another through one large and several small hydrophilic loops (see Fig. 3). Both the N-terminal and the C-terminal segment, and the hydrophilic loops, appear to be cytoplasmically oriented (Doan et al. 1996) in accordance with the lack of a signal peptide. This model of presenilin structure was derived (among other techniques) from an analysis of the sel-12 protein of *Caenorhabditis elegans* (Li and Greenwald 1996), which possesses approximately 50% homology to the presenilins and which presumably has a closely related function in the "notch" signal pathway, as discussed further below (Levitan and Greenwald 1995).

Analysis of the expression of PS1 and PS2 shows that, like APP, both of these are ubiquitously expressed. Interestingly, PS1 mRNA is present in different tissues in very similar quantities, while the expression of PS2 is markedly tissue dependent (Kovacs et al. 1996; Levy-Lahad et al. 1995; Rogaev et al. 1995; Sherrington et al. 1995). The expression of PS1 and that of PS2 in the brain are quite similar and most intense in neurons (Kovacs et al. 1996).

7.2
Cell Biology and Function

The presenilins have been detected in transfected mammalian cells as full-length molecules (MW of PS1, approx. 43,000–45,000; of PS2, approx. 53,000–55,000; Kim et al. 1997a; Podlisny et al. 1997; Thinakaran et al. 1996; Walter et al. 1996). In contrast, few or no full-length presenilin molecules have been found to date in native cell lines or in vivo, because a saturable, constitutive endoproteolytic processing takes place (Kim et al. 1997a; Thinakaran et al. 1996). Proteolytic cleavage occurs in the region of the hydrophilic loops, chiefly at position 292 (Podlisny et al. 1997), resulting in the creation of two fragments, a longer N-terminal fragment (MW, approx. 27,000–28,000) and a shorter C-terminal fragment (MW, approx. 17,000–18,000); the latter, in the case of PS1, seems to be an in vivo substrate for protein kinase C (Seeger et al. 1997; Walter et al. 1997). When the presenilins are overexpressed, the surplus full-length presenilin molecules are degraded by an alternative metabolic pathway (Thinakaran et al. 1997), possibly with the participation of the proteasome; this was proposed for PS2 at least (Kim et al. 1997a). It is suspected that the proteolytic fragments of PS1 form heterodimers (Capell et al. 1998), in accordance with the observation that overexpression of the N-terminal PS1 fragment also leads to accumulation of the C-terminal fragment (Lee et al. 1997).

In addition to this first-described mode of proteolytic splicing, an "alternative" proteolytic splicing of both presenilins at a distal position was recently described (Grunberg et al. 1998; Kim et al. 1997b; Loetscher et al. 1997). This occurs with the assistance of one of the caspases, a gene family of cysteine proteases that play an important role in apoptotic cell death. The presenilins thus appear to be proteolytic substrates in programmed cell death, which indicates that apoptosis-associated presenilin fragments may have proapoptotic properties (for reviews of the cellular biology of the presenilins, see Haass 1997; Hardy 1997; Kim and Tanzi 1997).

The results of in vitro studies according to which FAD presenilin has a stronger apoptosis-inducing effect than wild-type presenilin (Guo et al. 1996; Wolozin et al. 1996) may be relevant in this connection. Both FAD-PS1 and FAD-PS2 were found to be associated with an increased sensitivity to Aβ-induced apoptosis, which, in the case of PS1, might be due to a disturbance of calcium homeostasis (Guo et al. 1996);

in the case of PS2, however, it was described as being mediated by G-protein (Wolozin et al. 1996). Alternative models of FAD pathogenesis other than the amyloid cascade hypothesis may be derived from these experiments (for a review, see Mattson et al. 1998).

At the subcellular level, the presenilins are located mainly in the endoplasmic reticulum and, in lower concentrations, in the Golgi apparatus. Accordingly, the presenilins are found mainly in the somatodendritic compartment of in vitro cultured neurons (Capell et al. 1997; Cook et al. 1996; Kovacs et al. 1996; Walter et al. 1996).

This localization of the presenilins corresponds to the neuronal compartments in which Aβ42/43 appears to be formed, as discussed above (Cook et al. 1997; Hartmann et al. 1997; Wild-Bode et al. 1997). As all hitherto studied FAD mutations in the presenilin genes are associated with increased production of Aβ42/43 both in vivo and in vitro (Borchelt et al. 1996; Citron et al. 1997; Duff et al. 1996; Scheuner et al. 1996), this localization indicates that the presenilins evidently play a role in the cleavage of APP and, in particular, that they influence the activity of γ-secretase (for reviews, see Haass 1997; Hardy 1997; Sandbrink and Beyreuther 1996; Selkoe 1997).

It is not yet clear in detail how the FAD mutations in the presenilin genes induce increased formation of Aβ42/43. It has been shown both for wild-type PS2 and for PS2 FAD mutants that presenilin in fact co-precipitates in vitro with APP, which implies a direct interaction between APP and the presenilins, or at least PS2 (Weidemann et al. 1997). These results were recently confirmed by other authors and were also confirmed for PS1 as well (Xia et al. 1997), but are currently the subject of intense discussion. Another possibility is that the presenilins may influence APP transport and thereby the amyloidogenic processing of APP.

Recent studies of neurons from PS1 knockout mice ($PS1^{-/-}$) in primary cell culture reveal that the absence of PS1 is associated with a dramatic increase in transmembrane C-terminal fragments of APP, the result of lessened γ-secretase activity in the presence of unchanged α- and β-secretase activity (De Strooper et al. 1998). While the secretory derivatives of APP (sAPP) were unchanged, the formation of Aβ, in both Aβ40 and Aβ42/43 forms, was greatly reduced. These findings show that the γ-secretases cut sequentially after α- and β-secretase and that the activity of the γ-secretases is strongly regulated by PS1. A possible explanation would be, for example, that the presenilins bind to the C-terminal fragments of APP as chaperones, and only this complex is accessible to proteolytic cleavage by the γ-secretases.

7.3 Biological Function

Major evidence for a possible biological function of the presenilins also comes from its close structural resemblance to sel-12 (50% homology). Sel-12 is a protein of the nematode *C. elegans* that serves as a coreceptor for lin-12, a notch receptor. Loss-of-function mutations of the sel-12 gene lead in recessive fashion to constitutive activation of lin-12, which results in a defect in egg-laying (Levitan and Greenwald 1995). With the aid of transgenic *C. elegans* lines, it was shown that this phenotype can be reversed not only by wild-type sel-12, but also, to a large extent, by human PS1 and PS2 cDNA (Baumeister et al. 1997; Levitan et al. 1996).

This implies a functional equivalence of sel-12 and the presenilins. The presenilins are thus thought to play a role in fetal development and/or cellular differentiation in mammals, too, through participation in the notch signal pathway. This hypothesis was supported by an analysis of the above-mentioned PS1 knockout mice ($PS1^{-/-}$), which die before or very shortly after birth and have defects in the axial skeleton and in somite segmentation. In the central nervous system, these mice display a loss of neurons and a pronounced disposition toward hemorrhages (Shen et al. 1997; Wong et al. 1997). Similar defects in somite formation have been described for both notch1 knockout mice ($notch1^{-/-}$) (Conlon et al. 1995) and dll1 knockout mice ($dll1^{-/-}$) (Hrabe de Angelis et al. 1997) (dll1 is the mouse gene homologous to δ and thus codes for a ligand of the notch receptor). It is unknown, however, whether and in what way a function of the presenilins in the notch signal pathway might contribute to the pathogenesis of Alzheimer's disease.

8 tau Protein

In 1975, Weingarten et al. described a heat-stable protein (tau) which is apparently able to induce the formation of microtubules (Weingarten et al. 1975). Microtubules are linear polymers in the form of a hollow tube, assembled from tubulin units, and tubulin is a heterodimer of two globular proteins. Microtubules are important components of the neuronal cytoskeleton and play a role in major neuronal functions, such as neurite growth and the transport of cytoplasmic components. A number of proteins, called microtubule-associated proteins (MAP), are bound to microtubules, and tau is one of these.

The tau protein is found primarily in neurons and is predominantly localized in the axon (Binder et al. 1985). In 1986, several groups demonstrated that tau is the major component of paired helical filaments (PHF), and thereby also of the neurofibrillary tangles (NFT) that have been a histopathological hallmark of Alzheimer's disease as mentioned above (Grundke Iqbal et al. 1986a,b; Kosik et al. 1986; Wood et al. 1986). Because of these findings and those of further studies (Lee et al. 1991; Hasegawa et al. 1992), it was concluded that the tau protein of neurofibrillary tangles predominantly has an abnormally hyperphosphorylated form (so-called PHF-tau) (for reviews, see Billingsley and Kincaid 1997; Goedert 1995; Trojanowski and Lee 1995).

Many studies have revealed that tau is actually a family of proteins arising from a single gene by alternative splicing. In addition to the six isoforms of tau that are expressed in the brain (Goedert et al. 1989), there are also other isoforms, including the so-called big tau, that are expressed in the peripheral nervous system and will not be discussed here. Depending on alternative splicing of exon 10, the tau isoforms expressed in the brain contain, at their C-terminal end, either three or four similar repeats of a segment of either 31 or 32 amino acids, which is responsible for the binding of tau to microtubules. At their N-terminal end, the cerebral isoforms of tau have either an insert encoded by exons 2 and 3 (58 amino acids), an insert encoded by exon 2 only (29 amino acids), or no insert at all (Goedert et al. 1989).

The tau protein displays a total of 17 serine–threonine–proline motifs, as well as several more serine–threonine residues, and is therefore a potential substrate for protein kinases of the corresponding specificity. It is phosphorylated in vitro by many different protein kinases, including calcium- and calmodulin-dependent protein kinase II (CaMKII; Steiner et al. 1990), casein kinase II (Greenwood et al. 1994), cAMP-dependent protein kinase (cAMP-PK; Litersky and Johnson 1992), mitogen-activated protein kinase (MAP kinase), also known as ERK2 (Drewes et al. 1992), a neuronal cdc2-like protein kinase (cdk5/p35; Paudel et al. 1993), glycogen synthase kinase 3 (GSK3; Mandelkow et al. 1992), and microtubule affinity-regulating kinase (MARK; Drewes et al. 1995).

The dephosphorylation of tau is catalyzed in vitro by several protein phosphatases, including phosphatase 1 (PP1; Gong et al. 1994a), phosphatase 2A (PP2 A; Goedert et al. 1992; Gong et al. 1994b), and calcium- and calmodulin-dependent phosphatase, also known as phosphatase 2B (PP2B) or calcineurin (Goto et al. 1985). It is not yet known which of these protein kinases and phosphatases are active in vivo. It has been suggested that, under normal in vivo conditions, there is an equilibrium between protein kinases and phosphatases, leading to a stable phosphorylation of "adult" tau at (probably) five epitopes (Watanabe et al. 1993). As early as 1980, there was evidence that the degree of phosphorylation of tau and other microtubule-associated proteins affects their microtubule-stabilizing activity (Jameson et al. 1980). In the case of tau, this activity has been traced in recent studies to very specific phosphorylation sites, of which Ser262 is considered especially important, as this residue is the only one localized within the microtubule-binding domains.

An at least partial hyperphosphorylation of tau was found at 19 different epitopes in the postmortem analysis of the brain of a patient with Alzheimer's disease (Grundke Iqbal et al. 1986b; Lee et al. 1991; Matsuo et al. 1994; Morishima Kawashima et al. 1995). Even if many of these epitopes also seem to be partly phosphorylated in so-called "fetal" tau, the phosphorylation of PHF-tau is generally much more extensive (Morishima Kawashima et al. 1995). The same apparently holds for soluble tau in the brains of patients with Alzheimer's disease (Kopke et al. 1993). It is thought that this finding particularly reflects the protein kinase activity of GSK3 and cdk5/p35 (Morishima Kawashima et al. 1995; Watanabe et al. 1993), and of MARK as well, which induces extensive phosphorylation of Ser262 in vitro (Drewes et al. 1995). Independently of the state of phosphorylation, sulfated glycosaminoglycans such as heparin and heparin sulfate apparently induce the formation of aggregates similar to PHF-tau (Goedert et al. 1996).

Fibrillary Aβ, but not soluble Aβ, has also been found in vitro to induce the phosphorylation of tau at several serine residues in primary cultures of human and rat neurons (Busciglio et al. 1995). In this hyperphosphorylated form, tau was no longer associated with microtubules in the neurons studied and was unable to bind to microtubules in vitro either, but could do so again after renewed dephosphorylation (Busciglio et al. 1995). Thus it is possible that in vivo Aβ formation is responsible for the hyperphosphorylation of tau, and hence for the tau pathology observed in vivo, in Alzheimer's disease.

A recently published study yielded the surprising finding that a homogenate of biopsy material contained demonstrable reactivity for antibodies previously thought to be specific for PHF-tau and fetal tau (Matsuo et al. 1994). This implies a rapid dephosphorylation of this epitope in human adult tau in the postmortem brain by phosphatases, probably including PP2A and/or PP2B. The conclusion must then be drawn that the hyperphosphorylation of tau found in the postmortem brain affected by Alzheimer's disease is primarily the consequence of lessened activity of these phosphatases (Trojanowski and Lee 1995).

In any case, a faulty regulation of the protein kinase–phosphatase system in Alzheimer's disease must be assumed, perhaps mediated by Aβ (Busciglio et al. 1995). In combination with an increase in oxidative stress and in intracellular calcium concentration, perhaps also mediated by Aβ, this could create conditions contributing to neuronal dysfunction in Alzheimer's disease by way of a disturbance of the microtubule-stabilizing (and other) function or functions of tau protein, resulting in damage to the neuronal cytoskeleton (for reviews, see Behl 1997; Smith et al. 1996). Neuronal damage might lead not only to increased APP expression and Aβ formation, which might be pathogenetically relevant in the form of a positive feedback mechanism, but also to increased release of tau into the interstitial fluid. This is thought to explain the frequently replicated observation that the cerebrospinal fluid concentration of tau is significantly higher in patients with Alzheimer's disease, or other neurodegenerative illnesses, than in control subjects (Vandermeeren et al. 1993).

It was recently shown that mutations of the tau gene can produce another neurodegenerative illness known as frontotemporal dementia with parkinsonism of chromosome-17 type (FTDP-17). Individuals suffering from this autosomal dominant hereditary illness have atrophy of the frontal and temporal cortex, the basal ganglia, and the substantia nigra, associated with neuron loss and gliosis. Furthermore, tau deposits are found which, unlike those of Alzheimer's disease, are located both in neurons and in glial cells and also have different structural and biochemical properties. A number of groups have now been able to show that point mutations (missense mutations and splicing mutations) of the tau gene cause the disease FTDP-17 (Hutton et al. 1998; Poorkaj et al. 1998; Spillantini et al. 1998). These findings prove that a disturbance of the structure and function of tau protein suffices for the production of a neurodegenerative disorder leading to dementia, which, however, is not the same as Alzheimer's disease.

9 Apolipoprotein E

Mature apolipoprotein E (apoE) is a protein with 299 amino acid residues whose three most important variants, E2, E3, and E4, differ at two positions (112 and 158; see above). The N-terminal segment of apoE contains a domain that binds to the lipoprotein receptor, and the C-terminal end contains the most important lipoprotein (lipid)-binding domain (for a review, see Weisgraber and Mahley 1996). The three apoE variants apparently differ with respect to the types of lipoprotein to which they preferentially bind; while apoE2 and apoE3 bind preferentially to high-density lipoproteins (HDL), apoE4 binds primarily to VLDL. This is regarded as the reason for the elevated plasma cholesterol and LDL concentrations that have been found in carriers of the APOE-ε4 gene and their ensuing elevated risk of developing atherosclerosis. It is suspected that the influence of APOE genotype on the risk of Alzheimer's disease may also be related to this differential effect on lipid metabolism. This hypothesis was supported by the Rotterdam study referred to above, in which carriers of the APOE-ε4 gene have an even higher risk of developing Alzheimer's disease when arteriosclerotic changes are also present (Hofman et al. 1997). For this reason, the role of apoE in cerebral lipid metabolism is now being intensively studied.

In the brain, apoE is produced by astrocytes. The lipoprotein–apoE particles, bound to VLDL, are taken up by neurons with the aid of either the LDL receptor or the so-called HSPG-LRP pathway, which involves binding to heparin sulfate proteoglycans (HSPG) at the cell surface, followed by internalization via the LDL receptor-related protein (LRP) receptor. By this pathway, apoE might conceivably reach the cytoplasm (Han et al. 1994), although the mechanism for this is not yet adequately understood. In any case, the uptake of apoE4 into neurons has been found to be of much lesser magnitude than that of apoE3 (Nathan et al. 1994); this might be a consequence of differing affinities for the HPSG-LRP receptor. These findings might explain the observation that apoE3, but not apoE4, promotes neurite growth in vitro, presumably by way of binding to this receptor (Holtzman et al. 1995; Nathan et al. 1994; Nathan et al. 1995).

The inhibitory effect of apoE4 on neurite growth may also be explained by its much lower binding affinity to tau, compared with that of apoE3, as observed in vitro (Strittmatter et al. 1994a,b). As the phosphorylation of tau seems to reduce its binding to apoE3, it has been proposed that apoE3, by binding to tau, prevents the hyperphosphorylation of tau and thereby protects microtubules from destabilization (Strittmatter et al. 1994).

Another hypothesis concerning the role of apoE in the pathogenesis of Alzheimer's disease emerges from the in vitro observation that the different variants of apoE have different affinities for Aβ. Lipid-free apoE4 makes a much more stable complex with Aβ than apoE3 (Strittmatter et al. 1993b), although this effect seems to be reversed in the presence of lipids (Ladu et al. 1995). The different affinities of the different apoE variants may explain the multiply replicated observation of an association of the APOE-ε4 allele with the number and density of Aβ deposits in the brain (Rebeck et al. 1993; Schmechel et al. 1993). The findings of these and other studies (Castano et al.

1995; Evans et al. 1995) have led many current authors to hypothesize, as mentioned above, that the Alzheimer's disease-promoting effect of apoE4 is due not to altered production of Aβ protein, as in FAD, but to an increased aggregation or decreased degradation of Aβ.

The recent findings that not only the APOE genotype, but also the intensity of gene expression, influence the risk of developing Alzheimer's disease have already been mentioned (Bullido et al. 1998; Lambert et al. 1997, 1998). It is postulated that formation of larger amounts of apoE is associated with a greater risk of developing Alzheimer's disease, regardless of APOE genotype (Bullido et al. 1998; Lambert et al. 1997). This hypothesis gains further support from the observation that APOE knockout mice ($APOE^{-/-}$) that are simultaneously transgenic for the human V717I-APP-FAD gene (see above) develop amyloid pathology much more slowly than transgenic control animals, although they have the same amount of Aβ in the brain (Bales et al. 1997). These findings indicate that apoE promotes Aβ aggregation in vivo, but to an extent that differs depending on the particular apoE variant or variants present.

10 NAC Precursor/α-Synuclein

In 1993, Ueda and colleagues analyzed an amyloid preparation from the brain of a patient with Alzheimer's disease and identified two previously unknown peptides, whose deposition in neuritic and diffuse plaques, as well as in vascular amyloid, was demonstrated with immunohistochemical methods (Ueda et al. 1993). They were then able to isolate the cDNA of a protein, 140 amino acids in length, that contained these amyloid sequences in two segments of a hydrophobic domain. The authors designated the peptide of these amyloid sequences (MW, approx. 35,000) non-Aβ component of Alzheimer's disease amyloid, or NAC, and its precursor, i.e. NAC precursor (NACP) (Ueda et al. 1993). Analyses of secondary structure revealed that NAC has a marked tendency to assume the structure of a β-pleated sheet, which is presumably the basis of its association with Aβ.

Sequence comparisons revealed that NACP is the human homologue of synuclein, a synaptic protein previously described in the torpedo and in the rat (Campion et al. 1995). There are two synucleins encoded by different genes, α-synuclein (=NACP) and the closely related β-synuclein, both of which are specifically expressed in the brain, are localized in presynaptic terminals, and bind strongly to Aβ (Jensen et al. 1997; Yoshimoto et al. 1995).

It was recently shown in several families that a missense mutation (A53T) in the α-synuclein gene on chromosome 4 can produce an autosomal dominant hereditary form of Parkinson's disease (PARK1; Polymeropoulos et al. 1997). A further mutation (A30P) was identified very recently (Kruger et al. 1998). Aside from the early onset of disease (at an average age of 46 years), the first family to be identified with this disease has a typical and well-documented phenotype of Parkinson's disease, including the formation of histopathologically recognizable Lewy bodies in the brains of the affected individuals.

Lewy bodies are intracytoplasmic inclusion bodies found in approximately 20% of cases of dementia that come to autopsy (McKeith et al. 1996). They have been shown to contain NACP/α-synuclein in aggregate form (Spillantini et al. 1997). Lewy bodies are found not only in Parkinson's disease, but also in related neurodegenerative illnesses, which are currently designated as dementia with Lewy bodies (McKeith et al. 1996). Illnesses of this class include so-called diffuse Lewy body disease and the Lewy body variant of Alzheimer's disease, which lies in a region of transition to Alzheimer's disease and involves not only the deposition of Lewy bodies in the cerebral cortex, but also the typical histopathological abnormalities of Alzheimer's disease (neuritic plaques and NFT). The Lewy body variant of Alzheimer's disease is now increasingly interpreted as a coexistence of Alzheimer's disease and dementia with Lewy bodies (McKeith et al. 1996).

The demonstration of α-synuclein in both Lewy bodies and senile plaques indicates a possible role for NACP/α-synuclein in the pathogenesis not only of Parkinson's disease, but also of Alzheimer's disease (for a discussion, see Heintz and Zoghbi 1997). An earlier study of 26 EOFAD patients revealed no mutation in the α-synuclein gene (Campion et al. 1995), but another study yielded certain pieces of evidence for an interaction of a particular allele of the α-synuclein gene with apolipoprotein E with regard to the risk of developing Alzheimer's disease (Xia et al. 1996). This will certainly require further careful investigation in larger studies.

11 Implications for Diagnosis

The diagnosis of Alzheimer's disease is, above all, a clinical diagnosis and must always be accompanied by a differential diagnostic evaluation for other potential causes of dementia. The accuracy of clinical diagnosis can be improved by the ancillary use of modern diagnostic imaging techniques. To increase diagnostic

accuracy still further, making use of additional biochemical and molecular biological studies of various "biological markers" of the disease may be considered (see Chap. 5, this volume, Part 2).

A molecular genetic study with respect to one of the FAD mutations may be used to confirm, although not to exclude, a clinically suspected diagnosis of FAD if the pedigree data support an autosomal dominant pattern of inheritance and if the age of onset is no higher than 50–55 years (with unambiguous pedigrees, the age of onset may be higher). Even if the age of onset is in the range of 50–55 years, it must first be established whether the APOE-ε4/ε4 genotype is present, in which case the probability of FAD is considerably lower.

A molecular biological investigation with respect to a FAD mutation necessarily begins with the study of the PS1 gene. Because FAD mutations are generally missense mutations, and because such mutations are distributed throughout the coding region of the PS1 gene (exons 3–12), the diagnostic assessment may begin with screening for the presence of such a mutation or may proceed directly to sequencing. If no PS1 mutation is found, a directed assessment with regard to the known APP and PS2 mutations can then be performed. If no FAD mutation is identified in this way, the presence of FAD still cannot be ruled out, not least because it has to be assumed that FAD genes exist that have not yet been identified.

The children of a FAD patient generally have a 50% risk of developing Alzheimer's disease at a comparable age. Thus, as soon as the diagnosis is clinically suspected, the affected patient or caregiver should be informed of the opportunity of receiving thorough human genetic counseling, which should, of course, be available to other family members as well, if they wish. It goes without saying that the appropriate ethical guidelines should be applied in such cases. In Germany, guidelines for the genetic counseling of families of patients with (as yet) untreatable neurodegenerative illnesses have been developed for Huntington's disease by the Professional Association for Medical Genetics (*Berufsverband für medizinische Genetik*). For patients' relatives desiring predictive assessment, these guidelines recommend a neurological examination and accompanying psychotherapeutic care, as well as an adequate period for reflection before the drawing of blood for molecular genetic study (Konsortium zur molekulargenetischen Diagnostik der Huntington-Krankheit 1996).

Similarly, with respect to APOE gene typing, the recommendations of existing consensus statements should be followed, e.g. that of the National Institute of Aging/Alzheimer's Association Working Group (NIA/AA 1996). This consensus statement emphasizes that the sensitivity and specificity of APOE gene typing in asymptomatic persons are not sufficient to allow a recommendation of such gene typing for predictive diagnosis. For the estimation of an individual's risk of developing Alzheimer's disease, too, just as for predictive diagnosis, the data currently available do not justify a recommendation that gene typing be performed in view of the lack of therapeutic consequences of such a determination and the manifold social, legal, and ethical implications that it would have.

In the presence of clinical manifestations of dementia, however, APOE evaluation may be ordered by the physician as a supplementary diagnostic technique, in addition to other diagnostic studies. In such cases, the demonstration of positivity for APOE-ε4 raises the probability that a suspected diagnosis of Alzheimer's disease is correct. This was shown in a recently published study, in which the clinical diagnoses of 2188 patients with dementia were compared with their (later) pathological diagnoses, and the addition of APOE genotypic information was found to enhance the specificity of the clinical diagnosis (Mayeux et al. 1998).

12 Therapeutic Approaches

If intraneuronal Aβ indeed plays a decisive role in the pathogenesis of neurodegeneration in Alzheimer's disease, this still does not rule out an important pathogenetic influence of Aβ-amyloid plaques as well, e.g. through the activation of microglia or the inhibition of axonal transport with consequent accumulation of APP. Assuming that Aβ does play a major pathogenetic role, the Aβ plaques clearly bear a close relation to the pathogenetically relevant parameter, whatever it may be.

Studies concerning the occurrence of Aβ deposition in the general population are, therefore, of great significance. Immunohistochemical studies of postmortem brains in various age categories have revealed, for example, that about 20% of individuals between 50 and 60, and 80% of those between 80 and 90, show Aβ-positive changes (Rumble et al. 1989). If we consider that approximately one in five individuals between the ages of 80 and 90 has Alzheimer's disease, it follows that approximately 30 years pass between the appearance of the earliest Aβ deposits and the onset of manifestations of Alzheimer's disease.

More recent studies further showed that the formation of Aβ plaques is an even earlier sign of Alzheimer's disease; in accordance with the special pathological significance of Aβ42/43, these plaques seem to contain predominantly this longer form of Aβ (Iwatsubo et al. 1994; Tamaoka et al. 1995). If intra-

neuronal processes, as suspected, indeed precede the extracellular deposition of β-amyloid, then the preclinical phase of Alzheimer's disease must be still longer than this. This hypothesis accords with the observation that the phosphorylation of tau protein and the formation of neurofibrillary tangles also considerably precede the clinical onset of the disease, by as long as four decades, according to recent studies (Braak and Braak 1995).

These studies reveal that the preclinical phase of Alzheimer's disease stretches over at least three decades before the first signs of the disease can be recognized. By then, however, the irreparable loss of neurons and synapses in certain regions of the brain has already occurred to the extent of more than 90%. The goal of any truly effective therapy must, therefore, be to delay entry into the disease phase by preventive measures instituted during the preclinical phase. The discoveries of basic research discussed in this chapter yield clues as to how this goal might be achieved.

Six strategies may be considered that either directly address the process of generation of Aβ or else are intended to block the consequences of Aβ formation:

1. Lowering the rate of synthesis of APP itself.
2. Modulating the rate of conversion of APP into Aβ protein, i.e. lowering the rate of formation of Aβ, e.g. by inhibition of the β- or γ-secretases; γ-secretase activity, in particular, might be influenced by modulation of the presenilins, for example.
3. Promotion of Aβ degradation or – more generally – elimination of Aβ from the brain.
4. Inhibition of the formation of the β-pleated-sheet structure of Aβ (by alteration of conformation) and inhibition of Aβ aggregation.
5. Inhibition of a pathological activation of microglia, which might critically reinforce the toxic effect of Aβ or act to mediate this toxicity.
6. Protection of neurons on many levels of the Aβ cascade, e.g. protection against the consequences of Aβ production, of β-pleated-sheet formation, of Aβ aggregation and deposition, and of the pathological activation of microglia, if such indeed occurs.

At present, the modulation of amyloid formation is the most important target for rational therapeutic research. The best opportunity of prolonging the preclinical phase of the disease is thought to lie in the inhibition of Aβ production or aggregation. Beyond this, other approaches include the use of anti-inflammatory substances, which might lessen neuronal damage by counteracting the activation of microglia by amyloid deposition, or the formation of radicals secondary to the accumulation of APP. The most promising concept appears to be that of a combination of measures, which, it is hoped, will lessen neuronal damage, improve neuronal function, and thereby prevent the emergence of the clinical symptoms of Alzheimer's disease within the natural lifetimes of at least some of the patients at risk.

There are increasing signs that the prospect of diagnosing Alzheimer's disease in an early enough stage to enable successful therapy will in fact be achieved. The use of genetic and biological (i.e. phenotypic) markers, in combination with sensitive diagnostic imaging techniques, may one day make this hope a reality. These techniques, however, should not be used for diagnostic purposes in the preclinical phase of the disease until effective prevention or treatment is available. This holds especially for the investigation of susceptibility genes, whose cumulative occurrence might predict a very high risk of developing Alzheimer's disease (for a discussion, see Roses 1998).

13 References

**Alzheimer A (1907) Über eine eigenartige Erkrankung der Hirnrinde. Allg Z Psychiatr Psych Gerichtl Med 64: 146–148

Auld DS, Kar S, Quirion R (1998) Beta-amyloid peptides as direct cholinergic neuromodulators: a missing link? Trends Neurosci 21: 43–49

Bales KR, Verina T, Dodel RC et al (1997) Lack of apolipoprotein E dramatically reduces amyloid beta-peptide deposition. Nat Genet 17: 263–264

Banati RB, Gehrmann J, Czech C et al (1993) Early and rapid de novo synthesis of Alzheimer beta A4-amyloid precursor protein (APP) in activated microglia. Glia 9: 199–210

Banati RB, Gehrmann J, Wiessner C, Hossmann KA, Kreutzberg GW (1995) Glial expression of the beta-amyloid precursor protein (APP) in global ischemia. J Cereb Blood Flow Metab 15: 647–654

Barger SW, Mattson MP (1996) Induction of neuroprotective kappa B-dependent transcription by secreted forms of the Alzheimer's beta-amyloid precursor. Mol Brain Res 40: 116–126

Baumeister R, Leimer U, Zweckbronner J, Jakubek C, Grünberg J, Haass C (1997) The sel-12 mutant phenotype of *C. elegans* is rescued independent of proteolytic processing by wt but not mutant presenilin. Genes Function 1: 149–159

Behl C (1997) Amyloid beta-protein toxicity and oxidative stress in Alzheimer's disease. Cell Tissue Res 290: 471–480

Behl C, Davis JB, Lesley R, Schubert D (1994) Hydrogen peroxide mediates amyloid beta protein toxicity. Cell 77: 817–827

Billingsley ML, Kincaid RL (1997) Regulated phosphorylation and dephosphorylation of tau protein: effects on microtubule interaction, intracellular trafficking and neurodegeneration. Biochem J 323: 577–591

Binder LI, Frankfurter A, Rebhun LI (1985) The distribution of tau in the mammalian central nervous system. J Cell Biol 101: 1371–1378

Blacker D, Haines JL, Rodes L et al (1997) ApoE-4 and age at onset of Alzheimer's disease: the NIMH genetics initiative. Neurology 48: 139–147

Blanchard BJ, Konopka G, Russell M, Ingram VM (1997) Mechanism and prevention of neurotoxicity caused by beta-amyloid peptides: relation to Alzheimer's disease. Brain Res 776: 40–50

Borchelt DR, Thinakaran G, Eckman CB et al (1996) Familial Alzheimer's disease-linked presenilin 1 variants elevate A beta 1–42/1–40 ratio in vitro and in vivo. Neuron 17: 1005–1013

Borchelt DR, Ratovitski T, Vanlare J et al (1997) Accelerated amyloid deposition in the brains of transgenic mice coexpressing mutant presenilin 1 and amyloid precursor proteins. Neuron 19: 939–945

*Braak H, Braak E (1995) Staging of Alzheimer's disease-related neurofibrillary changes. Neurobiol Aging 16: 271–278

Breitner JCS (1996) The role of anti-inflammatory drugs in the prevention and treatment of Alzheimer's disease. Annu Rev Med 47: 401–411

Bullido MJ, Artiga MJ, Recuero M et al (1998) A polymorphism in the regulatory region of APOE associated with risk for Alzheimer's dementia. Nat Genet 18: 69–71

Busciglio J, Lorenzo A, Yeh J, Yankner BA (1995) Beta-amyloid fibrils induce tau phosphorylation and loss of microtubule binding. Neuron 14: 879–888

Bush AI, Multhaup G, Moir RD et al (1993) A novel zinc(II) binding site modulates the function of the beta A4 amyloid protein precursor of Alzheimer's disease. J Biol Chem 268: 16109–16112

Buxbaum JD, Gandy SE, Cicchetti P et al (1990) Processing of Alzheimer beta/A4 amyloid precursor protein: modulation by agents that regulate protein phosphorylation. Proc Natl Acad Sci USA 87: 6003–6006

Buxbaum JD, Ruefli AA, Parker CA, Cypess AM, Greengard P (1994) Calcium regulates processing of the Alzheimer amyloid protein precursor in a protein kinase C-independent manner. Proc Natl Acad Sci USA 91: 4489–4493

*Cai XD, Golde TE, Younkin SG (1993) Release of excess amyloid β protein from a mutant amyloid β protein precursor. Science 259: 514–516

Campion D, Martin C, Heilig R et al (1995) The NACP/synuclein gene: chromosomal assignment and screening for alterations in Alzheimer disease. Genomics 26: 254–257

Capell A, Saffrich R, Olivo JC et al (1997) Cellular expression and proteolytic processing of presenilin proteins is developmentally regulated during neuronal differentiation. J Neurochem 69: 2432–2440

Capell A, Grunberg J, Pesold B et al (1998) The proteolytic fragments of the Alzheimer's disease-associated presenilin-1 form heterodimers and occur as a 100–150-kDa molecular mass complex. J Biol Chem 273: 3205–3211

Caporaso GL, Gandy SE, Buxbaum JD, Ramabhadran TV, Greengard P (1992) Protein phosphorylation regulates secretion of Alzheimer beta/A4 amyloid precursor protein. Proc Natl Acad Sci USA 89: 3055–3059

Castano EM, Prelli F, Wisniewski T, Golabek A, Kumar RA, Soto C, Frangione B (1995) Fibrillogenesis in Alzheimer's disease of amyloid beta peptides and apolipoprotein E. Biochem J 306: 599–604

Chartier Harlin MC, Crawford F, Houlden H et al (1991) Early-onset Alzheimer's disease caused by mutations at codon 717 of the beta-amyloid precursor protein gene. Nature 353: 844–846

*Citron M, Oltersdorf T, Haass C et al (1992) Mutation of the β-amyloid precursor protein in familial Alzheimer's disease increases β-protein production. Nature 360: 672–674

Citron M, Westaway D, Xia W et al (1997) Mutant presenilins of Alzheimer's disease increase production of 42-residue amyloid beta-protein in both transfected cells and transgenic mice. Nat Med 3: 67–72

Clark RF, Hutton M, Fuldner RA et al (1995) The structure of the presenilin 1 (S182) gene and identification of six novel mutations in early onset AD families. Nat Genet 11: 219–222

Conlon RA, Reaume AG, Rossant J (1995) Notch1 is required for the coordinate segmentation of somites. Development 121: 1533–1545

Cook DG, Sung JC, Golde TE et al (1996) Expression and analysis of presenilin 1 in a human neuronal system: localization in cell bodies and dendrites. Proc Natl Acad Sci USA 93: 9223–9228

*Cook DG, Forman MS, Sung JC et al (1997) Alzheimer's A beta (1–42) is generated in the endoplasmic reticulum/intermediate compartment of NT2 N cells. Nature Med 3: 1021–1023

Corder EH, Saunders AM, Strittmatter WJ et al (1993) Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. Science 261: 921–923

*Cruts M, van Duijn CM, Backhovens H et al (1998) Estimation of the genetic contribution of presenilin-1 and -2 mutations in a population based study of presenile Alzheimer disease. Hum Mol Genet 7: 43–51

Davis RE, Miller S, Herrnstadt C et al (1997) Mutations in mitochondrial cytochrome c oxidase genes segregate with late-onset Alzheimer disease. Proc Natl Acad Sci USA 94: 4526–4531

*De Strooper B, Saftig P, Craessaerts K et al (1998) Deficiency of presenilin-1 inhibits the normal cleavage of amyloid precursor protein. Nature 391: 387–390

Doan A, Thinakaran G, Borchelt DR et al (1996) Protein topology of presenilin 1. Neuron 17: 1023–1030

Drewes G, Lichtenberg Kraag B et al (1992) Mitogen activated protein (MAP) kinase transforms tau protein into an Alzheimer-like state. EMBO J 11: 2131–2138

Drewes G, Trinczek B, Illenberger S et al (1995) Microtubule-associated protein microtubule affinity-regulating kinase (p110(mark)) – a novel protein kinase that regulates tau-microtubule interactions and dynamic instability by phosphorylation at the Alzheimer-specific site serine 262. J Biol Chem 270: 7679–7688

Duff K, Eckman C, Zehr C et al (1996) Increased amyloid-beta 42(43) in brains of mice expressing mutant presenilin 1. Nature 383: 710–713

Dyrks T, Weidemann A, Multhaup G et al (1988) Identification, transmembrane orientation and biogenesis of the amyloid A4 precursor of Alzheimer's disease. EMBO J 7: 949–957

Dyrks T, Dyrks E, Hartmann T, Masters C, Beyreuther K (1992) Amyloidogenicity of βA4 and βA4-bearing amyloid protein precursor fragments by metal-catalyzed oxidation. J Biol Chem 267: 18210–18217

Eckman CB, Mehta ND, Crook R et al (1997) A new pathogenic mutation in the APP gene (1716 V) increases the relative proportion of A beta 42(43). Hum Mol Genet 6: 2087–2089

Efthimiopoulos S, Felsenstein KM, Sambamurti K, Robakis NK, Refolo LM (1994) Study of the phorbol ester effect on Alzheimer amyloid precursor processing: sequence requirements and involvement of a cholera toxin sensitive protein. J Neurosci Res 38: 81–90

Efthimiopoulos S, Punj S, Manolopoulos V, Pangalos M, Wang GP, Refolo LM, Robakis NK (1996) Intracellular cyclic AMP

inhibits constitutive and phorbol ester-stimulated secretory cleavage of amyloid precursor protein. J Neurochem 67: 872–875

Elkhoury J, Hickman SE, Thomas CA, Cao L, Silverstein SC, Loike JD (1996) Scavenger receptor-mediated adhesion of microglia to beta-amyloid fibrils. Nature 382: 716–719

*Esch FS, Keim PS, Beattie EC et al (1990) Cleavage of amyloid β peptide during constitutive processing of its precursor. Science 248: 1122–1124

Etcheberrigaray R, Ito E, Kim CS, Alkon DL (1994) Soluble beta-amyloid induction of Alzheimer's phenotype for human fibroblast K^+ channels. Science 264: 276–279

Evans KC, Berger EP, Cho CG, Weisgraber KH, Lansbury PT Jr (1995) Apolipoprotein E is a kinetic but not a thermodynamic inhibitor of amyloid formation: implications for the pathogenesis and treatment of Alzheimer disease. Proc Natl Acad Sci USA 92: 763–767

Evin G, Beyreuther K, Masters CL (1994) Alzheimer's disease amyloid precursor protein (AßPP): proteolytic processing, secretases and βA4 amyloid production. Int J Exp Clin Invest 1: 263–280

Farrer LA, Cupples LA, Haines JL et al (1997) Effects of age, sex, ethnicity on the association between apolipoprotein E genotype and Alzheimer disease: a meta-analysis. JAMA 278: 1349–1356

Feany MB, Dickson DW (1996) Neurodegenerative disorders with extensive tau pathology: a comparative study and review. Ann Neurol 40: 139–148

Frautschy SA, Yang FS, Irrizarry M, Hyman B, Saido TC, Hsiao K, Cole GM (1998) Microglial response to amyloid plaques in APPsw transgenic mice. Am J Pathol 152: 307–317

Gabuzda D, Busciglio J, Chen LB, Matsudaira P, Yankner BA (1994) Inhibition of energy metabolism alters the processing of amyloid precursor protein and induces a potentially amyloidogenic derivative. J Biol Chem 269: 13623–13628

*Games D, Adams D, Alessandrini R et al (1995) Alzheimer-type neuropathology in transgenic mice overexpressing V717F β-amyloid precursor protein. Nature 373: 523–527

*Goate A, Chartier Harlin MC, Mullan M et al (1991) Segregation of a missense mutation in the amyloid precursor protein gene with familial Alzheimer's disease. Nature 349: 704–706

Goedert M (1995) Molecular dissection of the neurofibrillary lesions of Alzheimer's disease. Arzneimittelforschung 45(1): 403–409

*Goedert M, Spillantini MG, Potier MC, Ulrich J, Crowther RA (1989) Cloning and sequencing of the cDNA encoding an isoform of microtubule-associated protein tau containing four tandem repeats: differential expression of tau protein mRNAs in human brain. EMBO J 8: 393–399

Goedert M, Cohen ES, Jakes R, Cohen P (1992) p42 MAP kinase phosphorylation sites in microtubule-associated protein tau are dephosphorylated by protein phosphatase 2A1. Implications for Alzheimer's disease. FEBS Lett 312: 95–99

Goedert M, Jakes R, Spillantini MG, Hasegawa M, Smith MJ, Crowther RA (1996) Assembly of microtubule-associated protein tau into Alzheimer-like filaments induced by sulphated glycosaminoglycans. Nature 383: 550–553

Gong CX, Grundke Iqbal I, Damuni Z, Iqbal K (1994a) Dephosphorylation of microtubule-associated protein tau by protein phosphatase-1 and -2C and its implication in Alzheimer disease. FEBS Lett 341: 94–98

Gong CX, Grundke Iqbal I, Iqbal K (1994b) Dephosphorylation of Alzheimer's disease abnormally phosphorylated tau by protein phosphatase-2A. Neuroscience 61: 765–772

Goodman Y, Mattson MP (1994) Secreted forms of beta-amyloid precursor protein protect hippocampal neurons against amyloid beta-peptide-induced oxidative injury. Exp Neurol 128: 1–12

Goto S, Yamamoto H, Fukunaga K, Iwasa T, Matsukado Y, Miyamoto E (1985) Dephosphorylation of microtubule-associated protein 2, tau factor, and tubulin by calcineurin. J Neurochem 45: 276–283

Greenwood JA, Scott CW, Spreen RC, Caputo CB, Johnson GV (1994) Casein kinase II preferentially phosphorylates human tau isoforms containing an amino-terminal insert. Identification of threonine 39 as the primary phosphate acceptor. J Biol Chem 269: 4373–4380

Grilli M, Ribola M, Alberici A, Valerio A, Memo M, Spano PF (1995) Identification and characterization of a kappa B/Rel binding site in the regulatory region of the amyloid precursor protein gene. J Biol Chem 270: 26774–26777

Grunberg J, Walter J, Loetscher H, Deuschle U, Jacobsen H, Haass C (1998) Alzheimer's disease associated presenilin-1 holoprotein and its 18–20kDa C-terminal fragment are death substrates for proteases of the caspase family. Biochemistry 37: 2263–2270

Grundke Iqbal I, Iqbal K, Quinlan M, Tung YC, Zaidi MS, Wisniewski HM (1986a) Microtubule-associated protein tau. A component of Alzheimer paired helical filaments. J Biol Chem 261: 6084–6089

Grundke Iqbal I, Iqbal K, Tung YC, Quinlan M, Wisniewski HM, Binder LI (1986b) Abnormal phosphorylation of the microtubule-associated protein tau (tau) in Alzheimer cytoskeletal pathology. Proc Natl Acad Sci USA 83: 4913–4917

Guo Q, Furukawa K, Sopher BL et al (1996) Alzheimer's PS-1 mutation perturbs calcium homeostasis and sensitizes PC12 cells to death induced by amyloid beta-peptide. Neuroreport 8: 379–383

*Haass C (1997) Presenilins: genes for life and death. Neuron 18: 687–690

Haass C, Selkoe DJ (1993) Cellular processing of beta-amyloid precursor protein and the genesis of amyloid beta-peptide. Cell 75: 1039–1042

Haass C, Koo EH, Mellon A, Hung AY, Selkoe DJ (1992a) Targeting of cell-surface beta-amyloid precursor protein to lysosomes: alternative processing into amyloid-bearing fragments. Nature 357: 500–503

*Haass C, Schlossmacher MG, Hung AY et al (1992b) Amyloid beta-peptide is produced by cultured cells during normal metabolism. Nature 359: 322–325

Haass C, Hung AY, Selkoe DJ, Teplow DB (1994) Mutations associated with a locus for familial Alzheimer's disease result in alternative processing of amyloid beta-protein precursor. J Biol Chem 269: 17741–17748

Han SH, Einstein G, Weisgraber KH et al (1994) Apolipoprotein E is localized to the cytoplasm of human cortical neurons: a light and electron microscopic study. J Neuropathol Exp Neurol 53: 535–544

*Hardy J (1997) Amyloid, the presenilins and Alzheimer's disease. Trends Neurosci 20: 154–159

*Hardy J, Allsop D (1991) Amyloid deposition as the central event in the aetiology of Alzheimer's disease. Trends Pharmacol Sci 12: 383–388

Haring R, Gurwitz D, Barg J et al (1995) NGF promotes amyloid precursor protein secretion via muscarinic receptor activation. Biochem Biophys Res Commun 213: 15–23

Harris ME, Hensley K, Butterfield DA, Leedle RA, Carney JM (1995) Direct evidence of oxidative injury produced by the Alzheimer's beta-amyloid peptide (1–40) in cultured hippocampal neurons. Exp Neurol 131: 193–202

*Hartmann T, Bergsdorf C, Tienari P et al (1995) Alternative splicing of APP influences polarised sorting and βA4 production. Soc Neurosci Abstr 21: 504

Hartmann T, Bieger SC, Bruhl B et al (1997) Distinct sites of intracellular production for Alzheimer's disease A beta 40/42 amyloid peptides. Nat Med 3: 1016–1020

Hasegawa M, Morishima Kawashima M, Takio K, Suzuki M, Titani K, Ihara Y (1992) Protein sequence and mass spectrometric analyses of tau in the Alzheimer's disease brain. J Biol Chem 267: 17047–17054

Heintz N, Zoghbi H (1997) Alpha-synuclein – a link between Parkinson and Alzheimer diseases? Nat Genet 16: 325–327

Hendriks L, van Duijn CM, Cras P et al (1992) Presenile dementia and cerebral haemorrhage linked to a mutation at codon 692 of the β-amyloid precursor protein gene. Nat Genet 1: 218–221

Hirano M, Shtilbans A, Mayeux R, Davidson MM, Dimauro S, Knowles JA, Schon EA (1997) Apparent mtDNA heteroplasmy in Alzheimer's disease patients and in normals due to PCR amplification of nucleus-embedded mtDNA pseudogenes. Proc Natl Acad Sci USA 94: 14894–14899

*Hofman A, Ott A, Breteler MM et al (1997) Atherosclerosis, apolipoprotein E, prevalence of dementia and Alzheimer's disease in the Rotterdam Study. Lancet 349: 151–154

Holcomb L, Gordon MN, McGowan E et al (1998) Accelerated Alzheimer-type phenotype in transgenic mice carrying both mutant amyloid precursor protein and presenilin 1 transgenes. Nat Med 4: 97–100

Holtzman DM, Pitas RE, Kilbridge J, Nathan B, Mahley RW, Bu GJ, Schwartz AL (1995) Low density lipoprotein receptor-related protein mediates apolipoprotein E-dependent neurite outgrowth in a central nervous system-derived neuronal cell line. Proc Natl Acad Sci USA 92: 9480–9484

Houlden H, Crook R, Backhovens H et al (1998) ApoE genotype is a risk factor in nonpresenilin early-onset Alzheimer's disease families. Am J Med Genet 81: 117–121

Hoyer S (1998) Is sporadic Alzheimer disease the brain type of non-insulin dependent diabetes mellitus? A challenging hypothesis. J Neural Transm 105: 415–422

Hrabe de Angelis M, McIntyre JN, Gossler A (1997) Maintenance of somite borders in mice requires the Delta homologue DII1. Nature 386: 717–721

Hsiao KK, Borchelt DR, Olson K et al (1995) Age related CNS disorder and early death in transgenic FVB/ N mice overexpressing Alzheimer amyloid precursor proteins. Neuron 15: 1203–1218

Hsiao K, Chapman P, Nilsen S et al (1996) Correlative memory deficits, A beta elevation, amyloid plaques in transgenic mice. Science 274: 99–102

Hung AY, Selkoe DJ (1994) Selective ectodomain phosphorylation and regulated cleavage of beta-amyloid precursor protein. EMBO J 13: 534–542

Hung AY, Haass C, Nitsch RM et al (1993) Activation of protein kinase C inhibits cellular production of the amyloid beta-protein. J Biol Chem 268: 22959–22962

*Hutton M, Hardy J (1997) The presenilins and Alzheimer's disease. Hum Mol Genet 6: 1639–1646

Hutton M, Busfield F, Wragg M et al (1996) Complete analysis of the presenilin 1 gene in early onset Alzheimer's disease. Neuroreport 7: 801–805

*Hutton M, Lendon CL, Rizzu P et al (1998) Association of missense and 5′-splice-site mutations in tau with the inherited dementia FTDP-17. Nature 393: 702–705

Ida N, Hartmann T, Pantel J et al (1996) Analysis of heterogeneous beta A4 peptides in human cerebrospinal fluid and blood by a newly developed sensitive Western blot assay. J Biol Chem 271: 22908–22914

Ikezu T, Okamoto T, Komatsuzaki K, Matsui T, Martyn JAJ, Nishimoto I (1996) Negative transactivation of cAMP response element by familial Alzheimer's mutants of APP. EMBO J 15: 2468–2475

Irizarry MC, McNamara M, Fedorchak K, Hsiao K, Hyman BT (1997) APP(Sw) transgenic mice develop age-related A beta deposits and neuropil abnormalities, but no neuronal loss in CA1. J Neuropathol Exp Neurol 56: 965–973

Iversen LL, Mortishiresmith RJ, Pollack SJ, Shearman MS (1995) The toxicity in vitro of beta-amyloid protein. Biochem J 311: 1–16

Iwatsubo T, Odaka A, Suzuki N, Mizusawa H, Nukina N, Ihara Y (1994) Visualization of Aβ42(43) and Aβ40 in senile plaques with end-specific Aβ monoclonals: evidence that an initially deposited species is Aβ42(43). Neuron 13: 45–53

Iwatsubo T, Mann DMA, Odaka A, Suzuki N, Ihara Y (1995) Amyloid β protein (Aβ) deposition: Aβ42(43) precedes Aβ40 in Down syndrome. Ann Neurol 37: 294–299

Jacobsen JS, Spruyt MA, Brown et al (1994) The release of Alzheimer's disease beta amyloid peptide is reduced by phorbol treatment. J Biol Chem 269: 8376–8382

Jameson L, Frey T, Zeeberg B, Dalldorf F, Caplow M (1980) Inhibition of microtubule assembly by phosphorylation of microtubule-associated proteins. Biochemistry 19: 2472–2479

Jensen PH, Hojrup P, Hager H et al (1997) Binding of A beta to alpha- and beta-synucleins: identification of segments in alpha-synuclein/NAC precursor that bind A beta and NAC. Biochem J 323: 539–546

Jin LW, Ninomiya H, Roch JM, Schubert D, Masliah E, Otero DA, Saitoh T (1994) Peptides containing the RERMS sequence of amyloid beta/A4 protein precursor bind cell surface and promote neurite extension. J Neurosci 14: 5461–5470

Kaltschmidt B, Baeuerle PA, Kaltschmidt C (1993) Potential involvement of the transcription factor NF-kappa B in neurological disorders. Mol Aspects Med 14: 171–190

Kaltschmidt B, Uherek M, Volk B, Baeuerle PA, Kaltschmidt C (1997) Transcription factor NF-kappa B is activated in primary neurons by amyloid beta peptides and in neurons surrounding early plaques from patients with Alzheimer disease. Proc Natl Acad Sci USA 94: 2642–2647

*Kang J, Lemaire HG, Unterbeck et al (1987) The precursor of Alzheimer's disease amyloid A4 protein resembles a cell-surface receptor. Nature 325: 733–736

Kar S, Seto D, Gaudreau P, Quirion R (1996) Beta-amyloid-related peptides inhibit potassium-evoked acetylcholine release from rat hippocampal slices. J Neurosci 16: 1034–1040

Kibbey MC, Jucker M, Weeks BS, Neve RL, Van Nostrand WE, Kleinman HK (1993) Beta-amyloid precursor protein binds to the neurite-promoting IKVAV site of laminin. Proc Natl Acad Sci USA 90: 10150–10153

Kidd M (1963) Paired helical filaments in electron microscopy of Alzheimer's disease. Nature 197: 192–193

Kim TW, Tanzi RE (1997) Presenilins and Alzheimer's disease. Curr Opin Neurobiol 7: 683–688

Kim TW, Pettingell WH, Hallmark OG, Moir RD, Wasco W, Tanzi RE (1997a) Endoproteolytic cleavage and proteasomal degradation of presenilin 2 in transfected cells. J Biol Chem 272: 11006–11010

Kim TW, Pettingell WH, Jung YK, Kovacs DM, Tanzi RE (1997b) Alternative cleavage of Alzheimer-associated presenilins during apoptosis by a caspase-3 family protease. Science 277: 373–376

Kitaguchi N, Takahashi Y, Tokushima Y, Shiojiri S, Ito H (1988) Novel precursor of Alzheimer's disease amyloid protein shows protease inhibitory activity. Nature 331: 530–532

Konig G, Monning U, Czech et al (1992) Identification and differential expression of a novel alternative splice isoform of the βA4 amyloid precursor protein (APP) mRNA in leukocytes and brain microglial cells. J Biol Chem 267: 10804–10809

Konsortium zur molekulargenetischen Diagnostik der Huntington-Krankheit (1996) Informationsblatt zur molekulargenetischen Diagnostik der Huntington-Krankheit. Med Genet 8: 208–209

Kopke E, Tung YC, Shaikh S, Alonso AC, Iqbal K, Grundke Iqbal I (1993) Microtubule-associated protein tau. Abnormal phosphorylation of a non-paired helical filament pool in Alzheimer disease. J Biol Chem 268: 24374–24384

Kosik KS, Joachim CL, Selkoe DJ (1986) Microtubule-associated protein tau (tau) is a major antigenic component of paired helical filaments in Alzheimer disease. Proc Natl Acad Sci USA 83: 4044–4048

Kovacs DM, Fausett HJ, Page KJ et al (1996) Alzheimer-associated presenilins 1 and 2: neuronal expression in brain and localization to intracellular membranes in mammalian cells. Nat Med 2: 224–229

Kruger R, Kuhn W, Muller T et al (1998) Ala30Pro mutation in the gene encoding alpha-synuclein in Parkinson's disease. Nat Genet 18: 106–108

Ladu MJ, Pederson TM, Frail DE, Reardon CA, Getz GS, Falduto MT (1995) Purification of apolipoprotein E attenuates isoform-specific binding to beta-amyloid. J Biol Chem 270: 9039–9042

Lambert JC, Pereztur J, Dupire MJ et al (1997) Distortion of allelic expression of apolipoprotein E in Alzheimer's disease. Hum Mol Genet 6: 2151–2154

Lambert JC, Pasquier F, Cottel D, Frigard B, Amouyel P, Chartier-Harlin MC (1998) A new polymorphism in the APOE promoter associated with risk of developing Alzheimer's disease. Hum Mol Genet 7: 533–540

Lautenschlager NT, Cupples LA, Rao VS et al (1996) Risk of dementia among relatives of Alzheimer's disease patients in the MIRAGE study: what is in store for the oldest old? Neurology 46: 641–650

Lee VM, Balin BJ, Otvos L Jr, Trojanowski JQ (1991) A68: a major subunit of paired helical filaments and derivatized forms of normal Tau. Science 251: 675–678

Lee RKK, Wurtman RJ, Cox AJ, Nitsch RM (1995) Amyloid precursor protein processing is stimulated by metabotropic glutamate receptors. Proc Natl Acad Sci USA 92: 8083–8087

Lee MK, Borchelt DR, Kim G et al (1997) Hyperaccumulation of FAD-linked presenilin 1 variants in vivo. Nat Med 3: 756–760

Lehmann DJ, Johnston C, Smith AD (1997) Synergy between the genes for butyrylcholinesterase K variant and apolipoprotein E4 in late-onset confirmed Alzheimer's disease. Hum Mol Genet 6: 1933–1936

Levitan D, Greenwald I (1995) Facilitation of lin-12-mediated signalling by sel-12, a Caenorhabditis elegans S182 Alzheimer's disease gene. Nature 377: 351–354

Levitan D, Doyle TG, Brousseau D et al (1996) Assessment of normal and mutant human presenilin function in Caenorhabditis elegans. Proc Natl Acad Sci USA 93: 14940–14944

*Levy E, Carman MD, Fernandez MI et al (1990) Mutation of the Alzheimer's disease amyloid gene in hereditary cerebral hemorrhage, Dutch type. Science 248: 1124–1126

*Levy-Lahad E, Wasco W, Poorkaj P et al (1995) Candidate gene for the chromosome 1 familial Alzheimer's disease locus. Science 269: 973–977

Lezoualc'h F, Behl C (1998) Transcription factor NF-kappaB: friend or foe of neurons? Mol Psychiatry 3: 15–20

Lezoualc'h F, Sagara Y, Holsboer F, Behl C (1998) High constitutive NF-kappaB activity mediates resistance to oxidative stress in neuronal cells. J Neurosci 18: 3224–3232

Li XJ, Greenwald I (1996) Membrane topology of the C-elegans SEL-12 presenilin. Neuron 17: 1015–1021

Litersky JM, Johnson GV (1992) Phosphorylation by cAMP-dependent protein kinase inhibits the degradation of tau by calpain. J Biol Chem 267: 1563–1568

Loetscher H, Deuschle U, Brockhaus et al (1997) Presenilins are processed by caspase-type proteases. J Biol Chem 272: 20655–20659

Lowenberg K, Waggoner R (1934) Familial organic psychosis (Alzheimer's type). Arch Neurol Psychiatr 31: 737–754

Luo YQ, Hirashima N, Li YH, Alkon DL, Sunderland T, Etcheberrigaray R, Wolozin B (1995) Physiological levels of beta-amyloid increase tyrosine phosphorylation and cytosolic calcium. Brain Res 681: 65–74

Luo Y, Sunderland T, Wolozin B (1996) Physiologic levels of beta-amyloid activate phosphatidylinositol 3-kinase with the involvement of tyrosine phosphorylation. J Neurochem 67: 978–987

Mahdi F, Vannostrand WE, Schmaier AH (1995) Protease nexin-2/amyloid beta-protein precursor inhibits factor Xa in the prothrombinase complex. J Biol Chem 270: 23468–23474

Mandelkow EM, Drewes G, Biernat J, Gustke N, Van Lint J, Vandenheede JR, Mandelkow E (1992) Glycogen synthase kinase-3 and the Alzheimer-like state of microtubule-associated protein tau. FEBS Lett 314: 315–321

Maruyama K, Kametani F, Usami M, Yamao Harigaya W, Tanaka K (1991) "Secretase", Alzheimer amyloid protein precursor secreting enzyme is not sequence-specific. Biochem Biophys Res Commun 179: 1670–1676

Matsuo ES, Shin RW, Billingsley ML, Van deVoorde A, O'Connor M, Trojanowski JQ, Lee VM (1994) Biopsy-derived adult human brain tau is phosphorylated at many of the same sites as Alzheimer's disease paired helical filament tau. Neuron 13: 989–1002

Mattila KM, Forsell C, Pirttilä T et al (1998) The Glu318Gly mutation of the presenilin-1 gene does not necessarily cause Alzheimer's disease. Neurobiol Aging 19 [Suppl 4S] (abstr no 362)

*Mattson MP (1997) Cellular actions of beta-amyloid precursor protein and its soluble and fibrillogenic derivatives. Physiol Rev 77: 1081–1132

Mattson MP, Barger SW, Cheng B, Lieberburg I, Smith Swintosky VL, Rydel RE (1993a) Beta-amyloid precursor protein metabolites and loss of neuronal Ca^{2+} homeostasis in Alzheimer's disease. Trends Neurosci 16: 409–414

Mattson MP, Cheng B, Culwell AR, Esch FS, Lieberburg I, Rydel RE (1993b) Evidence for excitoprotective and intraneuronal calcium-regulating roles for secreted forms of the beta-amyloid precursor protein. Neuron 10: 243–254

Mattson MP, Guo Q, Furukawa K, Pedersen WA (1998) Presenilins, the endoplasmic reticulum, and neuronal apoptosis in Alzheimer's disease. J Neurochem 70: 1–14

*Mayeux R, Saunders AM, Shea S et al (1998) Utility of the apolipoprotein E genotype in the diagnosis of Alzheimer's disease. N Engl J Med 338: 506–511

McKeith IG, Galasko D, Kosaka et al (1996) Consensus guidelines for the clinical and pathologic diagnosis of dementia with Lewy bodies (DLB): report of the Consortium on DLB International Workshop. Neurology 47: 1113–1124

McKhann G, Drachman D, Folstein M, Kargman R, Price D, Stadlan EM (1984) Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of the Department of Health and Human Services Task Force on Alzheimer's Disease. Neurology 34: 939–944

Milward EA, Papadopoulos R, Fuller SJ, Moir RD, Small D, Beyreuther K, Masters CL (1992) The amyloid protein precursor of Alzheimer's disease is a mediator of the effects of nerve growth factor on neurite outgrowth. Neuron 9: 129–137

Monning U, Sandbrink R, Weidemann A, Banati RB, Masters CL, Beyreuther K (1995) Extracellular matrix influences the biogenesis of amyloid precursor protein in microglial cells. J Biol Chem 270: 7104–7110

Montoya SE, Aston CE, Dekosky ST, Kamboh MI, Lazo JS, Ferrell RE (1998) Bleomycin hydrolase is associated with risk of sporadic Alzheimer's disease. Nat Genet 18: 211–212

Morishima Kawashima M, Hasegawa M, Takio K, Suzuki M, Yoshida H, Titani K, Iharae Y (1995) Proline-directed and non-proline-directed phosphorylation of PHF-tau. J Biol Chem 270: 823–829

Motter R, Vigopelfrey C, Kholodenko D et al (1995) Reduction of beta-amyloid peptide(42), in the cerebrospinal fluid of patients with Alzheimer's disease. Ann Neurol 38: 643–648

*Mullan M, Crawford F, Axelman K, Houlden H, Lilius L, Winblad B, Lannfelt L (1992) A pathogenic mutation for probable Alzheimer's disease in the APP gene at the N-terminus of beta-amyloid. Nat Genet 1: 345–347

Müller-Hill B, Beyreuther K (1989) Molecular biology of Alzheimer's disease. Annu Rev Biochem 58: 287–307

Multhaup G (1994) Identification and regulation of the high affinity binding site of the Alzheimer's disease amyloid protein precursor (APP) to glycosaminoglycans. Biochimie 76: 304–311

Multhaup G, Schlicksupp A, Hesse L, Beher D, Ruppert T, Masters CL, Beyreuther K (1996) The amyloid precursor protein of Alzheimer's disease in the reduction of copper(II) to copper(I). Science 271: 1406–1409

Murrell J, Farlow M, Ghetti B, Benson MD (1991) A mutation in the amyloid precursor protein associated with hereditary Alzheimer's disease. Science 254: 97–99

Narindrasorasak S, Lowery D, Gonzalez DeWhitt P, Poorman RA, Greenberg B, Kisilevsky R (1991) High affinity interactions between the Alzheimer's beta-amyloid precursor proteins and the basement membrane form of heparan sulfate proteoglycan. J Biol Chem 266: 12878–12883

Nathan BP, Bellosta S, Sanan DA, Weisgraber KH, Mahley RW, Pitas RE (1994) Differential effects of apolipoproteins E3 and E4 on neuronal growth in vitro. Science 264: 850–852

Nathan BP, Chang KC, Bellosta S, Brisch E, Ge NF, Mahley RW, Pitas RE (1995) The inhibitory effect of apolipoprotein E4 on neurite outgrowth is associated with microtubule depolymerization. J Biol Chem 270: 19791–19799

**NIA/AA (1996) Apolipoprotein E genotyping in Alzheimer's disease. National Institute on Aging/Alzheimer's Association Working Group. Lancet 347: 1091–1095

Nishimoto I, Okamoto T, Matsuura Y, Takahashi S, Okamoto T, Murayama Y, Ogata E (1993) Alzheimer amyloid protein precursor complexes with brain GTP-binding protein G(o). Nature 362: 75–79

Nitsch RM, Slack BE, Wurtman RJ, Growdon JH (1992) Release of Alzheimer amyloid precursor derivatives stimulated by activation of muscarinic acetylcholine receptors. Science 258: 304–307

Nitsch RM, Farber SA, Growdon JH, Wurtman RJ (1993) Release of amyloid beta-protein precursor derivatives by electrical depolarization of rat hippocampal slices. Proc Natl Acad Sci USA 90: 5191–5193

O'Neill LA, Kaltschmidt C (1997) NF-kappa B: a crucial transcription factor for glial and neuronal cell function. Trends Neurosci 20: 252–258

Okamoto T, Takeda S, Giambarella U et al (1996) Intrinsic signaling function of APP as a novel target of three V642 mutations linked to familial Alzheimer's disease. EMBO J 15: 3769–3777

Paliga K, Peraus G, Kreger S et al (1997) Human amyloid precursor-like protein 1 – cDNA cloning, ectopic expression in COS-7 cells and identification of soluble forms in the cerebrospinal fluid. Eur J Biochem 250: 354–363

Pangalos MN, Efthimiopoulos S, Shioi J, Robakis NK (1995a) The chondroitin sulfate attachment site of appican is formed by splicing out exon 15 of the amyloid precursor gene. J Biol Chem 270: 10388–10391

Pangalos MN, Shioi J, Robakis NK (1995b) Expression of the chondroitin sulfate proteoglycans of amyloid precursor (Appican) and amyloid precursor-like protein 2. J Neurochem 65: 762–769

Paresce DM, Ghosh RN, Maxfield FR (1996) Microglial cells internalize aggregates of the Alzheimer's disease amyloid beta-protein via a scavenger receptor. Neuron 17: 553–565

Paudel HK, Lew J, Ali Z, Wang JH (1993) Brain proline-directed protein kinase phosphorylates tau on sites that are abnormally phosphorylated in tau associated with Alzheimer's paired helical filaments. J Biol Chem 268: 23512–23518

Pedersen WA, Kloczewiak MA, Blusztajn JK (1996) Amyloid beta-protein reduces acetylcholine synthesis in a cell line derived from cholinergic neurons of the basal forebrain. Proc Natl Acad Sci USA 93: 8068–8071

Perez RG, Zheng H, Vanderploeg LHT, Koo EH (1997) The beta-amyloid precursor protein of Alzheimer's disease enhances neuron viability and modulates neuronal polarity. J Neurosci 17: 9407–9414

Podlisny MB, Citron M, Amarante P et al (1997) Presenilin proteins undergo heterogeneous endoproteolysis between Thr291 and Ala299 and occur as sTable N- and C-terminal fragments in normal and Alzheimer brain tissue. Neurobiol Dis 3: 325–337

Polymeropoulos MH, Lavedan C, Leroy E et al (1997) Mutation in the alpha-synuclein gene identified in families with Parkinson's disease. Science 276: 2045–2047

Ponte P, Gonzalez DP, Schilling J et al (1988) A new A4 amyloid mRNA contains a domain homologous to serine proteinase inhibitors. Nature 331: 525–527

Poorkaj P, Bird TD, Wijsman E et al (1998) Tau is a candidate gene for chromosome 17 frontotemporal dementia. Ann Neurol 43: 815–825

Querfurth HW, Selkoe DJ (1994) Calcium ionophore increases amyloid beta peptide production by cultured cells. Biochemistry 33: 4550–4561

Rebeck GW, Reiter JS, Strickland DK, Hyman BT (1993) Apolipoprotein E in sporadic Alzheimer's disease: allelic variation and receptor interactions. Neuron 11: 575–580

Refolo LM, Salton SR, Anderson JP, Mehta P, Robakis NK (1989) Nerve and epidermal growth factors induce the release of the Alzheimer amyloid precursor from PC 12 cell cultures. Biochem Biophys Res Commun 164: 664–670

*Rogaev EI, Sherrington R, Rogaeva EA et al (1995) Familial Alzheimer's disease in kindreds with missense mutations in a gene on chromosome 1 related to the Alzheimer's disease type 3 gene. Nature 376: 775–778

Roses AD (1996) Apolipoprotein E alleles as risk factors in Alzheimer's disease. Annu Rev Med 47: 387–400

*Roses AD (1998) Alzheimer diseases: a model of gene mutations and susceptibility polymorphisms for complex psychiatric diseases. Am J Med Genet 81: 49–57

Rossor MN, Fox NC, Beck J, Campbell TC, Collinge J (1996) Incomplete penetrance of familial Alzheimer's disease in a pedigree with a novel presenilin-1 gene mutation. Lancet 347: 1560

*Rumble B, Retallack R, Hilbich C et al (1989) Amyloid A4 protein and its precursor in Down's syndrome and Alzheimer's disease. N Engl J Med 320: 1446–1452

Salbaum JM, Weidemann A, Lemaire HG, Masters CL, Beyreuther K (1988) The promoter of Alzheimer's disease amyloid A4 precursor gene. EMBO J 7: 2807–2813

Sandbrink R, Beyreuther K (1996) Unraveling the molecular pathway of Alzheimer's disease: research about presenilins gathers momentum. Mol Psychiatry 1: 438–444

Sandbrink R, Masters CL, Beyreuther K (1994a) βA4-amyloid protein precursor mRNA isoforms without exon 15 are ubiquitously expressed in rat tissues including brain, but not in neurons. J Biol Chem 269: 1510–1517

Sandbrink R, Masters CL, Beyreuther K (1994b) Similar alternative splicing of a non-homologous domain in βA4-amyloid protein precursor-like proteins. J Biol Chem 269: 14227–14234

Sandbrink R, Masters CL, Beyreuther K (1995) APP gene family: alternative splicing generates functionally related isoforms. Ann NY Acad Sci 777: 281–287

*Sandbrink R, Hartmann T, Masters CL, Beyreuther K (1996a) Genes contributing to Alzheimer's disease. Mol Psychiatr 1: 27–40

Sandbrink R, Zhang D, Schaeffer S, Masters CL, Bauer J, Förstl H, Beyreuther K (1996b) Missense mutations of the PS-1/S182 gene in German early-onset Alzheimer's disease patients. Ann Neurol 40: 265–266

**Scheuner D, Eckman C, Jensen M et al (1996) Secreted amyloid β-protein similar to that in the senile plaques of Alzheimer's disease is increased in vivo by the presenilin 1 and 2 and APP mutations linked to familial Alzheimer's disease. Nature Med 2: 864–870

Schmechel DE, Saunders AM, Strittmatter WJ et al (1993) Increased amyloid β-peptide deposition in cerebral cortex as a consequence of apolipoprotein E genotype in late-onset Alzheimer disease. Proc Natl Acad Sci USA 90: 9649–9653

Schubert D, Jin LW, Saitoh T, Cole G (1989) The regulation of amyloid beta protein precursor secretion and its modulatory role in cell adhesion. Neuron 3: 689–694

Schubert W, Prior R, Weidemann A, Dircksen H, Multhaup G, Masters CL, Beyreuther K (1991) Localization of Alzheimer beta A4 amyloid precursor protein at central and peripheral synaptic sites. Brain Res 563: 184–194

Seeger M, Nordstedt C, Petanceska S et al (1997) Evidence for phosphorylation and oligomeric assembly of presenilin 1. Proc Natl Acad Sci USA 94: 5090–5094

Selkoe DJ (1994) Cell biology of the amyloid β-protein precursor and the mechanism of Alzheimer's disease. Annu Rev Cell Biol 10: 373–403

*Selkoe DJ (1997) Alzheimer's disease: genotypes, phenotypes, and treatments. Science 275: 630–631

*Seubert P, Vigo Pelfrey C, Esch F et al (1992) Isolation and quantification of soluble Alzheimer's beta-peptide from biological fluids. Nature 359: 325–327

Shen J, Bronson RT, Chen DF, Xia W, Selkoe DJ, Tonegawa S (1997) Skeletal and CNS defects in Presenilin-1-deficient mice. Cell 89: 629–639

*Sherrington R, Rogaev EI, Liang Y et al (1995) Cloning of a gene bearing missense mutations in early-onset familial Alzheimer's disease. Nature 375: 754–760

*Shoji M, Golde TE, Ghiso J et al (1992) Production of the Alzheimer amyloid beta protein by normal proteolytic processing. Science 258: 126–129

Siman R, Card JP, Nelson RB, Davis LG (1989) Expression of beta-amyloid precursor protein in reactive astrocytes following neuronal damage. Neuron 3: 275–285

Simons M, Ikonen E, Tienari PJ, Cidarregui A, Monning U, Beyreuther K, Dotti CG (1995) Intracellular routing of human amyloid protein precursor: axonal delivery followed by transport to the dendrites. J Neurosci Res 41: 121–128

Small DH, Nurcombe V, Reed G, Clarris H, Moir R, Beyreuther K, Masters CL (1994) A heparin-binding domain in the amyloid protein precursor of Alzheimer's disease is involved in the regulation of neurite outgrowth. J Neurosci 14: 2117–2127

Smith MA, Perry G, Richey PL, Sayre LM, Anderson VE, Beal MF, Kowall N (1996) Oxidative damage in Alzheimer's. Nature 382: 120–121

Spillantini MG, Schmidt ML, Lee VM, Trojanowski JQ, Jakes R, Goedert M (1997) Alpha-synuclein in Lewy bodies. Nature 388: 839–840

Spillantini MG, Murrell JR, Goedert M, Farlow MR, Klug A, Ghetti B (1998) Mutation in the tau gene in familial multiple system tauopathy with presenile dementia. Proc Natl Acad Sci USA 95: 7737–7741

Sprecher CA, Grant FJ, Grimm G, O'Hara PJ, Norris F, Norris K, Foster DC (1993) Molecular cloning of the cDNA for a human amyloid precursor protein homolog: evidence for a multigene family. Biochemistry 32: 4481–4486

Steiner B, Mandelkow EM, Biernat J et al (1990) Phosphorylation of microtubule-associated protein tau: identification of the site for Ca2(+)-calmodulin dependent kinase and relationship

with tau phosphorylation in Alzheimer tangles. EMBO J 9: 3539–3544

*Strittmatter WJ, Saunders AM, Schmechel D, Pericak Vance M, Enghild J, Salvesen GS, Roses AD (1993a) Apolipoprotein E: high-avidity binding to β-amyloid and increased frequency of type 4 allele in late-onset familial Alzheimer disease. Proc Natl Acad Sci USA 90: 1977–1981

Strittmatter WJ, Weisgraber KH, Huang DY et al (1993b) Binding of human apolipoprotein E to synthetic amyloid β peptide: isoform-specific effects and implications for late-onset Alzheimer disease. Proc Natl Acad Sci USA 90: 8098–8102

Strittmatter WJ, Saunders AM, Goedert M et al (1994a) Isoform-specific interactions of apolipoprotein E with microtubule-associated protein tau: implications for Alzheimer disease. Proc Natl Acad Sci USA 91: 11183–11186

Strittmatter WJ, Weisgraber KH, Goedert M et al (1994b) Hypothesis: microtubule instability and paired helical filament formation in the Alzheimer disease brain are related to apolipoprotein E genotype. Exp Neurol 125: 163–171

Sturchlerpierrat C, Abramowski D, Duke M et al (1997) Two amyloid precursor protein transgenic mouse models with Alzheimer disease-like pathology. Proc Natl Acad Sci USA 94: 13287–13292

*Suzuki N, Cheung TT, Cai XD et al (1994) An increased percentage of long amyloid β protein secreted by familial amyloid β protein precursor (βAPP717) mutants. Science 264: 1336–1340

Tamaoka A, Sawamura N, Odaka A, Suzuki N, Mizusawa H, Shoji S, Mori H (1995) Amyloid beta protein 1–42/43 (A beta 1–42/43) in cerebellar diffuse plaques: enzyme-linked immunosorbent assay and immunocytochemical study. Brain Res 679: 151–156

Tanzi RE, McClatchey AI, Lamperti ED, Villa KL, Gusella JF, Neve RL (1988) Protease inhibitor domain encoded by an amyloid protein precursor mRNA associated with Alzheimer's disease. Nature 331: 528–530

Thinakaran G, Slunt HH, Sisodia SS (1995) Novel regulation of chondroitin sulfate glycosaminoglycan modification of amyloid precursor protein and its homologue, APLP2. J Biol Chem 270: 16522–16525

Thinakaran G, Borchelt DR, Lee MK et al (1996) Endoproteolysis of presenilin 1 and accumulation of processed derivatives in vivo. Neuron 17: 181–190

*Thinakaran G, Harris CL, Ratovitski T et al (1997) Evidence that levels of presenilins (PS1 and PS2) are coordinately regulated by competition for limiting cellular factors. J Biol Chem 272: 28415–28422

Thomas T, Thomas G, McLendon C, Sutton T, Mullan M (1996) Beta-amyloid-mediated vasoactivity and vascular endothelial damage. Nature 380: 168–171

Thomas T, Sutton ET, Hellermann A, Price JM (1997) Beta-amyloid-induced coronary artery vasoactivity and endothelial damage. J Cardiovasc Pharmacol 30: 517–522

Tienari PJ, Destrooper B, Ikonen E et al (1996) The beta-amyloid domain is essential for axonal sorting of amyloid precursor protein. EMBO J 15: 5218–5229

*Tienari PJ, Ida N, Ikonen E et al (1997) Intracellular and secreted Alzheimer beta-amyloid species are generated by distinct mechanisms in cultured hippocampal neurons. Proc Natl Acad Sci USA 94: 4125–4130

Trojanowski JQ, Lee VM (1995) Phosphorylation of paired helical filament tau in Alzheimer's disease neurofibrillary lesions: focusing on phosphatases. FASEB J 9: 1570–1576

Tysoe C, Whittaker J, Xuereb J et al (1998) A presenilin-1 truncating mutation is present in two cases with autopsy-confirmed early-onset Alzheimer disease. Am J Hum Genet 62: 70–76

Ueda K, Fukushima H, Masliah E et al (1993) Molecular cloning of cDNA encoding an unrecognized component of amyloid in Alzheimer disease. Proc Natl Acad Sci USA 90: 11282–11286

*Van Broeckhoven C, Haan J, Bakker E et al (1990) Amyloid β protein precursor gene and hereditary cerebral hemorrhage with amyloidosis (Dutch). Science 248: 1120–1122

van Duijn CM, Clayton D, Chandra V et al (1991) Familial aggregation of Alzheimer's disease and related disorders: a collaborative re-analysis of case-control studies. Int J Epidemiol 20[Suppl 1]: S13–S20

Van Gool WA, Evenhuis HM, van Duijn CM (1995) A case-control study of apolipoprotein E genotypes in Alzheimer's disease associated with Down's syndrome. Ann Neurol 38: 225–230

Van Nostrand WE, Farrow JS, Wagner SL, Bhasin R, Goldgaber D, Cotman CW, Cunningham DD (1991) The predominant form of the amyloid beta-protein precursor in human brain is protease nexin 2. Proc Natl Acad Sci USA 88: 10302–10306

Vandermeeren M, Mercken M, Vanmechelen E, Six J, Van de Voorde A, Martin JJ, Cras P (1993) Detection of tau proteins in normal and Alzheimer's disease cerebrospinal fluid with a sensitive sandwich enzyme-linked immunosorbent assay. J Neurochem 61: 1828–1834

von Koch CS, Zheng H, Chen H et al (1997) Generation of APLP2 KO mice and early postnatal lethality in APLP2/APP double KO mice. Neurobiol Aging 18: 661–669

Wallace DC, Stugard C, Murdock D, Schurr T, Brown MD (1997) Ancient mtDNA sequences in the human nuclear genome: a potential source of errors in identifying pathogenic mutations. Proc Natl Acad Sci USA 94: 14900–14905

Walter J, Capell A, Grunberg J et al (1996) The Alzheimer's disease-associated presenilins are differentially phosphorylated proteins located predominantly within the endoplasmic reticulum. Mol Med 2: 673–691

Walter J, Grunberg J, Capell A et al (1997) Proteolytic processing of the Alzheimer disease-associated presenilin-1 generates an in vivo substrate for protein kinase C. Proc Natl Acad Sci USA 94: 5349–5354

Wasco W, Bupp K, Magendantz M, Gusella JF, Tanzi RE, Solomon F (1992) Identification of a mouse brain cDNA that encodes a protein related to the Alzheimer disease-associated amyloid beta protein precursor. Proc Natl Acad Sci USA 89: 10758–10762

Wasco W, Gurubhagavatula S, Paradis MD et al (1993) Isolation and characterization of APLP2 encoding a homologue of the Alzheimer's associated amyloid β protein precursor. Nat Genet 5: 95–100

Watanabe A, Hasegawa M, Suzuki M et al (1993) In vivo phosphorylation sites in fetal and adult rat tau. J Biol Chem 268: 25712–25717

*Weidemann A, Konig G, Bunke D, Fischer P, Salbaum JM, Masters CL, Beyreuther K (1989) Identification, biogenesis, localization of precursors of Alzheimer's disease A4 amyloid protein. Cell 57: 115–126

*Weidemann A, Paliga K, Durrwang U, Czech C, Evin G, Masters CL, Beyreuther K (1997) Formation of stable complexes between two Alzheimer's disease gene products: Presenilin-2 and beta-amyloid precursor protein. Nature Med 3: 328–332

Weingarten MD, Lockwood AH, Hwo SY, Kirschner MW (1975) A protein factor essential for microtubule assembly. Proc Natl Acad Sci USA 72: 1858–1862

Weisgraber KH, Mahley RW (1996) Human apolipoprotein E: the Alzheimer's disease connection. FASEB J 10: 1485–1494

Whitson JS, Selkoe DJ, Cotman CW (1989) Amyloid beta protein enhances the survival of hippocampal neurons in vitro. Science 243: 1488–1490

Wild-Bode C, Yamazaki T, Capell A, Leimer U, Steiner H, Ihara Y, Haass C (1997) Intracellular generation and accumulation of amyloid beta-peptide terminating at amino acid 42. J Biol Chem 272: 16085–16088

Wisniewski T, Dowjat WK, Buxbaum JD et al (1998) A novel Polish presenilin-1 mutation (P117L) is associated with familial Alzheimer's disease and leads to death as early as the age of 28 years. Neuroreport 9: 217–221

Wolozin B, Iwasaki K, Vito P et al (1996) Participation of Presenilin 2 in apoptosis: enhanced basal activity conferred by an Alzheimer mutation. Science 274: 1710–1713

Wong PC, Zheng H, Chen H et al (1997) Presenilin 1 is required for Notch1 and DII1 expression in the paraxial mesoderm. Nature 387: 288–292

Wood JG, Mirra SS, Pollock NJ, Binder LI (1986) Neurofibrillary tangles of Alzheimer disease share antigenic determinants with the axonal microtubule-associated protein tau (tau). Proc Natl Acad Sci USA 83: 4040–4043

Wragg M, Hutton M, Talbot C et al (1996) Genetic association between intronic polymorphism in presenilin-1 gene and late-onset Alzheimer's disease. Lancet 347: 509–512

Xia Y, Desilva HAR, Rosi BL et al (1996) Genetic studies in Alzheimer's disease with an NACP/alpha- synuclein polymorphism. Ann Neurol 40: 207–215

Xia WM, Zhang JM, Perez R, Koo EH, Selkoe DJ (1997) Interaction between amyloid precursor protein and presenilins in mammalian cells: implications for the pathogenesis of Alzheimer disease. Proc Natl Acad Sci USA 94: 8208–8213

Yamatsuji T, Matsui T, Okamoto T et al (1996) G protein-mediated neuronal DNA fragmentation induced by familial Alzheimer's disease-associated mutants of APP. Science 272: 1349–1352

Yan SD, Chen X, Fu J et al (1996) RAGE and amyloid-beta peptide neurotoxicity in Alzheimer's disease. Nature 382: 685–691

*Yankner BA (1996) Mechanisms of neuronal degeneration in Alzheimer's disease. Neuron 16: 921–932

Yankner BA, Duffy LK, Kirschner DA (1990) Neurotrophic and neurotoxic effects of amyloid beta protein: reversal by tachykinin neuropeptides. Science 250: 279–282

Yoshimoto M, Iwai A, Kang D, Otero DAC, Xia Y, Saitoh T (1995) NACP, the precursor protein of the non-amyloid beta/A4 protein (A beta) component of Alzheimer disease amyloid, binds A beta and stimulates A beta aggregation. Proc Natl Acad Sci USA 92: 9141–9145

*Younkin SG (1995) Evidence that A beta 42 is the real culprit in Alzheimer's disease. Ann Neurol 37: 287–288

Zheng H, Jiang MH, Trumbauer ME et al (1995) beta-Amyloid precursor protein-deficient mice show reactive gliosis and decreased locomotor activity. Cell 81: 525–531

CHAPTER

8

P. Martinez-Lage, V.C. Hachinski

Vascular Cognitive Impairment and Dementia

1 Introduction

Vascular dementia (VD)[1] has been known, diagnosed, and investigated for more than 100 years and still many questions remain unanswered. The titles of two recent reviews on this disease are highly demonstrative of the confusion and disagreement that surround the field of cognitive decline due to cerebrovascular disease: *Vascular Dementia: Persisting Controversies and Questions* (Moncayo and Bogousslavsky 1996) and *Vascular Dementia: Still a Debatable Entity?* (Loeb and Meyer 1996). As of today, no reliable, validated diagnostic criteria are available and its epidemiology remains unknown. Prevalence and incidence figures are extremely variable, and different studies cannot be compared. Clinical and neuroradiological manifestations are not fully understood and remain tied to old preconceptions. The definition of dementia, for instance, is uncritically derived from the clinical picture of Alzheimer's disease (AD), and memory impairment is still taken as the hallmark of dementia. Moreover, while it is generally accepted that VD is heterogeneous, its different clinical and etiopathogenic subtypes have not been clearly separated and patients with different diseases are still grouped together.

Words such as "treatable" or "preventable" are inseparable from VD. Paradoxically, the definition of dementia currently in use, which requires a degree of severity high enough to affect the social or occupational functioning of the individual, leaves little room for preventive or therapeutic actions. No early detection can be attempted if diagnostic criteria do not move away from this concept. Similarly, little prevention of VD can be made if its risk factors remain unknown and it is not established whether they are the same as those for stroke. Hachinski and Bowler (1993) suggested the concept of vascular cognitive impairment (VCI)[2] to encompass all different degrees of intellectual decline related to ischemic cerebrovascular disease, highlighting the importance of early detection and the compelling need for new diagnostic criteria based on prospectively collected experience and not on uncontestedly collected literature. Throughout this chapter, the authors will use the term VCI instead of VD whenever possible.

[1]As defined by current criteria, the diagnosis of dementia requires the development of deficits in memory functions and one or more other areas of cognition which are severe enough to impair social or occupational functioning and represent a significant decline from a previous level of performance. VD is dementia due to ischemic vascular lesions.

[2]Cognitive impairment is a decline in any cognitive domain regardless of its severity. Memory loss is not necessarily required. Dementia would represent the most severe stage of impairment. VCI is a term that would encompass all types and degrees of cognitive impairment, from early intellectual changes to dementia, due to ischemic cerebrovascular disease.

2 Epidemiology of Vascular Cognitive Impairment

2.1 Prevalence of Vascular Dementia

The prevalence of VD is difficult to establish, as it shares so many features with diseases such as stroke or AD and no reliable diagnostic criteria have been developed (see below). The frequency of VD varies depending on the sample under survey and the diagnostic tool applied, especially whether or not a neuroimaging technique is applied. Autopsy series have the highest diagnostic accuracy but are necessarily biased, and only series of patients with dementia and not general consecutive autopsies have been analyzed. In such studies, the frequency of VD varies from 0% to 19% (Jellinger et al. 1990). Interestingly, among patients with a clinical diagnosis of AD, 3% may receive a diagnosis of pure VD at autopsy. Frequency figures from hospital-based series or memory clinic registries are not adequate, as stroke patients will be overrepresented in the former and AD patients in the latter.

Population-based studies offer the best scenario to control for selection biases but at the expenses of diagnostic accuracy, as neuroimaging investigation cannot always be applied. Hébert and Brayne (1995) reviewed ten population-based studies from different countries that applied the same diagnostic criteria and used similar sampling frames. Prevalence figures for VD in these surveys varied between 1.2% and 5.6% of the population. Methodological issues rather than geographic differences appeared to explain such variation. The highest figures were obtained in studies that included a computed tomography (CT) scan as part of the evaluation. A significant bias leading to underestimation was introduced by the exclusion of institutionalized patients in some studies. Contrary to what has been traditionally thought, no significant differences in the prevalence of VD were observed in these surveys between men and women. In addition to diagnostic accuracy, another source of confusion when epidemiologists attempt to estimate the contribution of vascular causes to the prevalence

of dementia is the way in which patients with mixed dementia[3] are classified, as some studies include them with VD patients and others count them as AD patients. Prevalence rates may vary from 14% to 1% depending on this factor (Skoog et al. 1993; Ott et al. 1995). Unfortunately, the clinical diagnosis of mixed dementia does not differentiate patients in which stroke is just a coincidental association with AD from those patients that would not be demented had they not presented with a stroke.

2.2 Dementia After Stroke

It is now well established that having a stroke is a significant risk factor for dementia. Almost 30% of stroke patients develop dementia in the first 3 months after stroke. The risk is nine times higher than that in the general population. Perhaps surprisingly, the type of dementia that stroke patients develop is both VD and AD (mixed dementia?), and in many cases no new cerebrovascular events precede dementia (Tatemichi and Desmond 1996). In population-based longitudinal studies, the incidence of total dementia among patients with stroke at study entry is nine times higher and the incidence of AD is twice that in the general population (Kokmen et al. 1996). Some authors have suggested the term "poststroke dementia" to include all possible mechanisms of dementia after stroke.

2.3 Prevalence of Vascular Cognitive Impairment

As outlined in the previous paragraphs, epidemiological research has focused only on patients with dementia, i.e., those in which the severity of intellectual decline is sufficient to fulfill diagnostic criteria. However, the impact of vascular causes on cognition may be much more important if all degrees of cognitive decline are considered. Prevalence and incidence figures for VCI will remain undetermined until valid diagnostic criteria are developed and validated. However, the Canadian Study of Health and Aging has provided some data in this respect. In this survey, approximately 10,000 subjects older than 65 were screened for cognitive decline. Among those who screened positive, subjects with dementia were distinguished from those with cognitive decline not severe enough to be labeled as dementia. Vascular causes were found to be the second most frequent cause of cognitive decline without dementia after the condition known as age-associated memory impairment. If subjects with cognitive impairment due to vascular causes are grouped together with those diagnosed with VD and mixed dementia, it can be said that 4.8% of the Canadian population older than 65 had some degree of VCI (Rockwood et al. 1997). This number may, however, underestimate the prevalence of VCI, as no neuroimaging studies were performed in these subjects.

With regard to mild cognitive impairment, data from numerous studies have provided sufficient evidence to stress the importance of vascular factors. Hypertension and diabetes are significant negative predictors of cognitive function in nondemented subjects in population-based studies, and this effect appears to be independent of the presence of stroke (Martinez-Lage et al. 1996). The deleterious effect of hypertension on cognitive function has been known for years, and it has been related to the presence of white matter changes. Additional epidemiological studies have shown that the presence of electrocardiographic evidence of myocardial infarction, peripheral artery disease, or the presence of carotid atheromatous changes doubles the proportion of subjects scoring 23 or less in the Mini-Mental State Examination (MMSE). History of stroke increases this proportion by three (Breteler et al. 1994). In stroke-free subjects, diabetes and hypercholesterolemia were independent correlates of abstract reasoning/visuospatial deficits and memory dysfunction, respectively (Tatemichi and Desmond 1996).

2.4 Cognitive Impairment After Stroke

The appearance of cognitive impairment in patients with stroke is now receiving significant attention. Some degree of cognitive impairment may be detected in up to 60% of stroke patients. However, the proportion depends on which diagnostic tool is applied and the level at which cutoff points are placed. Pohjasvaara and colleagues (1997) found that 61.7% of 486 stroke patients had deficits in at least one cognitive domain, 34.8% in one or two domains, and 26.8% in three cognitive areas or more, as measured by the modified MMSE (3MS) and other brief tests. Using a comprehensive neuropsychological battery, Grace and colleagues (1995) found that 46% of their stroke patients were cognitively impaired (scored two standard deviations below the norm in at least two cognitive

[3]This term is applied to cases of dementia in which the presence of both degenerative and vascular changes is suspected on clinical or radiological grounds or demonstrated at autopsy. Unfortunately, no clinical or even pathological criteria have been developed to distinguish cases in which both lesions are required to produce dementia from those in which either degenerative or vascular lesions would have been sufficient to cause dementia had they occurred on their own.

domains). Interestingly, more than half of these patients had scored higher than 24 on the MMSE and 31% had obtained more than 79 on the 3MS, hence questioning the sensitivity of these cognitive scales. Tatemichi and Desmond (1996) used similar criteria and required failure (scoring below the fifth percentile of the control sample) in at least four of the 17 items of their neuropsychological battery, which evaluated memory, orientation, language, visuospatial ability, abstract reasoning, and attention. Cognitive impairment was detected in 35% of their 227 stroke patients. Remarkably, cognitive impairment had a significant effect on functional impairment and was a significant predictor of dependent living even after adjusting for physical disability.

In summary, the impact of vascular factors on cognitive impairment and dementia may prove to be more significant than considered traditionally. The presence of clinical stroke may aggravate this effect, but the effect may be mediated by other mechanisms. Longitudinal studies are urgently needed to determine whether these forms of cognitive decline associated with vascular risk factors, especially hypertension and diabetes, with or without clinical stroke herald dementia or not. The answer to this question will provide clues for the early detection of VD and will allow rational strategies for prevention and treatment to be designed. Unfortunately, all these cases of VCI will go undetected in the population if diagnostic criteria are not developed and validated.

2.5 Risk Factors for Vascular Cognitive Impairment

Risk factors for VD remain unclear despite intensive and extensive research in this field (Table 1). Investigation of risk factors for a particular disease requires, first, that patients be clearly separated from those not suffering from the disease and, second, that presence or absence of a specific factor be reliably established. The first condition is not fully met in investigations on VD. As will be discussed, current diagnostic criteria show deficiencies and their validity has not been neuropa-

Table 1. Risk factors significantly associated with vascular dementia in different studies

Reference	Risk factors	Sampling frame	Analysis
Compared to general population			
Meyer (1988)	Hypertension, cardiopathy, diabetes, smoking, carotid bruit	175 MID patients vs. 125 neurologically normal controls	Chi-squared
Katzman (1989)	Prior stroke, diabetes	15 MID/MIX patients vs. 350 unselected volunteer controls (stroke patients allowed)	Chi-squared
Lindsay (1997)	Hypertension, alcohol abuse, cardiopathy, pesticide exposure, low level of education, current use of aspirin	129 VD patients vs. 531 controls from a population-based study	Logistic regression
Yoshitake (1995)	Age, alcohol use, diabetes, stroke, systolic BP	828 subjects from a population-based study on dementia followed for 7 years; 50 developed VD	Logistic regression
Compared to nondemented stroke patients			
Ladurner (1982)	Hypertension	40 MID vs. 31 nondemented stroke patients	Chi-squared
Loeb (1988)	Hypertension, cardiopathy, associated hypertension, cardiopathy, and diabetes	40 MID vs. 30 normal and 30 stroke controls	Chi-squared
Tatemichi (1990)	Previous stroke, previous MI	726 patients from a stroke database; 116 were demented	Chi-squared
Gorelick (1993)	Hypertension, MI, age, low level of education, smoking, obesity (OR<1), systolic BP (OR<1)	61 MID vs. 86 nondemented patients with two or more strokes	Logistic regression
Tatemichi (1993)	Age, race (non-white), low level of education, diabetes, previous stroke	66 VD vs. 185 nondemented stroke patients	Logistic regression
Kokmen (1996)	Age, sex (male), second stroke, mitral valve prolapse	971 stroke patients followed for 6782 person-years (196 developed dementia)	Cox proportional hazards

MID, multi-infarct dementia; VD, vascular dementia; MI, myocardial infarction; BP, blood pressure; OR, odds ratio.

thologically contrasted. Moreover, given the degree of severity required by these criteria, a substantial proportion of patients with early cognitive changes will be excluded and considered as controls. Moreover, the absence of etiopathogenic classification forces the researcher to include patients with different diseases in the same group. With regard to the second requirement, the presence of risk factors has often been used to establish the diagnosis of VD, and the conclusion that vascular risk factors are indeed risk factors for VD may be significantly biased. The question of whether VD is associated with the same circumstances that determine the risk of stroke or whether there are additional factors remains unanswered. Remarkably, research in this area has been limited to traditionally known risk factors (hypertension, diabetes, hypercholesterolemia, cardiopathy). Other conditions such as increased fibrinogen, hypotension, hemodynamic states, immunological changes, or genetic factors are only recently starting to be acknowledged (Martinez-Lage and Hachinski 1998).

When patients with VD are compared with neurologically normal controls, diabetes, heart disease, hypertension, and carotid bruits are significantly associated with VD (Meyer et al. 1988). However, if stroke patients are allowed in a nonselected control group, only prior stroke and diabetes reach significance (Katzman et al. 1989). Studies on population-based surveys are scarce (Table 1). Using incident cases with a high rate of autopsy confirmation, Yoshitake and colleagues (1995) found age, systolic blood pressure, history of stroke, and alcohol abuse to be significant predictors on a multivariate analysis. Diabetes, diastolic blood pressure, and increased hematocrit were significant only on univariate analysis. History of hypertension, alcohol abuse, heart conditions, and exposure to pesticides/fertilizers or liquid plastics or rubbers as well as low educational level showed a significant effect in analysis performed in the Canadian Study on Health and Aging database (Lindsay et al. 1997). In this study, diagnosis of VD followed ICD-10 criteria, and no neuroimaging study was required. Patients with cognitive decline and no dementia were excluded from the control group, but also from the patient group.

Several hospital-based studies have attempted to analyze which are the factors that determine the development of dementia in some stroke patients but not in others (Table 1). In the series presented by Ladurner and colleagues (1982), only hypertension was found to be more frequent among demented stroke patients. Investigating patients with multiple infarcts, Loeb and colleagues (1988) reported a higher prevalence of hypertension, cardiopathy, and diabetes in those with dementia. In the Stroke Data Bank Cohort, only myocardial infarction and prior stroke were significantly associated with dementia (Tatemichi et al. 1990). Reports from the population-based study by Rochester have described the significant effect of age, male sex, recurrent stroke, and mitral valve prolapse on incident dementia (Kokmen et al. 1996). In a prospective study, Tatemichi and colleagues (1993) found a significant association of dementia and age, education, race, history of prior stroke, and diabetes. Gorelick and colleagues (1993) investigated patients with two or more infarcts and found that age, hypertension, proteinuria, myocardial infarction, smoking, and systolic blood pressure were significant predictors of dementia. Both hypertension and proteinuria lost significance when analysis were adjusted for educational level. Surprisingly, the presence of high systolic blood pressure showed a "protective" effect.

All these results are too varied to draw any conclusions. Methodological differences probably account for a significant proportion of such variability. In addition, different categories of VD have not been separately analyzed, and only traditional risk factors have been taken into consideration. Isolated reports have claimed the importance of other conditions such as increased platelet activation, increased fibrinogen, immunological abnormalities, or hypoxic–ischemic diseases. Some studies have claimed a role for allele 4 of apolipoprotein E as a genetic risk factor for VD, but anatomopathological studies have not confirmed this finding. Familial or genetic factors have received little attention in VD. Analysis of all these data and the fact that incidence of VD has not decreased significantly despite intensive control of vascular risk factors for several decades lead to the conclusion that further research is needed in this field.

3 Neuropathology of Vascular Cognitive Impairment

For many years, dementia of vascular origin, so-called arteriosclerotic dementia, was thought to be the consequence of chronic cerebral ischemia. The prevailing hypothesis was that aging would be accompanied by a gradual narrowing of brain arteries that would compromise cerebral blood flow and lead to a state of neuronal nutritional insufficiency. This would cause neuronal loss, atrophy, and dementia. In 1970, Tomlinson and colleagues clearly identified neurodegenerative and vascular conditions as the most frequent causes of dementia and put forward the notion that brain infarction, and not chronic ischemia, was the true mechanism of intellectual impairment. A thorough review of available pathological data and of several blood flow studies, including their own, led Hachinski et al. (1974) to

conclude that vascular disease causes dementia through the occurrence of multiple small or large cerebral infarcts and to propose the term multi-infarct dementia (MID).[4] This view has been confirmed by the observation that surgical interventions intended to increase cerebral blood flow (endarterectomy or extracranial–intracranial bypass) did not significantly change neuropsychological outcomes. Moreover, positron emission tomography (PET) studies showing that decreased blood flow in patients with AD or MID is not coupled with an increased oxygen extraction fraction have provided further evidence against chronic ischemia (Tatemichi et al. 1994). However, chronic ischemia should not be confused with brain infarction from a hemodynamic mechanism (hemodynamic dementia).[5] Infarcts in the boundary territory of two different cerebral arteries or ischemic necrosis in vulnerable regions may occur after sustained hypotension, and possibly in relation with severe carotid occlusive disease, but even in these cases it is the presence of focal damage and not chronic hypoperfusion that causes cognitive decline. Focal vascular damage to the brain is then considered the pathological substrate of VCI.

The neuropathologic expression of focal vascular damage to the brain in patients with VD is varied and heterogenous (Munoz 1991; Olsson et al. 1996). Different cerebral and blood vessel pathologies may be found alone or in combination in the same patient (Table 2). The most common changes consist of macroscopic arterial infarcts, lacunes, and white matter vascular damage.

Arterial infarcts occur in the territory of large or medium-sized arteries. They appear as areas of cavitated necrosis with abundant inflammatory cells, including activated astrocytes. They usually involve the cortex, although the underlying white matter and basal ganglia may be affected as well. Location, number, and total volume of these type of lesions have been described as significant determinants for the presence of dementia, but data from different studies are controversial.

Lacunar infarcts are small, up to 1.5-cm cavitary lesions in the territory of deep penetrating arteries. They appear most frequently in the basal ganglia, thalamus, basis pontis, or the white matter. Lacunes

Table 2. Neuropathological findings in the brain and blood vessels in patients with vascular demetia (VD): potential etiopathogenic mechanisms

Lesion	Neuropathological findings
Brain vascular lesions	
Cortical infarcts	
Large/medium-sized arterial infarcts	Atherothrombosis Cardioembolism Coagulopathies
Cortical microinfarcts (granular cortical atrophy)	Small-vessel disease Amyloid angiopathy Multiple microemboli/ microthrombosis Angiitis obliterans
Watershed infarcts	Hypoperfusion states Severe carotid stenosis/occlusion (?)
Other ischemic lesions	Selective ischemic necrosis: Laminar necrosis Hippocampal sclerosis Incomplete infarction (?) Global anoxia Mild focal ischemia (?) Hypoperfusion states
Subcortical infarcts	
Lacunes	Arteriolosclerosis Microthrombosis
Deep infarcts	Atherothrombosis Embolism Amyloid angiopathy (familial?) Hereditary arteriopathies (CADASIL)
Watershed infarcts	Hypoperfusion/hypoxic states Severe carotid stenosis/occlusion (?)
White matter vascular damage	
Lacunes/infarcts, spongiosis, "incomplete infarcts", perivascular demyelination	Astrocytic reaction, small-vessel disease, ischemia/hypoxia, increased permeability, increased fibrinogen
Vessel wall disease	
Large-artery disease	Atherosclerosis (aortic, carotid, cerebral arteries) Giant cell or other inflammatory arterities
Small-vessel disease	Arteriolosclerosis Hypertensive Nonhypertensive – hereditary (CADASIL), diabetic (?), other Amyloid angiopathy Sporadic, familia AD-associated Inflammatory/noninflammatory arteritis

CADASIL, cerebral autosomic dominant arteriolopathy with subcortical strokes and ischemic leukoencephalopathy; AD, Alzheimer's disease.

[4]MID is a term that was originally defined to highlight the fact that dementia from vascular causes (all types of VD) is due to the accumulation of infarcts and not chronic ischemia. At present, the use of this term is restricted to cases of VD with multiple macroscopic cortical infarcts.

[5]Hemodynamic VD refers to cases of VD in which the cause of cerebral infarction is hemodynamic (hypofusional states, hypotension, global anoxia) and should not be confused with chronic ischemia.

have to be clearly distinguished from dilated perivascular Virchow-Robin spaces, which are characterized by their smooth, membrane-covered margins and the presence of a central vessel. In this respect, lacunar and cribiform states should not be taken as synonyms.

The presence of white matter rarefaction (leukoencephalopathy) was initially highlighted by Binswanger and Alzheimer (Mast et al. 1995) and is a frequent neuropathological and radiological finding in patients with VD. Histologically, white matter lesions include cavitated infarcts (lacunes), focal areas of astrocytosis (glial scars, noncavitated infarcts), demyelination, and diffuse areas of vacuolization (edema) of oligodendrocytes and neuroglia with decreased oligodendrocyte density. Sometimes these areas of vacuolization are localized and surround a blood vessel. In most cases, white matter changes are accompanied by arteriolosclerosis.

Together with these three basic types of lesions, other neuropathological findings are less conspicuously described. The term incomplete infarction is usually ascribed to white matter lesions, as Brun and Englund (1986) applied it to designate areas of myelin and axonal loss with glial reaction but no necrosis or cavitation. However, the concept of incomplete infarction was originally suggested by Lassen (1982) to describe ischemic lesions involving gray and white matter that did not show cavitation or emollision. Such lesions are characterized by neuronal death with little or no glial reaction (García et al. 1996).

Laminar necrosis might be interpreted as an example of selective ischemic necrosis and has been described in patients with dementia. It appears as a destructive lesion selectively involving neurons in the vulnerable layers of the cerebral cortex and seems to be common in the boundary zones of two different vascular territories.

In hippocampal sclerosis, selective neuronal destruction in segments CA_1 and the subiculum of the hippocampal formation is accompanied by severe gliosis. This has been reported as a frequent finding in aged subjects with dementia. Since these changes occur in patients with global cerebral anoxia–ischemia, a vascular mechanism has been claimed for this type of neuropathological findings, but whether this is true for all cases of hippocampal sclerosis is under discussion, as other neurodegenerative findings are associated in some cases (Corey-Bloom et al. 1997).

Other ischemic lesions described as the only finding in patients with dementia include cortical microinfarcts, which appear as small brownish gliotic scars that distort the cortical ribbon. The term granular cortical atrophy is used in these cases to describe the macroscopic appearance when cortical microinfarcts are multiple and widespread.

Vessel wall lesions in patients with VD include atherosclerosis of the large or medium-sized cerebral arteries and small-vessel disease (Olsson et al. 1996). Atherosclerotic changes with accumulation of lipids, extracellular matrix components, and fibroblast proliferation are mainly distributed in the extracranial trunks, the circle of Willis, and the leptomeningeal arteries. Aortic atheromatous changes are also gaining significance as a possible source of artery-to-artery emboli. Two different pathologic entities are encompassed under the term of small-vessel disease: arteriolosclerosis and amyloid angiopathy. Arteriolosclerosis is characterized by the loss of muscular and elastic components of small arteries and arterioles, which are substituted by a hyaline material and collagenous strands (hyalinosis, fibrohyalinosis). Lipid-laden macrophages (lipohyalinosis) and fibrinoid necrosis may be added in some vessels. This type of angiopathy has been classically described in chronic hypertension, but may be present in normotensive subjects. Almost identical changes are described in patients with hereditary forms of VD. In amyloid angiopathy, a proteinaceous material that stains with Congo red is deposited in the wall of meningocortical arterioles, inducing media degeneration. It has been described in both sporadic and familial cases and is a common finding in patients with AD. Other forms of inflammatory or noninflammatory angiopathies may involve small vessels in rare and selected cases of VD (Table 2).

4 Vascular Mechanisms of Cognitive Impairment

Two hypothesis have traditionally been held to explain how a focal process such as cerebrovascular disease affects a global capacity of the brain such as cognition or intelligence. The volumetric hypothesis was initially formulated by Tomlinson et al. (1970). These authors argued that there is a threshold of volume of infarcted brain that, once reached, would induce dementia. A second line of thought has identified the topography or location of vascular lesions as the crucial factor. The first explanation would be based in a model of functional reserve of the brain to sustain cognition, while the second stresses the extremely complex topic of the anatomy of intelligence. Topographical or anatomical correlations of dementia and vascular lesions have always been biased by the requirement of memory impairment in the definition of dementia, as patients with cognitive decline without memory impairment have been systematically excluded. Cognitive functions may be classified as localized or disperse, depending on whether they have a concrete neuronal representation in one hemisphere, or a part of hemisphere, or not. Language, praxis, calculation, or visuospatial and

visuoconstructional abilities are examples of localized functions that may be damaged by focal vascular lesions. On the other hand, disperse functions include attention/concentration, memory, and executive functions, the most genuine functions involved in the definition of VCI and dementia. If intelligence or cognition is defined as "the aggregate or global capacity of the individual to act purposefully, to think rationally, and to deal effectively with the environment," this can only be sustained by the brain working as a whole. Consequently, any focal lesion, apart from those involving primary motor or sensory areas or vegetative control centers, could potentially affect cognition. In this respect, it may be important to bear in mind the model of anatomical correlations of complex functions based on large-scale neurocognitive networks reviewed by Mesulam (1990). This model includes the following points:

1. Components of a single function are represented in different interconnected locations throughout the brain.
2. Individual cortical or subcortical areas may belong to different overlapping networks and participate in several complex functions.
3. Lesions confined to a single region may result in multiple deficits.
4. Impairment of an individual complex function usually requires simultaneous involvement of several regions to be severe and persisting.
5. The same cognitive function may be affected by a lesion in different brain areas. According to this simplified, but useful model, cerebrovascular disease can affect cognition by destroying neuronal components of the cognitive networks both in the cortex or subcortical relay nuclei as well as by lesioning their connections (white matter disease). Both size and location of infarcts are important, and no assumptions should be made as to what cognitive functions are to be impaired.

If infarct volume or infarct location are the only important determinants of cognitive decline and dementia from vascular causes, all efforts should be invested in the investigation of stroke alone. Defining new entities such as VCI or even VD would be unnecessary, confusing, and time-wasting. However, after decades of investigation, numerous studies on dementia and stroke-related variables have yielded contradictory results that are by no means sufficient to explain all the questions surrounding the etiopathogenesis and pathophysiology of cognitive decline due to vascular causes.

The prevalence of cognitive impairment among stroke patients is much higher than that in the general population, and stroke survivors develop dementia five times more frequently than normal controls (Tatemichi and Desmond 1996). Nevertheless, the factors that determine why some stroke patients present with cognitive decline while others do not remain unclear. Some neuropathologic and radiologic studies have emphasized the volume of infarcted tissue. Tomlinson et al. (1970), for instance, placed the critical threshold at 100 ml infarcted brain, but other studies have shown that dementia may appear with as little as 1% of infarcted brain (del Ser et al. 1990). Regarding infarct location, the presence of infarcts in both hemispheres has been highlighted by some authors (Erkinjuntti et al. 1987), while others claim the importance of dominant hemisphere involvement (Liu et al. 1992). Some studies have not found any significant effect of infarct location or even clinical severity on the subsequent rate of dementia (Kokmen et al. 1996). The role of the type of ischemic lesion has also been analyzed, and the relevance of subcortical lacunar infarcts versus cortical lesions has been reported in some studies, but not all. Data considering the presence or absence of white matter changes are also controversial. Moreover, risk factor profiles do not differentiate between stroke patients with or without cognitive impairment including dementia. The problem is even more complicated if we take into account the increasingly recognized fact that VD may ensue in the absence of any clinical history of stroke. All these data lead to the conclusion that, in addition to volume and location, there must be other factors by which focal vascular brain damage induces intellectual impairment (Martinez-Lage and Hachinski 1998). Perhaps a better understanding of the intrinsic cellular and molecular mechanisms underlying brain ischemia may help to solve this question. Recent advances in our knowledge of the molecular biology of ischemia allow us to interpret cerebral infarction as a concept that surpasses that of the classical necrotic, cavitary lesion.

As previously described, Lassen and his colleagues introduced the concept of incomplete infarction to describe ischemic lesions that did not show "emollision" or cavitation. The term was later applied by Brun and Englund (1986) to describe similar lesions in the white matter. This type of "selective ischemic necrosis" may occur in cases of global cerebral ischemia as well as in the periphery of classical cavitary infarcts. It has been suggested that episodes of short-lived mild ischemia might induce these changes without causing radiologically detectable lesions, i.e., they would not be taken into account in volumetric or topographic studies. Whether these lesions may be detected by functional neuroimaging and whether they contribute to cognitive impairment is a matter for further research (García et al. 1996). Another form of neuronal death that may be induced by ischemia is that of apoptosis. There is enough

experimental evidence that ischemic injuries may cause delayed "programmed" neuronal death, thus opening the question of whether such a process might become uncontrolled or self-perpetuating and lead to progressive cell loss (Bredesen 1995). In this respect, the profound changes that ischemia induces at the gene expression level may become particularly relevant (Matsushima et al. 1996). Immediate-early genes, the amyloid precursor protein gene, and genes coding for proteins involved in stress or inflammatory responses are upregulated in ischemic conditions (Higashi et al. 1995). Again, dysregulation of these responses might induce further radiologically undetectable vascular brain damage and contribute to cognitive decline. The role of non-neuronal cell populations should also be considered. Astrocytes and microglia may become rapidly activated by ischemia and contribute to exaggerated acute-phase or inflammatory responses with overproduction of nitric oxide, cytokines, and free radicals that could mediate postischemic brain damage (Gehrmann et al. 1995). Remarkably, all these changes might be induced by mild short-lived ischemic insults that could be insufficient to cause clinical symptoms or radiologically detectable infarcts. Oligodendrocytes, for instance, are highly susceptible to ischemia and may be affected by mild focal or global hypoxic–ischemic injuries, with the subsequent white matter damage and dysfunction (Yamanouchi 1991).

In summary, the pathophysiology and etiopathogenesis of ischemic brain damage is multiple and varied. Cognitive impairment must be the result of the destruction of brain areas that sustain intellectual functions, but also of those regions that allow for compensation. It is possible that the instruments or methods currently used to quantify, add, or locate vascular brain damage are not accurate enough. A similar situation has occurred in AD. Counts of what could be easily seen (plaques and tangles) correlated poorly with cognitive decline or dementia severity. It was synaptic density reduction that best explained intellectual deficits. Apart from established stroke, the clinical and pathogenic relevance of short-lived or mild episodes of ischemia insufficient to cause symptoms or cavitary infarcts – whether embolic, thrombotic, or hemodynamic in origin – might be more important than traditionally believed. The value and applicability of new diagnostic tools, such as functional magnetic resonance imaging (MRI), receptor-oriented PET or single photon emission computed tomography (SPECT), and the relevance of analytical determinations of acute-phase proteins or inflammation mediators in blood or cerebrospinal fluid (CSF) should be further investigated, as they might well detect cellular and subcellular changes induced by ischemia.

5
Categorization of Vascular Cognitive Impairment

The unquestionable evidence that VD has multiple etiologies is one of the big paradoxes of the literature on vascular neurology. On the one hand, almost any text on VD includes a section on the classification of the different types of VD, but on the other, all sets of diagnostic criteria ignore this question (Bowler and Hachinski 1996). The confusion that this creates is easy to understand using the following analogy. It is difficult to imagine a study in which patients with dementia due to hypothyroidism and patients with vitamin B_{12} deficiency are analyzed together under a diagnosis of metabolic dementia. However, a similar, if not identical, situation is occurring in the field of VD. It is beyond doubt that clinical manifestations, neuroimaging features, risk factors, and treatment in patients with dementia due to multiple cardioembolic infarcts must be different from those concurring in patients with dementia due to multiple lacunes. Surprisingly, all these patients are grouped together with a diagnosis of VD in epidemiological, clinical, and radiologic studies and even in clinical trials. There is little doubt that this leads to uncertainty and prevents progress in the understanding of cognitive impairment and dementia due to vascular causes.

Several proposals to classify and categorize VD are available in the literature. Unfortunately, anatomic and pathophysiologic criteria are confusingly mixed, and only a few authors have proposed clinically recognizable categories. To be clinically useful, any classification of VCI should be based on etiology. The different categories should be supported by pathologic data and, more importantly, should be clinically recognizable. The first step consists in the characterization of neuropsychological deficits and their combination into clinical syndromes. A systematic definition of the clinical manifestations together with neuroradiological findings will allow the topography of vascular lesions to be determined and a diagnosis of the type of infarcts and subsequently their etiology to be approximated (Martinez-Lage and Hachinski 1998).

Two main types of clinical situations may be distinguished: (1) those in which cognitive deficits are restricted to one cognitive area and (2) those involving two or more intellectual functions (Table 3).

5.1
Syndromes Restricted to One Area of Cognition

Syndromes restricted to one area of cognition usually include all those patients with deficits in localized cognitive functions, i.e., functions with a concrete

Table 3. Categorization of vascular cognitive impairment

Clinical picture	Neuropathology/ neuroradiology findings
Circumscribed cognitive deficits	
Localized functions	Cortical infarcts
	Cortical microinfarcts
	Watershed infarction
	Laminar necrosis (?)
Disperse functions	
Memory	Selective vulnerability (hippocampal sclerosis)
	PCA infarcts
	Subcortical infarcts (thalamus, basal forebrain)
	Lacunes (thalamus, basal forebrain?)
Executive functioning	Cortical infarcts/ microinfarcts (frontal)
	Lacunes, subcortical infarcts (frontal white matter, basal ganglia, thalamus, capsular genu)
	White matter disease
Multiple cognitive deficits	
With focal lesions	
Angular gyrus	Cortical infarct
PCA infarcts	Cortical infarct
Capsular genu	Lacunes/subcortical infarcts
Basal forebrain	Lacunes/subcortical infarcts
Thalamus	Lacunes/subcortical infarcts
Multiple lesions	
Cortical VCI	Multiple cortical infarcts/ microinfarcts
	Watershed infarction
	Selective vulnerability (laminar necrosis?)
Subcortical VCI	Lacunes (lacunar state)
	Lacunes (lacunar state) + white matter disease
	Watershed infarction
	Subcortical infarcts
White matter VCI	Vascular white matter disease

PCA, posterior cerebral artery; VCI, vascular cognitive impairment.

neuronal representation (Freeman et al. 1989). Disturbances of language (aphasia), reading/writing (alexia/agraphia), praxis (apraxia), and recognition (agnosia, neglect) will usually respond to cortical lesions. Exceptions include cases of transcortical aphasia or hemineglect syndromes caused by subcortical lesions involving the thalamus or basal ganglia. In both cases, atheroembolic/cardioembolic infarcts will be the underlying pathogenic mechanism in the majority of cases. Some cases of isolated transcortical aphasia or visuospatial disturbance may occur in patients with watershed infarcts with a possible hemodynamic mechanism.

Isolated impairment of memory functions is rare in cerebrovascular disease (Bowler and Hachinski 1996). Thalamic lesions and basal forebrain infarcts cause severe memory impairment, but are usually accompanied by other deficits. Episodes of global anoxia–ischemia may damage both hippocampi and produce amnesia. A similar mechanism has been claimed to explain a new neuropathological entity known as hippocampal or subicular sclerosis. This condition, characterized by profuse neuronal loss and astrocytosis in both hippocampal formations, has been described in old patients with dementia and has been hypothetically ascribed to ischemia. However, previous history of episodic anoxia cannot always be elicited, and some patients present other neuropathological findings that may suggest a neurodegenerative disease (senile plaques, neurofibrillary tangles, ballooned neurons, argyrophilic grains).

Executive functions include tasks such as attention, planning, programming, anticipation, set shifting, or memory search strategies, so that describing an "isolated" compromising of executive functions may be imprecise. Syndromes of executive dysfunction may result from different lesions with different etiopathogenic mechanisms: cortical infarcts in the territory of the anterior cerebral artery or the anterior branches of the middle cerebral artery; lacunar infarcts in the subcortical frontal white matter, thalamus, capsular genu; deep infarcts in the caudate; or basal forebrain infarcts (Mahler and Cummings 1991). Together with the syndromes of executive dysfunction, disturbances of motivation or drive and profound personality changes may characterize the clinical picture.

5.2 Focal Lesions Causing Syndromes Involving More than One Cognitive Domain

Whimsically placed infarcts (i.e., rare infarcts that occur in places in the brain where anatomical substrates for different cognitive functions coincide) may cause significant impairment in several cognitive areas, regardless of whether or not dementia (strategic infarct dementia[6]) ensues. Typical examples include the angular gyrus syndrome (aphasia, alexia, visuoconstructional disturbances, and Gerstmann syndrome) due to atheroembolic occlusion of the posterior branches of

[6]This term is applied to describe cases in which a single infarct, whimsically placed, causes a clinical picture of dementia. Infarcts in different locations, such as the angular gyrus, thalamus, basal forebrain, or capsular genu, which respond to different etiopathogenic mechanisms (lacunar, atheroembolic, or cardioembolic) are included.

the middle cerebral artery; paramedian unilateral or bilateral thalamic lacunar or embolic infarcts that may cause severe amnesia, apathy, and executive deficits, with or without aphasia; basal forebrain lesions secondary to anterior communicating aneurysm rupture or surgery that cause severe amnesia and personality changes; lacunar infarcts in the caudal capsular genu that may cause severe apathy, personality changes, and amnesia (Tatemichi et al. 1994). Similarly, embolic or atherothrombotic infarcts in the territory of the posterior cerebral artery may cause different combinations of visual agnosia, aphasia, reading or writing disturbances, and memory loss. Most proposed classifications of VD have considered strategic infarct dementia as a separate entity; however, as can be seen, the mechanisms leading to these type of lesions are varied and the category should disappear if etiopathogenic criteria are to prevail in the categorization of VCI.

Most of the conditions included in the two previous categories of VCI are clinically recognizable and their etiopathogenic implications may be easily interpreted. One of the most important aspects of this proposed classification is to detect these cases of focal or localized cognitive syndromes as cases of VCI in which cognitive changes may progress and worsen in the direction of dementia. Further research is needed to understand the mechanisms and risks factors by which such potential progression takes place.

5.3 Syndromes Involving Multiple Cognitive Areas

Patients with multiple cognitive syndromes which cannot be ascribed to a single lesion will probably represent the most frequent type of VCI in daily clinical practice. Detailed studies are still needed to establish the different categories following clinical and etiological criteria. Meanwhile, it can be hypothesized that most patients will have one of the following three clinical forms according to the pattern of neuropsychological deficits: cortical, subcortical, or white matter forms.

5.3.1 Cortical Forms

Patients with cortical forms of VCI will present with deficits in cognitive areas with a clear cortical representation such as language, praxis, memory, visuospatial and visuoconstructional functions, or calculation. Such deficits would be expected in patients with multiple cortical infarcts that usually respond to atherothrombotic or embolic mechanisms. Cortical damage (granular cortical atrophy) may also occur in the context of small-vessel disease. Amyloid angiopathy may potentially cause cortical microinfarcts, and it might be recognized in the absence of other cardiovascular processes, especially if there is positive family history or concomitant history of lobar hemorrhage. However, the clinical pattern in these patients may be dominated by those deficits due to an associated AD. Moreover, amyloid angiopathy may also cause subcortical deficits due to white matter ischemic changes. Multiple microembolic or thrombotic occlusion of cortical vessels has been described associated with cardiac surgery, ischemic heart disease, and thromboangiitis obliterans (Erkinjuntti and Hachinski 1993). Hemodynamic disorders and distal field ischemia may also lead to selective ischemic necrosis of vulnerable cortical layers and induce cortical cognitive syndromes in the absence of radiologically detectable infarcts. In such cases, blood pressure monitoring and functional neuroimaging techniques may prove useful for clinical diagnosis.

5.3.2 Subcortical Forms

Symptoms and signs of executive deficits, disturbances of motivation and drive, attentional deficits, personality changes, and affective disorders are the predominant features of subcortical cognitive syndromes (Cummings and Benson 1992). Memory loss may also be prominent but, in contrast to cortical forms of amnesia, patients may show good performance in cued recall or assisted learning tasks. Disturbances of motor control, including gait disturbances, slowness, rigidity, tremulousness, or pseudobulbar syndrome, are usually present. Subcortical cognitive syndromes have traditionally been regarded identical to either lacunar state or Binswanger's disease (Loeb and Meyer 1996). This remark needs further consideration. On the one hand, they probably represent the same clinical and radiological entity. Binswanger's disease is characterized by the presence of white matter disease and small subcortical lacunar infarcts, while lacunar states that cause cognitive decline are frequently accompanied by white matter disease. On the other hand, a diagnosis such as Binswanger's disease may be misleading in terms of etiology and pathogenesis. There are several processes that may cause subcortical cognitive deficits with white matter changes and small subcortical infarcts or lacunes. Hypertension is probably the most frequent one, but other entities such as amyloid angiopathy, hereditary forms of small-vessel disease, or coagulation disorders with increased fibrinogen should be taken into consideration, especially in normotensive patients (Caplan 1995). Hypotension and other processes leading to global brain ischemia or anoxia have also been associated with subcortical cognitive deficits (Brun 1994). Among the hereditary forms, the condition known as cerebral autosomic dominant arteriolopathy with subcortical strokes and

ischemic leukoencephalopathy (CADASIL) may prove to be more frequent than initially considered (Chabriat et al. 1995). It is characterized by the presence of subcortical strokes and dementia at unusually early ages as well as migraine with aura, psychiatric disorders such as depression, or symptomatic/asymptomatic white matter changes (leuko-araiosis, L-A) detected by CT or MRI in members of the same family. The responsible mutation has been located in the Notch-3 gene on chromosome 19.

Subcortical forms of VCI should be identified as clinical syndromes and subsequently classified by etiology. Terms such as Binswanger's syndrome or subcortical VCI are acceptable if they are followed by a statement about possible or probable etiopathogenic mechanisms. A distinction between hypertensive and nonhypertensive forms could be a first useful step toward a better categorization and understanding of this frequent form of VCI and its potential treatment and prevention.

5.3.3 White Matter Forms

There is enough evidence to support the hypothesis that white matter disease may occur in the absence of lacunes or other small infarcts. Some studies in which patients with discrete lesions are excluded have shown that white matter disease is associated with deficits in attention, speed of mental processing, and frontal lobe functions, while memory, verbal, and visuospatial abilities may be preserved. It may be interpreted that isolated white matter changes may induce similar but more restricted or less severe deficits than subcortical nuclei lesions. The clinical distinction is important if mechanisms leading to isolated white matter disease are different from those leading to white matter changes with white matter infarcts (Martinez-Lage and Hachinski 1998).

In summary, classification and categorization of VCI is a complex and laborious task. However, avoidance of this crucial question may lead to confusion and misunderstanding. If patients are not grouped together according to possible or probable etiologic criteria, no conclusions can be reached regarding pathophysiology or detection of risk factors. If patients with different diseases are mixed, no efficient preventive or therapeutic measures can be adequately tested.

6
Neuroimaging in Vascular Cognitive Impairment: The Matter of White Matter

That VCI can not be diagnosed in the absence of evidence of cerebrovascular disease requires little discussion. Such evidence must be obtained from any of two sources, clinical and/or radiological evaluation. In a patient with a sudden onset of hemiplegia and corticospinal signs on examination few clinicians would doubt the vascular etiology, and a CT scan would be requested to detect a hemorrhagic or an ischemic lesion. Even if brain CT shows no changes the diagnosis of ischemic stroke would be maintained if symptoms or signs lasted more than 24 hours. In patients with cognitive complaints, usually in the form of poor memory or disorientation, things are not that simple. First, clinical information is difficult to collect. The onset of symptoms is usually poorly recalled by patients or relatives and when it is, the fact that it is not abrupt does not rule out a vascular origin. With regard to examination, the importance of focal motor, sensory, visual signs has received significant attention but the fact that cognitive deficits, if well characterized, constitute focal signs themselves has been overlooked. Examination of cognition requires time and experience and the pattern of intellectual deficits is not always easy to establish. In such situation it is difficult to ascribe symptoms and signs to a particular neuroradiological finding. The role of neuroimaging techniques in detecting vascular lesions in patients with cognitive decline is evident, but information provided by clinical and radiological sources must be carefully contrasted.

Together with the inherent difficulty of establishing clinico-topographic correlations, controversies and disparities in the field of neuroimaging and VCI are easily explainable. To date, only patients fulfilling current criteria for VD, that is, patients with memory impairment and with advanced and severe cognitive decline (sufficient to impair social or occupational functioning), have been analyzed. If part of the cognitive pattern (memory) and the degree of severity are pre-established, conclusions are necessarily biased. In addition, given the degree of severity required, enough time may have been allowed for different vascular mechanisms to concur in the same patient and the specific role of each one of them can not be discerned. This is the other indispensable role of neuroimaging in the diagnosis of VCI: to assist in the etiopathogenic classification of patients.

6.1
Brain Infarction

Neuroimaging studies in patients with VCI or VD may show any kind of infarcts which appear as hypodense areas on CT and hyperintense/hypointense areas on T_2-weighted/T_1-weighted MRI images (Cummings and Benson 1992). The size and distribution of lesions,

together with clinical manifestations, may help in the differential etiologic diagnosis. Medium-sized or large infarcts involving the cortical areas in an end-artery distribution usually respond to atherothrombotic or cardioembolic mechanisms. The presence of multiple wedge-shaped cortical infarcts or evidence of hemorrhagic transformation (which can be demonstrated in the chronic phase by the low signal intensity of ferritin and hemosiderin on T_2-weighted MRI images) may reinforce the latter. Evidence of hemorrhagic infarction may also suggest an underlying amyloid angiopathy. Cortical infarcts involving the posterior frontal lobe and posterior parieto-occipital regions are compatible with watershed or border-zone infarction and might suggest a hemodynamic origin. Subcortical lesions include lacunar infarcts, but also atheroembolic and border-zone infarcts, which may be distinguished according to size, distribution, and other clinical and laboratory data. Lacunar infarcts appear as small lesions, less than 1.5 cm in size, and are distributed in the territory of deep penetrating arteries (white matter, thalami, basal ganglia, brainstem). Dilated Virchow-Robin spaces, which show similar behavior to the CSF on the different MRI sequences, should not be confused with lacunes, which appear hyperintense with respect to the CSF on proton density images. All these data, adequately combined with appropriately collected clinical and neuropsychological information as outlined in the previous section, will surely help to classify patients with VCI according to possible and probable etiologies, and those at risk of developing further cognitive decline will be better identified and characterized. In this respect, two additional neuroimaging findings are relevant: atrophy and white matter changes.

6.2 Atrophy

Cortical and/or subcortical atrophy is a frequent finding in patients with cerebral infarcts and dementia when compared to nondemented stroke patients (Gorelick et al. 1992; Chiu et al. 1992). This has been expressed by different parameters, such as generalized atrophy, atrophy of the third ventricle, ventricle to brain ratio, mean ventricular volume, lateral ventricle enlargement, total ventricular and subarachnoid space areas, or ventricular index. The importance of this finding resides in the fact that it may be a significant predictor of future dementia in patients with stroke, as was demonstrated in the longitudinal follow-up of the Stroke Data Bank Cohort (Tatemichi et al. 1990). Both previous stroke and cortical atrophy at stroke onset, as detected by direct visual analysis of CT scan, were the most important predictors of incident dementia. Other parameters, such as atrophy of the corpus callosum, have been reported as radiological correlates of cognitive impairment of vascular origin.

6.3 White Matter Vascular Damage

The introduction of CT scanning in the diagnostic evaluation of neurological patients allowed the detection of changes in the white matter in a substantial proportion of the adult population. This white matter rarefaction, which appears as a hypodense area surrounding the ventricles and spreading into the subcortical white matter (centrum semiovale) in the most severe cases, was soon related to the presence of cognitive decline, dementia, and a higher incidence of vascular risk factors (Meyer et al. 1992). Many authors hurried to diagnose patients with this finding as Binswanger's disease patients, and an "epidemic" of this previously rare disease invaded the literature (Román 1987). Pathologic studies described lacunar infarcts, areas of demyelination, and arteriolosclerosis in many cases, but also dilated perivascular spaces, hydrocephalus, and ependimitis granularis as the histologic substrate. To emphasize this heterogeneity and avoid any presumptions about etiology, the term leuko-araiosis (L-A) was introduced and widely used (Hachinski et al. 1987). The introduction of MRI brought more confusion and misunderstanding. Due to its high sensitivity, MRI detected some changes in the white matter in up to 80%–90% of healthy subjects older than 60. It was not surprising that L-A detected on MRI did not correlate with cognitive dysfunction or vascular risk factors (Chimowitz et al. 1989). Only when periventricular and deep white matter changes were distinguished and semiquantitative or quantitative grading approaches were applied were the clinical significance and the vascular origin of L-A identified (Schmidt et al. 1993). Pathological studies confirmed that small areas of L-A surrounding the frontal horns corresponded to a normal anatomical structure, the subcallosal fascicle. Similarly, thin areas adjacent to the body of the lateral ventricle ("rims" or "halos") and the frontal/occipital horns ("caps") were the expression of ependimits granularis, caused by ependimal disruption and periventricular gliosis. However, when periventricular changes extended into the subcortical white matter vascular, changes such as arteriolosclerosis and ischemic demyelination were present. Deep white matter changes were classified as punctate, patchy, patchy–confluent, and diffuse and were related to perivascular demyelination, spongiosis, gliosis, decreased oligodendrocyte densities, and lacunar or other infarcts. Small-vessel disease was present

in most cases. In some instances, punctate lesions corresponded to dilated perivascular Virchow-Robin spaces (Hachinski and Munoz 1991). With regard to VD, several studies showed that a moderate to severe degree of L-A was present in 50%–80% of patients as compared to 20% of healthy controls or 25%–30% of AD patients. In addition, it is important to emphasize that the presence of L-A on CT has been associated with an extra risk of future stroke in patients with transient ischemic attacks (TIA) or minor stroke, as well as with cerebrovascular progressive disease in patients with symptomatic lacunes (Van Zagten et al. 1996).

The etiopathogenesis of L-A remains a matter of open debate. Together with ischemia and infarction, some authors have suggested that small-vessel disease could lead to increased permeability of the vessel wall and leakage of serum proteins to the surrounding white matter (Munoz 1991; Erkinjuntti et al. 1996). This could lead to an astrocytic reaction, gliosis, spongiosis, and further ischemia. The role of hypotension and short-lived episodes of ischemia leading to white matter "incomplete infarction" has been emphasized by some groups (Brun 1994). In this respect, it is well known that oligodendrocytes are highly vulnerable to mild ischemia. White matter changes are also frequently associated with amyloid angiopathy, which typically affects meningocortical and not white matter vessels. A mechanism of distal field ischemia has been suggested to explain these cases. In summary, there may be different mechanisms leading to white matter changes, but all of them share a vascular origin. Subjects with moderate to severe L-A, especially if involving the deep white matter, show cognitive deficits in executive functions and speed of mental processing as well as a higher incidence of vascular risk factors such as hypertension and less frequently diabetes or cardiac disease. Increased fibrinogen or factor VIIc activity (Breteler et al. 1994), hypotension (Tarvonen-Schröder et al. 1996), and genetic factors (CADASIL and other hereditary angiopathies) have also been implicated and require further investigation.

CT neuroimaging criteria for the diagnosis of VD have been proposed (Pullicino et al. 1996). According to number and volume of infarcts, severity of L-A, and brain atrophy (ventricular index), the presence of cerebrovascular disease is graded on a scale from 0 to 3. These criteria were developed to improve the reliability of National Institute of Neurological Disorders and Stroke–Association Internationale pour la Recherche et l'Enseignement en Neurosciences (NINDS-AIREN) criteria. This was the main outcome measure to which they were compared, and they were subsequently subjected to the limitations that apply to the use of these diagnostic rules.

7
Clinical Diagnosis of Vascular Cognitive Impairment: Usefulness of Current Diagnostic Criteria for Vascular Dementia

There are two types of diagnostic criteria sets available for clinical practice or research purposes. On the one hand, the criteria from the *Diagnostic and Statistical Manual IV* (DSM-IV; American Psychiatric Association 1994) and those from the *International Classification of Diseases 10* (ICD-10; World Health Organization 1993) are included as general diagnostic tools and outline general principles for the diagnosis of dementia and its different types, including AD, VD, and others. In both sets of criteria, dementia is defined by the presence of memory impairment plus deficits in other cognitive areas, and these deficits must be sufficiently severe to interfere with social or occupational functioning. With regard to VD, history of TIA or stroke, presence of focal signs, and laboratory data would provide evidence of underlying cerebrovascular disease. How this relates etiologically to the cognitive deficits is left to the clinician and is a subjective matter. On the other hand, the NINDS-AIREN criteria (Román et al. 1993) and the criteria from the State of California AD Diagnostic and Treatment Centers (ADDTC; Chui et al. 1992) were specifically designed and operationalized for the diagnosis of VD.

The NINDS-AIREN criteria add little regarding the definition of dementia, which is described as a "decline from a [...] level of functioning manifested by impairment of memory and of two or more cognitive domains." Again, the degree of severity is required to be serious enough to interfere with activities of daily living. Once dementia is considered to be present, a diagnosis of probable VD requires the presence of focal signs consistent with stroke (with or without history) and (not or) radiological evidence of cerebrovascular disease (including large arterial infarcts, multiple lacunes, or extensive white matter lesions). Once dementia and cerebrovascular disease are present, they are considered etiologically related only on a pathochronological basis, i.e., if dementia appears 3 months after a clinical stroke or if cognitive decline starts abruptly or follows a fluctuating or stepwise course. The NINDS-AIREN criteria define a category of "possible VD" for patients with dementia and focal signs when brain imaging studies to confirm vascular disease or the pathochronological criteria are missing. The ADDTC criteria offer two original and interesting modifications. First, no specific patterns of cognitive impairment are required as far as it is not "isolated to a single narrow category"; in other words, presence of

memory impairment is not compulsory for the diagnosis of dementia. Second, the diagnosis of VD can be established when there is history of two clinical strokes and radiological evidence of at least one infarct outside the cerebellum. Clinical history of stroke or presence of focal signs is not considered necessary in all cases if there is evidence of two or more ischemic strokes on neuroimaging studies. The temporal coincidence of clinical stroke and dementia onset is considered relevant in patients with only one clinical stroke but is not necessary, and no specific chronological pattern of cognitive decline is required. Those patients with dementia and a single stroke not clearly related to the onset of dementia and those with a Binswanger's syndrome who also show vascular risk factors and extensive white matter changes may be diagnosed as possible VD patients.

The applicability of all these criteria in daily clinical practice is seriously limited. With regard to the definition of dementia, most of them require the presence of memory loss as well as an advanced degree of severity. Hence patients with VCI who do not have amnesic deficits and, most importantly, those early cases in which preventive measures may prove more efficient are left undiagnosed. Moreover, the requirement of clinical symptoms of stroke, presence of focal signs, and a specific pattern of temporal relationship between stroke and dementia onset or a stepwise progression is based on generally accepted hypotheses that have never been specifically tested (Bowler and Hachinski 1996). Epidemiological studies have shown that a large proportion of patients with VD have never had symptoms of stroke or TIA (Yoshitake et al. 1995). The important question of white matter changes on CT or MRI is treated too superficially by the summary of the NINDS-AIREN criteria. The way this matter is discussed in the text of the paper is acceptable, but when summarized in the operational criteria leads to confusion. White matter changes are considered a matter of future research by the ADDTC criteria (published in 1992). Finally, none of the criteria attempts to propose a classification of the different etiopathogenic types of VD. By avoiding this matter, entities as different as dementia due to cardioembolic infarcts and that due to multiple lacunes or watershed infarcts are all considered a single condition. None of the criteria sets has been validated in neuropathologically confirmed patients, and the inter-rater reliability of some of them has been questioned. It is also important to note that the different sets are not interchangeable and the frequency of VD in the same population may vary from 6% to 25% depending on the applied criteria (Pohjasvaara et al. 1997).

The alternative concept of VCI will probably shed light in this respect. New criteria to diagnose VCI should be the result of prospective studies on patients with early cerebrovascular disease (from those with only risk factors present to those with established stroke) in which functional, neuropsychological, and neuroimaging data are gathered systematically and analyzed without preformed ideas. A selected comprehensive neuropsychological battery administered to these patients would allow cognitive impairment to be defined according to consentaneous cutoff values in each test without emphasizing memory or any other cognitive domain. Short screening tests would be developed and validated from these neuropsychological tests. The MMSE is widely used, but its sensitivity in detecting cognitive decline in cerebrovascular patients may be too low (Grace et al. 1995). Patients could be classified according to possible or probable etiopathogenic criteria. Patterns of cognitive deficits could then be defined for each group of cerebrovascular disease. Patients with cognitive decline, from early changes to dementia, could be then compared to those without such decline in terms of clinical and neuroimaging features. Results from such an analysis could lead to the formulation of rational and reliable criteria for the diagnosis, early detection, and differentiation of VCI from other forms of cognitive decline such as AD. Using the current criteria, a differential diagnosis between AD and VD is theoretically impossible on a clinical basis, as dementia is identically defined in both cases. However, if patients are detected early and cognitive deficits are thoroughly characterized, both diseases should be clearly distinguished. AD patients may show impairment of orientation, memory, and language, while VD patients tend to be more impaired in executive functions such as attention, self-regulation, planning, and fine motor coordination (Villardita 1992).

8 Cerebrovascular Pathology in Alzheimer's Disease: New Concepts on Mixed Dementia

Mixed dementia is frequently mentioned as the third most frequent cause of dementia after AD and VD in numerous studies. To a significant extent, its study and characterization have been systematically avoided. This is understandable if the label of mixed dementia is only applied to those patients in which AD and vascular lesions coincide but both are severe enough to cause dementia on their own. This is how Tomlinson and colleagues originally defined the term. However, data reported in the last years on the association of vascular pathology and AD suggest that, in a proportion of patients, there might be patho-

physiologic interactions in the production of dementia and even a common etiopathogenic origin of the two processes. Epidemiologic data have shown that hypertension (Skoog et al. 1996), diabetes (Ott et al. 1996), and atrial fibrillation are significant risk factors for AD. Moreover, as mentioned, stroke patients develop AD more frequently than the general population. Numerous autopsy series have demonstrated the presence of ischemic infarcts in up to 30% of patients with a pathologic (not clinical) diagnosis of AD (not mixed dementia) (Olichney et al. 1995). Interestingly enough, in a group of patients with similar AD changes at autopsy, those who also had small infarcts in the thalamus, basal ganglia, or frontal white matter had symptoms dementia while they were alive much more frequently than those without infarcts. The risk of having symptoms of dementia increased by a factor of 20 in association with ischemic infarcts (Snowdon et al. 1997). Other studies show that, when infarcts are present, the density of neurofibrillary tangles is significantly lower and, contrary to what happens in "pure" AD, cognitive scores do not correlate with this pathological marker (Nagy et al. 1997). Another difference, from the clinical point of view, involves the age of onset, which is delayed at least 6 years in patients with vascular lesions. There is therefore enough neuropathological and clinical evidence to suggest that some cases of AD with vascular lesions behave differently to pure AD. Hypothetically, these patients might have remained free of dementia had they not had any cerebrovascular disease. Further research is required to solve the question of whether the association of the two pathologic processes is merely coincidental or whether they are etiopathogenically related. AD is usually accompanied by amyloid angiopathy, and this increases the risk of having an infarct at autopsy, especially if hypertension is present. Amyloid angiopathy may also increase the incidence of white matter ischemic changes. On the other hand, cerebral ischemia may upregulate the expression of the amyloid precursor protein gene and facilitate the deposition of β-amyloid in a similar way to brain trauma.

9 Conclusions and Outlook

Cerebrovascular disease with or without clinical stroke is a frequent cause of cognitive impairment and dementia. VD has traditionally been considered treatable and preventable, but many questions are still unclear regarding its risk factors. More importantly, early detection is not possible as no diagnostic criteria exist. Current diagnostic criteria for VD are only fulfilled by patients with advanced cognitive decline that matches a specific neuropsychological pattern. For these reasons, the concept of VCI was formulated as a necessary new approach to this important medical challenge. VCI should not be considered a new entity, but rather a change of perspective. Cognitive decline from vascular causes should be identified in its initial stages, diagnosed on the basis of reliable clinical and radiologic criteria and classified according to its multiple etiologies. As has been suggested, an international database of prospectively collected and longitudinally followed patients will certainly be the best scenario to achieve this task. Clinical data on age, symptoms and mode of onset, history of cerebrovascular events, and pattern of progression as well as demographic data regarding level of education, employment, and physical and intellectual activities should be systematically collected. Results from clinical neurological and cardiovascular examinations and measures of functional and physical disability should be gathered on standardized forms. Information on risk factors should be simple and complete, including traditional and putative risk factors as well as information on whether they are treated and effectively controlled. Exhaustive questionnaires on family history are essential, and quantitative and comparable data on cognitive functions are mandatory. While valid short screening tests are selected or developed, a comprehensive battery of neuropsychological tests would be desirable. Neuroimaging data should provide information on number, type, and location of infarcts, and white matter changes should be classified according to type, location, and degree of severity. Simple consentaneous measures of brain atrophy should be obtained. Genetic testing for apolipoprotein E genotypes has been recommended for any research project on cognitive decline and dementia. DNA should be stored as new susceptibility genes will soon be discovered. Laboratory tests including SPECT, echocardiogram, carotid ultrasound, Holter-ECG, blood pressure monitoring, autonomic function tests, tests for sleep apnea, vasculitis screening, and others may prove useful in selected cases. The combination of clinical and laboratory data would allow patients to be classified according to possible or probable etiopathogenic mechanisms. By doing so, it will be possible to discern what cognitive deficits are caused by what cerebrovascular process, and reliable diagnostic criteria will be formulated. Longitudinal follow-up will help to detect and understand the factors that determine further progression of cognitive deficits and ultimately dementia. The rationale for future therapy and prevention of VCI will then be established.

10 References

American Psychiatric Association (1994) Diagnostic and statistical manual of mental disorders, 4th edn. American Psychiatric Press, Washington, DC

Bowler J, Hachinski VC (1996) History of the concept of vascular dementia: two oppsing views on current definitions and criteria for vascular dementia. In: Prohovnik I, Wade J, Knezevic S, Tatemichi T, Erkinjuntti T (eds) Vascular dementia: current concepts. Wiley, Chichester, p 1

Bredesen DE (1995) Neural apoptosis. Ann Neurol 38: 839–851

Breteler MMB, Claus JJ, Grobbee DE et al (1994) Cardiovascular disease and distribution of cognitive function in elderly people: the Rotterdam study. Br Med J 308: 1604–1608

Brun A (1994) Pathology and pathophysiology of cerebrovascular dementia: pure subgroups of obstructive and hypoperfusive etiology. Dementia 5: 145–147

Brun A, Englund E (1986) A white matter disorder in dementia of the Alzheimer type: a pathoanatomical study. Ann Neurol 19: 253–262

Caplan LR (1995) Binswanger's disease-revisited. Neurology 45: 626–633

Chabriat H, Vahedi K, Iba-Zizen MT et al (1995) Clinical spectrum of CADASIL: a study of 7 families. Lancet 346: 934–939

Chimowitz MI, Awad IA, Furlan AJ (1989) Periventricular lesions on MRI. Facts and theories. Stroke 24: 7–12

Chui HC, Victoroff JI, Margolin D et al (1992) Criteria for the diagnosis of ischemic vascular dementia proposed by the State of California Alzheimer's disease diagnostic and treatment centers. Neurology 42: 473–480

Corey-Bloom J, Sabbagh MN, Bondi MW et al (1997) Hippocampal sclerosis contributes to dementia in the elderly. Neurology 48: 154–160

Cummings JL, Benson DF (1992) Vascular dementias. In: Cummings JL, Benson DF (eds) Dementia: a clinical approach. Butterworth Heinemann, Boston, pp 153–176

Del Ser T, Bermejo F, Portera A et al (1990) Vascular dementia. A clinicopathological study. J Neurol Sci 96: 1–17

Erkinjuntti T, Hachinski VC (1993) Rethinking vascular dementia. Cerebrovasc Dis 3: 3–23

Erkinjuntti T, Ketonen L, Sulkava R et al (1987) CT in the differential diagnosis between Alzheimer's disease and vascular dementia. Acta Neurol Scand 75: 262–268

Erkinjuntti T, Benavemte O, Eliasziw M et al (1996) Diffuse vacuolization (spongiosis) and arteriolosclerosis in the frontal white matter occurs in vascular dementia. Arch Neurol 53: 325–332

Freeman R, Bear D, Greenberg MS (1989) Behavioral disturbances in cerebrovascular disease. In: Toole JF (ed) Handbook of clinical neurology, vol 11 (55). Vascular diseases, part III. Elsevier, Amsterdam, pp 137–150

Garcia JH, Lassen NA, Weiller C et al (1996) Ischemic stroke and incomplete infarction. Stroke 27: 761–765

Gehrmann J, Banati RB, Wiessnert C et al (1995) Reactive microglia in cerebral ischaemia: an early mediator of tissue damage? Neuropathol Appl Neurobiol 21: 277–289

Gorelick PB, Chatterjee A, Patel D et al (1992) Cranial computed tomographic observations in multi-infarct dementia. Stroke 23: 804–811

Gorelick PB, Brody J, Cohen D et al (1993) Risk factors for dementia associated with multiple cerebral infarcts. A case-control analysis in predominantly African-American hospital-based patients. Arch Neurol 50: 714–720

Grace J, Nadler JD, White DA et al (1995) Folstein vs modified mini-mental state examination in geriatric stroke. Stability, validity, and screening utility. Arch Neurol 52: 477–484

Hachinski VC, Bowler JV (1993) Vascular dementia. Neurology 43: 2159–2160

Hachinski VC, Munoz DG (1991) Leuko-araiosis: an update. Bull Clin Neurosci 56: 24–33

Hachinski VC, Lassen NA, Marshall J (1974) Multi-infarct dementia: a cause of mental deterioration in the elderly. Lancet 2: 207–210

Hachinski VC, Potter P, Merskey P (1987) Leuko-araiosis: an ancient term for a new problem. Arch Neurol 44: 21–23

Hébert R, Brayne C (1995) Epidemiology of vascular dementia. Neuroepidemiology 14: 240–257

Higashi T, Nishi S, Nakai ZA et al (1995) Regulatory mechanisms of stress response in mammalian nervous system during cerebral ischaemia or after heat shock. Neuropathol Appl Neurobiol 21: 471–483

Jellinger K, Danielczyk W, Fischer P et al (1990) Clinicopathological analysis of dementia disorders in the elderly. J Neurol Sci 95: 239–258

Katzman R, Aronson M, Fuld P et al (1989) Development of dementing illnesses in an 80-year-old volunteer cohort. Ann Neurol 25: 317–321

Kokmen E, Whisnant JP, O'Fallon WM et al (1996) Dementia after ischemic stroke: a population-based study in Rochester, Minnesota. Neurology 46: 154–159

Ladurner G, Iliff LD, Lechner H (1982) Clinical factors associated with dementia in ischaemic stroke. J Neurol Neurosurg Psychiatry 45: 97–102

Lassen NA (1982) Incomplete cerebral infarction. Focal incomplete ischemic tissue necrosis not leading to emollision. Stroke 13: 522–523

Lindsay J, Hebert R, Rockwood K (1997) The Canadian Study of Health and Aging. Risk factors for vascular dementia. Stroke 28: 526–530

Liu CK, Miller BL, Cummings JL et al (1992) A quantitative MRI study of vascular dementia. Neurology 42: 138–143

Loeb C, Meyer JS (1996) Vascular dementia: still a debatable entity? J Neurol Sci 143: 31–40

Loeb C, Gandolfo C, Bino G (1988) Intellectual impairment and cerebral lesions in multiple cerebral infarcts. A clinical-computed tomography study. Stroke 19: 560–565

Mahler ME, Cummings JL (1991) Behavioral neurology of multiinfarct dementia. Alzheimer Dis Assoc Disord 5: 122–130

Martinez-Lage P, Hachinski VC (1998) Multi-infarct dementia. The vascular causes of cognitive impairment and dementia. In: Barnett HJM, Mohr JP, Stein BM, Yatsu FM (eds) Stroke: pathophysiology, diagnosis, and management, 3rd edn. Churchill Livingstone, New York

Martinez-Lage P, Manubens JM, Martinez-Lage JM et al (1996) Vascular risk factors and cognitive performance in a non-demented elderly population. Neurology 46 Suppl 1: A289

Mast H, Tatemichi TK, Mohr JP (1995) Chronic brain ischemia: the contribution of Ott Binswanger and Alois Alzheimer to the mechanisms of vascular dementia. J Neurol Sci 132: 4–10

Matsushima K, Schmidt-Kastner R, Hakim AM (1996) Genes and cerebral ischemia: therapeutic perspectives. Cerebrovasc Dis 6: 119–127

Mesulam MM (1990) Large-scale neurocognitive networks and distributed processing for attention, language, and memory. Ann Neurol 28: 597–613

Meyer JS, McClintic K, Rogers LG et al (1988) Aetiological considerations and risk factors for multi-infarct dementia. J Neurol Neurosurg Psychiatry 51: 1489–1494

Meyer JS, Kawamura J, Terayama Y (1992) White matter lesions in the elderly. J Neurol Sci 110: 1–7

Moncayo J, Bogousslavsky J (1996) Vascular dementia: persisting controversies and questions. Eur J Neurol 3: 299–308

Munoz DG (1991) The pathological basis of multi-infarct dementia. Alzheimer Dis Assoc Disord 5: 77–90

Nagy Z, Esiri MM, Jobst KA et al (1997) The effects of additional pathology on the cognitive deficit in Alzheimer disease. J Neuropathol Exp Neurol 56: 165–170

Olichney JM, Hansen LA, Hofstetter R et al (1995) Cerebral infarction in Alzheimer's disease is associated with severe amyloid angiopathy and hypertension. Arch Neurol 52: 702–708

Olsson Y, Brun A, Englund E (1996) Fundamental pathological lesions in vascular dementia. Acta Neurol Scand Suppl 168: 31–38

Ott A, Breteler MMB, Van Harskamp F et al (1995) Prevalence of Alzheimer's disease and vascular dementia: association with education. The Rotterdam Study. Br Med J 310: 970–973

Ott A, Stolk RP, Hofman A et al (1996) Association of diabetes mellitus and dementia: the Rotterdam Study. Diabetologia 39: 1392–1397

Pohjasvaara T, Erkinjuntti T, Vataja R, Kaste M (1997) Dementia three months after stroke. Baseline frequency and effect of different definitions of dementia in the Heklsinki Stroke Aging Memory Study (SAM) Cohort. Stroke 28: 785–792

Pullicino P, Benedict RHB, Capruso DK et al. (1996) Neuroimaging criteria for vascular dementia. Arch Neurol 53: 723–728

Rockwood K, Ebly E, Hachinski H et al (1997) Presence and treatment of vascular risk factors in patients with vascular cognitive impairment. Arch Neurol 54: 33–39

Román GC (1987) Senile dementia of the Binswanger type. A vascular form of dementia in the elderly. JAMA 258: 1782–1788

Román GC, Tatemichi TK, Erkinjuntti T et al (1993) Vascular dementia: diagnostic criteria for research studies. Report of the NINDS-AIREN International Workshop. Neurology 43: 250–260

Schmidt R, Fazekas F, Offenbacher H et al (1993) Neuropsychologic correlates of MRI white matter hyperintesities: a study of 150 normal volunteers. Neurology 43: 2490–2494

Skoog I, Nilsson R, Palmertz B et al (1993) A population-based study of dementia in 85-year-olds. N Engl J Med 328: 153–158

Skoog I, Lernfelt B, Landahl S et al (1996) 15-year follow-up study of blood pressure and dementia. Lancet 347: 1141–1145

Snowdon DA, Greiner LH, Mortimer JA et al (1997) Brain infarction and the clinical expression of Alzheimer's disease. JAMA 277: 813–817

Tarvonen-Schröder S, Röytä M, Räihä I et al. (1996) Clinical features of leuko-araiosis. J Neurol Neurosurg Psychiatry 60: 431–436

Tatemichi TK, Desmond DW (1996) Epidemiology of vascular dementia. In: Prohovnik I, Wade J, Knezevic S, Tatemichi T, Erkinjuntti T (eds) Vascular dementia: current concepts. Wiley, Chicvhester, p 40

Tatemichi TK, Foulkes MA, Mohr JP et al (1990) Dementia in stroke survivors in the Stroke Data Bank Cohort. Prevalence, incidence, risk factors, and computed tomographic findings. Stroke 21: 858–866

Tatemichi TK, Desmond DW, Paik M et al (1993) Clinical determinants of dementia related to stroke. Ann Neurol 33: 568–575

Tatemichi TK, Sacktor N, Mayeux R (1994) Dementia associated with cerebrovascular disease, other degenerative diseases, and metabolic disorders. In: Terry RD, Katzman R, Bick KL (eds) Alzheimer's disease. Raven, New York, p 123

Tomlinson BE, Blessed G, Roth M (1970) Observations on the brains of old demented people. J Neurol Sci 11: 205–242

Van Zagten M, Boiten J, Kessels F, Lodder J (1996) Significant progression of white matter lesions and small deep (lacunar) infarcts in patients with stroke. Arch Neurol 53: 650–655

Villardita C (1992) Alzheimer's disease compared with cerebrovascular dementia. Neuropsychological similarities and differences. Acta Neurol Scand 87: 299–308

World Health Organization (1993) The ICD-10 Classification of mental and behavioural disorders. Diagnostic criteria for research. World Health Organization, Geneva

Yamanouchi H (1991) Loss of white matter oligodendrocytes and astrocytes in progressive subcortical vascular encephalopathy of Binswanger type. Acta Neurol Scand 83: 301–305

Yoshitake T, Kiyohara W, Kato I et al (1995) Incidence and risk factors of vascular dementia and Alzheimer's disease in a defined elderly Japanese population: the Hisayama Study. Neurology 45: 1161–1168

CHAPTER

9

H.-J. Gertz

Dementias in Other Brain Diseases

Translator: S. Johnson

1 Introduction

Dementia can essentially result from any illness or insult to the brain which causes a certain critical quantity of brain damage. Thus damage to the brain may involve morphologically visible loss of brain substance or may consist of reduction in specific populations of neurones. It may also involve specific transmitter systems or originate in disturbances of metabolism. The list of diseases which may be associated with dementia is therefore long.

The modern concept of dementia, as formulated in ICD-10 and DSM-IV, is essentially based on the pattern of symptoms found in Alzheimer's disease. In a variety of other dementias, the ICD-10 and DSM-IV criteria are not so straightforwardly applicable. This applies especially to the group of conditions defined as subcortical dementias, where pathology mainly involves the basal ganglia, mid-brain and brainstem and there is a prominent disturbance of the retrieval of stored memory contents. For a range of conditions whose characteristic clinical picture is traditionally designated by the term dementia, no appropriate methods of examination with standardised psychiatric instruments are available.

A further characteristic of both ICD-10 and DSM-IV is that the critical threshold for a diagnosis of dementia is strongly tied to the area of activities of daily living. Most of the dementia syndromes discussed in this chapter are associated with a cluster of serious general medical and neurological symptoms, which are generally pointers to the diagnosis of the underlying illness. These medical and neurological illnesses are often in themselves associated with substantial impairments in daily living activities. Thus in this situation the usefulness of making classifications on the basis of limitations in social and occupational role functioning, as ICD-10 requires, is questionable.

The characteristics of the dementia syndromes described in association with different illnesses tend to be rather non-specific, and it is therefore rarely possible to make a convincing judgement regarding differential diagnosis on the basis of psychopathology. The observation of the early twentieth century German psychiatrist Karl Bonhoeffer about the great diversity of underlying illnesses and the contrasting similarity between psychopathological pictures is also applicable to the dementias.

Classification in terms of aetiology or pathogenesis is more likely to be possible in the earlier phases of illness. However, differentiation between patterns of psychopathology in the forms of dementia described here is important, as psychopathology provides clues to the localisation of pathological processes, giving significant insight into the mechanisms of pathogenesis in the dementing syndromes.

2 Degenerative Dementias

Degeneration is a summary term for a large number of aetiologically and pathogenetically unclear processes, in which morphological appearances suggest neither an inflammatory nor a vascular basis. Degeneration is often used as a synonym for the process of atrophy. Historically, the expression "degeneration" was closely linked to heritability. That the concept of degeneration is used to refer to processes which are actually very heterogeneous is indicated by the fact that Creutzfeldt-Jakob disease and the leucodystrophies, previously believed to be classical examples of the neurodegeneration process localised in the grey or white matter, now have to be classified in the group of infectious and metabolic illnesses.

In addition to the examples discussed more fully below, there are a variety of very rare degenerative dementias. Examples include primary progressive aphasia, regarded by some authors as closely related to Alzheimer's disease or Pick's disease, and progressive subcortical gliosis. The clinical picture in progressive subcortical gliosis resembles that of the fronto-temporal dementias. Its name describes the main histological process which characterises it, slowly advancing gliosis involving not only the cortex and basal ganglia but also the brainstem (Civil et al. 1993).

Other forms of dementia have thus far been defined exclusively in neuropathological terms, such as dementia lacking distinctive histological features (Knopman et al. 1990) or dementia with cortical and subcortical argyrophilic grains (Braak and Braak 1989), which may be a variant of Alzheimer's disease. A variety of so-called systemic degenerations, such as striatonigral degeneration, olivopontine cerebellar atrophy and Friedrich's ataxia, are sometimes associated with dementing syndromes, but these have so far been characterised only very imprecisely.

2.1 Fronto-temporal Dementias

The term fronto-temporal dementia is used here to refer both to Pick's disease in the narrower sense and to frontal lobe degeneration of non-Alzheimer type (Brun et al. 1994). The psychopathology characterising motor neurone diseases with dementia is similar.

Whereas motor neurone disease with dementia can be distinguished on the basis of the associated neurological symptoms, Pick's disease and frontal lobe degeneration of non-Alzheimer type are differentiated on histopathological grounds. In ICD-10, Pick's disease in the narrower sense and frontal lobe degeneration of non-Alzheimer's type are subsumed under the expression "dementia in Pick's disease", which is thus defined very broadly. Historically, the ICD-10 definition is close to the intentions of Pick, whose initial description was simply of dementias with a tendency towards focal atrophy, a categorisation which he considered nosologically non-specific. Neary et al. (1998) have suggested that fronto-temporal dementia, progressive aphasia and semantic dementia should all be referred to as subtypes within an overall "fronto-temporal lobar degeneration" category.

Because of the great lack of international consensus on nosology and definition of syndromes, it is very difficult to make generalisations from prevalence figures. Neary et al. (1988) suggest that fronto-temporal dementias make up around 20% of presenile dementias. Gustafson and Brun (1997) report an average age of onset for Pick's disease and frontal lobe degeneration of non-Alzheimer type of 51–54. The duration of illness is usually around 8 years, but courses lasting up to 17 years have been reported.

The early stage of the illness is characterised by a variety of changes in personality. These include disturbances of impulse control, such as highly irritable or aggressive reactions, psychomotor agitation and restlessness. Patients lose their spontaneity, neglect themselves and become impassive, insensitive and apathetic. A coarsening of social behaviour develops, in which usual social rules and norms are heeded less and less. Transient forms of Klüver-Bucy Syndrome may be observed (Cummings and Duchen 1981). Intellectual functions are often preserved for a long time, especially recent memory abilities and orientation in space. Verbal utterances diminish progressively, become monotonous and eventually consist largely of stereotyped, repetitive utterances. Finally, a general poverty of speech develops, which may be associated with total lack of expression and finally progresses to a mute state. Neary et al. (1998) have undertaken the task of trying to bring together the various psychopathological symptoms described in operational diagnostic criteria.

Currently, there seems little point in trying to distinguish between the psychopathologies in Pick's disease and in frontal lobe degeneration of non-Alzheimer type, since this classification has to be made on the basis of histopathology (Gustafson and Brun 1997).

Computed tomography (CT) and magnetic resonance imaging (MRI) show variable frontal or fronto-temporal atrophy, which in some cases may be extremely severe. Single photon emission CT (SPECT) shows diminished frontal blood flow. An unremarkable electroencephalogram (EEG) is a further characteristic finding in fronto-temporal dementia and can thus contribute to making a differential diagnosis from Alzheimer's disease. Even when slow waves appear, the α-rhythm is generally better preserved than in Alzheimer's disease (Neary et al. 1988).

The characteristic histopathological findings of frontal lobe degeneration of non-Alzheimer type and of Pick's disease are loss of cortical cells, gliosis and spongiform changes in the frontal and temporal cortices. The gliosis may also extend into the subcortical white matter. Typically, structures such as the amygdala, the hippocampus and the basal nucleus of Meynert, which are highly abnormal in Alzheimer's disease, are affected only to quite a limited extent.

The round argentophilic intraneuronal inclusion bodies called Pick bodies, together with ballooned Pick's cells, are indicators of Pick's disease in the narrower sense. Neither of these changes features in definitions of frontal lobe degeneration of non-Alzheimer type. At an ultrastructural level, Pick bodies consist of straight filaments, paired helical filaments and also endoplasmic reticulum. In terms of immunohistochemistry, they are characterised by antibodies against ubiquitin and against the tau protein. Thus they have antigenic characteristics which are similar to those of the neurofibrils of Alzheimer's disease. Mutations of the tau gene, coded on chromosome 17, have recently been found in numerous families with fronto-temporal dementia (Gertz et al. 1999).

In contrast to Alzheimer's disease, cortical choline acetyltransferase is not reduced in Pick's disease or frontal lobe degeneration of non-Alzheimer type, while reports regarding muscarinic receptors in the cortex are inconsistent (Hansen et al. 1988).

The aetiology of the illness is unknown. A number of cases of autosomal dominant inheritance have been described. The similarity of antigenic characteristics between Pick bodies and Alzheimer fibrils and occasional descriptions of co-existence of the two types of change in one and the same cell would be compatible with the conclusion that there may be a common metabolic defect or related responses to differing aetiological stimuli.

2.2 Progressive Supranuclear Palsy

Progressive supranuclear palsy is characterised by an initial paralysis of vertical gaze, pseudobulbar paralysis, dysarthria and rigidity (Steele et al. 1964). The illness is rare: only one to four cases are observed per

100,000 people (Globe et al. 1988). Dementia occurs in 20%–60% of patients.

Progressive supranuclear palsy illustrates the principle that standard operationalised psychiatric criteria are not generally applied to subcortical dementias. Most authors refer to the definition proposed by Albert et al. (1974). Typically, forgetfulness is limited, but slowing of thinking and changes in personality and mood are observed. In some cases, extreme slowing of thinking may be mistaken for memory disturbance. Milberg and Albert (1989) found no differences in memory performance in patients with progressive supranuclear palsy compared with control subjects. In the majority of patients, disturbed affect is manifested as indifference, apathy or depressiveness. Occasionally, there is increased irritability and/or euphoric mood. In a third of patients, there is intermittent, unprovoked forced laughter or crying. Podoll et al. (1991) related deficits in speech performance, as well as visual dyslexia, constructional dysgraphia and a raised rate of failure in object naming, to a primary disturbance in visual perception. The consensus is that primary aphasic symptoms are not normally present.

CT scans and MRI show mid-brain atrophy early in the course of the illness. As the illness progresses, brainstem atrophy also becomes apparent. Examinations with positron emission tomography (PET) and SPECT have shown some indications of reduced glucose metabolism in the subfrontal area and also the brainstem and basal ganglia (Karbe et al. 1992). However, these functional changes are not accompanied by neuropathologically demonstrable degeneration of the frontal lobes. They are regarded much more as a manifestation of disordered transmission between the mid-brain and the frontal cortex.

Neuropathologically, the illness is characterised by nerve cell loss, gliosis and globular Alzheimer fibrils, especially in the subthalamic nucleus, globus pallidus, substantia nigra and superior colliculus. Neurofibrils are also found in the neocortex in occasional cases. In contrast to the neurofibrils of Alzheimer's disease, the globular neurofibrils in progressive supranuclear palsy are made up of 15 nm thick, straight filaments, without the coiled filaments characteristic of Alzheimer's disease (Rewcastle 1991).

These typical neuropathological findings have also been described in some patients without gaze paralysis. It is interesting that nerve cell loss has been observed in the basal nucleus of Meynert in progressive supranuclear palsy, as memory impairments in Alzheimer's are attributed to cell loss in this region, yet such impairments are absent or very limited in progressive supranuclear palsy. It may be that this region becomes involved only in very advanced progressive supranuclear palsy.

Nothing is known about the aetiology of this rare disease. It occurs sporadically.

The major neurochemical changes in progressive supranuclear palsy are a fall in dopamine levels, reduced numbers of D_2 receptors in the striatum and a drop in choline acetyltransferase levels.

Dopamine agonists in low doses are thought to treat the motor disturbances more effectively than in idiopathic Parkinson's syndrome. Physostigmine is thought to improve visual attention and memory impairments. Gheka et al. (1991) reported therapeutic success with Idazoxan, a selective alpha-2 blocker, which intensifies noradrenergic transmission.

2.3 Parkinson's Disease with Dementia and Dementia with Lewy Bodies

In terms of clinical neurology, Parkinson's disease is characterised by rigidity, tremor and hypokinesis. Diagnostic indicators at a neuropathological level are Lewy bodies in the substantia nigra and some other brainstem nuclei. A dementia may develop in the course of the illness. This is termed Parkinson's disease with dementia. However, where the dementia develops concurrently with or only slightly after the specific movement disturbance, this is referred to as dementia with Lewy bodies. In such cases, Lewy bodies are also found in the cerebral cortex. The significance attributed to dementia with Lewy bodies has recently increased. The question of whether discrete illnesses can really be clearly differentiated here is still open, as is the question of how dementia with Lewy bodies relates to Alzheimer's dementia.

2.3.1 Parkinson's Disease with Dementia

The idiopathic form of Parkinson's disease is one of the commonest neurological illnesses. For every 100,000 people in the general population, there are around 80–100 cases of this illness (Mayeux et al. 1992). Risk increases with increasing age.

The risk of developing dementia is substantially increased in individuals with Parkinson's disease. Rajput et al. (1987) found a cumulative probability of dementia of 21% among subjects with Parkinson's disease, compared with 5.7% in control subjects of the same age. Mayeux et al. (1988) estimated a cumulative incidence of dementia of 65% by the 85th year of life in people with Parkinson's disease.

Risk factors for developing dementia in an established Parkinson's syndrome include advanced age and also the first occurrence of motor disturbances after the age of 70. Patients with a Parkinson's syndrome in

which tremor is particularly prominent are less likely to develop dementia than those in whom akinesis and marked gait disturbances dominate the picture. Dementia is more frequent where a depressive syndrome is already present (Stern et al. 1993). There seems to be no connection between risk of dementia and neurological impairments which are more severe unilaterally.

Bradyphrenia is a characteristic early intellectual disturbance in individuals with Parkinson's disease. Bradyphrenia is believed to indicate a slowing of the thinking process, which constitutes an intellectual equivalent of hypokinesis. A clear relationship between the occurrence of bradyphrenia and hypokinesis has not, however, been demonstrated, and it remains an open question as to how far the presence of bradyphrenia is a predictor for the later development of a dementia. The fully developed dementia syndrome should be regarded as rather uncharacteristic in Parkinson's disease (Pillon et al. 1991). People with Parkinson's rarely develop disturbances of the cortical association areas, such as aphasias or agnosias, and in the early stages tend to fit the clinical picture of frontal or subcortical types of dementia. However, cortical dementia syndromes do occur.

MRI findings in Parkinson's disease with dementia cannot be differentiated from those in whom the idiopathic Parkinson's syndrome is confined to motor manifestations.

The most significant neuropathological findings in Parkinson's disease are cell loss and gliosis in the pars compacta of the substantia nigra and in some other brainstem nuclei such as the locus coeruleus and the dorsal motor nucleus of the vagus, as well as the basal nucleus of Meynert in the basal forebrain. Lewy bodies are also found in these nuclei, confirming the diagnosis of idiopathic Parkinson's syndrome.

Lewy bodies are intraneuronal eosinophilic inclusion bodies. They occur predominantly in cell bodies but also in neuritic processes. The Lewy bodies found in the subcortical areas are denser centrally, with a lighter periphery. Immunocytochemically, Lewy bodies can be demonstrated with antibodies to neurofilaments and to ubiquitin. However, ubiquitin positiveness is not specific, but can also be demonstrated in Pick's bodies and in Alzheimer neurofibrils.

There has been considerable controversy as to whether the aetiological basis of the dementia syndrome in idiopathic Parkinson's disease is the same as that for the motor disturbances, whether it represents a variant of dementia with Lewy bodies or whether it should be related to concurrent Alzheimer's disease (Paulus and Jellinger 1991). The heterogeneity of clinical presentations of the dementia syndrome in Parkinson's disease permits a variety of aetiological models to be used by way of explanation.

Treatment with medication for dementia in idiopathic Parkinson's syndrome is not very satisfactory. The co-occurrence of cholinergic deficit and relative cholinergic overactivity makes administration of either cholinomimetic or anticholinergic substances rather problematic. Assuming that a common pathological process underlies both cognitive and motor disturbances, use of deprenyl may be a promising avenue (Korczyn 1995).

2.3.2 Dementia with Lewy Bodies

The first reports of a dementing syndrome with cortical Lewy bodies came from Japan. In a series of recent clinical autopsy investigations, the prevalence of Lewy body dementia has been given as 15%–25% of all dementing syndromes in old age. This figure suggests that dementia with Lewy bodies is the second commonest cause of dementia after Alzheimer's disease.

People with dementia with Lewy bodies have difficulties in retrieving memory contents. As well as reduced verbal fluency, disturbances in executive functions and problem-solving are also considered characteristic, as tested in the Wisconsin Card-Sorting Test or in the Trail-Making Test. Thus the clinical picture fits with a subcortical form of dementia. However, as dementia advances, these characteristics become less prominent.

In addition to cognitive impairments similar to those of progressive supranuclear palsy, fluctuating course, hallucinations and the motor symptoms of Parkinson's syndrome are characteristic of the illness. Both cognitive abilities and attention and conscious level tend to fluctuate. Although hallucinations in a variety of modalities occur, optical hallucinations are especially common. They are not generally present for extended periods, but often occur when alertness is reduced or also in the presence of impaired vision. A delusional elaboration of these hallucinations is rare. Accompanying motor symptoms are rigidity and hypokinesis, while tremor is less often seen. Other Parkinson's symptoms such as mask-like expression, hypophonic speech and gait disturbances also occur. Typically, motor and psychopathological disturbances begin around the same time.

Extrapyramidal motor symptoms may, however, precede the psychopathological disturbance or else follow it only after a gap. Recently, operationalised diagnostic criteria have been proposed for dementia with Lewy bodies (McKeith et al. 1996).

In contrast to subcortical Lewy bodies, cortical Lewy bodies stained with haemotoxolyn-eosin dye appear homogenous and evenly pink, and a very careful search is therefore required to find them.

It can be substantially easier to detect them by using ubiquitin antibodies. Cortical Lewy bodies primarily occur in the cingulate gyrus and the entorhinal cortex. Within the neocortex, they are found more often in the temporal than in the frontal or parietal lobes (Kosaka 1990). Ubiquitin-positive degenerated neurites also occur in area CA2/3 of the hippocampus, in the basal nucleus of Meinert and in some brainstem nuclei.

In a small group of "pure" cases of dementia Lewy bodies, there is quite a high density of Lewy bodies in the favoured neocortical sites, without additional Alzheimer's pathology being demonstrable (Kosaka 1990). However, in the majority of patients, the presence of cortical Lewy bodies is accompanied by the appearance of senile plaques and Alzheimer neurofibrils. This "common" form of dementia with Lewy bodies resembles Alzheimer's disease in that it is associated with a raised apolipoprotein E-epsilon 4 allele frequency, which is found neither in "pure" dementia with Lewy bodies nor in idiopathic Parkinson's syndrome.

Establishing the boundary between the "common" form of dementia with Lewy bodies and Alzheimer's disease raises considerable difficulties. However, in group comparisons, senile plaques and Alzheimer's neurofibrils are numerous enough to differentiate individuals with dementia with Lewy bodies from non-demented controls. Thus the Alzheimer-type pathology in dementia with Lewy bodies cannot be understood as a coinciding "normal" ageing process.

On the other hand, the Alzheimer-type pathology in dementia with Lewy bodies is less extensive than in Alzheimer's disease. Patients with dementia with Lewy bodies have a lower density of Alzheimer neurofibrils in the entorhinal cortex, the hippocampus and the neocortex than patients with Alzheimer's (Hansen 1994). The quantity of senile plaques is similar in dementia with Lewy bodies to that in Alzheimer's disease, but with amyloid plaques making up a higher percentage of the diffuse plaques in dementia with Lewy bodies (Dickson et al. 1987).

Interestingly, the activity of neocortical enzymes of cholinergic transmission, such as choline acetyltransferase and acetylcholinesterase, is more clearly reduced in dementia with Lewy bodies than in Alzheimer's disease (Perry et al. 1994).

CT and MRI show a predominantly frontal atrophy in some cases (Förstl et al. 1993). SPECT findings show a similar pattern to Alzheimer's disease.

No specific treatment is known for the cognitive impairments of dementia with Lewy bodies, although attempts at treatment with acetylcholinesterase inhibitors seem worth pursuing. An important consideration is that typical neuroleptics, which may be given in response to hallucinations occurring in dementia with Lewy bodies, exacerbate movement disturbances. They are therefore contraindicated in dementia with Lewy bodies.

2.4 Huntington's Chorea

The neurological clinical picture in Huntington's chorea is characterised by a hypotonic hyperkinetic syndrome. According to Harper (1992), the prevalence is between 4 and 8 cases per 100,000 people. This autosomal dominant illness is believed to occur in all populations, but seems to occur relatively infrequently in Finland, Japan and Central Africa. A raised rate has been reported from Venezuela and Zimbabwe. In most cases, the onset of illness is between the ages of 35 and 50 years. Manifestations before the age of 20 have also been reported, for which the prognosis is especially poor.

Cognitive impairment is frequently found in Huntington's chorea. There are grounds for a diagnosis of a dementing syndrome in two thirds of sufferers (Pillon et al. 1991). Various authors have classified dementia in Huntington's chorea as a subcortical type. This seems justified in view of the absence of aphasia and agnosia even at advanced stages. However, there are a number of differences from dementia in progressive supranuclear palsy, an important example being the absence of slowing down. Verbal recall and recognition are impaired to equal extents in people with Huntington's chorea and those with Alzheimer's disease (Brandt et al. 1992). Hodges et al. (1990) found that, compared with people with Alzheimer's disease, people with Huntington's chorea had greater difficulties in verbal fluency tasks, word recognition and copying of geometric figures. Word-finding difficulties are seen by these authors as the result of perceptual disturbances and not as an impairment in semantic speech organisation.

With CT and MRI, a flattening and shrinking of the caudate nucleus and the putamen can be shown, in the area where these form the ventrolateral wall of the lateral ventricles. There is generally also external brain atrophy, which in advanced cases can be substantial. On PET scans, diminished glucose metabolism has been shown primarily in the striatum, and to a lesser extent it is also seen in the frontoparietal and temperooccipital regions. A greater severity of dementia is associated with diminishing rates of cortical metabolism (Kuwert al. 1990). Reduced glucose consumption in the striatum can also be demonstrated in asymptomatic patients. Leenders et al. (1986) used the selective D_2-ligand *N*-methyl-spiperon to show a reduction in D_2 receptors in the striatum. On the other hand, uptake of ^{18}F-labelled levodopa did not

differentiate individuals with Huntington's chorea from normal controls. This suggests intact presynaptic and impaired postsynaptic functions.

In 30%–80% of patients, the illness process leads to low voltages on EEG, with amplitudes below 10 μV. This reduction in amplitude is thought to be correlated with changes in the caudate nucleus, but it only becomes evident with the fully established clinical syndrome (Fenton 1994).

A significant histopathological finding is the loss in the striatum of small neurones containing γ-aminobutyric acid (GABA), encephalin and substance P. Neurones containing somatostatin and neuropeptide Y are relatively preserved, as are large cholinergic neurones. Even though changes in the corpus striatum dominate the pathological picture, changes can also be found in the cerebral cortex. The cortex shows a volume reduction of 21%–29%, the white matter 29%–34%, and the thalamus 28% (De la Monte et al. 1988). Nerve cell loss can be observed especially in the upper frontal cortex and the cingulate gyrus. The cell populations of the basal nucleus of Meynert, the locus coeruleus and the dorsal raphe nuclei do not appear to be involved.

Huntington's chorea is an autosomal dominant illness with full penetrance. It is caused by a mutation on the short arm of chromosome 4. The gene codes for a protein referred to as Huntingtin, whose function is unknown. The mutation consists of the repetition more than 38 times of the trinucleotide cytosine-adenine-guanine (CAG) (Huntington's Disease Collaborative Research Group 1993). This trinucleotide codes for the amino acid glutamine. In the German population, this trinucleotide is normally repeated 11 to 32 times. In juvenile cases of Huntington's chorea, up to 180 repetitions may be found. The abnormal expansion of this trinucleotide repeat and thus of glutamine is reflected in a mutation in Huntingtin, as a result of which it appears to take on amyloid and therefore cytotoxic properties (Scherzinger et al. 1997).

The direct evidence now available concerning the nature of the mutation in Huntington's chorea allows DNA diagnosis in clinically asymptomatic people. This is feasible as early as the first trimester of pregnancy with the aid of chorionic villous sampling. However, it is essential to consider the serious ethical and psychological problems associated with presymptomatic diagnosis in a late-manifesting and incurable illness.

3 Infectious Diseases

Other illnesses which may underlie dementing syndromes include meningitis and encephalitis of viral, bacterial or protozoal origin. Important examples of viral encephalitides are herpes simplex encephalitis, the subacute sclerosing panencephalitis caused by the measles virus and progressive multifocal leucencephalopathy caused by papovavirus (McArthur et al. 1993). Neurosyphilis and tuberculous meningitis are among the most significant of the bacterial illnesses which can cause dementing syndromes. Among protozoal infections, toxoplasmosis has taken on a new significance among adults as a result of the human immunodeficiency virus (HIV; Ashe et al. 1993).

3.1 Acquired Immunodeficiency Syndrome

Infection with the HIV-1 virus generally leads to development of acquired immunodeficiency syndrome (AIDS) after a 10-year latency period. It is thought that 40%–70% of individuals with AIDS develop neuropsychiatric abnormalities. In up to 20% of patients, neuropsychiatric disturbances appear to be the first manifestation of the illness. Navia et al. (1986) introduced the term "AIDS dementia complex", which by the authors' definition encompasses the triad of "cognitive, motor and behavioural disturbance".

On the basis of this concept of AIDS dementia complex, very high prevalence rates have been reported for the dementia. However, a dementia as defined by ICD-10 is observed in only 4% of HIV-positive patients (Naber 1993).

The most frequent early neuropsychological disturbances are impairments of concentration and memory, slowing, apathy and reduced spontaneity, generally also accompanied by motor slowing. Discrete neuropsychological abnormalities in those with HIV infection should not be uncritically accepted as indicators of central nervous system (CNS) manifestations of the illness. In at least part of this population, neuropsychological abnormalities are present as a result of a pre-existing or concurrent neuropsychiatric illness (Pakesch et al. 1992). Paranoid psychotic syndromes with hallucinations and severe depressive episodes are rare as manifestations of HIV infection. With such syndromes, accurate differential diagnosis again requires pre-existing or concurrent separate illnesses to be distinguished from the manifestations of HIV infection.

CT and MRI may show discrete areas of atrophy. MRI, especially in T_2-weighted views, shows more clearly than CT that there are focal or confluent areas of high signal intensity in the white matter (Keiburtz et al. 1990). In asymptomatic patients, there tend at most to be smaller focal lesions in the white matter (Post et al. 1991).

In autopsy series, 75% of people dying of AIDS have neuropathological changes and 30% show multiple

CNS changes (Janssen et al. 1992). Where macroscopic appearances indicate diffuse atrophy, on microscopic examination circumscribed zones of demyelination are found, especially in the subcortical white matter. Axonal degeneration and the development of microglial nodules and multinuclear giant cells are also observed in these regions. By comparison, the cerebral cortex is affected relatively little (Budka 1991). Numerous opportunistic infections and tumours can complicate the picture.

The mechanisms behind the CNS abnormalities which HIV-1 can cause in the absence of superimposed infection are unclear. It is certain that the HIV-1 virus penetrates the blood–brain barrier at an early stage in the illness and replicates itself in the brain tissue, using mono- and multinuclear macrophages as host cells. Whether the virus causes the CNS pathology by direct or indirect means has not so far been established. The virus capsule protein gt 120 and quinolic acid have been discussed as possible candidates for involvement in this process (Heyes 1995).

It is hoped that recently developed combination therapies may reduce the incidence of neuropsychiatric symptoms (Vitkovic and Tardieu 1998). However, the CNS does constitute a virus reservoir, which for pharmacokinetic reasons is only accessible to pharmacological interventions to a very limited extent (Schrager and D'Souza 1998). Zidovudine is believed to improve the neuropsychiatric symptoms in AIDS at least transiently (Tozzi et al. 1993). This effect may be confined to cases which are already advanced. There have been individual case reports of a positive effect on bradyphrenia by psychostimulants (Fernandez and Levy 1991).

3.2 Creutzfeldt-Jakob Disease

Creutzfeldt-Jakob disease is one of the transmissible spongiform encephalopathies. Kuru, Gerstmann-Sträussler-Scheinker syndrome and the animal diseases scrapie and bovine spongiform encephalopathy (BSE) also belong to this group. The illness occurs world-wide with a prevalence of 1–2 cases per 1 million people. The highest rate of 42 per million is found among Jews of Libyan origin (Zilber et al. 1991).

In 39% of sufferers, the initial symptoms are non-specific prodromal ones such as tiredness, apathy, disturbances in concentration, attention and memory, increased irritability, depressed mood, dizziness or headaches. In 37% of patients, the first symptoms are neurological. The dementing syndrome which soon appears is distinguished not so much by a specific psychopathological form as by its dramatic course. It is accompanied by primitive reflexes, spasticity and pyramidal tract signs and extrapyramidal motor disturbances, which can take the form of hypokinesis and rigidity or of dystonic or choreiform movements. In around half of the patients, a cerebellar ataxia occurs, preceding the appearance of pyramidal and extrapyramidal disturbances. Impairments of vision in the form of partial or total cortical blindness occur in around a quarter to a third of those with the illness. The frequency of myoclonic jerks, which are held to be especially characteristic, has been quoted as 56%–86%. The end-stage is often akinetic mutism. As a rule, the illness leads to death after 6–20 months, although courses of illness lasting many years have also been reported (Brown et al. 1984). While for the sporadic form the average age of onset is around 64 years, new-variant Creutzfeldt-Jakob disease, which was first described in Britain, has a much earlier average onset around the age of 26. The symptom pattern of the new variant has been characterised as beginning with psychiatric symptoms, such as anxiety, depression and behavioural abnormalities.

The only signs of Creutzfeldt-Jakob disease on CT scanning are some indications of internal and external atrophy. However, MRI, particularly using T_2-weighted views, shows changes in the cortex and basal ganglia with greater sensitivity. A series of examinations have shown localisation of raised signal intensity in T_2-weighted views to correspond to the pattern of neuropathological damage (Gertz et al. 1988a; Röther et al. 1992).

Initially, EEG shows non-specific changes, such as a gradual slowing of the background rhythm. Periodic eruptions of steep bi- and triphasic waves have been regarded as characteristic of Creutzfeldt-Jakob disease (Brown et al. 1979). An almost unvarying interval between these eruptions of around 1 s has often been described. Reports of how frequently such periodic complexes are observed in proven Creutzfeldt-Jakob disease lie between 50% and 90%. The need for serial measurements has repeatedly and rightly been emphasised, so that these typical changes, often crucial for making a clinical diagnosis, may be captured.

Histopathologically, the illness is characterised by the pathognomonic triad of nerve cell destruction, spongiform vacuolisation and proliferation of astroglia. Various subtypes have been identified on the basis of the distribution of morphological changes, including simple polydystrophic, thalamic, cerebellar and panencephalopathic forms (Mitzutani and Shiraki 1985). The last of these is believed to be very rare in Western Europe, but in contrast appears to occur relatively frequently in Japan (Gertz et al. 1988b).

In diagnostically doubtful cases, the immunohistochemical demonstration of proteinase-resistant prion protein in the grey matter of the brain may help to provide clarification. The sensitivity of the corresponding antibodies appears high. Use of brain biopsy

is recommended only where treatable differential diagnostic alternatives are seriously being considered (Budka et al. 1995). New-variant Creutzfeldt-Jakob disease is characterised and distinguished from the sporadic form by the appearance of amyloid plaques in the cerebral cortex and cerebellum.

Most cases of Creutzfeldt-Jakob disease are sporadic. Cases of spongiform encephalopathies have, however, occurred after administration of contaminated growth hormone, use of contaminated deep EEG electrodes, corneal transplantation and also general surgical operations. The illness can be transmitted to primates. In 5%–10% of patients, there appears to be autosomal dominant inheritance with variable penetrance. New-variant Creutzfeldt-Jakob disease, which has so far predominantly been observed in Great Britain, has been connected with BSE. Evidence for this has been produced on neuropathological, molecular biology and epidemiological levels. A direct route of infection from cattle to humans has not been proved with certainty, but is probable (Almond 1998).

The illness is thought to be transmitted and triggered through so-called prions. Prions are thought to be cellular proteins which have pathogenic variants that possibly induce their own formation in a previously unknown way (Prusiner et al. 1989). Whether there are virus-like polynucleotides within the prion has not been conclusively established. It seems certain that the formation of pathological, i.e. amyloid proteinase-resistant prion protein requires the presence of physiological proteinase-resistant prion protein. Mice in which gene technology has been used to remove the proteinase-resistant prion protein gene do not become ill after inoculation with various infectious scrapie isolates. These findings also prove that the spongiform encephalopathies do not originate from loss of a normal isoform of proteinase-resistant prion protein.

The human proteinase-resistant prion protein gene is situated on the short arm of chromosome 20. Point mutations occur in this region as well as repetition four to nine times of an octapeptide on codon 53 (DeArmond and Prusiner 1995). These mutations can be demonstrated by molecular genetic analysis, which can be used as a diagnostic procedure. The infectious agent in Creutzfeldt-Jakob disease is resistant to all conventional sterilisation procedures; 1.2% solution of chlorhexidine or 10% solution of stericol are suitable for disinfecting. No treatment is known.

4 Reversible Dementias

Dementias caused by metabolic disturbances or toxic conditions may be reversible through successful treatment of the underlying illness. Medications prescribed by doctors are especially prominent among causes of toxic conditions, particularly diuretics, cardiac drugs and insulin (Larson et al. 1986). Where such toxic conditions have led to dementia, it can be assumed in older patients that some brain damage has occurred.

Endocrinological disorders such as hypothyroidism or hyperparathyroidism rarely lead to a full-blown dementia syndrome. Vitamin B_1 or B_2 deficiency can be associated with a dementing syndrome. However, replacement often fails to reverse the psychopathological phenomena.

Hakim and Adams (1965) described the clinical picture of normal-pressure hydrocephalus, characterised by the triad of dementia, gait disturbance and urinary incontinence. The symptoms improve after insertion of a shunt. The frequency of normal-pressure hydrocephalus in dementias in adult life has been given as between 1.6% and 7.0%. Normal-pressure hydrocephalus is thought to account for a quarter of all reversible cases of dementia. The onset of illness generally occurs between the ages of 50 and 70.

The gait disturbance often precedes the incontinence and dementia. It is generally described as a gait apraxia or frontal ataxia. The patient walks slowly with small steps and a rather groping broad base. Electromyographic investigations show simultaneous contractions of agonists and antagonists and a permanent activation of the gravity-opposing muscles during walking (Knutsson and Lying-Tunnell 1985).

In psychopathological terms, the cognitive deficit tends to be of a subcortical type, with disturbance of attention, drive and frontal executive functions. Generally, speech is also slowed and has little spontaneity. In neuropsychological investigations, attention and short-term memory are impaired, as are writing and drawing. Functional disturbances associated with the cortical association areas, such as aphasia and agnosia, do not occur (Filley et al. 1989).

Typical signs on CT and MRI are markedly enlarged inner ventricles, especially in the anterior horns of the lateral ventricles, with little or no brain atrophy. More recently, an MRI finding (albeit a non-specific one) of periventricular signal hyperintensity has been reported (Bradley et al. 1991). On SPECT and PET scans, diminished frontal global metabolism has been reported.

A substantial diagnostic work-up is a prerequisite for a successful shunt insertion operation. The method of choice is intraventricular brain pressure measurement, which in typical cases shows intermittent plateaux or peaks in intracerebral pressure, termed Lundberg-B waves. Good results can be expected from shunt insertion in patients in which, during 48-hour recording, intraventricular pressure has been raised in the Lundberg-B wave form at least twice.

A simpler and less invasive investigation procedure is the removal of 20–50 ml of cerebrospinal fluid by lumbar puncture. Around 30–60 min after this procedure, the gait abnormality and psychometric test results should improve, an improvement which may be sustained for several months.

Around 70% of patients with normal-pressure hydrocephalus have a history of subarachnoid haemorrhage, brain injury or meningitis, which probably causes this clinical picture through the mechanism of impaired absorption of cerebrospinal fluid.

Vanneste et al. (1992) reported that, even after careful diagnosis, a convincing improvement after shunt insertion can be demonstrated in only 21% of all patients.

5 Others

Head injuries are the most frequent cause of neurological disturbances in young adults. Duration of unconsciousness and of post-traumatic illness appear to be quite robust predictors of the severity of later residual symptoms (Mendez 1993). A further important differential diagnosis for dementing syndrome in younger and middle-aged adults is multiple sclerosis. A total of 2%–7% of patients with multiple sclerosis have neuropsychological deficits sufficiently severe to be defined as dementia. In some cases, the dementing syndrome may initially be most prominent and may only later be accompanied by neurological deficits (Rao et al. 1989).

6 References

Albert M, Feldmann RG, Willis AL (1974) The "subcortical dementia" of progressive supranuclear palsy. J Neurol Neurosurg Psychiatry 37: 121–130

Almond JW (1998) Bovine spongiform encephalopathy and new variant Creutzfeldt-Jakob disease. Br Med Bull 54: 749–759

Ashe J, Rosen SA, McArthur JC, Davis LE (1993) Bacterial, fungal and parasitic causes of dementia. In: Whitehouse PJ (ed) Dementia. Davis, Philadelphia, pp 276–305

Braak H, Braak E (1989) Cortical and subcortical argyrophilic grains characterise a disease associated with adult onset dementia. Neuropathol Appl Neurobiol 15: 13–26

Bradley WG, Whittemore AR, Kortman KE et al (1991) Marked cerebrospinal fluid void: indicator of successful shunt in patients with suspected normal-pressure hydrocephalus. Radiology 178: 459–466

Brandt J, Corwin J, Krafft L (1992) Is verbal recognition memory really different in Huntington's disease and Alzheimer's disease. J Clin Exp Neuropsychol 14: 773–784

Brown P, Cathala F, Sadowsky D, Gajdusek DC (1979) Creutzfeldt-Jakob disease in France. II. Clinical characteristics of 124 consecutive verified cases during the decade. Ann Neurol 6: 430–446

Brown P, Rodgers-Johnson P, Cathala F, Gibbs CJ, Gajdusek DC (1984) Creutzfeldt-Jakob disease of long duration: clinicopathological characteristics, transmissibility and differential diagnosis. Ann Neurol 16: 295–304

Brun A (1987) Frontal lobe degeneration of non-Alzheimer type. Neuropathology Arch Gerontol Geriatr 6: 193–208

*Brun A, Englund B, Gustafson L et al (1994) Clinical and neuropathological criteria for frontotemporal dementia. J Neurol Neurosurg Psychiatry 57: 416–418

Budka H (1991) Neuropathology of human immunodeficiency virus. Brain Pathol 1: 163–175

*Budka H, Aguzzi A, Brown P et al (1995) Neuropathological diagnostic criteria for Creutzfeldt-Jakob-disease (CJD) and other human spongiform encephalopathies (prion disease). Brain Pathol 5: 459–466

Civil RH, Whitehouse PJ, Lanska DJ, Mayeux R (1993) Degenerative dementias. In: Whitehouse PJ (ed) Dementia. Davis, Philadelphia, pp 167–214

Cummings JL, Duchen LW (1981) The Klüver-Bucy syndrome in Pick's disease. Neurology 31: 1415–1422

DeArmond SJ, Prusiner SB (1995) Prion disease. In: Bloom F, Kupfer DJ (eds) Psychopharmacology: the fourth generation of progress. Raven, New York, pp 1521–1530

De la Monte SM, Vonsattel JP, Richardson EP Jr (1988) Morphometric demonstration of atrophic changes in the cerebral cortex, white matter and neostriatum in Huntington's disease. J Neuropathol Exp Neurol 47: 516–525

Dickson DW, Davies P, Mayeux R, Crystal H, Horoupian DS, Thompson A, Goldman JE (1987) Diffuse Lewy body disease: neuropathological and biochemical studies of six patients. Acta Neuropathol 75: 8–15

Fenton GW (1994) Electroencephalography (EEG). In: Copeland JRM, Abou-Saleh MT, Blazer DG (eds) Principles and practice of geriatric psychiatry. Wiley, Chichester, pp 459–466

Fernandez F, Levy JK (1991) Psychopharmacotherapy of psychiatric syndromes in asymptomatic and symptomatic HIV infection. Psychiatr Med 9: 377–396

Filley CM, Franklin GM, Heaton RK, Rosenberg NL (1989) White matter dementia. Clinical disorders and implications. Neuropsychiatry Neuropsychol Behav Neurol 1: 239–254

Förstl J, Burns A, Luthert P, Cairns N, Levy R (1993) The Lewy body variant of Alzheimer's disease. Clinical and pathological findings. Br J Psychiatry 162: 385–392

Gertz HJ, Henkes H, Cervos-Navarro J (1988a) Correlation of MRI and neuropathologic findings. Neurology 38: 1481–1482

Gertz HJ, Stoltenberg G, Cruz-Sanchez F, Lafuente J, Schopol R (1988b) Der panencephalopathische Typ der Creutzfeldt-Jakob-Krankheit. Nervenarzt 59: 110–111

Gertz HJ, Arendt T (1999) Psychiatric disorders of the frontal lobe. Curr Opin Psychiatry 12: 321–324

Gheka J, Tennis M, Hoffmann E, Schoenfeld D, Growdon J (1991) Idazoxan treatment in progressive supranuclear palsy. Neurology 41: 986–991

Globe LI, Davis PH, Schoenberg BS, Duvoisin RC (1988) Prevalence and natural history of progressive supranuclear palsy. Neurology 38: 1031–1034

**Gustafson L, Brun A (1997) Fokal beginnende Hirnatrophie, "Morbus Pick". In: Förstl H (ed) Lehrbuch der Gerontopsychiatrie. Enke, Stuttgart, pp 278–290

Hakim S, Adams RD (1965) The special clinical problem of symptomatic hydrocephalus with normal cerebrospinal fluid pressure. Observations on cerebrospinal fluid hydrodynamics. J Neurol Sci 2: 307

Hansen LA (1994) Pathology of the other dementias. In: Terry RD, Katzman R, Bick K (eds) Alzheimer disease. Raven, New York, pp 167–177

Hansen LA, Deteresa R, Tobias H et al (1988) Neocortical morphometry and cholinergic neurochemistry in Pick's disease. J Pathol 131: 507–518

Harper PS (1992) The epidemiology of Huntington's disease. Hum Genet 89: 365–376

Heyes MP (1995) Potential mechanisms of neurologic disease in HIV infection. In: Bloom FE, Kupfer DJ (eds) Psychopharmacology: the fourth generation of progress. Raven, New York, pp 1559–1566

Hodges J, Salmon D, Butters N (1990) Differential impairment of semantic and episodic memory in Alzheimer's and Huntington's diseases: a controlled prospective study. J Neurol Neurosurg Psychiatry 53: 1089–1095

**Huntington's Disease Collaborative Research Group (1993) A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. Cell 72: 971–983

Janssen RS, Nwanyanwu OC, Selik RM, Stehr-Green JK (1992) Epidemiology of human immunodeficiency virus encephalopathy in the United States. Neurology 42: 1472–1476

Karbe H, Ground M, Huber M et al (1992) Subcortical damage and cortical dysfunction in progressive supranuclear palsy demonstrated by positron emission tomography. J Neurol 239: 98–102

Keiburtz KD, Ketonen L, Zettelmaier AE, Kido D, Caine ED, Simon JH (1990) Magnetic resonance imaging findings in HIV-1-cognitive impairment. Arch Neurol 47: 643–645

Knopman DS, Mastri AR, Frey WH et al (1990) Dementia lacking distinctive histologic features: a common non-Alzheimer degenerative dementia. Neurology 40: 251

Knutsson E, Lying-Tunnell U (1985) Gait apraxia in normal pressure hydrocephalus. Neurology 35: 155

Korczyn AD (1995) Parkinson's disease. In: Bloom FE, Kupfer DJ (eds) Psychopharmacology: the fourth generation of progress. Raven, New York, pp 1479–1484

Kosaka K (1990) Diffuse Lewy body disease in Japan. J Neurol 237: 197–204

Kuwert L, Lange HW, Langen KJ et al (1990) Cortical and subcortical glucose consumption measured by PET in patients with Huntington's disease. Brain 113: 1405–1423

Larson EB, Reifler BV, Suni SM, Canfield CG, Chinn NM (1986) Features of potentially reversible dementia in elderly outpatients. West J Med 145: 488–492

Leenders KL, Frackowiak RSJ, Quinn N, Marsden CD (1986) Brain energy metabolism and dopaminergic function in Huntington's disease measured in vivo using positron emission tomography. Mov Disord 1: 69–77

Mayeux R, Stern Y, Rosenstein R, Marder K, Hauser A, Cote L, Fahn S (1988) An estimate of the prevalence of dementia idiopathic Parkinson's disease. Arch Neurol 45: 260–262

Mayeux R, Denaro J, Hemenegildo N, Marder K, Tang MX, Cote LJ, Stern Y (1992) A population-based investigation of Parkinson's disease with and without dementia: relationship to age and gender. Arch Neurol 49: 492–497

McArthur JC, Roos RP, Johnson RT (1993) Viral dementias. In: Whitehouse PJ (ed) Dementia. Davis, Philadelphia

**McKeith IG, Galasko D, Kosaka K et al (1996) Consensus guidelines for the clinical and pathological diagnosis of dementia with Lewy bodies (DLB): report of the CDLB international workshop. Neurology 47: 1113–1124

Mendez MF (1993) Miscellaneous causes of dementia. In: Whitehouse PJ (ed) Dementia. Davis, Philadelphia, pp 337–358

Milberg W, Albert M (1989) Cognitive differences between patients with progressive supranuclear palsy and Alzheimer's disease. J Clin Exp Neuropsychol 11: 605–614

*Mizutani T, Shiraki H (1985) Clinicopathological aspects of Creutzfeldt-Jakob disease. Elsevier, Amsterdam

Naber D (1993) AIDS und ZNS. In: Schüttler R (ed) Tropon-Symposium, vol VIII: Organische Psychosyndrome, Springer, Berlin Heidelberg New York

Navia BA, Jordan BD, Price RW (1986) The AIDS dementia complex: I. Clinical features. Ann Neurol 19: 517–524

Neary D, Snowden JS, Northern BM, Goulding P (1988) Dementia of frontal lobe type. J Neurol Neurosurg Psychiatry 51: 353–361

Neary D, Snowden JS, Gustafson L et al (1998) Frontotemporal lobar degeneration. A consensus on clinical diagnostic criteria. Neurology 51: 1546–1554

Pakesch G, Loimer N, Grunberger J, Pfersmann D, Linzmayer L, Mayerhofer S (1992) Neuropsychological findings and psychiatric symptoms in HIV-1-infected and noninfected drug users. Psychiatry Res 41: 163–177

Paulus W, Jellinger K (1991) The neuropathologic basis of different clinical subgroups of Parkinson's disease. J Neuropathol Exp Neurol 50(6): 743–755

Perry EK, Haroutunian V, Davis KL et al (1994) Neocortical cholinergic activities differentiate Lewy body dementia from classical Alzheimer's disease. Neuroreport 5: 747–749

*Pillon B, Dubois B, Ploska A, Agid Y (1991) Severity and specificity of cognitive impairment in Alzheimer's, Huntington's and Parkinson's disease and progressive supranuclear palsy. Neurology 41: 634–643

Podoll K, Schwartz M, Noth J (1991) Language functions in progressive supranuclear palsy. Brain 114: 1457–1472

Post MJD, Berger JR, Quencer RM (1991) Asymptomatic and neurologically symptomatic HIV-1-seropositive individuals prospective evaluation with cranial MR imaging. Radiology 178: 131–139

Prusiner SB, Hsiao KK, Bredesen DE, DeArmond SJ (1989) Prion disease. In: Vinken PJ, Bruyn GW (eds) Handbook of clinical neurology, vol 12 North Holland, Amsterdam, pp 543–580

Rajput AH, Offord KP, Beard CM, Kurland LT (1987) A case-control study of smoking habits, dementia, and other illnesses in idiopathic Parkinson's disease. Neurology 37: 226–232

Rao SM, Leo GJ, Haughton VM, St. Aubin-Faubert P, Bernardin BS (1989) Correlation of magnetic resonance imaging with neuropsychological testing in multiple sclerosis. Neurology 39: 161–166

Rewcastle NB (1991) Degenerative diseases of the central nervous system In: David LR, Robertson DM (eds) Textbook of neuropathology. Williams and Wilkins, Baltimore, pp 904–961

Röther J, Schwartz A, Härle M, Wentz KU, Berlit P, Hennerci M (1992) Magnetic resonance imaging follow-up in Creutzfeldt-Jakob disease. J Neurol 239: 404–406

Scherzinger E, Lurz R, Turmaine M et al (1997) Huntington-encoded polyglutamine expansions form amyloid-like protein aggregates in vitro and in vivo. Cell 90: 549–558

Schrager K, D'Souza MP (1998) Cellular and anatomical reservoirs HIV-1 in patients receiving potent antiretroviral combination therapy. JAMA 1: 67–71

Steele JC, Richardson JC, Olszewski J (1964) Progressive supranuclear palsy. Arch Neurol 10: 333–359

Stern Y, Marder K, Tang MX, Mayeux R (1993) Antecedent clinical features associated with dementia in Parkinson's disease. Neurology 43: 1690–1692

Tozzi V, Narciso P, Galgani S et al (1993) Effects of zidovudine in 30 patients with mild to end-stage AIDS dementia complex. AIDS 7: 683–692

*Vanneste J, Augustijn P, Dirven C, Tan WF, Goednart ZD (1992) Shunting normal-pressure hydrocephalus: do the benefits outweigh the risks? A multicenter study and literature review. Neurology 42: 54–59

Vitkovic L, Tardieu M (1998) Neuropathogenesis of HIV-1 infection. Outstanding questions. CR Acad Sci 321: 1015–1021

Zilber N, Kahana E, Abraham M (1991) Creutzfeldt-Jakob disease in Israel: an epidemiological evaluation. Neurology 41: 17–25

CHAPTER 10

F.M. Reischies

Mild Cognitive Disorders

Translator: S. Henn

1 Introduction

If a person suffers impairment in cognitive function, it can mean serious disability in everyday life; this also applies to mild cognitive impairments, e.g. if a person works in an intellectually demanding career. Because of the new skills and re-training increasingly necessary in today's careers, an excellent ability to learn is essential, and people are frequently pensioned off as a result of decreased cognitive and mnemonic capability. Mild cognitive disorders are therefore becoming increasingly important.

The designations of organic brain syndromes or psycho-organic syndromes, previously frequently used, do not exclusively describe cognitive features of organic brain disorders. With the development of diagnostic classification systems, a distinction has been attempted to be made for cognitive and other psychopathological features; consequently, as an example, organic affective disorder and organic personality disorder are classified separately (see Chap. 12, Vol. 2, Part 2). The present article will describe those acquired disturbances of cognitive abilities which do not meet the quantitative dimensions and qualitative criteria of dementia, delirium and amnestic syndromes (see Chap. 11, Vol. 2, Part 2). The diagnostic category "mild cognitive disorder" does not pertain to a psychiatric illness or even to a psychiatric syndrome; instead, various psychological disorders with many diverse aetiologies, symptoms and syndrome developments are subsumed under this heading.

Cognitive impairments have specific causes and effects in each of three age-groups:

1. Perinatally acquired disorder with consequences for psycho-social development is already manifested in early childhood or adolescence and is constitutional or acquired. This group of mild cognitive impairments is excluded here or mentioned only in passing (see Chaps. 2 and 3, Vol. 2, Part 1). Diagnostic systems generally provide a category for borderline intellectual capabilities, i.e. for low intelligence. The intelligence quotient in this case is between 70 and 84, i.e. in the range of 1 to 2 standard deviations below the norm. This is also the range of test performance expected for individuals with mild cognitive disorder.
2. Mild cognitive disorders appear in middle and older age-groups, for example, as a result of brain insults with a reversible or persistent deficit.
3. In old age, more pronounced age-related changes of cognitive abilities have to be distinguished by differential diagnosis from early stages of degenerative dementia illnesses, e.g. Alzheimer's dementia.

The different nosological concepts will be addressed first. Two concepts are directly opposed to one another. On the one hand, there are psychiatric syndromes, which appear following clearly diagnosed brain diseases. They present as milder or atypical syndromes in contrast to other organic psychiatric illnesses. This is how they are viewed by both the ICD-10 (WHO 1992) and DSM-IV (APA 1994; see also below) diagnostic systems. The other case concerns those cognitive disorders in which brain injury can be suspected, but proof is lacking. This is especially the case for early stages of a dementia disease, in which the progressive decline of cognitive capabilities is striking, but the criterion for the diagnosis of dementia is not yet fulfilled and no clear evidence results from brain imaging investigations. Highly gifted individuals, in particular, remain in this unsatisfactory differential diagnostic situation for a long time due to their higher initial level of ability in intelligence and memory functions.

Some disease courses are progressive according to reports by the patient or relatives; they begin insidiously and later turn into a dementia syndrome. The problem of differentiating a dementia illness from normal age-related changes in cognitive function, which is especially difficult in the elderly, is presented in more detail below. Secondly, courses which are stable over time are observed, e.g. deficits following various brain insults. Frequently, brain injuries which result in impairment of already acquired cognitive functions show a plateau at a lower level following initial improvement. Thirdly, there are forms which disappear over time – whether following localized brain lesions or caused by a temporary disorder in brain function.

Similar diverse causes are possible for mild cognitive disorders as for dementia and for delirium (Gutierrez et al. 1994). Whether, for example, one of the progressive dementia illnesses develops in old age or a Wernicke encephalopathy or a brain trauma has led to a residual deficit state, clinical identification of the underlying brain disease is often difficult, although diagnosis is possible in specific cases.

2 Definitions

Mild cognitive disorder is a category which describes a clinical picture showing deficits in multiple cognitive functions quantitatively characterized as mild. Quantitatively, it has to be differentiated from the normal variance in cognitive abilities on the one hand and severe cognitive disorders, dementia, delirium and

amnestic syndrome on the other. Following the new international World Health Organization (WHO) classification (ICD-10; WHO 1992), the concept replaces the ICD-9 diagnosis 310.1, non-psychotic forms of intelligence disorders following brain insults (WHO 1978). The concepts such as "organic brain disorder of non-psychotic form" (ICD-9) were also common, but were not, however, limited to cognitive impairment.

2.1 Mild Cognitive Disorder According to ICD-10

Some cognitive features of dementia are listed in the diagnostic criteria for research of ICD-10 as criteria for mild cognitive disorder. The neuro-psychological dimensions of memory, attention, speed of mental processes, thinking, speech and visuo-spatial functions are essentially the same in both classifications. Although cognitive symptoms of dementia are named in the ICD-10 and DSM-IV criteria, the disorder is defined such that it does not reach the severity seen in dementia. Signs of severe cognitive dysfunctions, such as in orientation and judgement, are not accepted as criteria. Mild cognitive disorder is thus primarily a diagnostic category for a group of individuals who fall in the border area between cognitive health and dementia. ICD-10 does not specify how many symptoms are necessary to determine the diagnosis.

The ICD-10 definition of mild cognitive disorder requires objective proof of brain disease or systemic illnesses which lead to dysfunctions of the brain (see below). For example, these can appear on the basis of a cerebrovascular accident or intoxication. According to ICD-10, the diagnosis can only be reliably confirmed if the mild cognitive disorder is later reversed as the primary organic illness subsides (WHO 1992; Christensen et al. 1995). This additional requirement contributes to the considerable uncertainty present in the concept of mild cognitive disorder. A reason for this determination can be seen in that only when the beginning and end of the organic brain disorder coincide in time with that of the psychopathological symptoms can the cognitive impairment be considered to be caused by the organic brain event. However, if this requirement is strictly followed, then a relatively mild delirium in response to a traumatic event, for example, also falls under the definition. In the older literature, the term "acute exogenous reaction type" was used. Brain lesions and brain diseases usually lead to long-term consequences precisely in the cognitive area; however, with persistent brain injuries, the diagnosis of mild cognitive disorder cannot be determined with exactly the same degree of certainty as it can for a mild brain injury with reversible cognitive symptoms or temporary toxic-metabolic disorders.

A progressive illness at an early stage is ruled out if there is no objective proof of an organic origin of the cognitive disorder. However, if this proof is present, then the illness can be classified under the ICD-10 category of "mild cognitive disorder". The diagnostic criteria of delirium are described in detail in Chap. 11 (Vol. 2, Part 2).

In contrast to age-associated cognitive impairment (see below) and related concepts, ICD-10 does not limit the mild cognitive disorder to older age-groups; the American diagnostic system DSM-IV proceeds in the same way.

2.2 Mild Neurocognitive Disorder According to DSM-IV

In contrast to ICD-10, the corresponding DSM-IV category "mild cognitive disorder" (APA 1994) also includes cognitive impairment at an early stage of progressive dementia; however, damage to the brain has to be substantiated. This additional condition is currently problematic in many patients with early dementia diseases.

In contrast to ICD-10, DSM-IV sets a minimum for disturbed cognitive areas: two or more areas of disrupted cognitive ability are required. In addition, the cognitive dysfunction has to be confirmed by tests, which in ICD-10 are only demanded by the research criteria. DSM-IV lists essentially the same elements for cognitive disorder as ICD-10; however, it subsumes planning, organization, etc. under executive functions. Furthermore, DSM-IV does not include the subjective impairment of cognitive abilities (see Caine 1994). It is emphasized that mild neurocognitive disorder has to involve a clear suffering, but this can also refer to a reaction when confronted with psychosocial effects of the disorder (APA 1994).

According to DSM-IV, the cognitive disorder has to reach a degree of severity such that they result in an impairment of daily living. Heaton and Pendleton (1981) report relevant daily impairment with mild disorders of cognitive function. In addition to deficits in memory and in the area of speech, attention disorders and disorders involving abstraction and decision-making are also of direct social significance. The latter appear, for instance, following frontal brain injuries.

At this point, the significance of mild cognitive disorders in terms of the demands made by the social environment and a person's career need to be looked at. Slight impairments already have a marked effect on intellectually demanding careers. The same is also true for everyday life in culturally elite social circles. The demands of daily life can, however, generally be adapted to individual cognitive losses. If the patient copes without any problems in the adapted

environment, adaptation has also to be considered in assessing the impairment. The requirement that mild cognitive disorder should be related to performance of daily activities is certainly reasonable for psychiatric case definition and the classification as a "disorder", but it makes the differential diagnosis of dementia more difficult.

As Fig. 1 shows, with the differential diagnosis of mild cognitive disorder the quantitative dimension, i.e. the severity of cognitive impairment and the number of cognitive symptoms, has to be differentiated from the qualitative dimension, the additional criteria such as objective documentation of organic origin, the predisposition, or time of acquisition of cognitive impairment and the existence of non-organic psychiatric illnesses such as depression, schizophrenia and histrionic disorder. The amnestic syndrome is subsumed under severe disorders of cognitive abilities with the neuropsychological syndrome to which aphasia and apraxia belong.

2.3 Objective Proof of the Organic Origin

Both diagnostic classification systems demand proof of an organic cause, be it a brain disease or a systemic illness with effects on brain function. The criteria of the diagnostic classification systems take as a starting point the current diagnoses of organic brain diseases such as stroke, brain trauma or encephalitis. As a consequence of these illnesses, temporary or persistent mild disorders of cognitive abilities can be observed, which are then classified as mild cognitive disorder.

The neuropsychological symptoms in mild cognitive disorder are indicators of a primary brain disease, but can also be observed in psychiatric illnesses in which factors of brain dysfunction or neurophysiological correlates are not yet otherwise explained, for example. The demand for substantiation of the organic origin of cognitive function disorders can be understood such that the cognitive impairment should not be both a classified symptom and proof of the causation by a brain disease. Independent medical findings are called for.

There is still no consensus as to what constitutes valid objective evidence. Findings from imaging methods are frequently invoked in which the validity with regard to proof of organic brain injury is doubtful. For example, a finding by computer tomography may be cited showing "mild cortical atrophy" or "mild ventricular enlargement". A similar situation arises with patchy signal enhancement on magnetic resonance tomography in subcortical areas, findings which are

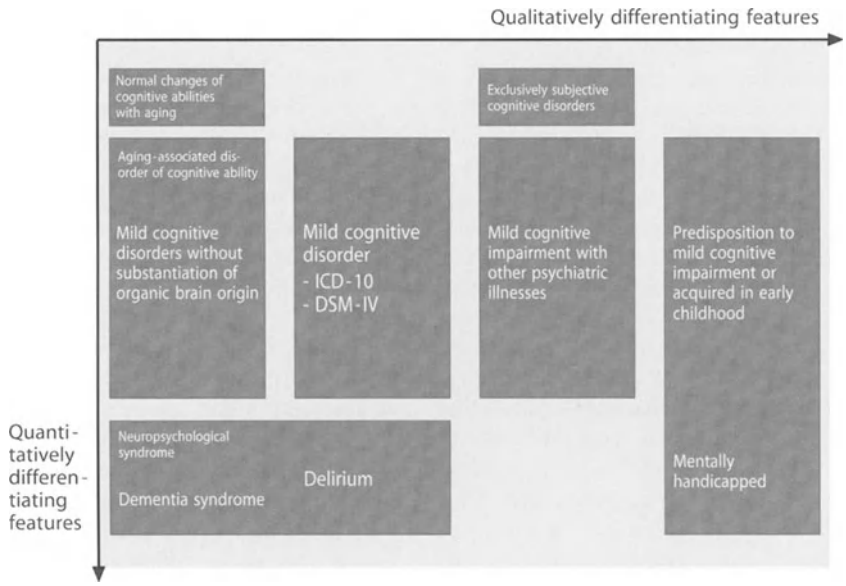


Fig. 1. Differential diagnosis of mild cognitive disorder in the quantitative dimension with regard to the severity of cognitive impairment and number of cognitive symptoms and in the qualitative dimension with regard to additional criteria such as objective proof of organic origin, predisposition, or the time of acquisition of cognitive impairment, and the presence of non-organic psychiatric illnesses such as depression, schizophrenia, and histrionic disorder. The exclusively subjective impairment of cognitive performance is separated here from the category of mild cognitive disorder (see text). Amnestic syndrome is subsumed under the severe disorders of cognitive performance, the neuropsychological partial syndromes to which aphasia and apraxia also belong.

aetio-pathogenetic, diagnostically non-specific and heavily age-dependent.

Since normal values for these findings are still frequently lacking for older age-groups and cut-off values are not validated, considerable research effort is needed in the area of neuroradiological validation of organic brain injuries in mild cognitive disorder. The difficulty that, depending on the state of diagnostic tests for organic brain illnesses, different numbers of patients fall into the group of mild cognitive disorder according to ICD-10 or DSM-IV is only mentioned in passing.

Laboratory findings which prove a dementia are presently not available. Genetic markers refer only to predisposition, but do not prove that the clinically recorded cognitive disorder marks the beginning of the development of a dementia syndrome, since other causes of cognitive disorder are not ruled out.

It is uncertain whether, for early dementia, clinical evidence of a typical time course, such as the typical sequence of appearance of neuropsychological disorders, for example, can also serve as required proof of an organic brain illness. It is also still unclear whether additional borderline values suffice if they complement one another with regard to the diagnosis of organic brain disease; examples worth mentioning include borderline signs of atrophy, a genetic load and the typical course of progression or characteristic sequence of symptoms.

2.4 Definition of Mild Cognitive Disorder Without Proof of an Organic Origin

In a large patient group, the diagnosis of mild cognitive disorder is, in fact, suggested by mild cognitive impairment itself and, where applicable, also by its course without, however, the diagnostically required brain damage being proven through additional medical findings. Practically speaking, this second concept of mild cognitive disorder is significant in the clearly ascertainable progressive decline of cognitive ability in a patient in whom an early dementia is suspected on clinical grounds and in whom the criteria for dementia syndrome are not yet fulfilled. The use of the term "mild cognitive disorder" appears justified here, even if the criteria of the diagnostic systems ICD-10 and DSM-IV are not fulfilled. A long series of terms have been proposed for the cases described, including "questionable dementia" or, to differentiate from normal cognitive ageing, "ageing-associated cognitive decline" (Henderson and Huppert 1984; Levy and Working Party of the International Psychogeriatric Association 1994; see also below). It should be emphasized here that ageing-associated cognitive decline and other suggested definitions require no everyday relevance of the cognitive change (Rediess and Caine 1996).

This second concept for mild cognitive disorder is described here because otherwise a large group of patients in psychiatric treatment and geriatric psychiatry would not be diagnosed. Since there is a heterogeneous category with the diagnosis "mild cognitive disorder", it is also necessary in this chapter to discuss patients who fall into this second concept. It must again be emphasized, however, that formal diagnosis according to the diagnostic systems does not allow for the classification of mild cognitive disorder in this case. An ICD-10 classification section for otherwise unclassified mild organic brain syndrome exists, which also includes mild cognitive impairment (ICD F. 07.8). Mild cognitive disorder without proof of organic origin will be discussed in the next section in its relationship to normal cognitive age-related changes. In the following, the common reference to both concepts of mild cognitive disorder with and without medical findings of organic origin is made, and only when a fact relates specifically to one of the two concepts is the specific concept referred to.

2.5 Mild Cognitive Disorder in Old Age

Because of the clinical significance of diagnostic differentiation between normal age-related changes in cognitive abilities, an accelerated decline of cognitive function in old age and a dementia syndrome, intensive geriatric psychiatry research work is being carried out with the aim of identifying an intermediate group of patients who certainly show cognitive abilities differing from the norm, but do not show dementia. A series of concepts have now been proposed in the literature which are distinguished in many different respects (see Zaudig 1995).

Kral (1962) already proposed a differentiation between age-related forgetfulness in the elderly with a good prognosis and that with a poor prognosis. The favourable type of age-related forgetfulness, benign senescent forgetfulness, is understood such that a further course of development of dementia is no more frequently observed than in non-forgetful individuals of the same age. It is described as stable over time; only normal age-related changes of cognitive abilities will gradually cause a further worsening. In a follow-up study after 4 years, Kral was able to show a clearly lower mortality rate in individuals classified as having the benign type of age-related forgetfulness. Features of benign senescent forgetfulness are that individuals forget relatively unimportant detailed information after a while, but important dates or appointments

are reliably stored and can be remembered. This characteristic was able to be confirmed as a normal age-related change in memory functions (Reischies et al. 1996). Furthermore, memory disorder not associated with dementia is restricted to intermittent problems, repeating names and dates from memory, and to slight temporary orientation disorders. Disorders which go beyond those of memory functions are less favourable or are diagnostic for the assessment of development of dementia. A further feature of benign senescent forgetfulness is that the patients themselves are aware of memory impairment and try to circumvent the gaps in memory or conceal them.

Benign senescent forgetfulness defines a boundary between mild ageing-dependent cognitive impairment on the one hand and cognitive-mnestic disorders which establish a dementia syndrome on the other. In recent years, additional results have been gathered which more closely describe the characteristics of the memory disorders belonging to the dementia syndrome, e.g. decreased ability to learn and reduced capability of increasing recall through semantic support (see Reischies et al. 1996). This confirms the possibility of distinguishing benign and prognostically unfavourable memory disorders on the basis of qualitative and quantitative features.

A disadvantage of the concept of benign senescent forgetfulness is the lack of definite criteria and quantitative cut-off values. In addition, studies on normal ageing of memory functions have only recently become available, enabling the criteria to be further developed. Another disadvantage of the concept is the restriction to memory disorders. These are certainly by far the most significant symptoms of early dementia illness for diagnostic purposes, but extending them to include cognitive impairments in the light of prognostically favourable and unfavourable features would be desirable. DSM-IV provides an additional diagnosis for mild cognitive impairments of normal ageing, age-related cognitive decline (780.9); this is intended essentially for the cases of even milder, but still age-related cognitive impairment. Since by definition it deals with normal ageing changes, a more fundamental consideration is necessary as to whether it is even justified to make a psychiatric diagnosis here.

In contrast to benign senescent forgetfulness, the concept of age-associated memory impairment (Crook et al. 1986) emphasizes the distinction between healthy ageing and mild cognitive impairment (see also Fig. 1). This concept therefore mainly attempts to clarify the transition by individuals with barely discernible impairment in cognitive abilities to the designated intermediate group. The essential feature of this concept is the classification of abilities below a constant test value as pathological, even at a very high age. The test values refer to healthy 50-year-olds. As a result, practically every elderly person over 90 years, for example, is given a diagnosis of an age-associated memory impairment. This circumstance contradicts the validity of the concept just discussed of differentiating prognostically favourable and unfavourable changes in cognitive abilities in old age. Within this concept, the cut-off value of the Mini-Mental State Examination (MMSE; Folstein et al. 1975) is additionally taken as the threshold for dementia.

Another kind of definition follows the typological approach. Using assessment scales, such as the Global Deterioration Scale (GDS; Reisberg et al. 1982), the conditions "healthy", "very mild cognitive deterioration" (intermediate group) or "demented" are described. The clinician can classify the patients according to their neuropsychological impairment and psychopathology. In the GDS, the cognitive ability of an older individual should be compared with a healthy person of the same age.

The proposed concepts are further differentiated by the separate elements of cognitive ability which are listed in the definition; in addition to memory disorders, these also include problem-solving, as found in the "questionable dementia" of the frequently used Clinical Dementia Rating (Hughes et al. 1982), or language comprehension disorders. In addition, the concepts are also differentiated in light of the question of whether or not a progressive deterioration of cognitive abilities is required.

Furthermore, there is a difference in treating the subjective cognitive impairment and awareness of the deficit. Some concepts depend only on objective impairments (e.g. the GDS), while others take into account the complaints made by the patients about an impairment or the fact that the patients themselves are aware of the impairment.

In summary, the heterogeneity of the diagnostic categories and concepts for mild cognitive impairments in the elderly must be emphasized. None of these concepts has been satisfactory in solving the difficult problem of differentiating a group of individuals with mild cognitive impairments, slow progression and without sure proof of brain lesions from normal cognitive ageing on the one hand and from dementias on the other hand, either theoretically or technically in terms of measurements and procedures. We will have to wait to see to what extent the diagnostic criteria as presented in ICD-10 and DSM-IV can be invoked with and without valid proof of brain injury in psychiatric diagnosis in the elderly.

3 Diagnosis

3.1 Differentiation from Normal Age-Related Changes in Cognitive Abilities

Performances on cognitive tests deteriorate in old age. This has been shown both in the course of longitudinal studies and in cross-sectional studies in comparing groups of individuals of varying ages (Schaie 1989; Salthouse 1985). All cognitive abilities appear to be affected by this age effect. It is more strongly evident in speed-dependent tasks. It has even been suspected that the general cognitive ageing is mediated by cognitive slowing (Salthouse 1985; Lindenberger et al. 1993).

Age effects on memory primarily concern the encoding and retrieval of information; however, the ability to store in itself is not decreased (see Reischies et al. 1996). In other words, if something is successfully stored, it will not be more quickly forgotten. Although it is more difficult to retrieve content from memory in old age, it can either be recalled with help or can be recognized from choosing among stimuli. There are clearly observable effects of ageing on the recall of learned material, but they do not in themselves inevitably lead to an impairment in everyday life (Reischies and Lindenberger 1996).

In the area of language abilities, finding a word becomes more difficult in old age, both in naming and in "fluency tasks" (see Table 1). The structure of semantic knowledge is apparently not altered in old age, but retrieval from semantic memory, i.e. lexical knowledge (Light 1988), is slowed down and runs less reliably (Light 1988).

Tests of visuo-spatial abilities (e.g. drawing, putting pieces of figures together) likewise show clear effects of old age (see Koss et al. 1991). This is especially true if the tasks are required to be done quickly.

The effect of old age presumably behaves linearly from about 70 years of age into the oldest age-range (Lindenberger et al. 1995). In practically all tested activities, a decrease was found in the test values in the range of 1–2 standard deviations when healthy 70- to 74-year-olds were compared with those over 95 years of age (Reischies et al. 1996).

In summary, the studies on cognitive performance of healthy older people show that practically no area of performance remains unaffected by impairment related to ageing. In other words, no test for the diagnosis of mild cognitive disorder can be selected which shows no ageing effect at all, and such a test presumably does not exist.

If a mild cognitive disorder is to be clinically identified as the early stages of a dementia disease, it has to be differentiated from the normal age effect of decline in cognitive performances. It should be pointed out here that the ability to substantiate an organic origin of cognitive performance disorders is for the most part lacking at this time for classification under the ICD-10 and the DSM-IV categories. The disorder corresponds to the second concept of mild cognitive disorders as discussed above. There are differences between individuals in the extent of cognitive ageing, and this has been discussed in terms of looking at dementia syndromes as illnesses that can be distinguished from normal cognitive ageing on the one hand or as extreme forms of cognitive ageing characterized only by intensity and rapidity of development on the other hand. The latter, of course, would suggest that everybody would eventually develop dementia if they only lived long enough (Drachman 1994). This would be the consequence of viewing the dementia syndrome as a non-specific symptom of the ageing brain. According to this thesis, all old age-associated cognitive impairments and mild cognitive disorders in old age could be regarded as outpost symptoms of a later dementia.

A differentiation between normal cognitive ageing and a typical development of dementia succeeds at least in many cases on the basis of the course. This means that, within the diagnostic category of mild cognitive disorder, assessing the more rapid progress of dementia diseases can distinguish them from other forms of mild cognitive disorder.

Table 1. Ranges of mild cognitive disorder between -1 SD and -2 SD in the distribution of test performances

Performance	-2 SD	-1 SD	Mean	+1 SD
Digit span[a]	4	5	6	7
Remember 15 words[b]	5.6	7.1	8.6	10.1
Name animals, 60 s[c]	8.4	13.2	18.0	22.8

[a]HAWIE (Tewes 1991): sequence of numbers is repeated (in correct order).
[b]Rey Auditory Verbal Learning Test (Lezak 1995): immediate recall without consideration of the order of presentation.
[c]"Fluency": to name as many animals as possible as quickly as possible; data for 68-year-olds, $n = 278$ (Morris et al. 1989).
SD, standard deviation.

3.2 Methodological Aspects of Diagnosis

Making certain of the diagnosis of mild cognitive disorder is not straightforward when differentiating it

from normal ageing and is not always satisfactory following brain injuries if the effect of injuries has to be distinguished from the normal range of cognitive performances in adults. This is related on the one hand to the basic defining concepts. On the other hand, a series of viewpoints, which will be presented briefly below, can be observed in assessing cognitive performances by means of tests. Since evaluation of subtle deficits in performance are involved in diagnosing mild cognitive impairments, it is probably considerably more crucial than is the case with other diagnoses to have neuropsychological testing methods in addition to the clinical examination. This testing can be carried out by neuropsychologists or clinical psychologists or by psychiatrists after appropriate training. The application of diagnostic test takes account of two quantitative limits which have already been mentioned, the lower limit for the dementia syndrome and the upper limit for the normal variance, for which the clinical validation of this limit presents an additional problem.

Although at first it appears to be clear what mild cognitive disorder means, difficulties arise with the details. With mild neuropsychological partial syndromes, e.g. mild memory disorders, a diagnostic dilemma frequently arises: a mild amnestic disorder is equated with mild dementia, i.e. the beginning of clinical dementia is suspected and it is then assumed that only in later stages will the remaining cognitive impairments be observed. An early progressive dementia illness usually first manifests itself by memory disorders. However, the amnesia syndrome alone may increase, with the other cognitive functions remaining undisturbed. Until disorders are seen in other neuropsychological areas, the dementia is more or less in a diagnostic grey area (Reischies 1996) or a diagnostic "detection zone" (Plassman and Breitner 1996).

The DSM-IV criteria of mild neurocognitive disorder take as a starting point a multidimensional neuropsychological impairment: deficits in at least two cognitive dimensions are required for the diagnosis. If on the other hand, as in ICD-10, it is not required that multidimensional performance disorders are present in the profile, then the question arises as to which critical cognitive functions have to be present for the diagnosis. Even with uniformly lower cognitive performance, the patient is particularly impaired with regard to everyday function only in some areas, such as attention or the ability to concentrate on a task. In old age, primarily memory disorders are noticed. This relates to the differential diagnosis of a development of early dementia. However, the consequences of a memory disorder are also directly relevant to everyday life. This cannot be claimed for many other partial performance disorders, e.g. impairment of musical achievements in a person who is not musically active. The mild cognitive disorder, if not present in all areas, must at least have an impact on critical areas of everyday life.

In child and adolescent psychiatry, attention is particularly paid to development disorders in language abilities (Stevenson et al. 1985). This is reasonable because of the prominent role of language in social contact and knowledge acquisition; language communication disorders are also probably underestimated in old age as the reason for inadequate social contact. The role of mild disorder of judgement after frontal lobe lesions is often neglected in psychiatric evaluation.

But how many people have a balanced achievement profile in the various cognitive dimensions, as shown in the standardized battery of tests? There is no satisfactory data about this in the literature, and there is no agreement about the question of interpretability of the test profile. Mitrushina and Satz (1995) found that only 3.8% of their sample of healthy 70-year-olds showed no conspicuous test result, i.e. most individuals had at least one abnormal subtest in the Wechsler Intelligence Test, which deviated by more than 3 points from the mean; a point difference of 3 points is viewed as clinically significant. The performances in the different domains of intelligence tests therefore usually show a considerable variance within the individual. The danger arises that, with the suspicion of mild cognitive disorder following an organic brain injury, chance deviations in the test profile will be judged to be evidence of the diagnosis.

The susceptibility of certain brain areas and corresponding particular cognitive abilities in comparison with non-specific brain injury is undisputed. The differentiation between fluid and crystallized intelligence can be cited in this connection (Horn 1982). The fluid abilities – with marked effect of ageing – include the time-dependent and visuo-spatial performances, while the crystallized abilities having higher persistence of achievement level following non-specific injury to the central nervous system include above all linguistic knowledge and motor skills. Thus it follows that, with non-specific brain injuries, it is more likely that impairments in fluid performances result.

The normal standard distribution of intelligence encompasses a wide range of cognitive capability. The range of the distribution of cognitive achievements is important for mild cognitive disorder. If we compare the groups of individuals with intelligence test results which are 1 standard deviation above or below the norm, the difference can be clearly seen, as is shown in Table 1. Education-specific norms for tests are uncommon and for the most part not available for old age.

At the same time, Table 1 shows the range in which individuals with a standard deviation between −1 and −2 perform and how those with mild cognitive disorder can be expected to perform. Performance in many

intelligence tests which require more complex and not everyday functions are too difficult for individuals with borderline abilities (IQ 71–85). This is especially true at higher ages.

The statistical frequency of individuals with mild cognitive performance impairment can also be seen from the normal standard distribution. About 2% of the population are below the range of 2 standard deviations, and about 16% below 1 standard deviation. It therefore follows that about 14% of the population fall between 1 and 2 standard deviations below the norm in cognitive performance. The reason for the lower test achievement may be slight limitation in cognitive achievements originating either constitutionally or in early childhood. These individuals do not fall under the diagnosis of mild cognitive disorder according to ICD-10 or DSM-IV. The appraisal of premorbid intelligence is sometimes difficult, and objective validation is most successful using vocabulary tests.

Clinicians can place only limited confidence in the test results which are referred to for assessment: if the reliability of a test is established by correlation of the test results in two stages, then it turns out that mostly 10%–40% of the performance tested twice is random variation. In addition to the random variations and fluctuations, there are all sorts of changes in temporary performance factors such as motivation, alertness and general state (Cohen 1992). The description of the complexity of test diagnostics makes it clear that the diagnosis of mild cognitive disorder, at least in problem cases, can often only be safely made in experienced specialized outpatient clinics such as memory clinics or neuropsychological departments.

Many patients cannot appropriately estimate their own cognitive achievements. This is shown by the comparison of subjective assessment with objective performances in tests. The correlation between subjective and objective disorder of cognitive achievements without exception turned out to be very low (Feehan et al. 1991; O'Connor et al. 1990). Many reasons may be given for this: the cognitive ability has to be observed by the patients; often, however, they lack a standard of comparison and probably also the ability to assess themselves appropriately. Demented individuals frequently minimize or do not admit to their deficits – out of shame or deficits in judgement. Furthermore, self-image also plays a role, both the image that a person has of him- or herself and the one which they want others to have of them. Expectations relating to the social effects of such a statement can therefore be observed. There is also a certain tendency to aggravation. It should be mentioned here that an objective diagnosis of mild cognitive disorder may be difficult if the patient wants to be underrated with respect to the performance level. This also holds true for neuropsychological test results. Often, only the consistency of the pattern of test results can support the validity of a low level of test performance.

Subjective cognitive impairment is frequent. Of those individuals who were not demented in the study by Livingstone et al. (1990), 23% complained of failing cognitive abilities, and Christensen et al. (1995) found subjective cognitive disorders in 17% of their probands. Subjective impairments alone are accepted by some authors and also by the ICD-10 classification of mild cognitive disorder as an essential component of a clinical syndrome diagnosis (Christensen et al. 1995). The suffering of individuals who perceive their own cognitive performances as disturbed, although this assumption is false, can certainly represent a symptom, e.g. of depression. Many researchers emphasize the strong dependence of subjective impairment of performance upon depressive symptoms, and in many patients a symptom of depression can be recorded, e.g. with the ICD-10 criteria of a mild cognitive disorder (Christensen et al. 1995).

According to results in the literature, subjective impairment cannot be used as a basis to predict a progressive course towards dementia (Reisberg et al. 1986; O'Brien et al. 1992). It is doubtful whether it is reasonable to accept purely subjective cognitive impairment as a sufficient criterion for mild cognitive disorder. As an additional feature, however, the everyday relevance of mild cognitive disorder appears justified in the criteria.

4 Differential Diagnosis

In addition to the differential diagnosis of dementia syndrome and ageing-related deterioration in performance discussed above, mild cognitive disorder also has to be differentiated from other organic brain syndromes such as delirium and amnestic syndrome (see Fig. 1 and Chap. 11, Vol. 2, Part 2). One of the important differential diagnostic features is the severity of cognitive disorder: first of all it is milder. In contrast to a delirium, additional differential diagnostic features include disturbance in consciousness (Gutierrez et al. 1994) and vegetative symptoms. There are fluctuations in the psychiatric symptoms of delirium. In benign senescent forgetfulness, however, inconsistent amnestic failures are not remembering a name once, whereas otherwise this name is confidently remembered, for example (see above). A need to differentiate with respect to amnestic syndrome arises; it was discussed in connection with the differential diagnosis of mild cognitive disorder and early dementia.

Mild disorder of cognitive performances in other psychiatric illnesses, e.g. affective or schizophrenic

psychoses, can be mentioned here only briefly. The validity of the results obtained in patients with these illnesses frequently remains unclear (Cohen 1992). On the one hand, an organic brain injury with mild cognitive impairment can produce a disposition for an affective or schizophrenic illness or at least can strengthen a constitutional disposition. One example is the connection which can be made between a mild form of frontal and mediotemporal brain dysfunction with schizophrenia (Bogerts et al. 1985; Weinberger et al. 1986). However, an impairment of test performance can also be caused by disorders of various pathophysiological or psychopathological levels of function in an affective or schizophrenic illness.

Reversible mild cognitive impairments in the context of a non-organic psychiatric illness are not subsumed under the diagnosis of mild cognitive disorder. A person who is depressed does not achieve optimal performance in cognitive tasks. However, the effect of depression on testing is overestimated, as has been shown by recent investigations (Burt et al. 1995; Reischies 1993).

More subjective than objective disorders of cognitive performance characterize a depressive illness (see above and Feehan et al. 1991). Patients with depression complain of inhibition. They compare themselves with the highest performances and thereby strengthen their negative self-image. In contrast, many patients with dementia tend to consider their test results to be better than they are, while depressed patients show surprisingly good test performance in contrast to their self-perception (O'Connor et al. 1990). This makes it clear that testing is necessary. It should, however, be brief and carried out with care, taking into consideration the low self-confidence of the patients.

Some patients with depression have memory disorders, are mildly disoriented and show a clear slowing down. Earlier, this would have been referred to as a depressive pseudo-dementia (Kiloh 1962; Lauter and Dame 1991), although clinically there is no problem in differentiating the syndrome from a fully developed dementia in most cases, but possibly with an early stage of dementia. In the subsequent course, in some of the patients with initially reversible disorders of cognitive abilities, it turns out that a dementia illness begins after a few years, although the proportion to which this unfavourable course applies has not yet been agreed on (Kral 1982; Sachdev et al. 1990). In investigations of progression in depressive patients with disorders of cognitive performance, no substantial improvement or normalization of test results usually occurs with recovery of mood (Abas et al. 1993; Reischies 1993). The question is raised for depression – as for schizophrenia – of the frequency of comorbidity with mild cognitive disorder. Results from imaging studies of these patients indicate organic brain-related neuropsychological disorders. A pathogenetic or a pathoplastic role of these mild cognitive impairments in depression needs to be further elucidated, i.e. whether or not the diagnosis of comorbidity of depression and mild cognitive disorder with or without a pathogenetic or pathoplastic aspect could better describe the facts.

5 Course and Prognosis

Studies which have followed the course in individuals with benign senescent forgetfulness or the classification of questionable dementia show that a rapid progression to a fully developed dementia is not common. Studies of the course of mild cognitive disorder lead to the conclusion that, especially in old age, a higher probability of transition to dementia can generally be established (Peterson et al. 1999). Nonetheless, the cognitive performance of a large group of patients remains constant for years without these patients developing dementia, and sometimes cognitive achievements may even improve or normalize (O'Connor et al. 1990; Copeland et al. 1992; Cooper et al. 1996). In investigations of course with initially detailed neuropsychological testing, it turns out that, in addition to memory deficits in reporting learned materials, a disorder in finding words – apparently a sign of a progressively developing aphasia – is also valuable in making a prognosis; it indicates a rapidly progressive course towards dementia (Storandt et al. 1992). Clusters of different patterns and time course of mild cognitive deficits have been described (Ritchie et al. 1996).

The decline in performance in patients with degenerative dementias is usually noticed by the family within 1 year; it can be recorded in tests for dementia. Thus Burns et al. (1991), for example, found a decrease of about 3.5 points in the MMSE per year in patients with Alzheimer dementia. The pure effect of ageing on dementia tests can be estimated from cross-sectional studies: Crum et al. (1993) found a decline in performance in the MMSE of about 1.3 points per 10 years, while Bleeker et al. (1988) found 0.8 points per 10 years. The data from the Berlin ageing study gave an ageing decline of about 1.6 points per 10 years (Reischies et al. 1996). In longitudinal studies of MMSE performance in healthy individuals, there turned out to be only an insignificantly small decline per year. So far, few studies have been devoted to reliability of measurement of changes using the MMSE (see Schmand et al. 1995). In summary, investigation of the progression of dementia in individuals in the grey area between normal ageing and dementia is very valuable. It is questionable, however, whether it can be

proved that, in addition to the distinction between the normal speed of cognitive ageing and the clearly more rapid cognitive decline in dementia, there is yet a third speed, that of an intermediate group with an age-related mild disorder of cognitive performance.

At present, an intensive search is being made for factors that can predict the further progress of cognitive disorders. An isoform of a protein involved in lipid metabolism, ApoE4, has been described as a predictor of deterioration in mild cognitive disorders (Petersen et al. 1995; Plassman and Breitner 1996). Major advances can be expected in this area in the coming years (see Chap. 6, Vol. 2, Part 2; Chaps. 8–12, Vol. 2, Part 1).

The frequency of mild cognitive disorder in the population is not yet well investigated using the new definitions. Christensen et al. (1995) found the disorder corresponding to the ICD-10 definition in 6% of those over 70 years of age. The authors describe a close relationship between this syndrome and a depressive syndrome. Within the epidemiological catchment area programme in the United States of America, mild to moderate cognitive impairment, including cases of various aetiologies, was found in 5.1% (Robins et al. 1991).

The diagnosis of mild cognitive disorders following brain lesions is significant for early retirement: among patients with psychiatric illnesses which lead to retirement, 1.3% are women and 4.1% are men who are unable to work because of mild cognitive impairments (Verband Deutscher Rentenversicherungsträger 1995). The significance of the diagnosis of mild cognitive disorder for early retirement will probably increase in the near future, as indicated in the discussion above. It is difficult to compile and assess the data for care needs following brain lesions, but, with the estimated 500,000 cases of brain lesion per year in Germany, which include various aetiologies (Kasten et al. 1997), significant numbers of mild cognitive disorders will certainly arise or remain. In addition, data from the Framingham study have shown that there is a higher mortality rate for individuals with lower cognitive performance (Liu et al. 1990).

6 Conclusion

Mild cognitive disorder comprises temporary symptoms or persistent residual symptoms of a disorder of brain function because of a brain disease or systemic disease with consequences of a disturbance of cerebral functions. It may appear following various kinds of brain lesions (e.g. brain trauma, infection, intoxication) or with various systemic illnesses. In old age, it is sometimes a precursor of clinical dementia syndromes. These aspects find different expression in the various diagnostic concepts, and a uniform and generally accepted, criterion-related diagnosis of mild cognitive disorder has therefore yet to be found. Mild cognitive disorder is neither a disease nor a syndrome, but rather a classification unit within the diagnostic classification systems. Mild cognitive disorder is, to our knowledge, quite a common disorder with different social consequences and an unfavourable prognosis in old age. The diagnosis of mild cognitive disorder according to ICD-10 and of mild neurocognitive disorder according to DSM-IV is substantiated by independent proof of an injury to or disorder in the brain coupled in time to the appearance of clinical symptoms. However, in many patients with a cognitive disorder exceeding that of normal ageing, a brain disease or brain lesion cannot be objectively verified; such patients were presented here under the category "mild cognitive disorder without proof of organic origin". At present, the most reliable criterion differentiating this disorder from the early stage of dementia is its relative stability over time in contrast to the rapid progression of dementia syndromes.

Exclusively subjective complaints of cognitive impairment appear not to be a reliable symptom of brain injury and to have little prognostic significance. Clinical research will show how this feature should be included in the diagnostic criteria of mild cognitive disorder. Further dimensions of psychopathological disorders following brain injuries, e.g. affective, compulsive or paranoid symptoms, are classified separately as comorbidity. Further research will be required to answer the question of whether or not a complex syndrome following brain injury might have clinical significance; one possibility is a syndrome that includes both cognitive and affective symptoms as well as personality changes.

7 References

Abas MA, Levy R, Sahakian BJ (1990) Neuropsychological deficits and CT scan changes in elderly depressives. Psychol Med 20: 507–520

APA (1994) Diagnostic and statistical manual of mental disorders, 4th edn (DSM-IV). American Psychiatric Association, Washington DC

Bleeker ML, Bolla Wilson K, Kawas C, Agnew J (1988) Age specific norms for the Mini-Mental State Examination. Neurology 38: 1565–1568

Bogerts B, Meertz E, Schönfeldt-Bausch R (1985) Basal ganglia and limbic system pathology in schizophrenia: a morphometic study of brain volume and shrinkage. Arch Gen Psychiatry 42: 784–791

Burns A, Jacoby R, Levy R (1991) Progression of cognitive impairment in Alzheimer's disease. J Am Geriatr Soc 39: 39–45

Burt DB, Zembar MJ, Niederehe G (1995) Depression and memory impairment: a meta-analysis of the association, its pattern, and specificity. Psychol Bull 117: 285–305

*Caine ED (1994) Should aging-associated memory decline be included in DSM-IV? In: Widiger TA, Frances AJ, Pincus HA, First MB, Ross R, Davis W (eds) DSM-IV sourcebook, vol 1. American Psychiatric Association, Washington, pp 329–337

**Christensen H, Henderson AS, Jorm AF, Mackinnon AJ, Scott R, Korten AE (1995) ICD-mild cognitive disorder: epidemiological evidence on its validity. Psychol Med 25: 105–120

Cohen R (1992) Probleme bei der Erfassung kognitiver Störungen bei endogenen Psychosen. In: Gaebel W, Laux G (eds) Biologische Psychiatrie Synopsis 1990/91. Springer, Berlin Heidelberg New York, pp 183–188

Cooper B, Bickel H, Schäufele M (1996) Early development and progression of dementing illness in the elderly: a general-practice based study. Psychol Med 26: 411–419

Copeland JRM, Davidson IA, Dewey ME et al (1992) Alzheimer's disease, other dementias, depression and pseudodementia: prevalence, incidence and three-year outcome in Liverpool. Br J Psychiatry 161: 230–239

Crook T, Bartus RT, Ferris SH, Whitehouse P, Cohen GD, Gershon S (1986) Age-associated memory impairment: proposed criteria and measures of clinical change. Dev Neuropsychol 2: 261–276

Crum RM, Anthony JC, Bassett SS, Folstein MF (1993) Population-based norms for the Mini-Mental State Examination by age and educational level. JAMA 269: 2386–2391

Drachman DA (1994) If we live long enough, will we all be demented? Neurology 44: 1563–1565

Feehan M, Knight RG, Partridge FM (1991) Cognitive complaint and test performance in elderly patients suffering depression or dementia. Int J Geriatr Psychiatry 6: 287–293

Folstein MF, Folstein SH, McHugh PR (1975) "Mini-Mental State": a practical method for grading the cognitive state for the clinician. J Psychiatr Res 12: 189–98

Gutierrez R, Atkinson JH, Grant I (1994) Mild neuro-cognitive disorder. In: Widinger TA, Frances AJ, Pincus HA, First MB, Ross R, Davis W (eds) DSM-IV sourcebook, vol 1. American Psychiatric Association, Washington, pp 287–317

Heaton RK, Pendleton MG (1981) Use of neuropsychological tests to predict adult patient's everyday functioning. J Consult Clin Psychol 49: 307–321

Henderson AS, Huppert FA (1984) The problem of mild dementia. Psychol Med 14:5–11

Horn JL (1982) The theory of fluid and crystallized intelligence in relation to concepts of cognitive psychology and aging in adulthood. In: Craik FIM, Trehub S (eds) Aging and cognitive processes. Plenum, New York, pp 237–278

Hughes CP, Berg, L, Danziger WL, Coben LA, Martin RL (1982) A new scale for the staging of dementia. Br J Psychiatry 140: 566–572

Kasten E, Eder R, Robra BP, Sabel BA (1997) Der Bedarf an ambulanter neuropsychologischer Behandlung. Z Neuropsychol 8: 72–85

Kiloh LG (1962) Pseudo-dementia. Acta Psychiatr Scand 37: 336–351

Koss E, Haxby JV, DeCarli C, Schapiro MB, Friedland RP, Rapoport SI (1991) Patterns of performance preservation and loss in healthy aging. Dev Neuropsychol 7: 99–113

Kral VA (1962) Senescent forgetfulness: benign and malignant. Can Med Assoc J 86:257–260

Kral VA (1982) Depressive Pseudodemenz und Senile Demenz vom Alzheimer-Typ. Nervenarzt 53: 284–286

Lauter H, Dame S (1991) Depressive disorders and dementia: the clinical view. Acta Psychiatr Scand Suppl 366: 40–46

*Levy R, Working Party of the International Psychogeriatric Association (1994) Aging-associated cognitive decline. Int Psychogeriatr 6: 63–68

Lezak MD (1995) Neuropsychological Assessment, 3rd edn. Oxford University Press, New York

Light LL (1988) Language and aging: competence versus performance. In: Birren JE, Bengtson VL (eds) Emergent theories of aging. Springer, Berlin Heidelberg New York, pp 177–213

Lindenberger U, Mayr U, Kliegl R (1993) Speed and intelligence in old age. Psychol Aging 8:156–164

Linn RT, Wolf PA, Bachman DL et al (1995) The 'preclinical phase' of probable Alzheimer's disease; a 13-year prospective study of the Framingham cohort. Arch Neurol 52: 485–490

Liu IY, LaCroix AZ, White LR, Kittner SJ, Wolf PA (1990) Cognitive impairment and mortality: a study of possible confounders. Am J Epidemiol 132: 136–143

Livingstone G, Hawkins A, Graham N, Blizard B, Mann A (1990) The Gospel Oak study: prevalence rates of dementia, depression and activity limitation among the elderly residents in inner London. Psychol Med 20: 137–146

Mitrushina M, Satz P (1995) Base rates of the WAIS-R inter subtests scatter and VIQ-PIQ discrepancy in normal elderly. J Clin Psychol 51: 70–78

Morris JC, Heyman A, Mohs RC et al (1989) The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). I: Clinical and neuropsychological assessment of Alzheimer's disease. Neurology 39: 1159–1165

O'Brien JT, Beats B, Hill K (1992) Do subjective memory complaints precede dementia? A 3 year follow up of patients presenting with supposed 'benign senescent forgetfulness'. Int J Geriatr Psychiatry 7: 481–486

O'Connor DW, Pollitt PA, Roth M, Brook CPB, Reiss BB (1990) Memory complaints and impairment in normal, depressed, and demented elderly persons identified in a community survey. Arch Gen Psychiatry 47: 224–227

Petersen RC, Smith GE, Ivnik RJ et al (1995) Apolipoprotein E status as a predictor of the development of Alzheimer's disease in memory-impaired individuals. JAMA 273: 1274–1278

Petersen RC, Smith GE, Waring SC, Ivnik RJ, Tangalos EG, Kokmen E (1999) Mild cognitive impairment – clinical characterization and outcome. Arch Neuro 56: 303–308

Plassman BL, Breitner JCS (1996) Apolipoprotein E and cognitive decline in Alzheimer's disease. Neurology 47: 317–320

Rediess S, Caine ED (1996) Aging, cognition, and DSM-IV. Aging Neuropsychol Cogn 3: 105–117

Reisberg B, Ferris SH, De Leon MJ, Crook T (1982) The global deterioration scale for assessment of primary degenerative dementia. Am J Psychiatry 139: 1136–1139

Reisberg V, Ferris SH, Shulman E et al (1986) Longitudinal course of normal ageing and progressive dementia of the Alzheimer's type: a prospective study of 106 subjects over 3.6 year mean interval. Progr Neuropsychopharmacol Biol Psychiatr 10: 571–578

Reischies FM (1993) Heterogeneity of the time course of cognitive performance of depressed patients. In: Bergener M, Belmaker RH, Tropper MS (eds) Psychopharmacotherapy for the

elderly – research and clinical implications. Springer, Berlin Heidelberg New York, pp 318–327

Reischies FM (1996) Normales Alter und leichte Demenz. In: Förstl H (ed) Lerhbuch der Gerontopsychiatrie. Enke, Stuttgart, pp 366–377

*Reischies FM, Lindenberger U (1996) Grenzen und Potentiale kognitiver Leistungen im hohen Alter. In: Mayer KU, Baltes PB (eds) Die Berliner Altersstudie. Akademie, Berlin, pp 351–377

Reischies FM, Schaub RT, Schlattmann P (1996) Normal ageing, impaired cognitive functioning, and senile dementia – a mixture distribution analysis. Psychol Med 26: 785–790

Ritchie K, Leibovici D, Ledesert B, Touchon J (1996) A typology of sub-clinical senescent cognitive disorder. Br J Psychiatry 168: 470–476

Robins LN, Locke BZ, Regier DA (1991) An overview of psychiatric disorders in America. In: Robins LN, Regier DA (eds) Psychiatric disorders in America. Free Press, New York, pp 328–366

Sachdev PS, Smith JS, Angus-Lepan H, Rodriguez P (1990) Pseudodementia twelve years on. J Neurol Neurosurg Psychiatry 53: 254–259

Salthouse TA (1985) A theory of cognitive aging. North Holland, Amsterdam

**Schaie KW (1989) Perceptual speed in adulthood: cross-sectional and longitudinal studies. Psychol Aging 4: 443–453

Schmand B, Lindebbom J, Launer L, Dinkgreve M, Hooijer C, Jonker C (1995) What is a significant score change on the Mini-Mental State Examination? Int J Geriatr Psychiatry 10: 411–414

Stevenson J, Richman N, Graham P (1985) Behavior problems and language abilities at three years and behavioral deviance at eight years. J Child Psychol Psychiatry 26: 215–230

Storandt M, Morris JC, Rubin EH, Coben LA, Berg L (1992) Progression of senile dementia of the Alzheimer type on a battery of psychometric tests. In: Bäckman L (ed) Memory function in dementia. Elsevier, Amsterdam, pp 207–226

Tewes U (1991) HAWIE-R, Hamburg-Wechsler Intelligenztest für Erwachsene, Revision 1991. Huber, Bern

Verband deutscher Rentenversicherungsträger (1995) VDR Statistik Rentenzugang. VDR, Frankfurt am Main

Weinberger DR, Berman KF, Zec RF (1986) Physiological dysfunction of the dorsolateral prefrontal cortex in schizophrenia. I: Regional blood flow evidence. Arch Gen Psychiatry 43: 114–124

WHO (1978) Mental disorders: glossary and guide to their classification in accordance with the 9th revision of the International Classification of Diseases. World Health Organisation, Geneva

WHO (1992) The ICD-10 classification of mental and behavioral disorders – clinical description and diagnostic guidelines. World Health Organisation, Geneva

*Zaudig M (1995) Demenz und "leichte kognitive Beeinträchtigung" im Alter. Huber, Bern

CHAPTER

11

L.G. Schmidt, H.J. Freyberger

Delirium, Amnesic Syndromes and Other Cognitive Disorders

Translator: S. Davison

1 Introduction

This chapter will discuss delirium, amnesic syndromes and other organic mental disorders characterised predominantly by cognitive dysfunction. These disorders are classified within the ICD-10 classification system under organic mental disorders (F0), except where they occur as a consequence of alcohol or substance misuse, in which case they are classified as mental and behavioural disorders due to psychoactive substance use (F1) (World Health Organisation 1992). The DSM-IV classification system uses the term amnestic disorder rather than amnesic syndrome (American Psychiatric Association 1994). Within DSM-IV, delirium, amnestic and other cognitive disorders, including dementia, are grouped within a single section, in which disorders are differentiated on the basis of their aetiology. The causes listed are general medical conditions, effects of a psychotropic substance (alcohol, drugs of abuse and medication side-effects are included under this heading) and multiple aetiologies.

2 Delirium

2.1 Symptomatology

The core symptom of delirium is a rapidly developing disturbance of consciousness accompanied by altered cognitive function. The term delirium was traditionally used within German psychiatry to describe a syndrome characterised by a disturbance of consciousness, manifest mainly as impaired wakefulness and orientation, in association with a number of other symptoms, namely, psychomotor agitation, florid perceptual disturbances such as visual hallucinations, and autonomic symptoms such as tremor, tachycardia and sweating. The disturbance of consciousness ranged along a continuum from reduced vigilance and fatigue through increased drowsiness and sleep to unconsciousness and coma. Disturbance of orientation likewise had a quantitative dimension ranging from a discrete disturbance of the ability to order time correctly through topographical uncertainty to a loss of orientation for person.

However, ICD-10 and DSM-IV have moved away from this narrow definition of delirium and broadened it to include those syndromes which previously were referred to as acute confusional states on the basis of disturbance of attention.

According to the clinical descriptions and diagnostic guidelines of ICD-10, delirium is described as follows: an aetiologically non-specific syndrome characterised by concurrent disturbances of consciousness and attention, perception, thinking, memory, psychomotor behaviour, emotion and the sleep–wake cycle. It may occur at any age but is most common after the age of 60 years. The delirious state is transient and of fluctuating intensity; most cases recover within 4 weeks or less. However, delirium lasting, with fluctuations, for up to six months is not uncommon, especially when arising in the course of chronic liver disease, carcinoma or subacute bacterial endocarditis. The distinction that is sometimes made between acute and subacute delirium is of little clinical relevance. The condition should be seen as a unitary syndrome of variable duration and severity ranging from mild to very severe. A delirious state may be superimposed on, or progress into, dementia (World Health Organisation 1992, pp. 57–58).

DSM-IV specifies the following diagnostic features of delirium:

- Criterion A: a disturbance of consciousness (i.e. reduced clarity of awareness of the environment) with reduced ability to focus, sustain or shift attention
- Criterion B: a change in cognition (such as memory deficit, disorientation, language disturbance that is not better accounted for by a pre-existing, established or evolving dementia
- Criterion C: develops over a short period of time (usually hours to days) and tends to fluctuate during the course of the day
- Criterion D: evidence from the history, physical examination or laboratory findings that the disturbance is caused by the direct physiological consequences of a general medical condition

DSM-IV describes disturbances of the sleep–wake cycle, psychomotor behaviour and emotions as associated features which are not essential to make the diagnosis. On the other hand, disturbances of the sleep–wake cycle and of psychomotor behaviour are considered essential diagnostic features in the ICD-10 research diagnostic criteria (World Health Organisation 1993). These criteria are particularly restrictive in their formulation in order to facilitate the grouping of homogeneous patient groups for research purposes. They emphasise the non-specificity of affective symptoms.

2.2 Epidemiology

Figures for the prevalence and incidence of delirium inevitably depend on the context in which they are measured. In childhood, delirium associated with

febrile illnesses predominates. In adulthood, delirium most commonly occurs as a result of alcohol withdrawal. In the elderly, delirium often presents within the context of dementing illnesses. In a review of existing studies, Lipowski (1990) concluded that about 10% of medical and surgical in-patients suffer from delirium at any one time. About 20% of patients in general hospitals that also admit elderly patients have been found to be suffering from delirium on admission or at some point during their stay. In surgical specialities, figures of 25%–50% have been reported. Because of this, delirium is one of the most common syndromes seen in neurology, general medicine, surgery and general practice (Hewer and Förstl 1994). These syndromes are commonly misdiagnosed or not recognised. Thus their prevalence tends to be underestimated.

2.3 Clinical Examination

Clinical assessment is based on three components: observation of behaviour, mental state examination and neuropsychiatric assessment. Any assessment must be repeated regularly because of the fluctuating course of delirium. A thorough assessment must also include taking a general history, including an account of the presenting complaints, and medical, developmental, family, social and personality histories.

A suspicion of delirium is justified if, on first meeting the patient, psychopathological symptoms of any sort appear to develop or change within a short space of time. This may be evident from a patient's speech and language, variable psychomotor speed, fluctuating affect or changeable social interactions.

The mental state examination must aim to elicit cognitive and affective abnormalities which are not immediately observable. Perceptual disturbances may present as symptoms of global cognitive dysfunction. These consist mainly of illusions, (visual) hallucinations and delusional percepts. Patients often maintain a delusional conviction that their hallucinations contradict reality and can react with intense emotion.

Finally, cognitive functions can be assessed in more detail by means of a neuropsychiatric assessment. Disturbance of consciousness and attention must each be observed accurately and repeatedly. For example, patients with impaired vigilance are noted to be easily fatigued. They generally show reduced interest in the outside world and are slower in their readiness to answer. Drowsy patients on the other hand keep falling asleep and can only be roused by strong external stimuli. The most severe degree of impaired vigilance is coma, in which the unconscious patient is no longer able to respond to stimuli from the outside world.

Impaired attention can also occur in wakeful patients. It is manifest as a decreased concentration span (e.g. questions have to be repeated to the patient), increased distractibility in response to external stimuli or perseveration, in which the patient repeats answers to earlier questions. It may sometimes be difficult or even impossible to conduct a conversation with the patient. Disorientation for time or place, but seldom for person, often presents as the first symptom.

Disorders of short-term memory are seen. However, abnormalities can extend to affect learning, encoding, recognition and recall of stored material. Memory is examined by giving the patient the names of several unrelated objects or a short sentence that they must repeat after about 5 min of distraction. Disorders may be observed in the comprehension of language and meaning. Language disturbance may also present as word-finding difficulties (nominal dysphasia) or writing disorders (dysgraphia). Sometimes speech is circumlocutory, loosened in its associations or even incoherent.

Delirium can be quantified by the use of standardised assessment instruments. The Mini-Mental State Examination (MMSE; Folstein et al. 1975) can be used to assess the severity of cognitive impairment. It covers in particular orientation, attention, higher speech, arithmetical and visuospatial functions. However, its use is limited because it does not directly measure the fluctuating vigilance that often characterises delirium. This can only be assessed indirectly via its effect on cognitive functions. There is less experience available in the use of the Delirium Rating Scale (Trezpacz et al. 1988) and the Confusion Assessment Method (Inouye et al. 1990), which have hitherto only been available in English. The Clinical Institute Withdrawal Assessment for Alcohol Scale (Sullivan et al. 1989) and the Mainz Alcohol Withdrawal Scale (Banger et al. 1992) are specialised instruments for the assessment of alcohol withdrawal.

Laboratory investigations are crucial in determining the aetiology of the delirium. Systemic or organ-specific diseases may be diagnosed by their characteristic test results. Routine investigations include full blood count and film, erythrocyte sedimentation rate (ESR), electrolytes (including serum calcium and phosphate), blood sugar, serum urea and creatinine, liver function tests and thyroid function tests. Where there is a clinical indication, more specialised investigations should be undertaken. These may include blood investigations, such as blood cultures, human immunodeficiency virus (HIV) and syphilis serology, and measurement of serum heavy metals, copper, caeruloplasmin, vitamin B_{12} and folic acid. They may include urine investigations, including culture and sensitivity, and screening for drugs and heavy metals. Finally, they may also include examination of

cerebrospinal fluid, including measurement of glucose and protein, cell count and cultures for bacteria, viruses and fungi. If intoxication is suspected, then blood or urine should be tested for alcohol and/or drugs, including medications.

Electroencephalography (EEG) is particularly important in the differential diagnosis of delirium. This non-invasive method can be used and repeated at any time. It has the advantage of high sensitivity, although specificity is lower. Evidence of a metabolic disorder (generalised EEG changes), a space-occupying lesion (focal changes) or complex partial epileptic seizures (spike-wave activity, 2–3/s) may be inferred from typical EEG abnormalities.

Computed tomography (CT) is particularly useful in the diagnosis of acute bleeds, whereas magnetic resonance imaging (MRI) is preferable for the investigation of pathological processes occurring within white and grey brain matter. However, the latter method may not produce definitive results in uncooperative or restless patients. Assessment of brain blood flow using single photon emission CT (SPECT) and of brain metabolism using positron emission tomography (PET) currently have no practical clinical use in the assessment of patients with delirium. However, these procedures should be used increasingly in research into the neurochemical and physiological bases of delirium.

2.4 Aetiology

Delirium is a manifestation of acute decompensation of cerebral functioning. It can usually be attributed to cerebral processes, general medical conditions, exogenous toxins or addictive withdrawal syndromes. The most common factors which affect cerebral functioning and may give rise to delirium are listed in Table 1. Older age has also been shown in many studies to be a general risk factor.

The most important aetiological medical disorders from a general medical point of view are cardiovascular disorders, respiratory disorders and infections. They affect central nervous system functioning via metabolic mechanisms with often unknown metabolic products or via hypoxic mechanisms in particularly sensitive brains. Similar pathogenetic processes are involved in primary renal, hepatic or endocrine metabolic disorders, gastrointestinal disorders and malnutrition. Delirium is not uncommon following heart and eye operations or after burns. Decreases in blood pressure, cardiac volume and partial pressure of oxygen are considered to be relevant in delirium occurring after cardiotomy; sensory deprivation and organic cerebral factors are thought to precipitate delirium following cataract operations; and loss of electrolytes and protein play a vital part in the decompensation of cerebral functioning following burns.

Table 1. Aetiological factors in delirium

Aetiology	Specification
Tumours	Cerebral
Trauma	Cerebral contusion Subdural haematoma
Infection	Cerebral (e.g. meningitis, encephalitis, syphilis, HIV) Systemic (e.g. sepsis, urinary tract infection, pneumonia)
Vascular	Cerebral (e.g. infarction, bleed, vasculitis) Cardiac (e.g. cardiac failure, shock)
Metabolic	Hypoxia, electrolyte disturbance, acidosis, alkalosis, liver or kidney failure, hypo- and hyperglycaemia, post-ictal phenomena Endocrine: thyroid or adrenal disorders Nutritional: thiamine or vitamin B_1 deficiency, pellagra
Intoxication	Poisons, heavy metals, therapeutic drugs
Withdrawal from substances	Alcohol, sedatives, hypnotics

HIV, human immunodeficiency virus.

As far as neurological syndromes are concerned, both pre-existing generalised brain disorders (such as vascular dementias, primary degenerative dementias or acquired immunodeficiency syndrome, AIDS) and localised disorders (such as tumours or trauma) may produce delirium. Delirium following brain injury usually presents in its typical form. However, psychopathological syndromes in the absence of pronounced disturbance of consciousness ("transit syndromes" as described by Wieck; see Scheid 1983, p. 64) may sometimes be encountered. Finally, in patients with multiple disorders and brain insults, different aetiological factors all act synergistically in precipitating delirium. In these patients, it is impossible to specify the relative importance of any single pathogenic factor. In all patients, the likelihood of delirious episodes increases sharply with increasing age and increasing severity of dementia.

Table 2 shows the medications which may precipitate delirium. Overdosage does not have to occur. Even at therapeutic doses, the individual sensitivity of a patient (which may often only first be recognised when delirium is precipitated) and a combination of medications (especially of psychotropic medications with anticholinergic side-effects) can both play crucial roles in the genesis of a delirium. Delirium is often due

Table 2. Medications with the potential to cause delirium

Type of medication	Drug
Antibiotics	Acyclovir Amphotericin B Cephalexin Chloroquine Chinolones Cycloserine
Drugs with anticholinergic effects	Tricyclic antidepressants Antihistamines Anti-Parkinsonian drugs (e.g. biperiden) Atropine Diphenhydramine Phenothiazines Clozapine Scopolamine Antispasmodics
Anticonvulsants	Phenobarbitone Phenytoin Sodium valproate
Anti-inflammatory agents	ACTH Corticosteroids Ibuprofen Indomethacin Phenylbutazone
Cytotoxic drugs	5-Fluorouracil
Antiparkinsonian agents	Amantidine Carbidopa Levodopa Bromocriptine
Anti-tuberculous drugs	Isoniazid Rimfampicin
Analgesics	Opiates Aspirin Synthetic opioids
Cardiac drugs	Beta blockers Clonidine Digoxin Lidocaine Methyldopa Quinidine Procainamide
Sedatives and hypnotics	Barbiturates Benzodiazepines Glutethimide
Sympathomimetics	Amphetamine Phenylephedrine
Miscellaneous	H2 blockers (e.g. cimetidine) Disulfiram Lithium Metronidazole Propylthiouracil Theophylline

ACTH, adrenocorticotrophic hormone.

to drugs with anticholingeric properties, but it can also occur after intake of L-dopa or amantadine, indicating dopaminergic mechanisms, or after benzodiazepines, indicating GABAergic mechanisms (which can also be observed after antibiotics of the chinolone type, which also interfere with GABAergic neurotransmission). Even glutamatergic mechanisms may be implied, as is indicated by cycloserine, which has properties of a glutamatergic agonist (Müller and Hartmann 1999). Delirium caused by medication is seen in about 1% of psychiatric in-patients treated with tricyclic antidepressants and medium-potency neuroleptics (phenothiazines). The risk of delirium is highest with clozapine (Schmidt et al. 1987).

Finally, alcoholic intoxication and withdrawal syndromes are often associated with delirium and are usually present with particularly prominent autonomic symptoms.

2.5 Clinical Presentation

A diagnosis of delirium should be suspected if there is a sudden onset or rapid fluctuation of psychiatric symptoms in patients with altered consciousness and prominent cognitive disturbances, particularly of attention and memory.

Prodromal phenomena such as anxiety, restlessness and irritability, so-called hyper-aesthetic emotional weakness disorders, as described by Bonhoeffer (1912, p. 37), may precede delirium. At the same time, lucid intervals become less frequent. Hyperactivity and hypoactivity alternate. Restlessness at night and tiredness during the day become more noticeable. The symptoms may stop at any time and resolve. This is then known as pre-delirium.

Full-blown delirium usually develops within a few hours or days, and the duration is usually cited as being anything from a few days to a few weeks. Alcohol withdrawal delirium most commonly begins 2–3 days after alcohol was last consumed, with symptoms reaching their maximal intensity after 4–5 days. In these patients, delirium is usually precipitated by a reduction or cessation of alcohol consumption, particularly in alcoholics who have been drinking excessively for 5–15 years and who develop a physical illness as a result of a road traffic accident, other trauma or an infection, for example.

Overall, the clinical course of delirium is characterised by great variability and fluctuation and is not usually predictable. The amount of alcohol consumption is a predictor of the severity of withdrawal. Exacerbations may occur at night and in unfamiliar surroundings because environmental cues can no longer be relied on for orientation.

The sense of time disintegrates and spatial orientation becomes uncertain. Retrograde and anterograde amnesia become apparent before orientation for person is lost. Confabulation may be observed in this phase.

Next, thinking becomes more loosely associated before becoming disjointed and incoherent. Illusions and visual hallucinations can occur and may be interpreted in a paranoid way. Unlike delusions seen in functional psychoses, those presenting in the context of delirium are typically moulded by the immediate situation and are less systematised and fully formed.

Affect is labile and very dependent on environmental stimuli. Some patients can seem emotionally indifferent and show diminished psychomotor activity. This is known as hypoactive delirium and is easily missed or misdiagnosed as depression. On the other hand, other patients are highly agitated and may display a tendency to become aggressive in response to anxiety or euphoria. This is referred to as hyperactive delirium. There is also a mixed variant. Thus a rapid change of the symptomatology may be a characteristic of a fully developed delirium. Aphasia, apraxia and agnosia may occur. In this context, they have no meaning in terms of cerebral localisation, as they arise as a result of generalised disturbances of cerebral functioning.

2.6 Neurobiological Background

Ever since the classical work by Engel and Romano (1959), it has been universally assumed that delirium arises out of a global disturbance of cerebral functioning. This is reflected in a slowing of background activity and an increase in the appearance of slow theta and delta waves on the EEG. A close correlation exists between slowing of background activity and the degree of disturbance of consciousness. Often, EEG analysis also reveals a flat, fast and irregular electrical picture. The earlier assumption that cerebral metabolism was universally decreased in delirium has so far not been confirmed. Despite the availability of functional imaging techniques such as SPECT and PET, there has hitherto been too little systematic research to enable the pathophysiology of delirium to be understood with regard to cerebral blood flow and oxygen consumption.

Functional neuroanatomical studies have shown that delirium may also occur in connection with focal lesions. In particular, disorders of the ascending reticular activating system in the brain stem may be associated with disturbed vigilance (reduced arousal). Severe disorders result in coma and unconsciousness. Lesions of medial brain structures such as the thalamus and hypothalamus have been blamed for changes in the level of consciousness (Wise and Brandt 1992). Furthermore, the posterior part of the left cerebral cortex is thought to be involved in conveying the contents of consciousness. It has been shown that disruption of the left posterior cerebral arterial supply to the temporo-occipital gyrus predisposes to delirium. Internal and cortical atrophy in the region of these association areas has frequently been demonstrated. It is assumed that the damage observed to these areas causes a separation of the neocortical association areas from the posterior hippocampus, thus disrupting the mobilising and sustaining of attention, thought to be a function of the non-dominant hemisphere.

The neurochemical changes associated with delirium are varied. Central to aetiology is a final common pathway characterised by disturbances of cerebral blood flow, oxygen consumption or glucose metabolism, resulting in a lack of energy-enriched phosphates in the CNS. Among other disturbances, this decrease leads to an imbalance of homeostasis, since the sodium-potassium ATPase is particularly vulnerable to disturbances of energy metabolism. These imbalances lead firstly to release of neurotransmitters due to increased calcium, which is particularly relevant for dopamine, noradreanaline and glutamate, and secondly to a decrease in reuptake of neurotransmitters, since active transport processes are driven by a gradient. Thus both mechanisms lead to an increase in neurotransmitters in the synaptic cleft. Interestingly, this is not true for acetylcholine, as synthesis and release of the neurotransmitter is reduced under hypoxia conditions (Müller and Hartmann 1999). Cholinergic neurons are also particularly vulnerable to alcohol. Thus post-mortem examination of brains of alcoholics have revealed a significant reduction (about 45%) in the number of cholinergic receptors (Freund and Ballinger 1988). These findings point to the importance of cholinergic neurons in the genesis of both delirium and dementia. Finally, the anticholinergic side-effects of different medications also suggest a similar mechanism in precipitating delirium. Noradrenergic overstimulation is responsible for the autonomic features of delirium and probably for the restlessness and agitation. Thus increased cerebral levels of noradrenaline have been found in patients with Parkinson's disease being treated with L-dopa. In addition, increased noradrenaline breakdown products (such as 3-methoxy-4-hydroxyphenylglycol, MHPG) have been found in the cerebrospinal fluid, blood and urine of patients with alcohol withdrawal delirium. Increased levels of corticotrophin-releasing hormone (CRH), adrenocorticotrophic hormone (ACTH) and cortisol are also found during alcohol withdrawal. These return to normal as the clinical symptoms subside, providing evidence of activation of the hypothalamo-pituitary-adrenal axis, the so-called stress axis (Hawley et al. 1994).

The psychotic symptoms such as hallucinations that are often observed in alcohol withdrawal may be related to dopamine binding to an increased number of receptors. This hyperdopaminergic condition might be a rebound phenomenon from the decreased numbers of dopaminergic receptors that have been demonstrated autoradiographically during chronic alcoholic intoxication. Likewise, reduced concentrations of the dopamine metabolite homovanillic acid (HVA) are found during withdrawal. The convulsions associated with complicated alcohol withdrawal are attributed in particular to the increased activity of the glutaminergic system as a result of increased numbers of *N*-methyl-D-aspartate (NMDA) receptors and of a reduced function of inhibitory mechanisms with slow restitution of γ-aminobutyric acid $(GABA)_A$ receptors (Rommelspacher et al. 1991). Finally, decreased levels of somatostatin and beta endorphin immunoreactivity have been found in the cerebrospinal fluid of delirious patients with differing causative illnesses (Koponen et al. 1994).

2.7 Differential Diagnosis

From a diagnostic point of view, delirium should be distinguished from other organic syndromes, in particular dementia. However, delirium and dementing syndromes can overlap. As a rule, the correct diagnosis depends on the history, the time course (e.g. whether it has an acute onset and whether it is progressive) and the results of laboratory investigations. Delirious patients are more likely to exhibit autonomic symptoms than those with dementia. Occasionally, a previously unrecognised dementing illness is diagnosed for the first time following an episode of delirium. Furthermore, delirium should be distinguished from psychotic disorders in which features of confusion may be present (e.g. confusion psychosis as described by Leonhard 1995, p. 72).

EEG can play an important role in distinguishing delirium from functional psychiatric disorders. A slowed frequency on the EEG can be a useful pointer towards a delirium, and a normal EEG more or less rules out the diagnosis.

2.8 Treatment

Delirium may precede other manifestations of a physical illness. Alternatively, it may mark the worsening of an already recognised clinical condition. Investigations must be carried out immediately to establish the cause of the delirium. However, this does not succeed in ascertaining the cause of the delirium in 20% of patients diagnosed with delirium (Lipowski 1990).

Alcohol withdrawal delirium is a potentially life-threatening condition. Death rates of about 20% used to be quoted where there were associated physical illnesses or multiple trauma. Newer statistics suggest a death rate closer to 1%. Overall, the mortality rates in patient groups with delirium are higher than those without delirium, whatever the underlying disease process. This is valid not only during the period of the illness, but also for a period afterwards. Death may be caused by a number of factors, including concurrent infections, fat embolisms, cardiac dysrhythmias, electrolyte imbalances, metabolic disturbances, fever or pancreatitis. Toxic delirium caused by anticholinergic medication is milder and without fatal complications. Whatever the cause, patients who develop delirium have, on average, longer hospital admissions than those who do not.

Management of delirium should be directed at treating the underlying medical condition. If there is pneumonia, then effective antibiotics and antipyretics should be administered. If blood pressure is pathologically elevated or there is papilloedema, anti-hypertensive treatment should be commenced immediately. In the case of an alcohol withdrawal delirium, thiamine therapy (100 mg i.m. daily) should be initiated along with an anti-delirium therapy. An appropriate first-line choice of treatment is chlormethiazole (as a liquid or in capsules). However, chlormethiazole should not be used in patients with multiple trauma because of the risk of respiratory depression associated with intravenous administration and of pneumonia as a result of the increased bronchial secretions. Such patients are therefore managed in intensive care settings with high doses of benzodiazepines, clonidine and haloperidol. In a delirium induced by medication, the responsible pharmacological agents should be stopped immediately. In cases where anticholinergics have caused the delirium, physostigmine is occasionally recommended as an antidote if stopping the medication is not sufficient on its own. The risk of serious side-effects such as cardiac dysrhythmias, bronchial asthma or seizures must always be considered.

In addition to these specific therapeutic approaches, there are a number of general measures which should be taken regardless of whether an underlying physical disorder is identified. Firstly, adequate nutrition and hydration should be provided with prophylaxis against embolisms and secondary infections. If necessary, there should be regular turning and early mobilisation as well as monitoring of vital parameters. Any electrolyte disturbances should be corrected.

In addition, the environment should be arranged such as to facilitate the patients' orientation. This might include placing a large clock, legible signs and

a calendar in their room and ensuring good lighting and a window for daylight. Irritation from loud and sudden noises and a lack of environmental stimuli should be avoided. As far as possible, it is desirable to have constancy of carers and nursing staff as well as regular contact with relatives. Emotional care should be reliable and language simple and not open to misunderstandings. Staff should be aware of the importance of avoiding the mismanagement of delirious patients.

The pharmacological treatment of delirium depends on the symptoms. A patient with a quiet, hypoactive delirium does not require any pharmacological intervention. However, in cases of intermittent or persistent psychomotor agitation, it may be necessary to sedate the patient. In order to do this, a medication should be chosen which does not compromise respiration, blood pressure or consciousness.

The high-potency butyrophenone neuroleptic haloperidol comes closest to meeting these requirements. A low dose (1 mg orally) should be chosen that can be administered several times a day if necessary. In cases of very high arousal, higher doses may also be administered, intravenously if necessary. However, the extrapyramidal side-effects then have to be dealt with, and these limit the use of haloperidol. Nevertheless, antimuscarinic agents should not automatically be administered, as they may exacerbate the delirium. Alternatively, benzodiazepines may be considered as they have advantages over the medium- and low-potency neuroleptics.

3 Amnesic Syndrome

3.1 Symptomatology

The key feature of the amnesic syndrome is the acquired impairment of the ability to store and retrieve new information in memory in clear consciousness. The ICD-10 clinical diagnostic guidelines define an organic amnesic syndrome as follows:

a syndrome of prominent impairment of recent and remote memory. While immediate recall is preserved, the ability to learn new material is markedly reduced and this results in anterograde amnesia and disorientation in time. Retrograde amnesia of varying intensity is also present but its extent may lessen over time if the underlying lesion or pathological process has a tendency to recover. Confabulation may be a marked feature but is not invariably present. Perception and other cognitive functions, including the intellect, are usually intact and provide a background against which the memory disturbance appears as particularly striking. The prognosis depends on the course of the underlying lesion (which typically affects the hypothalamic-diencephalic system or the hippocampal region); almost complete recovery is, in principle, possible (World Health Organisation 1992, p. 56).

DSM-IV uses the term amnestic disorder to refer to the same syndrome. It is defined as follows:

individuals with an amnestic disorder are impaired in their ability to learn new information or are unable to recall previously learned information (Diagnostic criterion A). The memory disorder must be sufficiently severe to cause marked impairment in social or occupational functioning and must represent a significant decline from a previous level of functioning (Diagnostic criterion B). The memory disturbance must not occur exclusively during the course of a delirium or dementia (Diagnostic criterion C) (American Psychiatric Association 1994, p. 156).

Criterion D states that there must be evidence from the history, physical examination or laboratory findings that the disturbance is the direct physiological consequence of a general medical condition (including physical trauma) for a diagnosis of amnestic disorder due to a general medical condition, or the persisting effects of a substance (e.g. a drug of abuse or a medication) for a diagnosis of substance-induced persisting amnestic disorder.

The DSM IV system then allows the amnestic disorder to be subclassified as transient if the memory disorder lasts for no more than 1 month or chronic if it lasts longer than 1 month.

3.2 Epidemiology

Amnesic syndromes are so rare that there are no reliable data to determine their point or lifetime prevalence. One recent study indicated that transient global amnesia may have an incidence of 5.2 cases per 100,000 population per year (Caine et al. 1995). It has been estimated that the Wernicke-Korsakoff syndrome accounts for 3% of all alcohol-related disorders (Franklin and Francis 1992).

3.3 Background: Learning and Memory

In order to facilitate the understanding of memory disorders, the function and organisation of normal memory must be described (Lezak 1995; Kandel et al. 1996; Squire and Alvarez 1995; Hodges and McCarthy 1995). According to current understanding, primary

memory is distinguished from secondary memory (see also Chap. 13, Vol. 1, Part 1).

Primary memory is also known as short-term memory and holds material for short time periods. It is also referred to as working memory, as it processes information cognitively. It is thought to consist of a central executive and of specialised verbal and visuospatial subsystems, in which the contents remain directly accessible in the conscious mind. Thus it enables problems to be solved and decisions to be reached about future behaviour. However, the capacity of primary memory is limited.

Secondary memory is the long-term memory. It is, in principle, unlimited in its capacity and can be further subdivided into explicit and implicit memory.

Explicit or declarative memory stores information about an individual's life in so-called episodic memory and about general knowledge in so-called semantic memory. The acquisition of these memory contents are dependent on their relating to primary memory. This enables cognitive processing such as conscious intake of information, evaluation, comparison and making inferences to occur. Next, incoming information is registered in the sensory store (in a matter of milliseconds). These processes are not yet the expression of actual memory function. Instead, they are related to those structures which process complex sensations. From here the information is taken into the working memory, where, depending on arousal, it is actively stored within seconds (immediate memory) and is associated with additional data in the relevant verbal or visuospatial subsystem within minutes. This storage is strengthened by rehearsal. These functions are associated with medial temporal lobes structures, including the hippocampus, entorhinal cortex (which communicates the majority of the incoming signals to the hippocampus), the subiculum (into which the majority of hippocampal projections project) and finally the parahippocampal cortical areas. Lesions in these areas disrupt the storage or coding of new information. The hippocampus is above all the area where contents are stored in the short term prior to transfer into long-term memory. The transfer of information to other brain areas, probably in the cortex, for long-term storage is known as consolidation. Impairment of these processes form the basis for anterograde amnesia. A particular neuronal representation in the inferotemporal cortex is responsible for the representation of faces. Lesions are associated with prosopagnosia. The contents of explicit memory are retrieved into consciousness by the conscious act of remembering something (retrieval, recall, decoding). If memory contents that were stored prior to brain damage can no longer be retrieved, this is known as a retrograde amnesia. This disorder is particularly associated with damage to the temporal neocortex. Recent PET imaging studies have shown that the right hippocampus is preferentially activated in particular priming processes, whereas the frontal lobes are activated in recall processes (Squire et al. 1992; Tulving et al. 1994).

Both the acquisition of memories into and retrieval of memories from so-called implicit (procedural) memory are more instinctual and automatic in nature. Neither are necessarily associated with conscious awareness or cognitive processes. This form of memory is built up slowly by many repetitions and is outwardly manifest by improved ability to perform different tasks. These include particular unconscious perceptual abilities ("priming"), motor skills and the learning of particular rules and ways of behaving, such as grammar. Implicit memory is built up by classical conditioning. The contents of implicit memory (such as learned vocabulary) can later be recalled automatically without requiring any particular effort. The implicit memory required for the solving of particular problems is tied to the sensory and motor systems involved in learning that task. Accordingly, the memory contents are stored within these functional systems. Implicit memory is not linked to the hippocampus. Instead, it is related to the cerebellum, amygdala and, in the case of very simple forms of learning, to the specific sensory and motor systems associated with the learning exercise. These memory functions are usually spared in patients with amnesia.

A few specific functional pathways relating to more complex forms of learning are understood, especially in relation to explicit memory (Kandel et al. 1996). A unique trisynaptic circuit in the hippocampus is activated by learning tasks in which storage is mediated largely by the phenomenon of long-term potentiation. This neuronal strengthening mechanism is excitatory in nature and based on glutaminergic transmission (via NMDA receptors) and the influx of calcium ions. It forms the basis for structural changes to the synapses and changes to neuronal gene expression. It can be postulated that long-term memory is based on the development of new synaptic connections (synaptic plasticity) and the biosynthesis of new proteins. These processes are thought to happen particularly during rapid eye movement (REM) or slow-wave sleep as part of consolidation. The molecular mechanisms underlying learning and memory are now being investigated with the aid of transgenic animals. In this way, it can be shown that the switching off of particular genes (gene knockout) leads to loss of long-term potentiation and disorders of spatial learning ability. It cannot yet be predicted whether drugs which modulate cholinergic functions and glutaminergic mechanisms will play an important role in the management of amnesic syndromes (Aigner 1995).

3.4 Clinical Examination

For clinical purposes, memory disorders are divided arbitrarily into immediate memory (i.e. ultrashort-term memory), recent memory and remote memory (Leszak 1995; Heilman and Valenstein 1993).

Immediate memory is activated when patients are asked to reproduce material consisting of about seven (plus or minus two) items. The digit span is an example of such an exercise. This falls within the patient's attention span and is therefore connected with registration. Performance in visual and verbal tasks of this sort may be differentially affected according to the disorder.

Recent memory is tested by asking patients to take in newly learnt material and repeat it after about 5–7 min of having their attention diverted by distracting stimuli. Anterograde amnesia is manifest as a disturbance of this type of memory whose processing occurred after the beginning of the illness.

Patients with amnesic syndromes are significantly disabled in their everyday functioning. Their learning and memory disorders can be observed in many ways. They cannot remember the names of staff or patients on the unit. They appear disoriented because they cannot memorise the area they are in or a particular route. However, there is often nothing remarkable about them intellectually during initial conversations or social interactions. The memory disorder only becomes evident when they are required to remember recent events. The impairment is also apparent when the patient is asked to repeat three words immediately to rule out perceptual or comprehension problems and is then unable or inadequately able to reproduce the three words after 5–7 min of intervening conversation.

When considering performance in everyday tasks, it appears to be particularly important to test ability to freely recall text that has been heard once. Patients who have difficulties in this task are subject to many limitations, e.g. in conversation, listening to the radio or watching television. The short-term ability to retain individual information can also be tested through free recall or recognition of complex geometrical patterns.

In addition to deficits of recent memory, amnesic patients often have disorders of remote memory. Remote memory is manifest as the ability to recall information from a long time ago, the storage of which preceded the onset of the illness causing the memory disorder. In this respect, so-called semantic memory includes knowledge related to language and so-called episodic memory includes autobiographical information.

In order to assess a retrograde amnesia, we can either ask about autobiographical details or about information in the public domain, such as historical events or political figures, e.g. former heads of state. Alternatively, patients can be presented with pictures of famous figures (Famous Faces Test) and asked for their names.

According to the Hodges and McCarthy (1995) cognitive model of memory, autobiographical memory is structured hierarchically. Planning processes are under frontal control. They relate to subordinate so-called thematic memory stores that are activated via thalamofrontal pathways. The corresponding verbal material from semantic memory is probably represented in the temporal and inferior parietal cortices. On the other hand, individual pieces of autobiographical information are stored in sensory and motor areas.

Three particular patterns of retrograde amnesia are recognised:

1. Time-limited disorders such as those seen in depressed patients treated with electroconvulsive therapy. These are restricted to one particular period of time.
2. Extensive disorders which extend beyond time-limited retrograde amnesia and affect contents that were acquired many years previously and had long been consolidated. Korsakoff's syndrome is an example of this type.
3. Global disorders which involve practically a complete loss of all memory contents, such as those seen in patients who have survived herpes encephalitis or who have advanced Huntington's disease. In these cases, extensive lesions have been found, in particular within the limbic system, but also in temporal, frontal, parietal and occipital cortical areas (Damasio and van Hoesen 1985; Heilman and Valenstein 1993).

Standardised tests allow a detailed assessment of the type and severity of the disorder. This can, however, only be achieved in the context of a comprehensive neuropsychological examination which assesses overall intellectual ability, language function, visuospatial skills, and motor and frontal "executive" functions, including self-assessment.

Patients with classical amnesia do not exhibit any gross deficits in the commonly used intelligence tests, such as the Wechsler Adult Intelligence Scale, or tests of concentration, as long as the information can be held within their actual attention span. On the other hand, amnesic patients can be identified using the Wechsler Memory Scale (WMS, with seven subtests; Wechsler 1945). The global memory quotient (MQ) that was derived from the subtests has been much criticised. However, the revised Wechsler Memory Scale (WMS-R) allows a differential diagnostic approach. For clinical purposes and in order to monitor the effectiveness of rehabilitative measures, it is

necessary to test the ability of patients to learn new material. Verbal learning tests are used to this end in which information is repeatedly presented and the improvement in learning after each stage is assessed. Examples of such tests include association tests using pairs of words or pairs of names and faces. The ability to learn and retain visual information, including simple and complex figures, can be assessed using the Benton Test (Benton 1968) or the *Diagnosticum für Cerebralschädigung* (Diagnosis for Cerebral Damage) (DCS; Weidlich and Lamberti 1980). Further details about neuropsychological assessment instruments and their uses can be found (in German) in von Cramon and Zihl (1988).

3.5 Aetiology

Amnesic syndromes are generally the consequences of pathological processes affecting, in particular, diencephalic structures (the thalamus and mamillary bodies), medial temporal lobe structures (the hippocampus and amygdala), the fornix, the basal forebrain and the prefrontal cortex.

Lesions are frequently bilateral, but one-sided lesions can also lead to deficits. Causes to consider include head trauma (blunt injuries and penetrating gunshot wounds), localised tumours, neurosurgical operations, infarcts in the area perfused by the posterior cerebral artery, hypoxia and encephalitis (especially due to herpes simplex).

Chronic alcohol misuse with associated thiamine deficiency is among the most common causes of amnesic syndromes. When amnesic syndromes are transient, they are usually attributable to cerebrovascular disease with changes in blood flow in the vertebro-basilar system, hypoxia (e.g. following cardiac arrest, resuscitation or suicide attempts), acute intoxication (e.g. during alcoholic blackouts, following ingestion of benzodiazepines, especially high-affinity substances such as triazolam, Van der Kroeff syndrome), poisoning (with carbon monoxide, isoniazid, lead and arsenic), cerebral convulsions (post-ictally or after electroconvulsive therapy) or other disorders which occur episodically.

3.6 Clinical Presentation

The inability to learn and reproduce new information (anterograde amnesia) or to remember past events that occurred before the onset of the brain damage (retrograde amnesia) is a central feature of an amnesic syndrome. Depending on the location of the lesion, the memory disorder may relate either to verbal or visual material. Patients are able to learn new motor skills without problem but are later unable to remember having done so. Transient and persistent disorders are distinguished from one another. However, there is some overlap. Retrograde amnesia following head trauma may abate with increasing recovery of the patient.

Transient global amnesia is a disorder with a sudden onset in which patients are unable to take in or reproduce any new information. Such patients are noticeable because, in the course of a normal conversation, they forget all information conveyed to them within a few minutes. In these circumstances, they repeat questions put to them. They appear to be aware of the disorder and present as baffled. As a result of typical questions such as " How did I get here?" and "What am I doing?", they can erroneously create the impression that they are suffering from an acute confusional state and are disoriented. However, communication and knowledge of personal identity remain unaffected, and more severe behavioural disturbances are not seen. The episode lasts for a period of hours, not longer than 1 day, and subsides suddenly. Afterwards, there are no detectable cognitive abnormalities other than the amnesic gap for the period of the episode. There is no specific single aetiology for the syndrome. Aetiological factors that have been described include epilepsy, emotional stress and migraine-related vasospasm. Transient disturbances of bilateral temporal blood flow to diencephalic and medio-temporal structures are the most likely cause. The syndrome typically occurs in later life (over 50). Episodes can re-occur.

Korsakoff's syndrome usually occurs as a consequence of many years of alcohol abuse in combination with nutritional deficits, in this case thiamine. It is occasionally seen in association with malabsorption syndromes. Often the syndrome starts with symptoms of delirium such as confusion and disorientation that may progress to stupor and coma. Typically, this is accompanied by eye movement disorders, ranging from nystagmus to complete ophthalmoplegia, as well as ataxia, particularly affecting the trunk. This clinical picture is called Wernicke's encephalopathy and it can develop into Korsakoff's syndrome. However, the onset of Korsakoff's syndrome can sometimes be insidious. The key characteristics of Korsakoff's syndrome are a persisting anterograde amnesia, often with severe disturbance of ability to remember things, a retrograde amnesia, in which the more recent the memory the greater the extent to which it is affected, a tendency to confabulate, particularly in the early stages and, in severe cases, disorientation. Not uncommonly, there are indications of frontal lobe involvement, such as apathy, poor judgement and a tendency to

perseverate, as well as of other functional disorders associated with neurotoxicity of the cerebellum or peripheral nervous system.

Amnesia following head trauma is characterised by the fact that the greatest functional deficits, both retrograde and anterograde, occur immediately after the trauma. Memory function can recover. Improvement is occasionally seen as long as 2 years after the trauma. Most commonly, some residual impairment persists in cases of severe injury.

Amnesia caused by cerebrovascular disorders usually presents suddenly, e.g. as a cerebrovascular accident, and can be attributed to occlusion of blood vessels in the area perfused by the posterior cerebral artery and branches to thalamic areas. Amnesia associated with disorders of the basal forebrain is usually caused by bleeding aneurysms of the anterior communicating artery or by surgical interventions in this area. Amnesic syndromes may also be caused by other cerebral processes, such as hypoxia, herpes simplex encephalitis and electroconvulsive therapy, as well as by limbic processes.

3.7 Neurobiological Background

Neuropathological research into Korsakoff's syndrome has demonstrated that both the memory disorder and disturbance of consciousness can be attributed to the damage observed to diencephalic structures, in particular of the ventral and lateral nuclei and the pulvinar of the thalamus, the mamillary bodies, the end of the fornix, the reticular activating system and the periaqueductal grey matter (Victor et al. 1971). In addition, there is diffuse cerebral damage (Lishman 1987). Such damage to the frontal lobes may account for the tendency to confabulate and the impaired judgement. Following herpes simplex encephalitis, hypoxia, abscesses or trauma, there are often extensive lesions to limbic and cortical structures.

The biochemical changes in amnesic disorder are thought to occur predominantly within the cholinergic system, as the synthesis of acetylcholine is particularly sensitive to hypoxia. Chronic toxic disorders of the noradrenergic and GABAergic systems have also been postulated. It is likely that disorders of the GABAergic system in the hippocampus occur as a result of the acute effect of certain high-affinity benzodiazepines such as triazolam. Where malnutrition also occurs, disorders of thiamine metabolism may be associated with genetically determined alterations in the activity of transketolase.

Neuroimaging techniques such as CT and MRI are able to demonstrate lesions in the medial temporal region, i.e. dilatation of the third ventricle and of the temporal horn of the lateral ventricle. Evidence can also be seen of cortical and subcortical atrophy. Angiography studies seldom provide evidence of disorders of perfusion, as these usually elude detection. SPECT studies have found evidence of reduced regional blood flow bilaterally in the temporal lobes and in the left thalamus and frontal cortex during transient global amnesia. EEG changes are usually not detectable even during a transient global amnesia.

3.8 Differential Diagnosis

The key features of amnesic syndromes, namely impairment of memory and learning, can occur to varying degrees and in varying contexts. Cognitive deficits are common in mild cognitive disorders such as in benign senescent forgetfulness or age-related cognitive decline. However, they are not severe impairments and are difficult to detect using quantitative measures (see Chap. 10, Vol. 2, Part 2). On the other hand, symptoms of amnesia can occur in the context of a number of other neuropsychiatric disorders, in particular dementia and delirium. When they present in association with impairment of consciousness and attention, the diagnosis is likely to be that of delirium.

Immediate memory is impaired in delirium but not in the amnesic syndrome. However, it can be difficult to distinguish the two, as the deficits in recent memory seen in patients with amnesia can make patients appear confused and disoriented. When amnesic symptoms present simultaneously with other neuropsychological symptoms such as aphasia, apraxia or agnosia, the diagnosis is likely to be that of dementia.

Many degenerative disorders, such as Pick's disease, Parkinson's disease and Huntington's disease, are associated with memory impairment. Patients with Alzheimer's disease are particularly likely to present initially with amnesic symptoms. Confabulation may be a feature of dementing disorders, delirium and amnesic syndromes. Because of the potential therapeutic benefit of an acute dose of thiamine in patients with Wernicke's encephalopathy, this diagnosis should always be considered. In patients with transient amnesic syndromes, the underlying cerebral processes should be investigated. If there is no evidence of a physical disorder, the possibility of a psychogenic origin should be considered.

Learning capacity and the reproduction of actual memory contents are usually not impaired in psychogenic amnesia. Instead, patients report that they cannot remember recently stored information or periods of their life. Such symptoms occur within the context of simulated amnesia of so-called multiple

personality disorder or in dissociative disorders, in which psychological trauma, intense affect and a particular excitement play a large role. Psychogenic amnesia may last days or weeks but seldom years. Episodes commonly end abruptly, usually in relation to experiences connected with the triggering event. Patients with psychogenic amnesia describe their memory difficulties in a very detailed and expressive manner. In the absence of any other severe cognitive deficits, the isolated forgetting of one's own name, occupation or address is only ever seen in psychogenic disorders.

3.9 Treatment

Where the underlying pathological cause of an amnesic syndrome is known or suspected, specific treatment of the pathology should be investigated without delay. This might include treatments such as thiamine, antiviral medication or aspirin.

There are currently no treatment measures that are reliably effective in reversing the memory deficits. Historically, a number of different rehabilitative approaches have developed using memory training in an effort to support patients in coping with everyday activities. However, these approaches have arisen from a pragmatic rather than a theoretical perspective (Heilmann and Valenstein 1993). Such methods include the repetitive exercising of memory with the help of games and simple exercises in the context of everyday clinical activity. However, there is nowadays scepticism about the likelihood of achieving a general improvement in memory functioning using this method.

On the other hand, internal memory aids are based on the use of mnemonic strategies. Such therapeutic approaches are based on the inclusion of mental imagery in memory training. The imagery serves to increase the memory capacity by linking verbal memory contents with visual engrams. The process has to be specially adapted, often with great effort, for each patient. This method appears to have proved useful in individual cases. However, further studies of its effectiveness are needed. Moreover, it is difficult to assess the extent to which learning is transferred from the laboratory to everyday situations. An alternative has been tried which consists of enabling patients who are sometimes able to learn particular procedural or sensorimotor skills to acquire a computer vocabulary with the help of a computer. In this way, successful learning can be quantified. Once again, the transfer to daily living is very limited.

External memory aids such as appointment diaries, shopping lists, individually programmable electronic notebooks have uses in appropriate cases (von Cramon and Zihl 1988). Finally, amnesic patients are often dependent on appropriate accommodation and comprehensive care from either relatives or medical staff.

4 Other Cognitive Disorders

Terms such as dementia, delirium or amnesia are often used categorically to describe illnesses. However, cognitive disorders are frequently observed in patients who do not fit the diagnostic criteria for dementia, delirium or amnesia. These cases are more easily conceptualised using a dimensional perspective of gradually decreasing cognitive function, as is seen in mild cognitive disorders (see Chap. 10, Vol. 2, Part 2). These disorders are grouped together in DSM-IV under the heading "cognitive disorders not otherwise specified".

5 References

*Aigner TG (1995) Pharmacology of memory: cholinergic-glutamatergic interactions. Curr Opin Neurobiol 5: 155–160

American Psychiatric Association (1994) The diagnostic and statistical manual of mental disorders, 4th edn. American Psychiatric Association, Washington, DC

Banger M, Benkert O, Röschke J, Herth T, Hebenstreit M, Philipp M, Aldenhoff JB (1992) Nimodipine in acute alcohol withdrawal state. J Psychiatr Res 26: 117–123

Benton AL (1968) Der Benton-Test. Huber, Berlin

Bonhoeffer K (1912) Die Psychosen im Gefolge von akuten Infektionen, Allgemeinerkrankungen und inneren Erkrankungen. In: Aschaffenburg G (ed) Handbuch der Psychiatrie. Deuticke, Leipzig (special part, Sect. 3, first half, pp 1–118)

Caine ED, Grossman H, Lyness JM (1995) Delirium, dementia, and amnestic and other cognitive disorders and mental disorders due to a general medical condition. In: Kaplan HI, Sadock BJ (eds) Comprehensive textbook of psychiatry, 6th edn, vol 1. Williams and Wilkins, Baltimore, pp 705–754

Damasio AR, van Hoesen GW (1985) The limbic system and the localization of herpes simplex encephalitis. J Neurol Neurosurg Psychiatry 48: 297–301

Engel GL, Romano J (1959) Delirium, a syndrome of cerebral insufficiency. J Chronic Dis 9: 260–277

Folstein MF, Folstein SE, McHugh PR (1975) Mini-Mental State: a practical method for grading the cognitive state for the clinician. J Psychiatr Res 12: 189–198

Franklin JE, Francis RJ (1992) Alcohol-induced organic mental disorders. In: Yudorfsky SC, Hales RE (eds) Textbook of neuropsychiatry. American Psychiatric Press, Washington, pp 563–583

Freund G, Ballinger WE (1988) Loss of cholinergic muscarinergic receptors in frontal cortex of alcohol abusers. Alcohol Clin Exp Res 12: 630–638

**Hawley RJ, Nemeroff CDB, Bissette G, Guidotti A, Rawlings R, Linnoila M (1994) Neurochemical correlates of sympathetic activation during severe alcohol withdrawal. Alcohol Clin Exp Res 18: 1312–1316

Hewer W, Förstl H (1994) Verwirrtheitszustände im höheren Lebensalter – eine aktuelle Übersicht. Psychiatr Prax 21: 131–138

Heilman KM, Valenstein E (1993) Clinical neuropsychology, 3rd edn. Oxford University Press, New York

Hodges JR, McCarthy RA (1995) Loss of remote memory: a cognitive neuropsychological perspective. Curr Opin Neurobiol 5: 178–183

Inouye S, van Dyck C, Alessi C (1990) Clarifying confusion: the confusion assessment method. Ann Intern Med 113: 941–948

Kandel ER, Schwartz JH, Jessell TM (eds) (1996) Neurowissenschaften. Spektrum, Heidelberg, pp 685–714

Koponen HJ, Leinonen E, Lepola U, Riekkinen PJ (1994) A long-term follow-up study of cerebrospinal fluid somatostatin in delirium. Acta Psychiatr Scand 89: 329–334

Leonhard K (1995) Aufteilung der endogenen Psychosen und ihre differenzierte Ätiologie, 7th edn. Thieme, Stuttgart

*Lezsak MD (1995) Neuropsychological assessment, 3rd edn. Oxford University Press, New York

Lipowski ZJ (1990) Delirium: acute confusional states. Oxford University Press, New York

Lishman WA (1987) Organic psychiatry, 2nd edn. Blackwell, Oxford

Müller WE, Hartmann H (1999) Pathogenese von Verwirrtheitszuständen. Münch Med Wochenschr 141: 84–87

Rommelspacher H, Schmidt LG, Helmchen H (1991) Pathobiochemie und Pharmakotherapie des Alkoholentzugssyndroms. Nervenarzt 62: 649—657

Scheid W (1983) Lehrbuch der Neurologie, 5th edn. Thieme, Stuttgart

Schmidt LG, Grohmann R, Strauss A, Spiess-Kiefer C, Lindmeier D, Müller-Oerlinghausen B (1987) Epidemiology of toxic delirium due to psychotropic drugs in psychiatric hospitals. Compr Psychiatry 28: 242–249

Shallice T (1988) From neuropsychology to mental structure. Cambridge Univ Press, New York

**Squire LR, Alvarez P (1995) Retrograde amnesia and memory consolidation: a neurobiological perspective. Curr Opin Neurobiol 5: 169–177

Squire LR, Ojemann JG, Miezin FM, Petersen SE, Videen TO, Raichle ME (1992) Activation of the hippocampus in normal humans: a functional anatomical study of memory. Proc Natl Acad Sci USA 98: 1837–1841

Sullivan JT, Sykora K, Schneiderman J, Naranjo CA, Sellers EM (1989) Assessment of alcohol withdrawal: the revised Clinical Institute Withdrawal Assessment for Alcohol Scale (CIWA-Ar). Br J Addict 84: 1353–1357

Trezpacz P, Baker R, Greenhouse J (1988) A symptom rating scale for delirium. Psychiatry Res 23: 89–97

*Tulving E, Kapur S, Craik FIM, Moscovitch M, Houle S (1994) Hemispheric encoding/retrieval asymmetry in episodic memory: positron emission tomography findings. Proc Natl Acad Sci USA 91: 2016–2020

*Victor W, Adams RD, Collins GH (1971) The Wernicke Korsakoff Syndrome. Blackwell, Oxford

von Cramon D, Zihl J (1988) Neuropsychologische Rehabilitation. Springer, Berlin Heidelberg New York

Wechsler DA (1945) A standardized memory scale for clinical use. J Psychol 19: 87–95

Weidlich S, Lamberti G (1980) DCS. Diagnosticum für Cerebralschädigung, 2nd edn. Huber, Bern

Wise MG, Brandt GT (1992) Delirium. In: Yudorfsky SC, Hales RE (eds) Textbook of neuropsychiatry. American Psychiatric Press, Washington, pp 291–310

World Health Organisation (1992) The ICD-10 classification of mental and behavioural disorders. Clinical descriptions and diagnostic guidelines. World Health Organisation, Geneva

World Health Organisation (1993) The ICD-10 classification of mental and behavioural disorders: diagnostic criteria for research. World Health Organisation, Geneva

CHAPTER 12

H.J. Freyberger, L.G. Schmidt

Organic Personality Changes

Translator: E. Taub

1 Development of Terminology

The traditional concept of "organic personality change" implies a lasting and irreversible change of personality structure, behavior, and affect in consequence of changes in the brain of whatever origin. In its clinical application, the term "organic" in this context refers to psychopathological manifestations produced by a morbid disturbance of brain function. Kraepelin (1909) and other authors validated this concept largely in the context of a single construct, that of the so-called epileptic personality change. According to the classical understanding, the epileptic personality change is characterized by (among other features) viscous, perseverative, circumstantial, and diffuse thinking; loss of initiative; certain affective abnormalities; and a dedifferentiation, or leveling out, of premorbid personality traits.

In Anglo-American psychiatry, much attention was paid to organic personality changes in the late nineteenth and early twentieth centuries in connection with research on the end stage of neurosyphilis, the epidemic of encephalitis in 1918, and the many traumatic brain injuries sustained in the course of the First World War (Popkin 1986). Many authors writing in German, including Huber, Scheid, Bleuler, von Baeyer, and Häfner, contributed to the conceptualization of this group of disorders, mainly through qualitative research methods.

Thus Huber (1972) divided the organic psychoses into acute forms (exogenous reaction types), transitional syndromes, and syndromes of chronic, irreversible form. He further subdivided the latter category into pseudoneurasthenic syndromes (without major impairment of intellect or memory, but with "irritable weakness") and "encephalopathies" (with "formal preservation of personality"). Huber denoted disturbances of these types as "organic personality changes" and stated that they are characterized by alterations of the dynamic components of personality (affective responsiveness and baseline affect, psychomotor speed, adaptive ability, formal volitional structure) without any change in the substance of the patient's character. In his conception, the impairment of brain functioning characteristic of dementia is absent in these disorders. In his later works, Huber (1976) stressed that brain damage incurred in early childhood is followed most frequently not by cerebral palsy or feeble-mindedness, but by so-called pseudopsychopathic syndromes, which manifest themselves in the affective rather than the intellectual sphere. This "exogenous psychosyndrome of early childhood," to use Harbauer's term, could then set the stage for neurotic developmental abnormalities.

According to Bleuler (1966), the local cerebral psychosyndrome is a "special form of an isolated organic personality change" that is associated with "episodic or persistent changes of mood, motivation, and individual vital drives." Local cerebral psychosyndromes arising on the basis of mild, diffuse brain atrophy are difficult to classify into frontal, diencephalic, and temporal psychosyndromes. Characteristic, but nonetheless nonspecific, psychopathologic features of these subtypes may be made out only in particularly marked individual cases, such as the "lack of spontaneity with preserved responsiveness to external stimuli" seen in frontal abulia.

Scheid (1980) described three types of "irreversible defect syndrome." One of these types was early childhood brain injury associated with a deficiency of personality and intelligence. With reference to Kurt Schneider's concept of "organic personality change," he further described a syndrome of deterioration of the personality and mental abilities, which he divided into two forms, mild (reduction of spontaneity, vitality, and initiative, or a tapering off of the personality) and severe (leveling off of the personality).

Although the construct of organic personality change, as described above, figures importantly in the European literature, particularly in works written in German, empirical studies have been rather rare, and those that have been performed have been fraught with a number of serious methodological problems. The predominant objects of study were clinical populations, which are inherently laden with a high prevalence of personality disorders or accentuations. The diagnostic problem of distinguishing personality disorders from organic personality changes remains largely unsolved, and the prominent position given to organic personality changes in the more recent classification systems – a throwback to the classical diagnostic rules – therefore has in fact no empirical justification.

There are as yet no large prospective studies on the effect of brain damage in early childhood on the later occurrence of personality disorders. Nevertheless, many authors stress its importance as a risk factor. Similarly, little empirical clarity exists concerning the relevance of the stage of cerebral development at the time the lesion is produced, e.g. after brain trauma in early childhood.

Research on organic personality changes has been performed almost exclusively in retrospective, cross-sectional studies employing instruments of assessment that, in part, fail to satisfy minimal test-theoretical requirements. Empirical longitudinal studies are almost entirely lacking. In the particular case of epileptic personality change, the type that has been most

thoroughly studied by far, the findings to date have been partly inconsistent, as will be shown below.

Despite these limitations regarding both content and method, the authors of both the ICD-10 (Dilling et al. 1994) and the DSM-IV (APA 1994) decided to incorporate some of the categories just discussed into their operational diagnostic manuals, not least because of the clinical obviousness of organic personality changes.

2 Symptomatology and Terminology in Recent Classification Systems

As Lauter (1988) pointed out, proper use of the current terminology, as employed in the recent classification systems, requires attention not only to the psychopathological state at the time of assessment, but also to a number of other important features. These include, among others, the course and prognosis of disease, the localization of injury, the degree of severity, the age of onset, and underlying physical conditions.

With regard to disease course, it should be emphasized that personality changes, encephalopathies, and so-called pseudoneurasthenic syndromes may all occur without dementia, in the sense of a deficiency of cerebral function. As for prognosis, dementias, organic personality changes, and pseudoneurasthenic syndromes were considered irreversible and were thereby held to be distinct from reversible organic psychoses, at least in the German-speaking countries. Although these prognostic principles worked out by Scheid and Wieck, among others, still possess a certain validity today, they must increasingly be modified in the light of new concepts of neural plasticity and neuronal regeneration (Spitzer and Casas 1997).

As for the localization of injury, Bleuler recognized that it bears little relevance to the type and severity of the clinical syndrome, at least in milder disturbances. Localization becomes significant only when the disturbance is massive, in which case certain marked neuropsychological syndromes also appear (e.g. the syndromes of the frontal, parietal, and temporal lobes). Previous classifications of organic psychic disorders often considered the degree of severity of cognitive functional impairment an important element of differential diagnosis, so that the diagnosis of dementia, for example, was reserved for particularly severe and irreversible forms. Today, however, the concept of dementia is broader. Furthermore, with regard to the age of onset of disease, it appears to be important whether an injury, of whatever type it may be, takes place in a developing or a mature brain. Concepts such as that of the exogenous psychosyndrome of early childhood, or of minimal cerebral dysfunction, are partly subsumed today under the developmental disorders, even though their significance as predisposing factors for the later development of (for example) complex, primarily nonorganic personality disorders, mainly of antisocial type, seems to be beyond doubt (Herpertz and Sass 1997). Much more research is required in this area if the organic contribution to the etiology of these disturbances is to be clarified.

As for the relevance of underlying physical conditions, Lauter (1988) and others have emphasized that the term "organic" does not imply that the psychopathological manifestations are determined exclusively by organic factors; rather, they may also be influenced by other moderating variables.

If we return to a consideration of the recent classification systems, it becomes quite clear that many of the traditional concepts of German-speaking authors, in particular, have been abandoned and replaced by social measures of assessment and internationally acceptable demands for reliability and validity. Nonetheless, even the categorical constructs chosen in these systems do not do justice to the dimensional character of personality changes. Furthermore, they largely neglect the possibility of variation of personality.

Both ICD-10 and DSM-IV posit the following basic requirements for the diagnosis of an organic personality disorder:

- Objective evidence or historical documentation of a cerebral illness, injury, or functional disturbance
- No impairment of consciousness and absence of a severe memory deficit
- Exclusion of other premorbid mental illnesses, including personality disturbances, as possible causes of the disorder

According to the Clinical Diagnostic Guidelines and Research Criteria of ICD-10, organic personality disorder (F07.0, Appendix A) is characterized by a major change from the patient's premorbid behavior, with especially marked involvement of the expression of affect, needs, and impulses. Cognitive functions may be impaired mainly with regard to the planning of activities and the anticipation of their individual and social consequences, as in the so-called frontal lobe syndrome. The Guidelines and the Research Criteria require the fulfillment of two or three diagnostic criteria, respectively, from a list of criteria involving the areas of endurance and gratification of needs, affectivity, impulse control, cognitive disturbances, language and speech, and sexual behavior (see Appendix A).

Furthermore, the post-encephalitic syndrome (F07.1) and the organic psychosyndrome after traumatic

brain injury (F07.2) are listed as further specific disorders to be classified under the organic behavioral changes. The post-encephalitic syndrome is said to be distinguished from organic personality change by being largely reversible, although no corresponding temporal intervals or other specific diagnostic criteria are given here (Appendix B). Similarly, a number of rather non-specific criteria are given for the organic psychosyndrome after traumatic brain injury (see Appendix B).

DSM-IV rejects a differentiation of this type and subsumes organic personality change under category 310.1, personality change because of a medical disease factor (Appendix C). Aside from characteristics appearing in the catalogue of criteria that have to do with etiology (criterion B) and with the diagnostic differentiation from other entities (criteria C and D), DSM-IV places emphasis on the personality disturbance (criterion A) and on the limitation of psychosocial functioning (criterion E). In a manner almost analogous to the ICD-10 criteria, the personality change is characterized by affective instability, inadequate impulse control, sudden, inappropriate outbursts of aggression or rage, apathy, suspiciousness, and paranoid ideation. Unlike ICD-10, DSM-IV provides a subclassification based on the most prominent clinical manifestations, which can then be considered in the context of the type and localization of the underlying pathological processes.

3 Epidemiology

Reliable epidemiological data on organic personality changes derived from representative population samples are not available. Published clinical studies do, however, contain estimates of the frequency of organic personality changes associated with certain underlying illnesses. Thus, depending on the population sampled, 8%–60% of patients with long-standing epilepsy will develop personality changes, and patients with temporal lobe epilepsy seem to be at highest risk (Taylor 1987; Mendez 1988, 1995). Motomura et al. (1988) report, in a study employing the DSM-III-R criteria, that approximately 33% of a consecutive series of patients whom they studied with right hemispheric injury later developed secondary organic personality changes. The occurrence of personality changes in this study, as in others, had a more than random association with neglect syndromes, anosognosia, extinction phenomena, constructive apraxia, and motor disturbances. The extent of the clinical manifestations of post-traumatic personality changes was correlated with the severity of the causative brain trauma (Levin et al. 1979).

A similar dependence of basic epidemiological data on the degree of severity of disease, in addition to considerable interindividual variation in disease course, complicates the provision of exact prevalence figures for organic personality disorder in the context of a number of other relevant illnesses, including human immunodeficiency virus (HIV) infection and other infectious diseases that affect the cerebrovascular system, endocrine diseases, and autoimmune diseases (Perkins et al. 1993; Popkin and Tucker 1994). In these diseases, clinical neurological findings, or other findings made with additional laboratory tests, are often not associated with the presence or the degree of severity of personality changes. Furthermore, questions of comorbidity, e.g. with disorders brought about by substance abuse, have been studied inadequately or not at all.

4 Techniques of Assessment

Organic personality disorders are difficult to diagnose, principally because they are complex disturbances of regulatory or integrative functions that may also be viewed as partial impairments of performance in the area of higher cognitive functioning. Beyond the characterization of syndromes and etiological and pathogenetic considerations, the diagnostic process must also make use of different sources of information as well as of longitudinal observation.

Alongside the determination of psychopathological findings and neurological and psychiatric diagnosis, the observation of behavior plays a particularly important role in clinical assessment, because the deficiencies of the patient's personality can often be diagnosed only in the context of a complex interaction. Some of the most important features of organic personality change, such as a disturbance of impulse control, an inability to plan one's activities in a goal-directed fashion, and inappropriate behavior in social relationships, may often be recognizable only in the framework of complex behavioral processes, while disturbances of affectivity and of the form and content of the patient's thinking are subjects for psychiatric exploration.

Furthermore, in order to render the diagnosis more certain, historical data should be obtained from informants other than the patient, so that it can be documented that the patient's problem indeed represents a change from an intact premorbid personality. Characterization of the personality change as persistent is possible only with an adequate historical or prospective period of observation; these means provide the otherwise lacking longitudinal perspective.

Structured or standardized diagnostic instruments may be very useful in this process. To be sure, there are no specific instruments for the diagnosis of organic personality changes at present, but a number of screening questionnaires and diagnostic interviews have been developed and empirically tested in the area of personality changes of nonorganic etiology, and these can be used to advantage in the provision of a diagnosis. With regard to screening questionnaires, the Personality Diagnostic Questionnaire (PDQ-R) deserves mention as a self-assessment scale of comparatively high specificity, but low sensitivity. As for structured diagnostic interview techniques, the International Personality Disorder Examination (IPDE; World Health Organization 1994), which is designed for the diagnosis of all personality disorders according to ICD-10 and DSM-IV, is most useful. A further interview technique, the Structured Clinical Interview for DSM-IV Personality Disorders (SCID-II), has been developed for DSM-IV (for a brief description, see Stieglitz and Baumann 1994).

Many mostly Anglo-American studies on the diagnosis of organic personality changes have made use of the Minnesota Multiphasic Personality Inventory (MMPI), which, however, has been repeatedly criticized on test-theoretical grounds. An instrument similar to the MMPI, the Bear-Fedio-Inventory (Bear and Fedio 1977), was developed to assess personality changes in the context of epilepsy. This instrument has not been sufficiently studied from the test-theoretical viewpoint.

The so-called frontal lobe tests in current use (Wisconsin Card Sorting Tests, Continuous Performance Tests, N-Back Tests) have no established role in routine psychiatric diagnosis.

An etiological diagnosis is provided on the basis of the primary medical condition that led to the organic personality change. Laboratory studies and imaging techniques, in the widest sense of term, may be used for this purpose. Electroencephalography (EEG) may reveal generalized abnormalities, focal signs, and epileptiform abnormalities, which, generally speaking, may be due to more than one type of underlying illness. Computed tomography (CT) may reveal acute hemorrhages and their residua as well as other structural defects, while magnetic resonance imaging (MRI) more sensitively detects pathological processes in the gray and white matter of the brain, particularly in the frontal area, which is of special relevance to personality disorders.

The assessment of cerebral perfusion with single photon emission computed tomography (SPECT) and of cerebral metabolism with positron emission tomography (PET) has found application mainly in scientific research; SPECT, in particular, may yield valuable diagnostic information about regional hypoperfusion, most often in the frontal lobes. The sensitivity of these techniques needs to be improved if they are to provide a more precise understanding of the "organically damaged background" than we have from the techniques available today (for a further discussion, see Chap. 11, Vol. 1, Part 1).

5 Etiology

Organic personality changes should be considered in the context of the primary medical conditions that cause them. We shall discuss a number of especially important conditions in this section.

5.1 Epileptic Personality Change

The construct of a characteristic epileptic personality change is a subject of controversy in the literature (Hermann and Whitman 1984; Popkin and Tucker 1994). Taylor (1987) and Mendez (1988, 1995) found that complex partial seizures and temporal lobe epilepsy, at least, are statistically significantly associated with signs of an organic personality change. Mendez et al. (1993) compared 42 epileptic patients who suffered from personality changes with epileptic controls who did not suffer from them and found that the group with personality changes had auras significantly more frequently, and generalized seizures significantly less frequently. No personality profile typical of epileptic patients has yet been revealed by psychological testing (Mendez 1995).

Borderline personality disorders or accentuations, whose characteristic features are instability of identity and impulsiveness, seem to occur most frequently in epileptic patients and are not uncommonly considered in the context of dissociative phenomena that may be traced back to an actual trauma in childhood. It must also be pointed out in this connection that antiepileptic pharmacotherapy, which is often continued for many years, may lead to functional impairment, particularly in the realm of cognition (Lang and Stefan 1990); thus an additional variable affecting the course of disease must be kept in mind.

In a number of studies, right-sided epileptic foci were found to be more strongly associated with affective thought disorders, extraversion, and positive self-esteem, while left-sided foci were more strongly associated with paranoid thought disorders and socially withdrawn behavior. Cummings (1985) found increased interest in religious and philosophical matters, graphomania, lessened sexual responsiveness, circumstantiality, and perseverative thinking.

5.2 Traumatic Brain Injury

A personality accentuation with features of a depressive syndrome, among which apathy and reduced emotional variability are the most prominent, has been repeatedly described in the context of traumatic brain injuries, particularly those involving the dorsomedial portion of the frontal lobe (Capruso and Levin 1995). Personality accentuations of a rather more pseudopsychopathic type, including egocentricity, disinhibition, and sexual deviance, are often attributed to lesions of the orbital frontal lobe. In the view of most authors, the likelihood of later development of an organic personality change rises in nearly linear fashion with the severity of the traumatic brain injury. Personality changes after traumatic brain injury must be distinguished from the anxiety disorders and depressive disorders that also occur more frequently in this group of patients (Starkstein et al. 1990).

Eames (1997) adds that traumatic brain injuries, and especially the so-called minor head injuries, which probably occur frequently in young adults, may be mistaken for organic disorders and then given a different diagnostic label (post-traumatic stress disorder, hypomanic or manic episodes, obsessive–compulsive disorder).

5.3 Frontal Lobe Injury and Frontal Lobe Lesions

Benson (1984) and Lishman (1987) emphasize in their review articles that frontal lobe injuries or lesions may be associated with personality changes, manifesting themselves primarily as changes in initiative and disturbances of the cognitive control of affective reactions. Moreover, disturbances of impulse control have a statistically significant association with frontal lobe lesions.

Three types of frontal lobe disorder may be distinguished on the basis of the location of injury (Cummings 1993):

1. The dorsolateral prefrontal brain syndrome is characterized by disturbance of conceptual recognition, lack of cognitive flexibility, inadequate cognitive planning, and disturbance of initiative.
2. The orbital frontal brain syndrome is characterized by loss of distance, indifference, lack of social control, attention deficit, and a tendency toward imitative behavior.
3. The anterior cingulate syndrome ranges from apathy, indifference, and perseveration to akinetic mutism.

Furthermore, the affective sequelae of frontal lobe lesions vary depending on the side of the injury. While left frontal lesions have a more than random association with depressive manifestations, right frontal injuries are more frequently associated with manic syndromes (Starkstein and Robinson 1991). Frontopolar lesions are associated with marked personality changes, even though there is frequently no cognitive deficit; personality changes may be absent when the lesion is more posterior (Eslinger and Dimasio 1985).

Sciutella and Feinberg (1997) refer to the close association of frontal lobe disturbances with impulsive behavior, which, in a number of studies, was found to be related to psychopathological features of obsessive–compulsive disorder and of aggressive and violent behavior. Herpertz and Sass (1997) discuss frontal lobe lesions as a predisposing factor for various disorders, particularly antisocial personality disorder.

5.4 Right and Left Hemispheric Brain Damage

Motomura et al. (1988), in the article cited above, reported a statistically significant association of organic personality changes with right brain damage, but these findings were not confirmed by longitudinal studies. An association with residual neurologic abnormalities was also revealed in numerous studies of affective changes after right and left brain damage (see, e.g. Starkstein and Robinson 1989), but methodologically adequate comorbidity studies are not yet available. The literature to date unequivocally shows that from 20% to 60% of clinically treated brain-damaged patients develop depressive disorders or, somewhat less frequently, anxiety disorders, in the postacute phase. In the setting of the cognitive deficits that are also present, these disorders may be clinically indistinguishable from organic personality changes (Castillo and Robinson 1994). Well over 50% of these disorders are no longer demonstrable on follow-up 1 year after injury (Steller and Schultz-Venrath 1995). Patients with right brain damage seem to suffer from depressive syndromes, in the broadest sense, approximately as often as patients with left brain damage, while manic syndromes have been observed to date only after right brain lesions (House et al. 1990).

5.5 Personality Changes as Early Manifestations of Dementing Disease

Many authors state emphatically that personality changes may appear as early manifestations of a dementing illness. Thus Heston and White (1983)

consider very mild personality changes as being among the early manifestations of Alzheimer's disease. Unlike the frontal brain degenerations, however, Alzheimer's disease is rarely accompanied by personality changes that impair psychosocial functioning (Cummings and Benson 1992). The course of illness is characterized by complex thought disturbances and, on the behavioral level, first by apathy and agitation, and later by depressed mood, irritability, and psychomotor restlessness (Mega et al. 1996).

In contrast to the situation in Alzheimer's disease, personality changes play an extremely important role in the focal brain degenerations, of which the most important type is frontotemporal degeneration (so-called frontal lobe dementia), which also includes Pick's disease. The characteristic findings here are said to include disinhibition, diminution of social adaptive ability and self-criticism, quarrelsomeness, impulsiveness, and a tendency toward neglect (Miller et al. 1991). This syndrome is contrasted in the literature with an apathetic form corresponding more closely to the dorsolateral prefrontal brain syndrome or to the anterior cingulate syndrome and is characterized by social isolation, loss of interests, and self-neglect (see Cummings and Benson 1992). A form of frontotemporal degeneration that spreads to involve the amygdala and surrounding cortex on both sides is clinically manifested as the Klüver–Bucy syndrome, which may also appear in the setting of herpes simplex encephalitis or after bilateral medial temporal infarction (Poeck 1985). Patients are emotionally indifferent and lacking in initiative, and they tend to put objects of practically any kind in their mouths; they also display hypersexual behavior, react immediately to external stimuli, and have visual agnosia.

Marotta and Perry (1989) showed that HIV-positive patients develop personality changes long before dementia becomes manifest. Perkins et al. (1993) showed, in a study of patients and HIV-negative controls, that HIV-positive homosexual men suffer from personality disorders more often than would be expected by chance. In view of the high degree of comorbidity of various forms of dementia and depressive disorders in this group of patients, further longitudinal studies of personality changes are clearly needed.

Patients suffering from a parkinsonian syndrome have approximately a twofold risk of developing dementia compared to healthy persons of the same age (Marder et al. 1995). Even in the early stage of Parkinson's disease, the patients have disturbances of executive functioning with impairment of the ability to plan and act and of initiative; these disturbances, if they persist, create the impression of personality changes (Levin et al. 1989). In other diseases associated with the development of dementia, such as Steele–Richardson–Olszewski syndrome and Huntington's disease, marked personality changes are seen before dementia becomes manifest (Cummings and Benson 1992).

6 Therapeutic Interventions

Therapeutic interventions must be directed, above all, against the underlying medical condition, as long as the brain has not been permanently damaged. Next in order of priority are attempts at psychopharmacological treatment, based on the predominant clinical manifestations. A few studies have shown a highly beneficial effect of lithium or carbamazepine, in the usual doses, on disturbances of impulse control. No specific cognitive training programs have been developed for the organic personality disorders.

Appendix A. Inclusion Criteria for Category F07 and Criteria for Organic Personality Disorder (Category F07.0) in ICD-10 (World Health Organization 1993)

Disorders of Personality and Behavior Caused by a Disease, Injury, or Functional Disturbance of the Brain (F07)

G1 Objective demonstration, by means of physical, neurological, and laboratory examination and/or history, of a cerebral disease, injury, or functional disturbance

G2 Absence of impairment of consciousness or severe memory disturbance

G3 No adequate and convincing documentation of another cause of the disorder of personality and behavior, which would justify classification in Chapter F6

Organic Personality Disorder (F07.0)

At least three of the following features must be present for a duration of 6 or more months:

1. Persistently reduced ability to carry out goal-directed activities, particularly over long periods of time and when gratification is delayed
2. One or more of the following affective changes:
   a) Emotional lability (uncontrolled, inconstant, and fluctuating expression of emotion)
   b) Euphoria and superficial, inappropriate facetiousness
   c) Irritability and/or outbreaks of anger and aggression

d) Apathy
3. Uninhibited expression of needs or impulses without regard for consequences or for social conventions; affected individuals may engage in antisocial activity such as stealing, inappropriate sexual advances, gluttonous eating, or extreme neglect of bodily hygiene
4. Cognitive disturbances, typically in the form of:
   a) Extreme suspiciousness and paranoid ideas
   b) Excessive preoccupation with a single topic, such as religion, or the strict classification of the behavior of others as "right" and "wrong"
5. Marked changes in language production and the flow of speech, including circumstantiality, vagueness, viscous thinking, and graphomania
6. Altered sexual behavior (hyposexuality or change of sexual preference)

Appendix B. Criteria for the Post-encephalitic Syndrome (F07.1) and the Organic Psychosyndrome After Traumatic Brain Injury (F07.2) in ICD-10 (World Health Organization 1993)

Post-encephalitic Syndrome (F07.1)

A. At least one of the following residual neurological deficits:
   1. Paralysis
   2. Deafness
   3. Aphasia
   4. Constructive apraxia
   5. Acalculia
B. The syndrome is reversible and rarely lasts more than 24 months.

Residual symptoms and behavioral changes after viral or bacterial encephalitis are nonspecific and do not justify the clinical diagnosis. These include a general feeling of illness, apathy or irritability, certain impairments of cognitive function (learning disabilities), disturbances of the sleep–wake cycle, and altered sexual behavior.

Organic Psychosyndrome After Traumatic Brain Injury (F07.2)

A. History of a traumatic brain injury with loss of consciousness, preceding the onset of clinical manifestations by no more than 4 weeks (there may be no objective evidence for brain injury on electroencephalography, imaging studies, and oculonystagmography)
B. At least three of the following features:
   1. Complaints of unpleasant sensations and pains such as headache, lightheadedness (usually without the typical features of vertigo), a general feeling of illness, severe exhaustion or sensitivity to noise
   2. Affective changes such as irritability and emotional lability, both easily provoked by emotional arousal and stress, depression and/or anxiety of a certain degree of severity
   3. Subjective complaints or difficulties in concentrating and in cognitive performance, and memory disturbances, without clear, objective demonstration of unequivocal impairment (e.g. by means of psychological tests)
   4. Sleep disturbances
   5. Reduced alcohol tolerance
   6. Preoccupation with the above symptoms and fear of permanent brain injury, rising to the level of hypochondria, and assumption of the role of an ill person

Appendix C. Criteria for Personality Change Caused by a Medical Disease Factor According to DSM-IV (Category 310.1) (see American Psychiatric Association 1994)

A. A persistent personality disturbance representing a change from the previously characteristic individual personality pattern. (In children, the disorder consists of a significant deviation from normal development or a significant change from the usual behavior pattern of the child lasting at least 1 year.)
B. There is evidence from the patient's history, physical examination, or laboratory tests that the disturbance is the direct physical consequence of a medical disease factor.
C. The disturbance cannot be better explained by another mental disorder (including another mental disorder caused by a medical disease factor).
D. The disturbance does not occur exclusively in the course of a delirium and does not fulfill the criteria for dementia.
E. The disturbance causes, to a clinically significant extent, suffering or impairments in social, professional, or other functional areas.

Subtypes:
- Labile type (affective lability)
- Disinhibited type (reduced impulse control)
- Aggressive type (aggressive behavior)
- Apathetic type (apathy and indifference)
- Paranoidtype (suspiciousness and paranoid ideas)
- Other type
- Combined type
- Nonspecific type

7 References

American Psychiatric Association (1994) Diagnostic and statistical manual of mental disorders, 4th edn (DSM-IV). American Psychiatric Association, Washington, DC

APA (1994) Diagnostic and statistical manual of mental disorders, 4th edn (DSM-IV). American Psychiatric Association, Washington, DC

*Bear DM, Fedio P (1977) Quantitative analysis of interictal behaviour in temporal lobe epilepsy. Arch Neurol 34: 454–467

Benson DF (1984) The neurology of human emotion. Bull Clin Neurosci 49: 23–42

Bleuler E (1966) Lehrbuch der Psychiatrie, 10th edn. Springer, Berlin, Heidelberg, New York

Capruso DX, Levin HS (1995) Neuropsychiatric aspects of trauma. In: Kaplan HI, Sadock BJ (eds) Comprehensive textbook of psychiatry, 6th edn, vol I. Williams and Wilkins, Baltimore, pp 207–220

*Castillo CS, Robinson RG (1994) Depression after stroke. Curr Opin Psychiatry 7: 87–90

Cummings JL (1985) Clinical neuropsychiatry. Grune and Stratton, Orlando

*Cummings JL (1993) Frontal-subcortical circuits and human behaviour. Arch Neurol 50: 873–880

Cummings JL, Benson DF (1992) Dementia. A clinical approach, 2nd edn. Butterworth Heinemann, Boston

*Eames P (1997) Traumatic brain injury. Curr Opin Psychiatry 10: 49–52

Eslinger PJ, Dimasio AR (1985) Severe disturbances of higher cognition after bilateral frontal lobe ablation: patient EVR. Neurology 35: 1731–1741

*Hermann BP, Whitman S (1984) Behavioural and personality correlates of epilepsy: a review, methodologic critique and conceptual model. Psychol Bull 95: 451–497

Herpertz S, Sass H (1997) Psychopathy and antisocial syndromes. Curr Opin Psychiatry 10: 436–440

Heston LL, White JA (1983) Dementia: a practical guide to Alzheimer's disease and related illness. Freeman, New York

House A, Dennis M, Warlow C, Hawton K, Molyneux A (1990) Mood disorders after stroke and their relation to lesion location. Brain 113: 1113–1129

Huber G (1972) Klinik und Psychopathologie der organischen Psychosen. In: Kisker KP, Meyer JE, Müller M, Stroemgren E (eds) Psychiatrie der Gegenwart. Forschung und Praxis, vol II, part 2. Springer, Berlin Heidelberg New York, pp 71–146

Huber G (1976) Lehrbuch der Psychiatrie. Schattauer, Stuttgart

Kraepelin E (1909) Psychiatrie. Ein Lehrbuch für Studierende und Ärzte, vols 1–4, 8th edn. Barth, Leipzig

Lang C, Stefan H (1990) Psychische Veränderungen bei Epilepsie. In: Hopf H, Poeck K, Schliack H (eds) Neurologie in Klinik und Praxis, vol III. Thieme, Stuttgart

Lauter H (1988) Die organischen Psychosyndrome. In: Kisker KP, Lauter H, Meyer JE, Müller C, Stroemgren E (eds) Psychiatrie der Gegenwart, vol 1. Springer, Berlin Heidelberg New York, pp 1–56

Levin DN, Grossman RG, Rose JE, Teasdale G (1979) Long-term neuropsychological outcome of closed head injury. J Neurol Neurosurg Psychiatry 50: 412–422

Levin BE, Llabre MM, Weiner WJ (1989) Cognitive impairments associated with early Parkinsons disease. Neurology 39: 557–561

Lishman WA (1987) Symptoms and syndromes with regional afflictions. In: Lishman WA (ed) The psychological consequences of cerebral disorder. Blackwell, Oxford, pp 21–77

Marder K, Tang M, Cote L, Stern Y, Mayeux R (1995) The frequency and associated risk factors for dementia in patients with Parkinson's disease. Arch Neurol 52: 695–701

Marotta R, Perry S (1989) Early neuropsychological dysfunction caused by HIV. J Neuropsychiatry 1: 225–234

Mega MS, Cummings JL, Fiorello T, Gornbein J (1996) The spectrum of behavioural changes in Alzheimer's disease. Neurology 46: 130–142

Mendez MF (1988) Psychopathology in epilepsy: prevalence, phenomenology and management. Int J Psychiatry Med 18: 193–210

*Mendez MF (1995) Neuropsychiatric aspects of epilepsy. In: Kaplan HI, Sadock BJ (eds) Comprehensive textbook of psychiatry, 6th edn, vol I. Williams and Wilkins, Baltimore, pp 198–206

Mendez MF, Doss RC, Taylor JL, Arguello R (1993) Relationship of seizure variables to personality disorders in epilepsy. J Neuropsychiatry Clin Neurosci 5: 283–286

Miller BL, Cummings JL, Villanueva-Meyer J et al (1991) Frontal lobe degeneration: clinical, neuropsychological, and SPECT characteristics. Neurology 41: 1374–1382

Mombour W, Zaudig M, Berger P et al (1996) International Personality Disorder Examination (IPDE). Interviewheft und Manual. Huber, Bern

Motomura U, Sawada T, Inoue N, Asaba H, Sakai T (1988) Neuropsychological and neuropsychiatric findings in right hemisphere damaged patients. Jpn J Psychiatry Neurol 42: 747–752

Perkins DO, Davidson EJ, Leserman J, Liao D, Evans DL (1993) Personality disorder in patients infected with HIV. A controlled study with implications for clinical care. Am J Psychiatry 150: 309–315

Poeck K (1985) The Klüver-Bucy syndrome in man. In: Fredericks JAM (ed) Handbook of clinical neurology, vol 45. Clinical neuropsychology. Elsevier, Amsterdam, pp 257–263

Popkin MK (1986) Organic brain syndromes with little or no cognitive impairment. In: Winokur G (ed) Medical psychiatry. Saunders, Philadelphia, pp 29–38

*Popkin MK, Tucker GJ (1994) Mental disorders due to a general medical condition and substance-induced disorders: mood, anxiety, psychotic, catatonic, and personality disorders. In: Widiger TA, Frances AJ, Pincus HA, First MB, Ross R, Davis W (eds) DSM-IV sourcebook, vol I. American Psychiatric Association, Washington, DC, pp 243–276

Scheid W (1980) Lehrbuch der Neurologie, 4th edn. Thieme, Stuttgart

*Sciutella A, Feinberg TE (1997) Focal behavioural syndromes in neuropsychiatry. Curr Opin Psychiatry 10: 53–58

Spitzer M, Casas B (1997) Project for a scientific psychopathology. Curr Opin Psychiatry 10: 395–401

Starkstein SE, Robinson RG (1989) Affective disorders and cerebrovascular disease. Br J Psychiatry 154: 170–182

Starkstein SE, Robinson RG (1991) The role of the frontal lobes in affective disorder following stroke. In: Lewin HS, Eisenberg HM, Benton AL (eds) Frontal lobe function and dysfunction. Oxford University Press, New York, pp 288–303

*Starkstein SE, Cohen BS, Fedoroff P, Parikh RM, Price TR, Robinson RG (1990) Relationship between anxiety disorders

and depressive disorders in patients with cerebrovascular injury. Arch Gen Psychiatry 47: 246–251

Steller U, Schultz-Venrath U (1995) Zerebrovaskuläre Erkrankungen. In: Ahrens S, Hasenbring M, Schultz-Venrath U, Strenge H (eds) Psychosomatik in der Neurologie. Schattauer, Stuttgart, pp 152–178

Stieglitz RD, Baumann U (1994) Psychodiagnostik psychischer Störungen. Enke, Stuttgart

Taylor MA (1987) DSM-III organic mental disorders. In: Tischler G, Cambridge MA (eds) Diagnosis and classification in psychiatry. Cambridge University Press, pp 147–174

World Health Organization (1993) International classification of diseases, Chap. V(F): mental and behavioural disorders. Diagnostic criteria for research. WHO, Geneva

World Health Organization (1994) International Personality Disorder Examination (IPDE). WHO, Geneva

CHAPTER

13

W. Hewer

Mental Disorders and Internal Medicine

Translator: E. Taub

1 Prevalence of Medical Comorbidity Among the Mentally Ill

Many studies have shown that a considerable percentage of the mentally ill suffer from concomitant physical illnesses.[1] According to a meta-analysis of published results, this percentage is approximately 50% (Felker et al. 1996); this figure refers to all accompanying illnesses, regardless of whether they bear a causal relationship to the mental disorder or are random, coincidental occurrences. The fact that such illnesses often go undiagnosed is of considerable importance to clinical practice. It has been estimated in several studies that up to half of all comorbid somatic illnesses are undiagnosed at the time of referral to a psychiatric institution (Koranyi and Potoczny 1992; Koranyi 1979).

It is not uncommon for patients with previously unrecognized physical illnesses to present with psychopathological phenomena (Marsh 1997). It has also been shown that, among patients obtaining care in psychiatric institutions, a high percentage of comorbid illnesses are not diagnosed, and previously diagnosed illnesses are often neglected; one or the other of these was the case in 47% of comorbid medical illnesses in 529 mentally ill patients in California, most of them less than 40 years old, who were investigated in a methodologically very careful study (Koran et al. 1989).

The reported percentages of mentally ill patients with comorbid physical illnesses are usually global and do not differentiate the illnesses according to medical specialty. These figures, therefore, should not be equated with the frequency of comorbid disorders in the field of internal medicine, although there is no doubt that they are by far the most common type represented (Felker et al. 1996). For example, disorders in the domain of internal medicine accounted for approximately two thirds of all physical illnesses in a study of a representative sample of patients undergoing acute inpatient treatment (Hewer et al. 1991).

As might be expected, physical comorbidity is particularly common among mentally ill patients of advanced age. Studies of psychogeriatric patients undergoing inpatient treatment have shown that they suffer from an average of two to five accompanying illnesses, of which approximately two thirds are within the domain of internal medicine. Cardiovascular disorders were most common, followed in order of frequency by metabolic and endocrine disorders, respiratory disorders, and gastrointestinal disorders (Hewer and Förstl 1998; Zubenko et al. 1997). The high prevalence of physical illnesses and disabilities among psychogeriatric patients was also demonstrated specifically for patients with dementia (Fichter et al. 1995) and depression (Lyness et al. 1996).

In the Berlin Old Age Study, the prevalence of depression among subjects with unequivocal evidence for significant physical impairment (e.g. multiple physical illnesses, immobility) was found to be two- to threefold higher than that among patients without such impairment, of whom 14.1% were depressed (Linden et al. 1998). It should be mentioned in this connection that chronic physical illnesses typically associated with persistent pain are not uncommonly a determinant of suicidal behavior in elderly patients; according to Summa (1988), this is the case in approximately 40% of patients admitted to hospitals because of attempted suicide.

A number of methodological problems complicate the simultaneous determination of physical and psychiatric morbidity. One such problem is that such studies are supposed to encompass the entire diagnostic spectrum of (internal) medicine, and it is quite difficult to find a satisfactory method for doing this. Furthermore, the question of psychiatric–medical comorbidity is inevitably fraught with the difficulty of deciding whether particular disease manifestations should be attributed to medical or to psychiatric causes; this gives rise to considerable diagnostic imprecision (Linden et al. 1998). In addition, the majority of studies in this field to date have failed to employ satisfactory methods of psychiatric and medical data collection, such as those used in the Berlin Old Age Study cited above (Linden et al. 1995).

Although a number of questions must remain open because of the methodological difficulties just mentioned, the available data, taken as a whole, suggest

[1] A remark on nomenclature: terms such as "medical comorbidity," "physical illness," "somatic illness," or "medical illness" are used synonymously in this chapter to refer to illness(es) in the domain of any medical specialty other than psychiatry or psychotherapy. These terms are approximately equivalent to "general medical condition," as used in DSM-IV (American Psychiatric Association 1994). The use of terms of this kind is not meant to imply that mental illnesses lack physical causes or correlates or that psychiatry and psychotherapy are not medical specialties.
When we use the terms mentioned above, especially "medical illness/disorder/condition" (mainly in the singular), we are referring – with regard to the topic of this chapter – in particular to comorbid physical disorders in the field of internal medicine. Beyond that, these terms may also, in certain contexts, relate to comorbid physical disorders in general, namely for two reasons: first, because it is often difficult to differentiate between diseases within the frame of internal medicine and those "belonging" to other medical specialties; second, because many of the interactions between medical and mental illness discussed below are not confined to internal medicine, but may also be found in relation to other medical branches, such as orthopedics and urology (see Chap. 24, Vol. 3, Part 2).

that mentally ill patients have a degree of physical impairment that, on average, is higher than that of the general population, at least for certain patient groups. Further evidence for a global link between psychiatric and physical morbidity was provided by a Danish epidemiological study of the frequency of hospital admission. Inpatient hospitalizations in general hospital and psychiatric wards were strongly correlated with each other: individuals requiring inpatient treatment in a general hospital at some point in the 8-year period of the study were four times more likely than others to require psychiatric hospitalization (Fink 1990).

In this context, we would like to refer to the numerous studies of the prevalence of psychiatric morbidity among general medical patients, especially those treated in hospital. Presently available information indicates that psychiatric disorders can be diagnosed in 30%–50% of these patients (Arolt et al. 1995; Saupe and Diefenbacher 1999).

Lastly, numerous studies have very consistently shown a significant elevation of mortality among the mentally ill, not only because of suicide, accidents etc., but also because of a higher rate of death from natural causes. According to the recent, extensive literature review by Harris and Barraclough (1998), approximately half of the excess mortality of patients suffering from affective disorders, for example, is due to a higher incidence of death from natural causes. For schizophrenic patients, the corresponding figure is approximately two thirds. Moreover, by evaluating the clinical course of more than 50,000 patients, these authors arrived at the conclusion that the mentally ill in general have a mortality rate that is twice as high as that of the general population.

2 Interactions Between Medical Illness and Mental Disorders

The observations in the previous section concerning the frequency of medical comorbidity are of importance primarily because physical and mental disorders interact with each other in many different ways. In the present section, various medical–psychiatric interactions will be discussed in a simplified manner, as these interactions are often extremely complex, as was shown by Linden et al. (1998) for interactions between medical morbidity and depressive illnesses in advanced age. It should also be noted that "medical" and "psychiatric" factors are differentiated from each other purely from a didactic viewpoint, and no particular philosophical stance toward the mind–body problem is thereby implied.

2.1 Medical Illness as the Immediate Cause of Mental Disorders

Many medical illnesses may lead to cerebral pathology or to brain dysfunction and thereby cause psychopathological syndromes of variable phenomenology in a physiologically explicable manner. The most important pathogenetic mechanisms are the following:

- Disturbances of brain function as a consequence of inadequate supply of oxygen and/or substrate (caused by a focal or generalized disturbance of cerebral perfusion, or a reduction of the oxygen or substrate content of the blood, e.g. in hypoxia or hypoglycemia)
- Disturbances of brain function by means of metabolic–endocrine or toxic processes or other humoral factors (e.g. renal or hepatic failure, disturbances of fluid and electrolyte balance, nutritional disorders, paraneoplastic processes, and infectious or other inflammatory diseases)

Examples of mental disorders as direct physiological consequences of medical illness will be given mainly in Sects. 3.1–3.5 of this chapter, in which dementia, delirium, and other mental disorders resulting from general medical conditions will be discussed from the point of view of internal medicine.

Further typical examples are provided by those medical disorders that may produce an organic amnestic syndrome. It is well known that the Wernicke-Korsakoff syndrome, which is caused by thiamine deficiency, occurs not only in alcoholics, but also (more rarely) in patients suffering from malnutrition, diseases of the upper gastrointestinal tract (such as gastric carcinoma), or protracted vomiting (as in hyperemesis gravidarum), when these entities cause a severe restriction of nutrient intake or absorption. Furthermore, post-anoxic states (such as that seen after cardiopulmonary resuscitation) and recurrent severe hypoglycemia may cause an amnestic syndrome. The particular vulnerability of the hippocampus to hypoxia and hypoglycemia seems to be the most important pathogenetic factor in this process (Förstl 1999).

2.2 Medical Illness as a Factor Influencing the Occurrence and Course of Mental Disorders

Medical illness may be an important risk factor for mental disorders; vascular dementia syndromes are an example. The available data imply that arterial hypertension is the most important treatable risk factor for

vascular dementia. Further risk factors include diabetes mellitus, disturbances of lipid metabolism, smoking, the presence of potential sources of cardiogenic emboli, intermittent hypotension, and syncopal events (Kloss et al. 1994).

Medical illness may also have unfavorable effects on the course of mental disorders. Several empirical lines of evidence indicate that the appearance of resistance to drug treatment in elderly, depressed patients may be influenced by an increased prevalence of comorbid medical disorders, among other factors (Bonner and Howard 1995).

Medical conditions may influence the occurrence and course of mental disorders by a variety of mechanisms. They may have various secondary effects on cerebral function, as is true of arterial hypertension (Lis and Gaviria 1997) and diabetes mellitus (Mooradian 1997). On the other hand, severe, chronic physical illness may also lead to various psychosocial problems, e.g. early retirement and loss of social support, which in turn render the patient more susceptible to related psychiatric disorders, such as adjustment disorders and depression (Ormel et al. 1997).

Finally, the treatment of medical illness may also cause mental disorders. Many primarily non-psychotropic drugs may have unwanted psychiatric effects. Even if such an association remains unproved for some agents, many drugs used in internal medicine have been shown empirically to have undesired psychotropic effects (Kasper and Jung 1995). Propranolol and reserpine, for example, can cause depression, while a broad spectrum of substances can produce delirium (see Sect. 3.2). Such undesired effects may also be produced by nonpharmacological therapeutic measures: mental disorders are found in association with the postoperative phase (Huber 1988), treatment in intensive care units, and organ transplantation (see Chap. 17, Vol. 2, Part 2).

2.3 Mental Disorders as Immediate Causes of Medical Illness

The physical harm resulting from abuse of, and dependence on, psychotropic substances is by far the most common problem in this category (Cherubin and Sapira 1993; Seitz et al. 1995). Further examples include the diverse medical complications of catatonia (Fricchione et al. 1997) and, lastly, the many medical problems brought about by self-destructive behavior, regardless of possible suicidal intent. Special problems are posed in this regard by the factitious disorders and the diagnostic uncertainties typically associated with them, which directly result from the characteristic behavior of the affected patients (Kapfhammer et al. 1998).

2.4 Mental Disorders as Factors Influencing the Occurrence and Course of Medical Illness

The presence of mental illness can negatively affect the patient's physical health in many ways. Empirical evidence suggests that mental disorders may be an important risk factor for the occurrence of medical illness. Impressive support for this concept comes from a longitudinal study by Vaillant (1998), in which the possible relation between psychopathological variables and the subjects' physical health was studied prospectively over a period of several decades. There are also indications that the risk of developing ischemic heart disease is elevated by the presence of depression. It has been shown, in relation to the same paradigm of ischemic heart disease and depression, that mental disorders can negatively affect the course of medical illness (see Chap. 14, Vol. 2, Part 2).

Negative effects on the course of medical illness are also to be expected because many mentally ill patients have a limited ability to cooperate with their medical treatment ("noncompliance"; Häfner and Bickel 1989; see also Sect. 6).

Dementia and *schizophrenia* are the most important examples of mental disorders that put the patient's physical health at increased risk:

- As *dementing illnesses* progress, the affected patients are increasingly subject to bronchopulmonary infections and to disturbances of nutritional and fluid balance. Furthermore, these patients' ability to express physical complaints is often limited, and major diagnostic difficulties may result (Hewer and Förstl 1998).
- Several behavioral features characteristic of patients with *schizophrenia* may have unfavorable effects on the occurrence and course of comorbid somatic illnesses (Adler and Griffith 1991; Vieweg et al. 1995). An alteration of pain perception is not uncommon; this may be due in part to antipsychotic treatment, but it was observed even in the preneuroleptic era (Jakubaschk and Böker 1991). Furthermore, the consequences of drastically elevated nicotine consumption, widespread abuse of psychotropic substances, and (less frequently) polydipsia must be kept in mind in this group of patients (Jeste et al. 1996). Furthermore, schizophrenic patients, like those suffering from dementia, carry an increased risk of bolus aspiration (Schmitt and Hewer 1993).

It should be remarked, for the sake of completeness, that the literature contains several reports of possible protective effects of physical illness on the patients' mental health. These effects, however, seem to be

rather less significant than the unfavorable effects discussed above (Häfner and Bickel 1989). For example, it has been observed in individual cases that severely depressed patients suffering from acutely life-threatening medical illness may obtain significant improvement of their psychopathological condition (Deahl 1990).

Psychotropic drugs are, of course, an important cause of medical problems. They are known to have negative effects on the functioning of many different organ systems (Kane and Lieberman 1992; Tueth 1994); nonetheless, according to the findings of a large study on the safety of medications used in psychiatry (the AMÜP Study, a drug surveillance program carried out in selected university departments of psychiatry in Germany; Grohmann et al. 1994), acute, severe medical problems appear to be relatively rare when these drugs are given in accordance with recommended practice and under appropriate supervision. Long-term use of psychotropic drugs has not been shown to be associated with elevated mortality (Brown 1997; Weeke 1987). It must not be forgotten, however, that most of the currently used drugs can cause weight gain with the well-known associated risks to the patients' health (Fritze et al. 1992).

2.5 Other Constellations

Coexistent medical and psychiatric illness may have a common etiology, as in certain hereditary metabolic disorders (e.g. metachromatic leukodystrophy, Fabry's disease), in which an inborn enzyme deficiency causes both systemic pathology and brain damage with resulting psychiatric disturbance (for a review, see Huber 1988).

Finally, it must be mentioned that a medical illness and a mental disorder may occur in the same patient by random coincidence. Statistical considerations lead us to expect this constellation to occur mainly with relatively common illnesses. As we will discuss further in Sect. 6, medical problems should not be neglected within the framework of psychiatric care even in these cases.

2.6 Frequency with Which Medical Disorders Induce or Aggravate Mental Disorders

Difficult methodological problems, as discussed in part in Sects. 1 and 3 of this chapter, complicate any attempt to answer the question of how often medical conditions induce or aggravate mental disorders (Popkin 1995). The major problem is that the judgment of whether a particular medical illness is a cause of, or exacerbating factor in, a mental disorder is nearly inevitably influenced by the subjective experiences and attitudes of the evaluator (Koran et al. 1989). Thus all reported figures on this topic must be interpreted with an appropriate degree of caution.

Nonetheless, in view of the clinical and scientific importance of the question, one figure will be cited: Felker et al. (1996), in their review article, conclude that medical disorders are a cause of or exacerbating factor in mental illness in approximately 20% of patients treated in psychiatric institutions.

When interpreting this figure, it should be borne in mind that it refers not only to the direct physiological consequences of somatic conditions, but also to indirect effects mediated by the patients' altered psychosocial situation. In other words, types of illness that would be classified as "nonorganic" according to ICD-10 are included in this figure, as long as evidence suggests that they are subject to significant influence from comorbid medical disorders.

3 Mental Disorders Resulting from General Medical Conditions

This section concerns the major types of psychopathology resulting from medical disorders with reference to the internationally accepted systems of classification, particularly ICD-10 (WHO 1992), but also to some extent the diagnostic criteria for mental disorders formulated in DSM-IV (American Psychiatric Association 1994). The organic mental disorders[2] that may arise in connection with medical conditions are listed in the Appendix.

The first three syndromes named in the Appendix differ from the others in that they involve a disturbance of consciousness and/or an impairment of higher cognitive function as a pathognomonic feature and are thereby clearly linked to an organic cause. The syndromes of dementia, amnesia, and delirium can therefore be designated as "psycho-organic syndromes of the first rank" (*psychoorganische Syndrome ersten Ranges*, Lauter 1988), or – in the nomenclature of DSM-IV – as cognitive disorders (American Psychiat-

[2]The use of the term "organic" in the sense of ICD-10 does not imply that "nonorganic disorders" have no cerebral substrate. "Organic mental disorders" according to ICD-10 correspond largely to the "mental disorders due to a general medical condition" as defined by DSM-IV (with the exception that mental disorders caused by nonpsychotropic medications are classified as "organic mental disorders" in ICD-10, while they are included under the heading "substance induced disorder" in DSM-IV.

ric Association 1994) (see also Chaps. 1, 2, 8–11, Vol. 2, Part 2).

The assignment of the attribute "organic" to the syndromes listed in the Appendix, which are coded under F06 in ICD-10, implies that they are "causally related to brain dysfunction" (WHO 1992). ICD-10 names four features that it considers to be prerequisites for the assignment of a clinical condition to one of these categories, which essentially correspond to the criteria formulated earlier by Kurt Schneider (1967):

1. Evidence of a potentially causative cerebral or systemic disease
2. Temporal relationship between the development of the underlying disease and the onset of the mental disorder
3. Recovery from the mental disorder following remission or improvement of the underlying disease
4. No evidence of an alternative cause of the psychiatric symptomatology (such as a strong family history or precipitating stress)

Conditions classified under F06 whose phenomenology resembles, or may even be identical to, that of nonorganic syndromes may be designated as "psycho-organic syndromes of the second rank" (*psychoorganische Syndrome zweiten Ranges*, Lauter 1988). Nevertheless, for a number of reasons that have been discussed in DSM-IV and elsewhere, it may be difficult or impossible to verify the "organic" origin of these syndromes in the individual case. The possibility, already mentioned above, that a mental disorder and a physical illness can occur together by chance must first be considered. Furthermore, certain syndromes accompanying serious physical illnesses – particularly depressive and anxious states – require differentiation from adjustment disorders that have similar manifestations; this may pose a difficult problem in the individual case (Lieb et al. 1997).

Mental disorders due to a general medical condition manifest themselves only by a limited spectrum of syndromes, as compared with a considerably greater variety of possible underlying illnesses (Table 1). It may therefore be asked what factors influence the occurrence and the manner of presentation of a mental disorder in the individual patient. In this context, it is useful to remember the principle of nonspecificity formulated by Bonhoeffer, according to which different medical illnesses may lead to identical psychopathological manifestations, just as a single underlying illness may lead to different types of organic mental disorder (Gross and Huber 1993). The psychopathological manifestations appearing in the individual case are determined not only by the causative disease process, but also by several other unrelated or only indirectly related factors (Gross and Huber 1993),

Table 1. Medical conditions that may be associated with organic mental disorders

Disease group	Examples
Infectious diseases	Syphilis, HIV infection
Neoplastic diseases	Paraneoplastic syndromes, e.g., limbic encephalitis in bronchial carcinoma
Endocrine disorders	Hyperthyroidism, hypothyroidism
Nutritional and metabolic disorders	Malnutrition, vitamin deficiencies, acute intermittent porphyria
Hematological diseases	Severe anemia
Cardiovascular diseases	Congestive heart failure, cardiac dysrhythmias
Respiratory diseases	Chronic obstructive pulmonary disease
Diseases of the gastrointestinal tract, the hepatobiliary system, and the pancreas	Malabsorption syndrome, severe hepatic diseases, acute pancreatitis
Diseases of the kidneys and urinary tract	Advanced renal failure
Rheumatic and autoimmune diseases	Systemic lupus erythematosus
Toxic disorders	Psychotropic effects of nonpsychotropic drugs, heavy metal poisoning

HIV, human immunodeficiency virus.

including those designated by Lauter (1988) as "secondary factors" (e.g. limitation of mobility and social activity because of disease) and "vulnerability factors" (e.g. biological predisposition, premorbid personality).

An illustration of these concepts is furnished by the systemic autoimmune diseases, some of which are relatively frequently associated with organic mental disorders (Table 2). The following conditions are more commonly seen:

- Paranoid–hallucinatory syndromes
- Affective disorders (especially depressive syndromes)
- Cognitive deficit syndromes (of variable degrees of severity, up to the level of dementia)
- Other disorders, e.g. organic personality and behavioral disorders, organic anxiety disorder

In systemic lupus erythematosus, for example, psychopathological abnormalities of these types are observed in approximately 20%–60% of patients (Lieb et al. 1997; Shannon and Goetz 1995). In the case of the systemic autoimmune diseases, the principle remains valid that each of the organic mental syndromes listed

Table 2. Frequency of organic mental disorders in systemic autoimmune diseases (after Lieb et al. 1997)

Disease	Associated mental disorders
Systemic lupus erythematosus	Common
Sjögren's syndrome	Common
Scleroderma	Very rare
Mixed collagenosis	Very rare
Primary CNS angiitis	Rare
Takayasu's arteritis	Rare
Temporal arteritis	Rare
Churg-Strauss syndrome	Rare
Wegener's granulomatosis	Rare
Polyarteritis nodosa	Rare
Microscopic polyangiitis	Rare
Antiphospholipid syndrome	Less common
Sneddon's syndrome	Common
Behçet's disease	Less common

CNS, central nervous system.

above can be associated with any underlying illness in this group, just as any underlying illness can give rise to any of these organic mental syndromes. A similar nonspecificity of psychopathological syndromes is associated with several other classes of illness, e.g. the endocrinopathies (see Sect. 4); the presence of a particular type of psychopathological manifestation cannot therefore be taken as a reliable clue to the underlying etiopathogenetic process.

Nonetheless, the principle of nonspecificity is not universally applicable. Taking up the example of systemic autoimmune diseases once again, we find that, even though diagnostic precision cannot be obtained in the individual case, there are nonetheless clear correlations between the type of underlying illness and the frequency of certain psychopathological syndromes. For instance, affective and paranoid–hallucinatory syndromes are most common in systemic lupus erythematosus and Sjögren's syndrome, while dementia is more characteristic of angiitis, antiphospholipid syndrome, and Sneddon's syndrome (Lieb et al. 1997). Furthermore, the nonspecificity paradigm has been criticized on the grounds that the manner in which psychopathological features and syndromes have been described to date may still be too undifferentiated to enable more specific correlations of the desired type (Lauter 1988). In view of the rapid progress now being made in the neurobiological understanding of mental illness, it is conceivable that future research will establish clearer correlations between disease processes in the field of internal medicine and the types of psychopathological manifestation that they cause.

Because limitations of space preclude a discussion of all of the conditions listed in the Appendix in relation to all possible causes in the field of internal medicine, the following subsections will deal with only a few of the more frequently occurring organic mental syndromes.

3.1 Dementia

A great variety of medical conditions may underlie a dementing process, including the following (Mumenthaler 1987; Lang 1994):

1. Cardiovascular and pulmonary diseases:
   - Arteriosclerotic and degenerative angiopathies
   - Arteritides (e.g. systemic lupus erythematosus, giant cell arteritis)
   - Major heart diseases (e.g. congestive heart failure, valvular heart disease, cardiac dysrhythmias)
   - Recurrent cerebral emboli
   - Chronic respiratory failure
   - Sleep apnea syndrome
2. Infectious diseases:
   - Human immunodeficiency virus (HIV) infection/acquired immunodeficiency syndrome (AIDS)
   - Lyme disease
   - Brucellosis
   - Whipple's disease
   - Malaria
3. Metabolic and endocrine diseases:
   - Endocrine disorders (e.g. hypothyroidism, hypoparathyrodism, hyperparathyroidism, hypoglycemia)
   - B_{12} and other vitamin deficiencies
   - Uremic encephalopathy, dialysis dementia, hyponatremia
   - Hepatic failure
   - Metabolic diseases (e.g. disorders of lipid metabolism, porphyria)
   - Malabsorption syndromes
4. Miscellaneous diseases:
   - Hematological diseases (e.g. polycythemia vera, paraproteinemias, coagulopathies)
   - Sarcoidosis
   - Paraneoplastic limbic encephalitis

It is important to establish the particular disease underlying dementia in the individual case because of the potential implications for treatment. It is estimated that 13%–15% of demented patients have a potentially reversible underlying illness, and 10%–12% an actually reversible one (Clarfield 1988; Weytingh et al. 1995).

Among possible types of treatable underlying illness, diseases within the domain of internal medicine (most importantly metabolic–endocrine disorders) are third in order of frequency. Approximately 2% of demented patients have an underlying medical problem whose directed treatment will result in an improvement of the cognitive deficits. Complete remission of dementia can be obtained in no more than one third of such patients, while dementia regresses partially in the remaining patients (Clarfield 1988; Weytingh et al. 1995).

Vascular processes are the second most frequent cause of dementia after Alzheimer's disease. Vascular dementia is generally caused by arteriosclerotic and degenerative changes of the extra- and intracranial cerebral vasculature, of both macro- and microangiopathic types; the latter type is currently thought to be of particular pathogenetic importance (Poeck and Hacke 1998). Rarer causes of vascular dementia include cardioembolic events, inflammatory angiopathies, and coagulopathies (Geldmacher and Whitehouse 1996; see also Chap. 8, Vol. 2, Part 2).

The diagnostic assessment of dementia must, therefore, always include a basic general medical assessment, whose findings may then necessitate further investigation (Hewer and Förstl 1998). From the therapeutic viewpoint, a question that frequently arises in clinical practice is whether a dementia syndrome caused by arteriosclerotic–degenerative vascular disease can be influenced by treatment. The therapeutic prospects become worse with increasing duration and severity of the clinical manifestations; the medical risk factors named in Sect. 2 should thus be treated at an early stage, if possible even before the appearance of dementia. The management of high blood pressure is particularly important. An abrupt decrease of blood pressure leading to hypotension must be avoided, however, as this may promote further progression of the cerebral parenchymal insult (Dettmers et al. 1997).

3.2 Delirium

Since Bonhoeffer originally described the "acute exogenous reaction" at the beginning of the twentieth century, it has become clear that a very wide variety of medical disorders can cause this syndrome, in addition to primary brain disorders and the use of psychotropic substances (Lipowski 1990). The occurrence of delirium can be understood as a threshold phenomenon (Jacobson 1997): the more intense a delirogenic factor is and the more such factors are involved, the more likely delirium is to develop. In a typical constellation, for example, acutely or subacutely emerging medical disorders, superimposed on a preexisting cerebral insult, may precipitate delirium. When the preexisting insult is severe, even relatively mild medical problems, such as an uncomplicated urinary tract infection, may lead to delirium.

The following are the more important medical conditions that may cause delirium (Jacobson 1997):

- Infections (e.g. pneumonia, urinary tract infection)
- Disturbances of fluid and electrolyte homeostasis (e.g. hypovolemia, acid–base disturbances, and disorders of sodium, potassium, calcium, and magnesium balance)
- Metabolic and endocrine disturbances (e.g. hypoglycemia, renal or hepatic failure, vitamin deficiencies, thyroid or parathyroid dysfunction)
- Cardiopulmonary diseases (e.g. myocardial infarction, congestive heart failure, cardiac dysrhythmias, pulmonary embolism, respiratory failure, hypertensive encephalopathy, shock)
- Severe anemia
- Undesired effects of medications used in internal medicine (e.g. antiarrhythmics, antibiotics, antirheumatic agents, corticosteroids, cytostatic agents)

Delirium may occur in conjunction with medical illnesses at any age, but in the industrial countries it tends to affect mainly the elderly, because they are most subject to multimorbidity, including neurodegenerative processes. It is estimated that at least 20% of elderly patients admitted to general hospitals are in a state of delirium at the time of admission or at some point in their hospital course; prevalence may even be higher for patient subgroups at particular risk (Inouye 1998).

The major prognostic significance of delirium becomes clear if we consider that there is a mortality rate of roughly 25% within 6 months; this may be explained by the fact that many of the diseases causing or precipitating delirium are life-threatening (Trzepacz 1996). Furthermore, the occurrence of delirium, at least in the elderly, is correlated with other prognostically unfavorable features (American Psychiatric Association 1999), such as prolonged hospital stays, frequent nursing-home admissions, and certain complications (e.g. falls, decubitus ulcers); this remains so even after appropriate statistical control for potential confounding variables, e.g. age and severity of illness (O'Keeffe and Lavan 1997).

Even though delirium is often associated with a negative overall prognosis, it generally responds to therapy, and there is a good chance that the patient's mental state can be restored to what it was before the appearance of the acute psychopathological disturbance. Early and specific treatment of the causative or contributing physical illnesses is therefore essential.

In view of the etiological and pathogenetic relationships discussed above, the diagnosis and treatment of states of delirium clearly require an interdisciplinary process involving both internal medicine and neuropsychiatry. The treatment of delirium includes not only the treatment of the underlying illness, but also all medical and nursing measures promoting physiological homeostasis (e.g. nutrition, fluid balance, mobility). Moreover, certain principles of therapeutic interaction with patients should be adhered to, including the provision of aids to orientation, the avoidance of excessive stimulation, and the greatest possible constancy of treating personnel (Förstl 1999; Rabins 1991). The indication for symptomatic psychopharmacological treatment should be reviewed in each case (Trzepacz 1996). It should be recalled that many medications have delirogenic effects (Inouye 1998), including a number that are frequently used in internal medicine and do not have primary psychotropic effects (Tune et al. 1992).

3.3 Mild Cognitive Disorder

According to the ICD-10 Research Criteria (WHO 1993), the essential, feature of mild cognitive disorder is an impairment of cognitive abilities (e.g. memory, attention, concentration) to a lesser degree than in dementia, delirium, or various other conditions. A duration of symptomatology of at least 2 weeks is also required (see also Chap. 10, Vol. 2, Part 2). As discussed in DSM-IV (research criteria for mild neurocognitive disorder), the cognitive deficit takes a variable course, depending on the dynamics of the underlying process. Stable, fluctuating, or remitting courses may be observed; the deficits may also progress and eventually rise to the level of dementia.

The position occupied by mild cognitive disorder among the other cognitive deficit syndromes is still a matter of discussion (Caine et al. 1995; WHO 1993). According to present conceptions, the following medical disorders may be considered as possible causes of mild cognitive disorder (American Psychiatric Association 1994; Egberts 1993; Elias 1998; Stern and Prange 1995; Strachan et al. 1997):

- Arterial hypertension
- Diabetes mellitus
- Endocrine disorders
- Milder forms of hepatic encephalopathy
- Severe metabolic imbalances
- HIV infection and other infectious diseases
- Hypoxic states

Various other medical disorders might also be responsible for this presumably fairly common disorder.

3.4 Organic Depressive Disorder

Depressive syndromes caused by medical disorders are classified in ICD-10 as organic depressive disorder (F06.32). As discussed in Sect. 2, above, a prerequisite for this diagnosis is that the depressive manifestations are caused by physiological processes associated with the underlying illness, rather than by the patient's emotional reaction to the knowledge that he or she has a serious illness. The spectrum of medical disorders that may cause organic depressive disorder includes the following, among others (Caine et al. 1995; Popkin 1995):

- Endocrine disorders (e.g. hormonal excess or deficiency states of the thyroid or parathyroid glands)
- Metabolic disorders (e.g. vitamin B_{12} deficiency and certain electrolyte disorders, such as hyper- and hypocalcemia)
- Cardiopulmonary diseases (e.g. congestive heart failure, respiratory failure)
- Hepatic or renal failure
- Systemic autoimmune diseases (e.g. systemic lupus erythematosus, temporal arteritis, rheumatoid arthritis)
- Cancer (e.g. pancreatic carcinoma)
- Infectious disease (e.g. infectious mononucleosis, hepatitis, influenza, HIV infection, syphilis)
- Medications (e.g. reserpine, propranolol, clonidine, glucocorticoids, nonsteroidal anti-inflammatory agents, H_2 inhibitors)

As discussed in Sect. 3, it may be difficult to determine in the individual case whether an organic depressive disorder is present or whether there is an alternative explanation, such as an adjustment disorder or the presence of depression as a second, independent disease. Comorbid medical illness as a precipitating factor in patients with preexisting vulnerability to depression is a further possibility to consider (Tölle 1990). Such an explanation is currently being discussed for the frequent observation, in earlier decades, of depression in patients under treatment with reserpine (Akiskal 1995).

3.5 Organic Delusional Disorder

Organic delusional or schizophrenia-like psychoses may be produced by the following medical conditions, among others (or by the undesired effects of medications used in internal medicine) (Davison and Bagley 1969; American Psychiatric Association 1994; Huber

1994; Lieb et al. 1997; Marsh 1997; Pearlson and Petty 1994):

- Endocrine diseases (hyper- or hypothyroidism, hyper- or hypocortisolism, hypoparathyroidism, hypopituitarism)
- Metabolic disorders (Vitamin B_{12}, folic acid, or thiamine deficiency, acute intermittent porphyria, Wilson's disease, hyponatremia, renal and hepatic insufficiency, hypoglycemia)
- Systemic autoimmune diseases (systemic lupus erythematosus, Sjögren's syndrome, Churg-Strauss syndrome)
- Infections (syphilis, HIV infection, malaria, subacute bacterial endocarditis)
- Miscellaneous disorders (cerebral anoxia, hypercapnia, carbon monoxide poisoning, sarcoidosis, polycythemia, thalassemia)
- Undesired effects of medications used in internal medicine (e.g. digitalis, certain antiarrhythmics, propranolol, β-sympathomimetic agents, cephalosporine and other antimicrobial agents, nonsteroidal anti-inflammatory agents, corticosteroids, anticholinergics, cimetidine)

Schizophreniform states are only rarely produced by medical illness, but the possibility is clinically important because of its therapeutic implications. For that reason, the guidelines for the treatment of schizophrenia recently issued by the German Psychiatric Association (Deutsche Gesellschaft für Psychiatrie, Psychotherapie und Nervenheilkunde 1998) contain explicit recommendations for basic medical assessment in schizophrenic conditions. In addition to its clinical importance, the occurrence of schizophreniform syndromes in certain medical illnesses is of basic scientific interest for the understanding of the neurobiology of schizophrenia.

4
Medical Illness Associated with Mental Disorders: Examples from Endocrinology

As can be seen from Table 1, a broad spectrum of medical illness may be associated with mental disorders. A number of the disease groups listed in Table 1 form the subject of individual chapters in Part 2 of this volume (Chaps. 14–17) and will therefore not be dealt with in detail here.

The major endocrine diseases will be discussed in the remainder of this section, in view of their common occurrence and frequent interactions with mental disorders. Although the principle of nonspecificity remains operative, as M. Bleuler (1954) pointed out, the individual endocrine diseases nonetheless have particular features that are worthy of consideration from a psychiatric point of view. In endocrinopathies, like other medical disorders, psychopathological manifestations may be the presenting symptoms (Heuser 1993); this fact should make clear once again that psychiatric diagnosis requires a background of experience in general medicine.

4.1
Disorders of Glucose Metabolism

Diabetes mellitus is the most common endocrine disease; its prevalence in the German population, for example, is approximately 5% (Hauner 1998). Patients with both insulin-dependent (IDDM) and non-insulin dependent diabetes (NIDDM) have a markedly elevated risk of developing depressive syndromes (Eiber et al. 1997; Gavard et al. 1993). IDDM patients also have an elevated prevalence of anxiety disorders (Eiber et al. 1997). Whether these mental disorders occur more frequently in diabetics because of a direct causative effect of the metabolic abnormality in diabetes or primarily because of nonspecific effects related to the presence of a chronic illness remains uncertain (Eiber et al. 1997). Their clinical importance results not only from the suffering they cause, but also from the unfavorable effect they may have on blood sugar control (Kovacs et al. 1996).

Diabetes mellitus is a major risk factor for vascular dementia in the elderly (Skoog 1998). As discussed in Sect. 3.3, multiple studies have shown that certain neuropsychological deficits, including memory impairment and cognitive slowing, occur more frequently in patients with NIDDM (Strachan et al. 1997); it remains unknown to what extent these deficits may be precursors of frank cognitive deficit syndromes that will develop later. Recent findings also suggest a possible association of diabetes mellitus with dementing processes of the Alzheimer type (Ott et al. 1996).

Hypoglycemia is most commonly caused by antidiabetic treatment with insulin or oral antidiabetic agents. Other causes of severe hypoglycemia, such as insulinoma, end-stage liver diseases, or severe malnutrition, are much rarer. Acute hypoglycenia may be accompanied by psychopathological manifestations of many different kinds, of which the most common are clouding of consciousness, psychomotor agitation, and anxiety (Lishman 1998). The characteristic vegetative signs of hypoglycemia – tremor, tachycardia, and severe hyperhidrosis – are not always seen (Cryer 1997). It follows that the possibility of hypoglycemia must always be considered in the differential diagnosis of acute psychopathological changes of unknown

origin. Chronic, recurrent hypoglycemia of any degree of severity may lead to irreversible cerebral damage, including dementia. This risk is especially prominent in cases of insulinoma that go unrecognized for years (Lishman 1998), but cases of cognitive deficit syndromes arising in diabetics as a consequence of severe hypoglycemia have also been reported (Langan et al. 1991).

4.2 Disorders of Thyroid Function

Disorders of thyroid function occupy a special position among all types of endocrine disease in the differential diagnosis of mental disorders, because they are common and may give rise to highly varied psychopathological manifestations. Alongside patients with known thyroid dysfunction who may require psychiatric treatment, there are also more than a few cases of thyroid dysfunction that are first diagnosed in psychiatric institutions. The proportion of patients with previously undiagnosed thyroid dysfunction among those undergoing acute psychiatric treatment is reported to be on the order of 1%–2% (White and Barraclough 1989; Lederbogen et al., in preparation).

Typical psychopathological phenomena produced by hyperthyroidism are emotional lability, psychomotor agitation, and anxious and depressive states (Heuser 1993). It is noteworthy that elderly patients with hyperthyroidism may also manifest apathetic–depressive syndromes while simultaneously losing weight; this constellation may bring a neoplastic process to mind. In other cases, manifestations resembling those of organic or schizophreniform psychosis may be produced (Lishman 1998). Interestingly, psychotic syndromes also seem to be precipitated by the treatment of hyperthyroidism in rare cases (Irwin et al. 1997); a rapid decline in hormone production or a transient hypothyroid state has been suggested as the explanation (Lishman 1998).

Affective abnormalities are often prominent in the clinical picture of hypothyroidism. Depressive mood alteration, with prominent apathetic–lethargic manifestations, is typical (Lishman 1998), while agitated depression may also be observed (Heuser 1993). Psychoses of variable phenomenology may occur, just as in hyperthyroidism (Lishman 1998). Acquired hypothyroidism is, lastly, one of the more important causes of potentially reversible dementia (Mumenthaler 1987), just as congenital hypothyroidism leads to an impairment of intelligence whose severity depends on the duration and severity of the hormone deficiency (Postellon and Abdallah 1986).

Under modern conditions of medical care in industrial countries, massive psychopathological abnormalities (such as psychosis) due to thyroid dysfunction are only rarely seen, presumably because thyroid hormone substitution or antithyroid therapy is generally initiated while the disease is still at a relatively early stage (Baumgartner 1993). With regard to the psychopathological effects of subclinical hyper- and hypothyroidism, varying opinions can be found in the literature. While some authors maintain that these disturbances can, in principle, evoke manifestations similar to, but less pronounced than, those of frank hyper- and hypothyroidism (Haggerty et al. 1993), others state that no association of this type has been conclusively demonstrated (Baumgartner 1993). There are also no generally accepted recommendations for the treatment of subclinical disorders of thyroid function (Stern and Prange 1995).

4.3 Disorders of Parathyroid Function

Psychopathological manifestations of hyper- and hypoparathyroidism are largely similar (Heuser 1993). Both conditions are frequently associated either with depressive mood disturbances, typically accompanied by fatigue and lack of initiative (Heuser 1993), or with cognitive deficit syndromes. The latter may be severe enough to qualify as dementia or delirium (Lishman 1998). Mental disorders are commonly seen in association with disorders of parathyroid function: Petersen (1967) showed, in a study of 54 patients suffering from hyperparathyroidism, that mild mental abnormalities were present in approximately one third, and severe mental abnormalities in a further third of the patient group.

It is presumed that the psychopathological manifestations of parathyroid dysfunction are directly caused not by the abnormally high or low peripheral concentration of parathyroid hormone, but rather by the resulting hyper- or hypocalcemia. The psychopathological manifestations of hyper- and hypocalcemia are largely independent of etiology, and their severity is significantly correlated with the magnitude of the deviation of the serum calcium concentration from the normal range (Petersen 1967).

Now that the determination of serum calcium concentration has become part of clinical routine, and an assay for intact parathyroid hormone is available, milder cases of hormone deficiency and excess are being diagnosed much more commonly than before. Because of this, the number of diagnosed

cases of primary hyperparathyroidism, in particular, has increased (Raue 1996). The serum calcium concentration is often only mildly or moderately elevated (i.e. less than 3 mmol/l). It has not yet been established whether milder degrees of parathyroid dysfunction can produce cause psychopathological abnormalities; the authors of a recent review of the literature conclude that they can (Okamoto et al. 1997).

4.4 Disorders of Adrenal Cortical Function

Mental disturbances are observed in 40% or more of patients with Cushing's disease, which is caused by excessive endogenous secretion of steroid hormone by the adrenal cortex (Heuser 1993; Stern and Prange 1995). Depressive syndromes are the type most commonly encountered (Kelly 1996). Affective disorders are also the most common psychopathological side effects of the therapeutic administration of glucocorticoids; in this situation, hypomanic syndromes with euphoria seem to be more common than in Cushing's disease (Kershner and Wang-Cheng 1989; Naber et al. 1996).

In principle, however, hypercortisolism of whatever etiology may produce psychopathological manifestations of any kind. The same is true of hypocortisolism, a condition characterized by fatigue, marked physical weakness, and marked hypotension (Heuser 1993).

4.5 Pheochromocytoma

Pheochromocytoma is a usually benign secretory tumor of the chromaffin cells of the sympathetic nervous system, located, in 90% of patients, in the adrenal medulla. Paroxysmal hypertension is the hallmark of the disease. More than 60% of patients have baseline hypertension, upon which paroxysms of still higher blood pressure are superimposed (Bornstein 1996). The massively elevated concentration of circulating catecholamines produces the typical psychopathological manifestations of pheochromocytoma: anxiety and the associated characteristic vegetative phenomena.

When due attention is paid to the patient's blood pressure, it is generally not difficult to differentiate cases of pheochromocytoma from the primary anxiety disorders (Rubin and King 1995). It should also be noted that there is a possibility of hypertensive encephalopathy in pheochromocytoma, which may be caused by the dramatically elevated blood pressure, with resulting clouding of consciousness and cognitive deficit syndromes (Lishman 1998).

5 Diagnostic Problems

As discussed in Sect. 1, a considerable percentage of the mentally ill suffer from undiagnosed somatic illnesses. This is so because of various problems that typically arise in the context of medical and psychiatric comorbidity and that may impede the recognition of medical illness. One such problem is that of mental disorders limiting the patient's capacity to perceive and communicate the presence of physical symptoms, as discussed in Sect. 2. Another is the occurrence, in many mental disorders, of diverse physical complaints that may lead treating physicians to suspect medical illness where there is none or, conversely, reinforce a pattern of behavior on the physicians' part in which physical symptoms are primarily regarded as expressions of mental illness (Thiel et al. 1998). Furthermore, the diagnosis of medical disorders in the mentally ill may be hindered by factors such as a lack of competence in internal medicine on the part of the psychiatrist, insufficient experience in dealing with the mentally ill on the part of the internist, or the unavailability of necessary diagnostic technology in psychiatric institutions (Vieweg et al. 1995).

Because many manifestations of illness are nonspecific, it is often difficult to determine whether they are the result of a medical or a psychiatric disorder. There may be difficulty not only with the diagnosis of medical illness in the mentally ill, but also with the diagnosis of psychiatric illness in the medically ill. An example of the latter situation is the frequently missed diagnosis of depression in medical patients (Popkin 1995). One of the reasons for this may be that certain manifestations of depression (e.g. weight loss, lack of initiative) may also be manifestations of medical illness, and their attribution to disorders of one or the other type is highly variable among physicians of different specialties (Linden et al. 1995).

6 Therapeutic Problems

6.1 Treatment of Medical Illness

A number of factors that may complicate the treatment of medical illness in the mentally ill are discussed in

the literature. Certain constellations of psychopathological symptomatology are commonly associated with noncompliance (Adler and Griffith 1991), while, conversely, a lack of acceptance of psychiatric patients on the part of medical institutions may also be observed (Vieweg et al. 1995). Furthermore, the mentally ill may be less able to make use of available medical resources (Felker et al. 1996), problems may arise in the multidisciplinary collaboration of internists and psychiatrists, and, as mentioned in Sect. 5, there may be a lack of expertise in internal medicine within psychiatric institutions.

The indication for medical treatment may result from the presence of a mental disturbance if a medical illness is found to be the underlying etiology. If, as discussed in Sect. 3, the etiology cannot be reliably established, this is usually only of limited relevance for the therapeutic approach. Largely regardless of whether it is assumed that there is an organic mental disorder associated with medical illness or whether a "nonorganic" disorder is diagnosed, the same principles of syndrome-related therapy are applied, as are the therapeutic measures indicated for each particular medical condition.

Problems may arise when the decision about whether a particular mental disorder is due to a general medical condition determines whether potentially harmful treatments are indicated. For example, the decision to treat systemic autoimmune disorders with corticosteroids (Lieb et al. 1997), or primary hyperparathyroidism with parathyroidectomy (White et al. 1996), may largely depend on whether the medical illness is considered the essential etiological factor for the concomitant psychopathological syndrome, or whether the latter is thought to be a causally unrelated "nonorganic" mental disorder. There are no generally applicable rules for such decisions, which may be very difficult. Close cooperation between the psychiatrist and the medical specialist is required in all such cases.

6.2
Treatment of Psychiatric Illness

Concomitant medical illness may have implications for the treatment of psychiatric illness, and these implications must always be taken into account, even when the medical and psychiatric illness are causally unrelated. Physical limitations with respect to mobility or sensory perception, for instance, may make it difficult or impossible to carry out certain behavioral treatment measures that would otherwise be indicated.

The administration of psychotropic medication to medically ill patients is particularly problematic. The numerous special considerations involved in the use of antidepressants by such patients are well known, particularly the effect of tricyclic antidepressants on the cardiovascular system (Glassman et al. 1993). The clinician faces a dilemma: patients with cardiovascular disease are at increased risk for depression (see Chap. 14, this volume), but the available antidepressants often are known to be contraindicated in such patients (tricyclic agents) or have been insufficiently studied in this regard (non-tricyclic antidepressants; Kapfhammer 1998). Recent findings suggest, however, that the serotonin reuptake inhibitors have a more favorable risk–benefit ratio than the tricyclic agents in this group of patients, although it is too early for definite conclusions to be drawn (Roose et al. 1998). Furthermore, the results of studies demonstrating the effectiveness of certain agents in medically healthy depressed patients are not necessarily applicable to patients with accompanying medical illness (Popkin 1995).

7
Conclusions and Future Prospects

Nearly half of all mentally ill patients suffer from comorbid physical illnesses, among which roughly two thirds are within the domain of internal medicine. Medical and psychiatric disturbances interact with each other in many ways, not all of which have been adequately studied to date, and each type of disturbance may, on occasion, be a cause of the other. It is common for medical illnesses to complicate mental disorders, and vice versa. The mentally ill require a considerable amount of general medical care, but the care that they receive is apparently inadequate in many respects (Felker et al. 1996; Jeste et al. 1996). This discrepancy is explained by aspects of the patients' behavior that are the consequence of mental illness as well as physician-related factors and certain shortcomings of psychiatric and medical institutions (Thiel et al. 1998).

The foregoing discussion should make it clear that steps must be taken to improve the general medical care of the mentally ill. These will involve not just ensuring that the required level of basic medical competence is available in psychiatric institutions, but also improving interdisciplinary cooperation between psychiatrists and other medical specialists and among the various services involved in the care of the mentally ill. An important prerequisite for achieving these goals should be the position of psychiatry and psychotherapy as an integral component of general medicine, which might even be reinforced in the future (Kathol et al. 1997).

Appendix. Classification of Organic Mental Disorders According to ICD-10[3]

F00–F03 Dementia
F04 Organic amnestic syndrome
F05 Delirium
F06 Other mental disorders due to brain damage or dysfunction and to physical disease
- F06.0 Organic hallucinosis
- F06.1 Organic catatonic disorder
- F06.2 Organic delusional (schizophrenia-like) disorder
- F06.3 Organic mood (affective) disorders
- F06.4 Organic anxiety disorder
- F06.5 Organic dissociative disorder
- F06.6 Organic emotionally labile (asthenic) disorder
- F06.7 Mild cognitive disorder
- F06.8/6.9 Other disorders, unspecified disorder

F07 Personality and behavioral disorders due to brain disease, damage and dysfunction
F09 Not otherwise specified organic or symptomatic mental disorders

8 References

Adler LE, Griffith JM (1991) Concurrent medical illness in the schizophrenic patient: epidemiology, diagnosis, and management. Schizophr Res 4: 91–107

Akiskal HS (1995) Mood disorders: clinical features. In: Kaplan HI, Sadock BJ (eds) Comprehensive textbook of psychiatry, vol 1, 6th edn. Williams and Wilkins, Baltimore, pp 1123–1152

American Psychiatric Association (1994) Diagnostic and statistical manual of mental disorders: DSM-IV, 4th edn. American Psychiatric Press, Washington, DC

American Psychiatric Association (1999) Practice guidelines for the treatment of patients with delirium. Am J Psychiatry 156 [Suppl 5]: 1–20

Arolt V, Driessen M, Bangert-Verleger A et al (1995) Psychische Störungen bei internistischen und chirurgischen Krankenhauspatienten. Nervenarzt 66: 670–677

Baumgartner A (1993) Schilddrüsenhormone und depressive Erkrankungen – Kritische Übersicht und Perspektiven, Teil I. Nervenarzt 64: 1–10

Bleuler M (1954) Endokrinologische Psychiatrie. Thieme, Stuttgart

Bonner D, Howard R (1995) Treatment resistant depression in the elderly. Int J Geriatr Psychiatry 10: 259–264

Bornstein SR (1996) Phäochromozytom. In: Allolio B, Schulte HM (eds) Praktische Endokrinologie. Urban & Schwarzenberg, Munich, pp 266–272

Brown S (1997) Excess mortality of schizophrenia: a meta-analysis. Br J Psychiatry 171: 502–508

*Caine ED, Grossman H, Lyness JM (1995) Delirium, dementia, and amnestic and other cognitive disorders and mental disorders due to a general medical condition. In: Kaplan HI, Sadock BJ (eds) Comprehensive textbook of psychiatry, vol 1, 6th edn. Williams and Wilkins, Baltimore, pp 705–754

*Cherubin CE, Sapira JD (1993) The medical complications of drug addiction and the medical assessment of the intravenous drug user: 25 years later. Ann Intern Med 119: 1017–1022

Clarfield AM (1988) The reversible dementias: do they reverse? Ann Intern Med 109: 476–486

Cryer PE (1997) Hierarchy of physiological responses to hypoglycemia: relevance to clinical hypoglycemia in type I (insulin dependent) diabetes mellitus. Horm Metab Res 29: 92–96

Davison K, Bagley C (1969) Schizophrenia-like psychoses associated with organic disorders of the central nervous system: a review of literature. In: Herrington RN (ed) Current problems in neuropsychiatry. Headly, Ashford (Br J Psychiatry, special publication 4: 113–184)

Deahl MP (1990) Physical illness and depression: the effects of acute physical illness on the mental state of psychiatric inpatients. Acta Psychiatr Scand 81: 83–86

Dettmers C, Hagendorff A, Lüderitz B, Hartmann A (1997) Progrediente zerebrale Parenchymschäden durch rezidivierende arterielle Hypotonien. Nervenarzt 68: 625–632

Deutsche Gesellschaft für Psychiatrie, Psychotherapie und Nervenheilkunde (1998) Praxisleitlinien in Psychiatrie und Psychotherapie. 1. Behandlungsleitlinie Schizophrenie. Steinkopff, Darmstadt

Eiber R, Berlin I, Grimaldi A, Bisserbe JC (1997) Diabète insulinodépendent et pathologie psychiatrique: revue générale clinique et épidémiologique. Encephale 23: 351–357

Egberts EH (1993) Hepatische Enzephalopathie. In: Schüttler R (ed) Organische Psychosyndrome. Springer, Berlin Heidelberg New York, pp 183–196

Elias MF (1998) Effects of chronic hypertension on cognitive functioning. Geriatrics 53[Suppl 1]:S49–S52

*Felker B, Yazel JJ, Short D (1996) Mortality and medical comorbidity among psychiatric patients: a review. Psychiatr Serv 47: 1356–1363

Fichter MM, Meller I, Schröppel H, Steinkirchner R (1995) Dementia and cognitive impairment in the oldest old. Br J Psychiatry 166: 621–629

Fink P (1990) Mental illness and admission to general hospitals: a register investigation. Acta Psychiatr Scand 82: 458–462

*Förstl H (1999) Organische (und symptomatische) psychische Störungen. In: Berger M (ed) Lehrbuch der Psychiatrie und Psychotherapie. Urban & Schwarzenberg, Munich, pp 259–344

Fricchione G, Bush G, Fozdar M, Francis A, Fink M (1997) recognition and treatment of the catatonic syndrome. J Intensive Care Med 12: 135–147

Fritze J, Schneider B, Lanczik M (1992) Vaskuläre Risikofaktoren bei affektiven Psychosen. Krankenhauspsychiatrie 4: 14–18

Gavard JA, Lustman PJ, Clouse RE (1993) Prevalence of depression in adults with diabetes. An epidemiological evaluation. Diabetes Care 16: 1167–1178

Geldmacher DS, Whitehouse PJ (1996) Evaluation of dementia. N Engl J Med 335: 330–336

[3]The categories F00–F09 exclude disorders induced by psychoactive substance use, which are listed under F10–F19.

Glassman AH, Roose SP, Bigger JT (1993) The safety of tricyclic antidepressants in cardiac patients. Risk-benefit reconsidered. JAMA 2673–2675

*Grohmann R, Rüther E, Schmidt LG (1994) Unerwünschte Wirkungen von Psychopharmaka. Springer, Berlin Heidelberg New York

*Gross G, Huber G (1993) Psychopathologie organischer Psychosyndrome. In: Schüttler R (ed) Organische Psychosyndrome. Springer, Berlin Heidelberg New York, pp 29–39

*Häfner H, Bickel H (1989) Physical morbidity and mortality in psychiatric patients. In: Öhman R, Freeman HL, Franck Holmkvist A, Nielzén S (eds) Interaction between mental and physical illness. Springer, Berlin Heidelberg New York, pp 29–47

Haggerty JJ, Stern RA, Mason GA (1993) Subclinical hypothyroidism: a modifiable risk factor for depression? Am J Psychiatry 150: 508–510

Harris CE, Barraclough B (1998) Excess mortality of mental disorder. Br J Psychiatry 173: 11–53

Hauner H (1998) Verbreitung des Diabetes mellitus in Deutschland. Dtsch Med Wochenschr 123: 777–782

*Heuser I (1993) Endokrine Psychosyndrome. In: Schüttler R (ed) Organische Psychosyndrome. Springer, Berlin Heidelberg New York, pp 53–67

*Hewer W, Förstl H (1998) Häufige internistische Probleme bei psychisch Kranken im höheren Lebensalter. In: Hewer W, Lederbogen F (eds) Internistische Probleme bei psychiatrischen Erkrankungen. Enke, Stuttgart, pp 13–28

Hewer W, Rössler W, Jung E, Fätkenheuer B (1991) Somatische Erkrankungen bei stationär behandelten psychiatrischen Patienten. Psychiatr Prax 18: 133–138

*Huber G (1988) Körperlich begründbare psychische Störungen bei Intoxikationen, Allgemein- und Stoffwechselstörungen, bei inneren und dermatologischen Erkrankungen, Endokrinopathien, Generationsvorgängen, Vitaminmangel und Hirntumoren. In: Kisker KP, Lauter H, Meyer JE, Müller C, Strömgren E (eds) Psychiatrie der Gegenwart, vol. 6, 3rd edn. Springer, Berlin Heidelberg New York, pp 197–252

Huber G (1994) Psychiatrie, 5th edn. Schattauer, Stuttgart

Inouye SK (1998) Delirium in hospitalized older patients. Clin Geriatr Med 14: 745–764

Irwin R, Ellis PM, Delahunt J (1997) Psychosis following acute alteration of thyroid status. Aust NZ J Psychiatry 31: 762–764

Jacobson S (1997) Delirium in the elderly. Psychiatr Clin North Am 20: 91–110

Jakubaschk J, Böker W (1991) Gestörtes Schmerzempfinden bei Schizophrenie. Schweiz Arch Neurol Psychiatr 142: 55–76

*Jeste DV, Gladsjo JA, Lindamer LA, Lacro JP (1996) Medical comorbidity in schizophrenia. Schizophr Bull 22: 413–430

Kane JM, Lieberman JA (eds) (1992) Adverse effects of psychotropic drugs. Guilford, New York

*Kapfhammer HP (1998) Internistische Aspekte der Behandlung mit Antidepressiva. In: Hewer W, Lederbogen F (eds) Internistische Probleme bei psychiatrischen Erkrankungen. Enke, Stuttgart, pp 51–87

Kapfhammer HP, Rothenhäusler HB, Dietrich E et al (1998) Artifizielle Störungen – Zwischen Täuschung und Selbstschädigung. Nervenarzt 69: 401–409

*Kasper S, Jung B (1995) Psychiatrisch relevante Nebenwirkungen der nichtpsychopharmakologischen Pharmakotherapie. Nervenarzt 66: 649–661

Kathol RG, Kick SD, Morrison MF (1997) Let's train psychiatric residents to use their medical skills to meet twenty-first century demands. Psychosomatics 38: 570–575

Kelly WF (1996) Psychiatric aspects of Cushing's syndrome. Q J Med 89: 543–551

Kershner P, Wang-Cheng R (1989) Psychiatric side effects of steroid therapy. Psychosomatics 30: 135–139

Kloss TM, Malessa R, Weiller C, Diener HC (1994) Vaskuläre Demenz im Wandel – eine Übersicht zur vaskulären Demenz von zurückliegenden zu neuen Konzepten. Fortschr Neurol Psychiatr 62: 197–219

Koran LM, Sox HC, Marton KI et al. (1989) Medical evaluation of psychiatric patients. Arch Gen Psychiatry 46: 733–740

Koranyi EK (1979) Morbidity and rate of undiagnosed physical illness in a psychiatric clinic population. Arch Gen Psychiatry 36: 414–419

Koranyi EK, Potoczny WM (1992) Physical illnesses underlying psychiatric symptoms. Psychother Psychosom 58: 155–160

Kovacs M, Mukerji P, Iyengar S, Drash A (1996) Psychiatric disorder and metabolic control among youths with IDDM. A longitudinal study. Diabetes Care 19: 318–323

*Lang C (1994) Demenzen: Diagnose und Differentialdiagnose. Chapman and Hall, London

Langan SJ, Deary IJ, Hepburn DA, Frier BM (1991) Cumulative cognitive impairment following recurrent severe hypoglycemia in adult patients with insulin-treated diabetes mellitus. Diabetologia 34: 337–344

*Lauter H (1988) Organische Psychosen. In: Kisker KP, Lauter H, Meyer JE, Müller C, Strömgren E (eds) Psychiatrie der Gegenwart, vol 6, 3rd edn. Springer, Berlin Heidelberg New York, pp 1–56

Lederbogen F, Hermann D, Hewer W, Henn FA. Prevalence of thyroid function disturbances in psychiatric inpatients (manuscript in preparation)

*Lieb K, Vaith P, Berger M, Bauer J (1997) Immunologische Systemerkrankungen als Differentialdiagnose in der Psychiatrie. Nervenarzt 68: 696–707

Linden M, Borchelt M, Barnow S, Geiselmann B (1995) The impact of somatic morbidity on the Hamilton Depression Rating Scale in the very old. Acta Psychiatr Scand 92: 150–154

Linden M, Kurtz G, Baltes MM et al (1998) Depression bei Hochbetagten: Ergebnisse der Berliner Altersstudie. Nervenarzt 69: 27–37

*Lipowski ZJ (1990) Delirium: acute confusional states. Oxford University Press, New York

Lis CG, Gaviria M (1997) Vascular dementia, hypertension, and the brain. Neurol Res 19: 471–480

*Lishman WA (1998) Organic psychiatry: the psychological consequences of cerebral disorder, 3rd edn. Blackwell Science, Oxford

Lyness JM, Bruce ML, Koenig HG et al (1996) Depression and medical illness in late life: report of a symposium. J Am Geriatr Soc 44: 198–203

Marsh CM (1997) Psychiatric presentations of medical illness. Psychiatr Clin North Am 20: 181–204

Mooradian AD (1997) Pathophysiology of central nervous system complications in diabetes mellitus. Clin Neurosci 4: 322–326

*Mumenthaler M (1987) Behebbare und vermeidbare Demenzen. Schweiz Med Wochenschr 117: 964–967, 1002–1008, 1040–1045

Naber D, Sand P, Heigl B (1996) Psychopathological and neuropsychological effects of 8-days' corticosteroid treat-

ment. A prospective study. Psychoneuroendocrinology 21: 25–31

Okamoto T, Gerstein HC, Obara T (1997) Psychiatric symptoms, bone density and non-specific symptoms in patients with mild hypercalcemia due to primary hyperparathyroidism: a systematic overview of the literature. Endocr J 44: 367–374

O'Keeffe S, Lavan J (1997) The prognostic significance of delirium in older hospital patients. J Am Geriatr Soc 45: 174–178

Ormel J, Kempen GI, Penninx BW et al (1997) Chronic medical conditions and mental health in older people: disability and psychosocial resources mediate specific mental health effects. Psychol Med 27: 1065–1077

Ott A, Stolk RP, Hofman A et al (1996) Association of diabetes mellitus and dementia: the Rotterdam Study. Diabetologia 39: 1392–1397

Pearlson GD, Petty RG (1994) Late-life-onset psychoses. In: Coffey CE, Cummings JL (eds) The American Psychiatric Press textbook of geriatric neuropsychiatry. American Psychiatric Press, Washington, DC, pp 261–277

Petersen P (1967) Die Psychiatrie des primären Hyperparathyreoidismus. Springer, Berlin Heidelberg New York

Poeck K, Hacke W (1998) Neurologie, 10th edn. Springer, Berlin Heidelberg New York

Popkin MK (1995) Consultation-liaison psychiatry. In: Kaplan HI, Sadock BJ (eds) Comprehensive textbook of psychiatry, vol 2, 6th edn. Williams and Wilkins, Baltimore, pp 1592–1605

Postellon DC, Abdallah A (1986) Congenital hypothyroidism: diagnosis, treatment, and prognosis. Compr Ther 12: 67–71

Rabins PV (1991) Psychosocial and management aspects of delirium. Int Psychogeriatr 3: 319–324

Raue F (1996) Primärer Hyperparathyreoidismus. In: Allolio B, Schulte HM (eds) Praktische Endokrinologie. Urban and Schwarzenberg, Munich, pp 280–288

Roose SP, Laghrissi-Thode F, Kennedy JS et al (1998) Comparison of paroxetine and nortriptyline in depressed patients with ischemic heart disease. JAMA 279: 287–291

Rubin RT, King BH (1995) Endocrine and metabolic disorders. In: Kaplan HI, Sadock BJ (eds) Comprehensive textbook of psychiatry, vol 2, 6th edn. Williams and Wilkins, Baltimore, pp 1514–1528

Saupe R, Diefenbacher A (1999) Konsiliarpsychiatrie und -psychotherapie. In: Berger M (ed) Lehrbuch der Psychiatrie und Psychotherapie. Urban and Schwarzenberg, Munich, pp 941–956

Schmitt MF, Hewer W (1993) Lebensbedrohliche Situationen durch Bolusaspiration bei stationär behandelten psychisch Kranken – Klinik, Risikofaktoren, Prophylaxe, Therapie. Fortschr Neurol Psychiatr 61: 313–318

Schneider K (1967) Klinische Psychopathologie, 8th edn. Thieme, Stuttgart

*Seitz HK, Lieber CS, Simanowski UA (1995) Handbuch Alkohol, Alkoholismus, Alkoholbedingte Organschäden. Barth, Leipzig

Shannon KM, Goetz CG (1995) Connective tissue diseases and the nervous system. In: Aminoff MJ (ed) Neurology and general medicine, 2nd edn. Churchill Livingstone, New York, pp 447–471

Skoog I (1998) Status of risk factors for vascular dementia. Neuroepidemiology 17: 2–9

Stern RA, Prange AJ (1995) Neuropsychiatric aspects of endocrine disorders. In: Kaplan HI, Sadock BJ (eds) Comprehensive textbook of psychiatry, vol 1, 6th edn. Williams and Wilkins, Baltimore, pp 241–251

Strachan MWJ, Deary IJ, Ewing FME, Frier BM (1997) Is type II diabetes associated with an increased risk of cognitive dysfunction? Diabetes Care 20: 438–445

Summa JD (1988) Körperliche Erkrankungen als Risikofaktoren von Suizidhandlungen im Alter. In: Böhme K, Lungershausen E (eds) Suizid und Depression im Alter. Roderer, Regensburg, pp 118–129

Thiel A, Nau R, Willers T (1998) Häufige internistische Probleme bei psychisch Kranken im jüngeren und mittleren Lebensalter. In: Hewer W, Lederbogen F (eds) Internistische Probleme bei psychiatrischen Erkrankungen. Enke, Stuttgart, pp 1–12

Tölle R (1990) Organisch bedingte Depressionen. Nervenarzt 61: 176–182

*Trzepacz PT (1996) Delirium: advances in diagnosis, pathophysiology, and treatment. Psychiatr Clin North Am 19: 429–448

Tueth MJ (1994) Emergencies caused by the side effects of psychiatric medications. Am J Emerg Med 12: 212–216

Tune L, Carr S, Hoag E et al (1992) Anticholinergic effects of drugs commonly prescribed for the elderly: potential means for assessing risk of delirium. Am J Psychiatry 149: 1393–1394

Vaillant GE (1998) Natural history of male psychological health. XIV. Relationship of mood disorder vulnerability to physical health. Am J Psychiatry 155: 184–191

Vieweg V, Levenson J, Pandurangi A, Silverman J (1995) Medical disorders in the schizophrenic patient. Int J Psychiatr Med 25: 137–172

Weeke A, Juel K, Vaeth M (1987) Cardiovascular death and manic-depressive psychosis. J Affect Disord 13: 287–292

Weytingh MD, Bossuyt PMM, van Crevel H (1995) Reversible dementia: more than 10% or less than 1%? J Neurol 242: 466–471

White AJ, Barraclough (1989) Benefits and problems of routine laboratory investigations in adult psychiatric admissions. Br J Psychiatry 155: 65–72

White RE, Pickering A, Spathis GS (1996) Mood disorder and chronic hypercalcemia. J Psychosom Res 41: 343–347

World Health Organization (1992) The ICD-10 classification of mental and behavioural disorders. Clinical descriptions and diagnostic guidelines. World Health Organization, Geneva

World Health Organization (1993) The ICD-10 classification of mental and behavioural disorders. Diagnostic criteria for research. World Health Organization, Geneva

Zubenko GS, Marino LJ, Sweet RA et al (1997) Medical comorbidity in elderly psychiatric inpatients. Biol Psychiatry 41: 724–736

CHAPTER

14

A.H. Glassman, E.G.V. Giardina

An Examination of the Linkage Between Ischemic Heart Disease and Depression

This work was supported in part by the Suzanne C. Murphy Foundation, Grant RR-00645 from the Research Resources Administration, Bethesda, Maryland, and a grant from Linda and Peter Nisselson. Parts of the material in this manuscript appeared in an article by Professor Glassman in the American Journal of Psychiatry in January 1998.

1 Introduction

Both literature and folklore have long acknowledged a relationship between depression and mortality. Even our language refers to the relationship between grief and the heart – in speech, poem, and song we describe people dying of a broken heart. However, in spite of this widespread common recognition, scientific evidence supporting these relationships has come only recently.

2 Depression and Mortality

The first scientific report of the relation between emotion and mortality did not occur until the early 1930s. At that time, Malzberg examined the mortality rate of involutional melancholic patients in the New York State civil hospital system and compared it with the mortality of an age-matched sample of the general population (Malzberg 1937). Not surprisingly, the rate was significantly elevated; however, Malzberg's findings did not receive widespread acceptance. The problem with Malzberg's data was that they confused the diagnosis of depression with being an inmate of the New York State asylums. It was not possible to tell which condition was responsible for the observed increase in mortality.

Further scientific inquiry into this question was delayed by the onset of the Second World War. After the war, attention focused not so much on diagnostic categories as on personality. In the United States, two cardiologists developed the concept of the "type A" personality (Rosenman et al. 1964). Early evidence supported the usefulness of this category of "time urgent" personality, and most of American psychological medicine pursued this concept. It was not until the late 1970s that investigators returned to examining the diagnostic concept of depression and its relationship to mortality. A Danish epidemiologist by the name of Weeke surveyed the extensive Danish National Registries. She compared the mortality of all patients admitted to the Danish mental health care system who received a diagnosis of either major depression or manic depressive disease between 1974 and 1978 with the mortality of the general population (Weeke 1979). This approach eliminated the scepticism that had haunted Malzberg's institutional study, because in the Weeke data the vast majority of the patients under observation were outpatients and resided at home. Although the risk of dying was not elevated to the same degree as in Malzberg's data, cardiovascular mortality was increased 50% above the general population.

The major criticism raised about the Weeke data and a number of similar studies is that, in the 40 years since Malzberg's observations, a number of antidepressant drugs had come into widespread use and it was possible that these pharmacological treatments rather than the diagnosis of depression were responsible for increased mortality. Weeke attempted to deal with these concerns in a second study published in 1987 (Weeke et al. 1987). Cardiovascular mortality among depressed patients before and after the introduction of either antidepressant drugs or lithium was compared. Cardiovascular mortality was elevated during both periods; however, the risk was actually higher during the predrug period.

In addition to the two studies by Weeke, seven other investigators examined the relative mortality among patients with major depression and/or bipolar disease and compared that mortality to the general population (Black et al. 1985; Norton and Whalley 1984; Rabins et al. 1985; Sharma and Markar 1994; Tsuang et al. 1980; Vestergaard and Aagaard 1991; Zilber et al. 1989). Eight of the nine studies found evidence of a significantly increased cardiovascular mortality rate. Only a single Israeli study that collected patients and compared their mortality to a nonpatient population failed to find an increased cardiac mortality rate (Zilber et al. 1989). That study by Zilber also found an increased mortality rate, but the increase was primarily from infectious, not cardiovascular, disease.

All nine of these studies collected patient samples from individuals in treatment at the time they were identified. As a result, the relationship that was observed between depression and cardiovascular mortality could not really be extended beyond patients with relatively serious depression who came to those treatment settings. These studies did not address the question of whether milder cases of depression would also be associated with any increased mortality. In addition, although neither the second study by Weeke nor any other available data set (Muller-Oerlinghausen et al. 1992; Avery and Winokur 1976) gave any evidence that antidepressant treatment accounted for the increased mortality rate, it was not possible to eliminate entirely this consideration. By going out into the community and collecting epidemiological, rather than clinical samples, research workers were able to resolve both of these issues.

In the late 1980s and early 1990s, data from two such epidemiological studies became available. Murphy examined the Sterling County data (Murphy et al. 1987). This data set followed 1000 individuals for 16 years, and specific causes of death were available. A diagnosis of depression was explicitly associated with cardiovascular mortality. In the second study of this

type, Bruce used the New Haven data set that had followed 3500 individuals for 9 years (Bruce et al. 1994). She also found that a diagnosis of depression was associated with an increase in natural deaths; however, no information was available as to the specific cause of these excess deaths.

Thus, in the 15 years between the late 1970s and the early 1990s, all 11 studies that had looked for an association between a diagnosis of depression and mortality had found a significantly increased risk of death. In nine of the 11 studies, the increased mortality rate was specifically demonstrated to be cardiovascular in nature. This increased risk had been seen in both unipolar and bipolar patients and in cases that would seem to be both mild and severe. It did not appear in any way to be related to treatment. At that point in time, this relationship appeared to be real. However, in the late 1980s, a serious potential confound became apparent. At that time, the powerful relationship between depression and cigarette smoking began to become apparent (Glassman et al. 1988, 1990). This relationship might readily explain the apparent association between depression and cardiovascular mortality. It could easily be that depression was associated with cigarette smoking and that smoking caused the increased risk of death. None of the 11 studies that had searched for a relationship between depression and cardiovascular disease had in any way attempted to control for smoking.

3
Depression and Heart Disease Controlled for Smoking

One of the early replications of the initial observation linking depression to smoking and to smoking cessation failure came from the Communicable Disease Center (CDC) of the United States Department of Health (Anda et al. 1990). Anda and coworkers realized that a second CDC data set actually contained information that would allow for an analysis of the association between depression and cause-specific mortality while controlling for smoking. The Medical Illness Follow-Up Study contained information on 2832 individuals over the age of 45 who were found to be free of any physical disease after initial physical and laboratory examinations (Anda et al. 1993). These individuals were then followed for a minimum of 12.5 years. In order to preclude the possibility that subjects may have had some awareness of symptomatology that prompted their depression even though it was not revealed by their history, physical, or laboratory examination, patients who died in the first 2.5 years of follow-up were excluded. Nevertheless, after controlling for known cardiovascular risk factors (gender, weight, physical activity, blood pressure, and cholesterol level), including cigarette smoking, apparently healthy individuals that had an elevated depression rating were more likely both to develop and to die of ischemic heart disease (IHD). This study by Anda and colleagues was the first to show that the association between depression and cardiovascular disease persisted even after controlling for smoking and the first to identify the dual risk of both developing and dying of ischemic disease.

In the last few years, six more epidemiological studies have taken a similar approach to different data sets. All studies controlled for smoking, and five of the six studies found a relationship between depression and cardiovascular disease. Ford reported on 1198 former Johns Hopkins medical students followed for a median of 35 years. Men with depression had a higher relative risk of myocardial infarction (MI) (RR, 1.68; 95% CI, 1.03–2.74) than men free of depression (Ford et al. 1994). On average, the first report of depression preceded the first report of cardiovascular disease by 10 years. Aromaa et al. (1994) reported on the Mini-Finland Health Survey follow-up of individuals 40–64 years of age. The study was based on 5355 individuals who had both medical and psychiatric evaluation at baseline and who were followed for an average of 6.6 years. The risks of both developing and dying of IHD were significantly elevated among depressed individuals after controlling for age, education, and traditional IHD risk factors, including smoking.

The Kuoppio Ischemic Heart Disease Study followed a separate sample of 2428 Finnish men over a 6-year period (Everson et al. 1996). Here, very much like the data obtained by Anda and colleagues in the United States, men without any prior history of angina or MI who had higher levels of depression were more likely to experience their first infarction even after controlling for a wide array of biological, behavioral, and social risk factors. There was again an elevated risk of cardiovascular mortality among men with higher depression scores, and again the investigators controlled for traditional cardiac risk factors, including smoking.

Another recently published long-term study that also controlled for smoking is the Glostrup cohort (Barefoot and Schroll 1996). This is a study of 730 individuals born in Glostrup, Denmark in 1915. They had both physical and psychological examinations in 1964 and again in 1974 and were followed for an average of 27 years. After controlling for both smoking and physical health, like the CDC and Finnish data, individuals with elevated depression scores were 65% more likely to develop IHD. Elevated depression scores were also associated with a significant increase in all natural causes of mortality; however, the question of

whether this increased mortality rate was specifically cardiovascular in nature was not addressed.

The most recently published data come from a 13-year follow-up of the Baltimore cohort of the Epidemiological Catchment Area Study (Pratt et al. 1996). Among the 1551 respondents apparently free of medical illness at baseline, a diagnosis of major depression increased the risk of MI more than fourfold after controlling for both medical risk factors and other psychiatric diagnoses. This study is interesting because it used a DSM-III diagnosis of depression, is the only study to control for other psychiatric diagnoses, and included the category of dysphoria in the same population. Individuals that were dysphoric but never met lifetime criteria for major depression were intermediate in their risk for MI compared to patients with DSM-III major depression and individuals totally free of depression.

Only one study that started with healthy individuals and controlled for all the usual medical risk factors failed to find a relation between depression and new IHD and/or cardiovascular mortality, and that is the 15-year follow-up of 2573 members of the Northwest region of Kaiser Permanente (Vogt et al. 1994). Only 1399 of these individuals were over the age of 45. Even here, the authors refer to a "possible" relationship between mortality and depression among men.

Thus, since 1993, six of the seven studies that started with healthy individuals and controlled for cardiovascular risk factors, including smoking, found an association between increasing levels of depression and the onset of new IHD and/or cardiovascular death. In addition to these seven studies that either started with healthy individuals or controlled for existing medical risks, there are also two studies with patients who were in treatment for hypertension. Wassertheil-Smoller and coworkers (1996) followed 4367 otherwise healthy hypertensive patients for 5 years, while Simonsick et al. (1995) observed 3561 hypertensive individuals for 6 years. Both groups were controlled for smoking. The depressed individuals in both studies were found to be at increased risk for cardiovascular death. Wassertheil-Smoller and colleagues also found an excess risk for both stroke and MI. Thus, since the need to control for cardiovascular risk factors, including smoking, became apparent in the early 1990s, eight of nine epidemiological studies have found a significant association between depression and cardiovascular disease or death.

4 Known Cardiovascular Disease and Depression

In 1988, Carney, working at Washington University in St. Louis, adopted a somewhat different strategy. He started not with individuals that were healthy or essentially healthy, but deliberately began with patients who had documented IHD (Carney et al. 1988). He performed structured psychiatric examinations on consecutive patients undergoing coronary angiography. Of the 52 patients whose studies confirmed the presence of coronary artery disease, just less than 20% met criteria for major depression. This depressed population was two and a half times more likely to develop a serious adverse cardiac complication over the next 12 months. Although Carney's sample size was too small to adequately control for other risk factors, this was corrected in subsequent studies using this strategy.

Ahern obtained baseline ratings of anxiety, anger, and depression in 350 patients with ventricular arrhythmia following a heart attack (Ahern et al. 1990). Those postinfarction patients who survived the first year had lower baseline depression scores than nonsurvivors after controlling for the usual medical and social risk factors. Neither anger nor anxiety scores were associated with survival. Several other investigators evaluating post-MI patients also found a higher mortality rate associated with depression (Frasure-Smith et al. 1993; Frasure-Smith 1991; Ladwig et al. 1991). The most convincing of these studies was by Frasure-Smith et al. (1993).

Frasure-Smith and Lesperance obtained a structured psychiatric examination in 222 patients 5–15 days after an MI and contacted them 6, 12, and 18 months after discharge from the hospital. Sixteen percent showed evidence of major depression while hospitalized for their index MI. This is consistent with other surveys that regularly show 15%–20% of postinfarction patients developing major depression (Hance et al. 1996; Schleifer et al. 1989). At 6 months, approximately 17% of the depressed patients had died compared to 3% of the nondepressed patients (Frasure-Smith et al. 1993). Controlling for other independent predictors of risk, the relative hazard for depressed patients was almost three and a half times greater. The strongest predictor of mortality after an infarct is generally heart failure, and similarly it is associated with a three and a half times greater mortality rate.

Although concern has been expressed about the reliability of a psychiatric diagnosis made in a medically ill patient, and especially in one who is post-MI, only 25% of the patients Frasure-Smith and Lesperance identified as depressed were free of major depression at both the 6- and 12-month follow-up interviews (Lesperance et al. 1996). This compares with 81% of those patients who had not originally received a diagnosis of depression remaining free of depression. Similarly, Hance et al. (1996) found that half the patients diagnosed with major depression in the immediate post-MI period either remained depressed or relapsed within 12 months.

In addition to a diagnosis of major depression, Frasure-Smith also measured symptomatic depression using the Beck Depression Inventory (Beck et al. 1961). At 6 months after MI, patients with elevated Beck scores but free of major depression had mortality rates much like those of patients free of any depression. However, by 12 months, patients with Beck scores above 10 were dying at a significantly higher rate than patients free of depression, and by 18 months their mortality rate was not much less than those with major depression (Frasure-Smith et al. 1995a). It is not clear whether increased mortality is really associated with lesser degrees of depression or whether high Beck scores are merely identifying a group at high risk for major depression. In the study by Hance and coworkers, 42% of patients with minor depression developed major depression over the subsequent 12 months (Hance et al. 1996).

Recently, Barefoot and coworkers at Duke University in Durham, North Carolina, published a 17-year follow-up of patients admitted to the cardiology service with a diagnosis of coronary artery disease. This study combines the strategy of using a high-risk population and long-term follow-up. Once again, depression was associated with diminished survival in patients with preexisting cardiovascular disease (Barefoot et al. 1996).

5
Anxiety and Cardiovascular Disease

Taken together, the studies described above constitute compelling evidence linking depression to an increased risk of IHD. However, a number of issues need to be clarified. One of these issues is whether the association between depression and cardiovascular disease is unique to depression or whether similar associations exist with all negative affective states. Although there is more information available about depression than any other negative affect, there are extensive data concerning anxiety. In thinking about the impact of any affective state, it is important to realize that the short-term effect of mood states, i.e. over hours or days, could be different from its effect over months or years. There are three long-term epidemiological studies that have examined the effects of anxiety (Kawachi et al. 1994a,b; Haines et al. 1987). However, all anxiety states may not have the same consequences, and all authors do not define anxiety in the same way. Examining the available data would suggest that, over the long term, being an anxious person or having a nonspecific anxiety diagnosis does not increase the likelihood of developing or dying of IHD. However, phobic anxiety or panic states do over the long term seem to be associated with an increased risk of death, particularly sudden death (Kawachi et al. 1994a; Haines et al. 1987). Over the short term, much fewer data are available. Frasure-Smith has examined the impact on survival of high anxiety scores following infarction over the subsequent 6–18 months. She found that high scores were associated with an increased likelihood of developing cardiac complications. However, in a meta-analysis of multiple "type A" personality studies following infarction, anxiety played a very minor role if it played any role at all (Booth-Kewley and Friedman 1987).

6
Anger and Cardiovascular Disease

The influence of anger in distinction to symptoms of anxiety or depression has been examined in a number of studies. Many of these studies suggest that increased anger ratings are associated with more rapid progression of IHD (Booth-Kewley and Friedman 1987; Kawachi et al. 1996; Julkunen et al. 1994). The problem is that almost none of the studies that carefully measured anger also obtained reliable measures of depression. This is particularly troublesome because anger and depression frequently coexist and measures of depression are so consistently associated with cardiovascular disease. To establish that anger is independently associated with either a more malignant course or an increased mortality rate in IHD would necessitate that both mood states be measured. Two studies that did measure both failed to find any independent contribution of anger; however, both of these studies specifically studied post-infarction patients and both followed patients for approximately 1 year (Frasure-Smith et al. 1995b; Ahern et al. 1990).

7
What Is It About Depression that Carries the Risk

Another unresolved issue is precisely what it is that carries the risk. In phenomenological terms, the studies that have been mentioned up to this point have identified populations at increased risk using either a diagnosis of major depression, various scales that measure symptoms of depression, or the hopelessness item of the Well-Being Scale (Dupuy 1974). The obvious question is whether the increased risk travels specifically with the diagnosis of major depression, whether it varies with the severity or chronicity of a depressive state, or whether it can be associated with symptoms of depression in the absence of a diagnosis.

There are only very limited data available to address these distinctions. In the short term following a heart attack, Lesperance et al. (1996) found that the risk of death is strongly associated with recurrent major depression, but that a single episode of major depression occurring for the first time after an MI was not associated with any increased risk of mortality. Because of the manner in which Frasure-Smith and Lesperance collected their follow-up data, it is not possible to say whether the excess deaths occurred among those patients who were persistently depressed after infarction, whether those who died had their depression return prior to their death, or whether the mere occurrence of depression after infarction in those with a history of prior depression marked a group at higher risk of death whether or not they continued to be depressed. In the long-term studies carried out by Anda et al. (1993), Everson et al. (1996), and Barefoot and Schroll (1996), it was found that the higher the symptom scores at baseline, the higher the risk both to develop IHD and to die of it. However, it is again not clear whether this increased risk is the result of these individuals merely having more severe symptoms of depression. Alternatively, these higher symptom scores may have identified individuals who were more likely to develop major depression and the risk is concentrated in those individuals who developed major depressive episodes. Everson et al. (1996) assert that hopelessness is a state that can be separated from depression and that this state confers a risk independent of depression. In general, these remain open issues.

In addition to the studies that have been described up to this point, a series of studies from the Netherlands have examined a condition described as "vital exhaustion." This state is characterized by a lack of energy, increased irritability, and feelings of demoralization (Appels et al. 1987). Vital exhaustion sounds similar to depression and seems to demonstrate a similar association with cardiovascular disease. In long-term prospective studies, it was associated with an increased relative risk of initial MI (Appels and Mulder 1988), and it is associated with an increase in adverse cardiac events in the year and a half following angioplasty (Appels et al. 1995). It would seem very likely that recurrent major depression and vital exhaustion tap the same underlying process, but the data available do not allow us to test that hypothesis.

Ultimately, whether these descriptive entities are measuring the same or different processes will be established by identifying the mechanism or mechanisms that lie behind these associations. In the Montreal study (Frasure-Smith et al. 1995a), the excess mortality among the patients with post-MI depression was due almost exclusively to sudden death. Because sudden death is almost always the result of ventricular arrhythmia, Frasure-Smith looked to see whether there was an interaction between depression and post-MI arrhythmia. There was, in fact, a striking increase in the mortality rate among those post-MI patients who had both depression and even mild baseline ventricular arrhythmia (Frasure-Smith et al. 1995a). There is considerable evidence that fluctuations in autonomic tone influence the risk of ventricular fibrillation and sudden death (Schwartz et al. 1992). Changes in autonomic tone have long been considered an integral part of serious depression, and the direction of those changes is such that one would anticipate an increase in sudden death (Carney et al. 1995; Dalack et al. 1992). This could easily explain a good part of the increased mortality associated with depression following infarction. However, it would be an unlikely explanation of the increased rate of new infarction seen in depressed individuals initially free of any cardiac disease.

8
Platelet Studies and Depression

A number of investigators have reported both increases in platelet serotonin ($5\text{-}HT_2$) binding and a decreased platelet 5-HT transporter site density in depressed patients (Nemeroff et al. 1994). These were seen as abnormalities associated with the biology of depression, but not as abnormalities that would influence the clotting function of the platelet. When Anda first noticed that depression was associated not just with cardiac mortality, but also with the onset of new cases of IHD, it occurred to him that these platelet abnormalities might actually have functional significance. As a result of Anda's suggestion, Musselman and colleagues compared platelet function in depressed patients and normal controls. In a particularly sophisticated study, they found that depressed patients had a heightened susceptibility to platelet activation (Musselman et al. 1996).

Musselman compared healthy depressed patients and healthy normal controls. Perhaps more surprisingly, Laghrissi-Thode et al. (1997) found evidence of increased platelet activation in a group of depressed patients comorbid for cardiovascular disease. This abnormality was observed in spite of the fact that many of these patients were already taking aspirin because of a prior MI. Laghrissi-Thode's methodology was not as sophisticated as Musselman's, but taken together their results strongly suggest that depression alters platelet function such that one would expect an increased rate of both heart attacks and strokes. In fact, there is considerable evidence that depression's effect is not limited to cardiovascular disease but

involves all vascular disease, including stroke (Morris et al. 1993a,b; Wassertheil-Smoller et al. 1996).

It would be easy to think that depression in association with cardiovascular disease is nothing more than a natural reaction to a stressful life event. Depressions are certainly common among medically ill patient populations. However, numerous epidemiological studies have shown that the onset of depression precedes cardiovascular disease, and in addition the absolute rate of depression seen in the post-infarction period is far greater than that seen in other medically stressful situations. A number of studies have examined the onset of symptoms of major depression in the first 10 days or so following MI (Hance et al. 1996; Schleifer et al. 1989). This 10-day rate is triple that ordinarily expected in a 6-month incidence rate and double that associated with chronic medical illness. In addition, in follow-up studies of post-MI depression, the depression either becomes recurrent or persists over the next 6 months in at least 50% of patients (Travella et al. 1994; Lesperance et al. 1996).

9 Conclusion

What has been presented to the reader is a series of studies that document two associations. First, depression is associated with an increased rate in the onset of new cases of IHD. Furthermore, depression is also associated with a higher rate of cardiac mortality among patients with preexisting IHD, especially those in the postinfarction period. It would be easy to think that depression causes this worsening. Increased platelet aggregation and the decrease in various measures of heart rate variability shown to be a finding in postinfarction death (Dalack et al. 1992) as well as the poor health behavior seen in depression all suggest a causal relationship. However, at this time, what has been proven is still only an association. It is conceivable that a genetic predisposition could put an individual at risk for both depression and arteriosclerotic disease. Evidence that some forms of late-onset depression may be secondary to arteriosclerotic disease in the brain support this view (Krishnan 1993). At this point, these issues need to be resolved. Although we suspect that certain characteristics of chronic depression contribute to the relationship with IHD, that relationship could be a two-way street and the process of arteriosclerosis in the brain could increase the risk for chronic depression.

The other question raised by the relationship between depression and mortality is whether successful treatment of depression would reduce cardiovascular mortality. However, even before we test whether treatment reduces mortality, we need to establish that treatment is safe and effective in a seriously ill cardiac population. While tricyclic antidepressants are not suitable in this patient sample (Glassman et al. 1993), serotonin reuptake inhibitors would seem to be more benign drugs in this population. To date, the data are inadequate to be certain about either their safety or efficacy, and evaluation of safety and efficacy are critical to this issue.

10 References

Ahern DK, Gorkin L, Anderson JL, Tierney C, Hallstrom A, Ewart C, Capone RJ, Schron E, Kornfeld D, Herd JA, Richardson DW, Follick MJ, Cardiac Arrhythmia Pilot Study (CAPS) Investigators (1990) Biobehavioral variables and mortality or cardiac arrest in the Cardiac Arrhythmia Pilot Study (CAPS). Am J Cardiol 66: 59–62

Anda RF, Williamson DF, Escobedo LG, Mast EE, Giovino GA, Remington PL (1990) Depression and the dynamics of smoking; a national perspective. JAMA 264: 1541–1545

**Anda RF, Williamson DF, Jones D, Macera C, Eaker E, Glassman AH, Marks J (1993) Depressed affect, hopelessness, and the risk of ischemic heart disease in a cohort of U.S. adults. Epidemiology 4: 285–294

Appels A, Mulder P (1988) Excess fatigue as a precursor of myocardial infarction. Eur Heart J 9: 758–764

Appels A, Hoppener P, Mulder P (1987) A questionnaire to assess premonitory symptoms of myocardial infarction. Int J Cardiol 17: 15–24

Appels A, Kop W, Bar F, de Swart H, Mendes de Leon C (1995) Vital exhaustion, extent of atherosclerosis, and the clinical course after successful percutaneous transluminal coronary angioplasty. Eur Heart J 16: 1880–1885

Aromaa A, Raitasalo R, Reunanen A, Impivaara O, Heliovaara M, Knekt P, Lehtinen V, Joukamaa M, Maatela J (1994) Depression and cardiovascular diseases. Acta Psychiatr Scand Suppl 377: 77–82

Avery D, Winokur G (1976) Mortality in depressed patients treated with electroconvulsive therapy and antidepressants. Arch Gen Psychiatry 33: 1029–1037

Barefoot JC, Schroll M (1996) Symptoms of depression, acute myocardial infarction, and total mortality in a community sample. Circulation 93: 1976–1980

Barefoot JC, Helms MJ, Mark DB, Blumenthal JA, Califf RM, Haney TL, O'Connor CM, Siegler IC, Williams RB (1996) Depression and long-term mortality risk in patients with coronary artery disease. Am J Cardiol 78: 613–617

Beck AT, Ward CH, Mendelson M, Mock J, Erbaugh J (1961) An inventory for measuring depression. Arch Gen Psychiatry 4: 561–571

Black DW, Warrack G, Winokur G (1985) Excess mortality among psychiatric patients: the Iowa record-linkage study. JAMA 253: 58–61

Booth-Kewley S, Friedman HS (1987) Psychological predictors of heart disease: a quantitative review. Psychol Bull 101: 343–362

Bruce ML, Leaf PJ, Rozal GPM, Florio L, Hoff RA (1994) Psychiatric status and 9-year mortality data in the New Haven

Epidemiologic Catchment Area Study. Am J Psychiatry 151: 716–721

Carney RM, Rich MW, Freedland KE, Saini J, Tevelde A, Simeone C, Clark K (1988) Major depressive disorder predicts cardiac events in patients with coronary artery disease. Psychosom Med 50: 627–633

Carney RM, Saunders RD, Freedland KE, Stein P, Rich MW, Jaffe AS (1995) Association of depression with reduced heart rate variability in coronary artery disease. Am J Cardiol 76: 562–564

Dalack GW, Roose SP, Glassman AH, Woodring S, Bigger JT Jr. (1992) Depression, cardiac regulation and sudden death. Proceedings of the American Psychiatric Association New Research Abstracts. American Psychiatric Association, Washington DC, p 193

Dupuy HJ (1974) A concurrent validational study of the NCHS General Well-Being Schedule. Vital and health statistics, series 2, no 73. DHEW publ no (HRA) 78-1347. U.S. Government Printing Office, Washington DC

Everson SA, Goldberg DE, Kaplan GA, Cohen RD, Pukkala E, Tuomilehto J, Salonen JT (1996) Hopelessness and risk of mortality and incidence of myocardial infarction and cancer. Psychosom Med 58: 113–121

Ford DE, Mead LA, Chang PP, Levine DM, Klag MJ (1994) Depression predicts cardiovascular disease in men: the precursors study. Circulation 90(4, part 2): I-614

Frasure-Smith N (1991) In-hospital symptoms of psychological stress as predictors of long-term outcome after acute myocardial infarction in men. Am J Cardiol 67: 121–127

**Frasure-Smith N, Lesperance F, Talajic M (1993) Depression following myocardial infarction: impact on 6-month survival. JAMA 270: 1819–1825

Frasure-Smith N, Lesperance F, Talajic M (1995a) Depression and 18-month prognosis after myocardial infarction. Circulation 91: 999–1005

Frasure-Smith N, Lesperance F, Talajic M (1995b) The impact of negative emotions on prognosis following myocardial infarction: is it more than depression? Health Psychol 14: 388–398

Glassman AH, Stetner F, Walsh BT, Raizman PS, Fleiss JL, Cooper TB, Covey LS (1988) Heavy smokers, smoking cessation, and clonidine: results of a double-blind, randomized trial. JAMA 259: 2863–2866

Glassman AH, Helzer JE, Covey LS, Cottler LB, Stetner F, Tipp JE, Johnson J (1990) Smoking, smoking cessation, and major depression. JAMA 264: 1546–1549

Glassman AH, Roose SP, Bigger JT Jr (1993) The safety of tricyclic antidepressants in cardiac patients: risk/benefit reconsidered. JAMA 269: 2673–2675

Haines AP, Imeson JD, Meade TW (1987) Phobic anxiety and ischemic heart disease. Br Med J 295: 297–299

Hance M, Carney RM, Freedland KE, Skala J (1996) Depression in patients with coronary heart disease: a 12-month follow-up. Gen Hosp Psychiatry 18: 61–65

Julkunen J, Salonen R, Kaplan GA, Chesney MA, Salonen JT (1994) Hostility and the progression of carotid atherosclerosis. Psychosom Med 56: 519–525

Kawachi I, Colditz GA, Ascherio A, Rimm EB, Giovannucci E, Stampfer MJ, Willett WC (1994a) Prospective study of phobic anxiety and risk of coronary heart disease in men. Circulation 89: 1992–1997

Kawachi I, Sparrow D, Vokonas PS, Weiss ST (1994b) Symptoms of anxiety and risk of coronary heart disease: the normative aging study. Circulation 90: 2225–2229

Kawachi I, Sparrow D, Spiro A, III., Vokonas P, Weiss ST (1996) A prospective study of anger and coronary heart diease: the normative aging study. Circulation 94: 2090–2095

*Krishnan KR (1993) Neuroanatomic substrates of depression in the elderly. J Geriatr Psychiatry Neurol 6: 39–58

Ladwig KH, Kieser M, Konig J, Breithardt G, Borggefe M (1991) Affective disorders and survival after acute myocardial infarction: results from the post-infarction late potential study. Eur Heart J 12: 959–964

*Laghrissi-Thode F, Wagner WR, Pollock BG, Johnson PC, Finkel MS (1997) Elevated platelet factor 4 and beta-thromboglobulin plasma levels in depressed patients with ischemic heart disease. Biol Psychiatry 42: 290–295

Lesperance F, Frasure-Smith N, Talajic M (1996) Major depression before and after myocardial infarction: its nature and consequences. Psychosom Med 58: 99–110

Malzberg B (1937) Mortality among patients with involution melancholia. Am J Psychiatry 93: 1231–1238

Morris PLP, Robinson RG, Andrzejewski P, Samuels J, Price TR (1993a) Association of depression with 10-year poststroke mortality. Am J Psychiatry 150: 124–129

*Morris PLP, Robinson RG, Samuels J (1993b) Depression, introversion and mortality following stroke. Aust N Z J Psychiatry 27: 443–449

Muller-Oerlinghausen B, Ahrens B, Grof E, Grof P, Lenz G, Shou M, Simhandl C, Thau K, Volk J, Wolf R, Wolf T (1992) The effect of long-term lithium treatment on the mortality of patients with manic-depressive and schizoaffective illness. Acta Psychiatr Scand 86: 218–222

Murphy JM, Monson RR, Olivier DC, Sobol AM, Leighton AH (1987) Affective disorders and mortality. A general population study. Arch Gen Psychiatry 44: 473–480

*Musselman DL, Tomer A, Manatunga AK, Knight BT, Porter MR, Kasey S, Marzec U, Harker LA, Nemeroff CB (1996) Exaggerated platelet reactivity in major depression. Am J Psychiatry 153: 1313–1317

Nemeroff CB, Knight DL, Franks J, Craighead WE, Krishnan KRR (1994) Further studies on platelet serotonin transporter binding in depression. Am J Psychiatry 151: 1623–1625

Norton B, Whalley LJ (1984) Mortality of a lithium-treated population. Br J Psychiatry 145: 277–282

Pratt LA, Ford DE, Crum RM, Armenian HK, Gallo JJ, Eaton WW (1996) Depression, psychotropic medication, and risk of myocardial infarction: prospective data from the Baltimore ECA follow-up. Circulation 94: 3123–3129

Rabins PV, Harvis K, Koven S (1985) High fatality rates of late-life depression associated with cardiovascular disease. J Affect Disord 9: 165–167

Rosenman RH, Friedman M, Straus R, Wurm M, Kositchek R, Hahn W, Werthessen NT (1964) A predictive study of coronary heart diease: the Western Collaborative Group Study. JAMA 189: 15–22

Schleifer SJ, Macari-Hinson MM, Coyle DA, Slater WR, Kahn M, Gorlin R, Zucker HD (1989) The nature and course of depression following myocardial infarction. Arch Intern Med 149: 1785–1789

Schwartz PJ, La Rovere MT, Vanoli E (1992) Autonomic nervous system and sudden cardiac death: experimental basis and clinical observations for post-myocardial infarction risk stratification. Circulation 85[Suppl I]: I77–I91

Sharma R, Markar HR (1994) Mortality in affective disorder. J Affect Disord 31: 91–96

Simonsick EM, Wallace RB, Blazer DG, Berkman LF (1995) Depressive symptomatology and hypertension-associated morbidity and mortality in older adults. Psychosom Med 57: 427–435

Travella JI, Forrester AW, Schultz SK, Robinson RG (1994) Depression following myocardial infarction: a one year longitudinal study. Int J Psychiatry Med 24: 357–369

Tsuang MT, Woolson RF, Fleming JA (1980) Premature deaths in schizophrenia and affective disorders: an analysis of survival curves and variables affecting the shortened survival. Arch Gen Psychiatry 37: 979–983

Vestergaard P, Aagaard J (1991) Five-year mortality in lithium-treated manic-depressive patients. J Affect Disord 21: 33–38

Vogt T, Pope C, Mullooly J, Hollis J (1994) Mental health status as a predictor of morbidity and mortality: a 15-year follow-up of members of a health maintenance organization. Am J Public Health 84: 227–231

Wassertheil-Smoller S, Applegate WB, Berge K, Chang CJ, Davis BR, Grimm R Jr, Kostis J, Pressel S, Schron E (1996) Change in depression as a precursor of cardiovascular events. Arch Intern Med 156: 553–561

Weeke A (1979) Causes of death in manic-depressives. In: Schou M, Stromgren E (eds) Origin, prevention and treatment of affective disorders. Academic, London, pp 289–292

Weeke A, Juel K, Vaeth M (1987) Cardiovascular death and manic-depressive psychosis. J Affect Disord 13: 287–292

Zilber N, Schufman N, Lerner Y (1989) Mortality among psychiatric patients – the groups at risk. Acta Psychiatr Scand 79: 248–256

CHAPTER

15

A. Tortorella, P. Monteleone

Psychiatric Syndromes in Infectious Diseases

1 Introduction

Infectious diseases of the central nervous system (CNS) are characterized mainly by neurological symptoms and signs. However, in several cases, psychiatric features not only accompany neurological manifestations, but even predominate in the clinical picture, being the early or even the only clinical manifestations of the disease. In this chapter, we will consider those infectious diseases of the CNS in which psychiatric manifestations are prominent.

Infectious agents can reach the CNS either by hematogenous spread from extracranial foci or by retrograde propagation of infected thrombi within emissary veins. Viruses and bacteria can invade different sites in the CNS (brain parenchyma, meninges, spinal cord) and peripheral nerves, producing inflammatory reactions (encephalitis, meningitis, meningoencephalitis, brain abscess) and/or degenerative processes.

The cellular composition of CNS, its sequestration from the rest of the body by the blood–brain barrier, its close confinement within rigid skeletal structures, and the presence of lymphatic system only within the epidural space greatly influence the course of brain infections (Greenlee 1995).

Many viruses, bacteria, and fungi do not present an elective tropism for specific brain regions; hence these agents induce a global impairment of brain functioning, which may be responsible for a nonspecific organic psychiatric syndrome. In contrast, other infectious agents may localize in some brain areas; in these cases, the functional specialization of the involved brain region can produce characteristic psychiatric symptoms. In some cases, the infectious agent appears to consist of proteins and is devoid of functional nucleic acid; such agents are designated as prions (i.e. proteinaceous infectious agents) in order to distinguish them from viruses.

Finally, systemic infectious diseases with no direct localization in the CNS may induce psychiatric manifestations as a consequence.

2 Nonspecific Psychiatric Syndrome Due to Infectious Agents

The generalized psychiatric syndrome which follows some infections of the CNS consists of a nonspecific set of acute symptoms often evolving in two stages (Sheld et al. 1991; Mandell et al. 1995).

The first stage is characterized by insomnia or hypersomnia, irritability, restlessness, forgetfulness, difficulty in thinking, inability to concentrate, and distractibility. These symptoms are frequently accompanied by fever, headache, meningism, and other neurological signs. If the infectious agent is brought under control, this clinical picture subsides in a few days. If the infection persists, the symptomatology evolves into the second stage.

The second stage is characterized by the clinical picture usually referred to as delirium, in which the most important feature is the impairment of consciousness, often varying in intensity through the day and being worse at night. Visual hallucinations and memory deficits frequently coexist and affect registration, retention, and recall. On recovery, there is usually amnesia for most of the illness. Normally, these manifestations disappear if a specific therapy is available and it is possible to control and overcome the infection.

In some cases, several signs can persist as sequelae. Patients can present personality changes, such as irritability, loss of spontaneity, and drive, reduced control of aggressive impulses, and/or persistent cognitive deficits, ranging from slight defects in intellectually demanding activities to a stable residual dementia.

3 Psychiatric Syndromes in Specific Infectious Diseases

Many infections of the CNS can present a more specific symptomatology, related either to the infectious agent and its elective tropism for certain brain areas and/or to the host reaction. They include neurosyphilis, Lyme disease, tuberculous meningitis, encephalitides, subacute sclerosing panencephalitis (SSPE), and brain abscess.

3.1 Neurosyphilis

Syphilis is a systemic infectious disease caused by the spirochete *Treponema pallidum*. It is transmitted by sexual contact, and the percentage of patients with CNS infection has been increasing since the advent of the human immunodeficiency virus (HIV) epidemic (Musher et al. 1990; Kinghorn 1993). The psychiatric signs of syphilis may be present at any time during the evolution of the disease, which evolves in three phases.

During the first phase, psychopathological signs are nonspecific and rare. In fact, depressive and anxious

reactions, and in some cases suicide, are usually due to the patient's awareness of his or her disease (Lukehart et al. 1988).

In the second (or septic) phase, *Treponema pallidum* spreads throughout the whole body. The most common psychiatric symptoms are sleep disturbances, distractibility, irritability, disorientation, confusion, depression, anxiety, and delusional and hypochondrial worries. This phase may also be characterized by periods of elevated mood, confabulation, and auditory and, to a lesser extent, visual hallucinations (Tramont 1995).

The third phase is characterized by the neurologic and psychiatric signs of neurosyphilis, which can assume the clinical presentation of meningovascular syphilis, tabes dorsalis, or general paresis. As a general rule, meningovascular (inflammatory) syndromes occur within a few years of infection, while parenchymatous (degenerative) syndromes such as tabes dorsalis and general paresis may have a latency of decades.

Patients with meningovascular syphilis may show the same symptomatology as in the second phase (e.g. insomnia, depression, anxiety), but the disease is more often characterized by intellectual weakness, followed by a subconfusional or confusional state. Irritability, mood disturbances, and memory impairment are usually present. Sometimes, this state can evolve with the appearance of oneiric, delusional agitation or stupor, which are frequently consequences of ictal events related to cerebral arteritis (Simon 1985; Hook 1991). Very occasionally, it is possible to observe focal neurological signs (hemiplegia, aphasia, epilepsy), amnesia, or a schizophrenic-like state.

Mental disturbances are not usually part of the clinical picture of tabes dorsalis. When they occur, the patients usually show a psychotic state with impairment of consciousness. Sometimes, they may present with dementia or chronic psychosis, with persecutory delusions related to clear sensory disturbances.

General paresis, although rare, is the most important psychiatric complication of neurosyphilis. It usually develops 10–15 years after the infection, and patients present with personality changes, dysphoric or elevated mood, and striking lapses in social functioning. Any other combination of psychiatric symptoms may occur (Rundell and Wise 1985; Emsley et al. 1988; Sirota et al. 1988; Roberts and Emsley 1992). As the disease progresses, it is possible to distinguish three main clinical pictures: (1) a megalomanic syndrome with grandiosity and elation without insight, (2) a depressive syndrome with nihilistic delusions and hypochondrial worries, and (3) a primitive and deep dementia, involving the patient's whole personality. Neurological signs such as paralysis, ataxia, and seizures may be also present. In untreated cases, death normally occurs within 4–5 years.

Although the incidence of all forms of syphilis has decreased in industrialized countries, there is now a rise in the number of cases in the groups at risk for AIDS (Temmerman et al. 1992). It may be predicted that, as the survival of HIV-infected subjects is prolonged by the new treatments that are becoming available, psychiatrists will face a recrudescence of mental complications of neurosyphilis.

3.2 Lyme Disease

Lyme disease is a multisystemic illness, caused by the tick-borne spirochete *Borrelia burgdorferi*. Because of its spirochetal etiology and its course, with early skin localization and rapid invasion of the CNS, Lyme disease has been believed to be similar to syphilis. It has been reported in several countries around the world and throughout the United States, where a dramatic increase in its incidence has recently occurred (Centers for Disease Control 1993).

Typical clinical manifestations of Lyme disease involve an initial erythematous annular rash at the site of the bite of a tick infected with the pathogenic agent. Within a few weeks after the skin reaction, the spirochetes disseminate to the CNS, where they may remain silent for months to years before producing symptoms. In 15%–40% of patients, neurological problems may be the presenting symptoms, involving initial headache, followed by meningitis, motor or sensory radiculitis, and encephalitis with fluctuating disturbances of mood, concentration, memory, and sleep (Reik 1992).

Later-stage illness generally affects the joints, eyes, and skin, and patients may present with an encephalopathy, characterized by subtle to severe cognitive changes, fatigue, sleep disturbance, extreme irritability or mood lability, and spatial disorientation (Fallon et al. 1992). At this stage, a distinction between an organic mood disorder and a concomitant major depression may be quite difficult to make. Severe cases of anxiety, panic, and obsessive-compulsive disorders have also been associated with Lyme disease.

3.3 Central Nervous System Tuberculosis

Tuberculous meningitis is the most common form of tuberculosis in the CNS. The immunosuppression induced by HIV makes more people susceptible to tuberculous infection and, in a West African study, tuberculous meningitis was found to be the cause of death in 11% of HIV-infected patients (Lucas et al.

1993). Less frequent psychiatric manifestations of tuberculosis are a subacute psychosis and a neurasthenic syndrome related to a latent infection.

The clinical picture of tuberculous meningitis is typically heterogeneous, with prodromes lasting from 1 week to several months characterized by irritability, sadness, anxiety, asthenia, thought slowness, and memory impairment. Most frequently, an atypical depression develops, but a delusional and hallucinatory syndrome with possible antisocial reactions can also occur and rapidly subside.

Subacute psychosis is a very rare complication of a developing tuberculosis. It is characterized by subacute mental confusion with either anxiety or euphoria, delusions, and multiple hallucinations. This clinical picture is normally transient, and regression can be obtained more rapidly with specific treatment.

The neurasthenic syndrome related to latent tuberculosis is characterized by asthenia, anergy, difficulty in thinking, inability to concentrate, and distractibility, probably secondary to a generalized toxic, infectious state.

3.4 Encephalitides

Encephalitis may be either the consequence of a primary viral infection of the brain or a complication of bacterial meningitis, general sepsis, or brain abscess. In the acute stage, the above-mentioned nonspecific psychiatric syndrome usually occurs. More relevant for psychiatrists are the complications that follow the acute episode, including dementia, prolonged depressive or anxiety states, and personality changes (Lishman 1987).

Among all the encephalitides, epidemic encephalitis, although nowadays rare and sporadic, deserves special mention because of the frequency of psychiatric manifestations (Gelder et al. 1996). Initially described by Von Economo at the Viennese Psychiatric Clinic in 1917, it has been etiopathogenetically linked to influenza viruses, although this link has never been proved conclusively.

The acute stage is characterized by sleep disturbances, with inversion of the sleep–wake rhythm and predominance of narcolepsy and cataplexy. The patient is unable to perform any movement during the waking or the sleep phase; hallucinations, anxiety, depersonalization, and depression are also frequent. In some cases, impulsivity and suicidal ideation are present.

The chronic sequelae are of psychiatric interest. Besides parkinsonism and oculogyric crises, severe psychotic syndromes may develop. It is possible to distinguish the following syndromes:

- A delusional–hallucinatory syndrome, with chronic paranoid delusions, hallucinations, and pseudohallucinations
- A hebephrenic–catatonic syndrome with negativism, eccentricity, and stereotypies
- Depressive episodes, either pure or associated with delusions and hallucinations
- A rapidly progressing dementia (rare)
- Personality changes possibly leading to antisocial behavior

3.5 Subacute Sclerosing Panencephalitis

SSPE is a slow viral infection of the CNS, accompanied by degeneration of gray matter and loss of white matter, likely due to the measles virus.

The clinical picture is initially characterized by behavioral and cognitive changes, such as forgetfulness and irritability, while neurologic signs and symptoms predominate as the disease progresses (Jabour et al. 1969). Patients with SSPE often present with a psychosis in the absence of impairment of consciousness, with bizarre behavior, hallucinations, and aphasia, suggesting the temporal lobe as the site of infection (Duncalf et al. 1989).

3.6 Brain Abscess

Clinical signs of brain abscess are caused by the space-occupying mass rather than the infection itself. Headache, papilledema, seizures, and other neurological focal signs may occur (Nielsen et al. 1982).

In some cases, a cerebral abscess may develop insidiously, mimicking a psychiatric syndrome. For example, depressive symptoms accompanied by mild confusion and fever should cause the physician to suspect the presence of a brain abscess. The clinical picture of a frontal lobe abscess is often dominated by drowsiness, distractibility, and a generalized deterioration of mental functioning. Impairment of consciousness, ranging from lethargy to coma, occurs in most patients (Samson and Clark 1973).

The primary focus of the infection is usually outside the brain, in the mastoid, middle ear, or nasal sinuses. Penetrating head injury may be a further cause (Tay and Garland 1987). The anaerobes most commonly encountered in brain abscess include *Bacteroides*, *Prevotella*, *Fusobacterium*, *Clostridiium*, and *Actinomycetes*. In the antibiotic era, abscesses due to *Staphylococci* are decreasing in frequency, whereas those due to Enterobacteriaceae are more prevalent (Dacey and Winn 1983).

4 Prion Diseases

4.1 General Remarks

Prion diseases, sometimes referred to as transmissible spongiform encephalopathies, include human CNS disorders characterized by an aberrant metabolism of the cellular prion protein (PrP^{C}), a normal soluble protein of the cell membrane, which undergoes a conformational change to an insoluble form. This form, denoted PrP^{SC} (prion protein, scrapie, as it was first identified as the agent which causes scrapie in sheep), is poorly metabolized and accumulates in the CNS, causing cell death with spongiform changes (Prusiner and Hsiao 1994).

Prion diseases are therefore quite different from typical infectious diseases which require genetic material to be transferred. Furthermore, in these diseases, incubation usually takes months or years, leading to the concept of "slow infection" of the CNS.

Classical prion diseases in humans comprise Creutzfeldt-Jakob disease (CJD), Kuru, Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, and the recently described new variant of CJD. Here, we focus on CJD and its new variant.

4.2 Creutzfeldt-Jakob Disease

CJD is a subacute prion disease with spongiform encephalopathy. Although its prevalence and incidence is only 1 per million of the population worldwide, hereditary, infectious, and sporadic types can be distinguished, with the latter accounting for 85%–95% of cases.

In the majority of patients, CJD begins in the late 50s with mental deterioration, rapidly progressing to a severe dementia accompanied by behavioral disturbances. The presence of myoclonus is another almost invariant feature of the disease. Frequently, visual, cerebellar, and extrapyramidal signs also occur. Characteristic triphasic complexes are found on electroencephalography (EEG) in about 80% of patients. The average duration of the illness from onset of symptoms to death is 7–9 months.

Recently, an apparently new variant of CJD has been identified in United Kingdom (Will 1996), with a single case also found in France (Zeidler et al. 1997). In contrast to typical CJD, onset of the new variant of CJD occurs at an unusually young age (with most patients being in their 20s), and psychiatric symptoms are present at initial presentation, with anxiety, depression, transient delusions, and auditory or visual hallucinations being the most prominent clinical features. Cognitive impairment and neurological signs, including ataxia and involuntary movements, generally occur relatively late.

This apparently new variant of CJD has attracted worldwide interest because of the possibility that it is causally linked to bovine spongiform encephalopathy (BSE) and thus possibly transmitted to humans through "infected" cow meat.

5 Chronic Fatigue Syndrome

Chronic fatigue syndrome (CFS) is a disorder of unknown etiology, characterized by debilitating fatigue associated with fever, myalgia, arthralgia, lymphoadenopathy, and headache. Photophobia, transient scotomata, forgetfulness, difficulty in thinking, inability to concentrate, excessive irritability, sleep disturbances (insomnia or hypersomnia), confusion, and depressed mood may also occur (Krupp et al. 1991).

A striking feature of CFS is its sudden onset following an acute infectious illness and the subsequent recurrent "flu-like" symptoms. Infectious agents, including the human herpes virus, *Candida albicans*, human spumavirus, enterovirus, Epstein-Barr Virus, and brucella, have been involved. CFS is also common after recovery from hepatitis A and B (Landay et al. 1991). In addition to infectious agents, metabolic factors have also been implicated. Although an infectious dysregulation of the immune system has been postulated in the pathogenesis of CFS, the mechanisms responsible for this syndrome are still unclear.

6 Management

The management of psychiatric syndromes due to infectious agents consists, for the most part, in the management of the specific infection. The pharmacologic treatment of psychiatric symptoms is similar to that for other acute or chronic organic brain syndromes. In the delirious patient, the general rule is to give as few drugs as possible and to avoid any that may increase impairment of consciousness. General measures, such as reducing disorientation and avoiding too much or too little sensory stimulation, are necessary to relieve distress and to prevent

behaviors potentially dangerous to the patient or to other people.

Apart from adequate nursing care, the patient should be given repeated explanations of his or her condition. Preservation of a familiar environment is important, although not always easy to achieve, mainly in those infectious diseases that necessitate isolation. If possible, relatives and friends should visit the patient frequently. It is good practice to advice them about how they may contribute to reassuring and orientating the patient.

7 References

Centers for Disease Control (1993) Lyme disease. United States 1991–1992. MMWR 42: 345–348

Dacey RGJ, Winn HR (1983) Brain abscess and perimeningeal infections. In: Stein JH, Cline MJ, Daly WJ (eds) Internal medicine. Little Brown, Boston, p 1213

Duncalf CM, Kent JNG, Harbord M et al (1989) Subacute sclerosing panencephalitis presenting as schizophreniform psychosis. Br J Psychiatry 155: 557–559

Emsley RA, Roberts MA, Higson EA et al (1988) Neurosyphilis and psychiatry. Br J Psychiatry 152: 573

Fallon BA, Nields JA, Parsons B et al (1992) The neuropsychiatric manifestations of Lyme borreliosis. Psychiatry Q 63: 95–115

Gelder M, Gath D, Mayou R et al (1996) Oxford textbook of psychiatry. Oxford University Press, Oxford

Greenlee JE (1995) Anatomic considerations in central nervous system infections. In: Mandell G, Douglas R, Bennet JE et al (eds) Principles and practice of infectious diseases, 4th edn. Churchill Livingstone, New York, pp 821–831

Hook EW (1991) Central nervous system syphilis In: Sheld WM, Whitley RJ, Durack DT (eds) Infections of the central nervous system. Raven, New York, pp 639–656

Jabbour JT, Garcia JH, Lemmi H et al (1969) Subacute sclerosing panencephalitis, a multidisciplinary study of 8 cases. JAMA 297: 2248–2254

Kinghorn G (1993) The re-emergence of syphilis. Br J Hosp Med 49: 683–685

Krupp LB, Mendelson WB, Friedman R (1991) An overview of chronic fatigue syndrome. J Clin Psychiatry 52: 403–410

Landay AL, Jessop C, Lennette ET et al (1991) Chronic fatigue syndrome: clinical condition associated with immune activation. Lancet 338: 707–712

Lishman WA (1987) Organic psychiatry. Blackwell, Oxford

Lucas SB, Hounnou A, Peacock C et al (1993) The mortality and pathology of HIV infection in a West African city. AIDS 7: 1569–1579

Lukehart SA, Hook EW, Baker-Zander SA et al (1988) Invasion of the central nervous system by Treponema pallidum: implications for diagnosis and treatment. Ann Intern Med 109: 855–862

Mandell G, Douglas R, Bennet JE et al (eds) (1995) Principles and practice of infectious diseases, 4th edn. Churchill Livingstone, New York

Musher DM, Hamill RJ, Baughn RE (1990) Effect of human immunodeficiency virus (HIV) infection on the course of syphilis and on the response to treatment. Ann Intern Med 113: 872–881

Nielsen H, Glydensted C, Harmsen A (1982) Cerebral abscess: Aetiology and pathogenesis, symptoms, diagnosis and treatment. Acta Neurol Scand 65: 609–622

**Prusiner SB, Hsiao KK (1994) Human prion diseases. Ann Neurol 35: 385–395

*Reik L (1992) Lyme disease and the nervous system. Thieme, New York

Roberts MC, Emsley RA (1992) Psychiatric manifestations of neurosyphilis. S Afr Med J 82: 335–337

Rundell JR, Wise MG (1985) Neurosyphilis: a psychiatric perspective. Psychosomatics 26: 287–295

Samson DS, Clark K (1973) A current review of brain abscess. Am J Med 54: 201–210

**Sheld WM, Whitley RJ, Durack DT (eds) (1991) Infections of the central nervous system. Raven, New York

Simon RP (1985) Neurosyphilis. Arch Neurol 42: 606–613

Sirota P, Eviatar J, Spivak B (1988) Neurosyphilis presenting as psychiatric disorders. Br J Psychiatry 155: 559–561

Tay JS, Garland JS (1987) Serious head injuries from lawn darts. Pediatrics 79: 261–263

Temmerman M, Ali FM, Ndinya-Achola J et al (1992) Rapid increase of both HIV-1 infection and syphilis among pregnant women in Nairobi, Kenya. AIDS 6: 1181–1185

Tramont EC (1995) Treponema pallidum (syphilis) In: Mandell G, Douglas R, Bennet JE et al (eds) Principles and practice of infectious diseases, 4th edn. Churchill Livingstone, New York, pp 2117–2132

Will RG (1996) Epidemiology of Creutzfeldt-Jakob disease. Br Med Bull 49: 960–970

**Zeidler M, Johnstone EC et al (1997) New variant of Creutzfeldt-Jakob disease: psychiatric features. Lancet 350: 908–910

CHAPTER

16

M. Maj, A. Tortorella

Mental Health Problems and Psychiatric Disorders in Subjects with Human Immunodeficiency Virus Infection

1 Introduction

According to the most recent estimates by the United Nations and the World Health Organization (UNAIDS 1996), there are currently approximately 22 million individuals with human immunodeficiency virus (HIV) living in the different areas of the world (about 21 million adults and almost 1 million children). More than 90% of them live in developing countries, and almost two thirds live in sub-Saharan Africa.

Heterosexual intercourse accounts for more than 70% of all adult infections to date, being the predominant mode of transmission in Africa, Asia, and the Caribbean; male homosexual intercourse accounts for 5%–10% of cases and remains the predominant mode of transmission in North America; sharing of contaminated needles by i.v. drug users accounts for a further 5%–10% of cases and is the predominant mode of transmission in some European countries; finally, transfusion of contaminated blood or blood products accounts for 3%–5% of adult infections to date.

Mother-to-child transmission of the infection occurs during pregnancy or delivery or through breast-feeding in 25%–35% of infants born to HIV-infected women. Over 85% of all infections in children have occurred up to now in sub-Saharan Africa.

In industrialized countries, about 60% of adults progress to acquired immunodeficiency syndrome (AIDS) within 12–13 years of becoming infected with HIV (UNAIDS 1996). Although no long-term cohort study has been completed, progression to AIDS may be more rapid in developing countries. The average survival time after the onset of AIDS is at present about 3 years in industrialized countries and is estimated to be less than 1 year in developing countries. The majority of AIDS cases occur before the age of 35, and over 90% of all AIDS deaths occur in individuals under the age of 50.

An epidemic of this proportion and with these characteristics can be expected to have profound mental health implications at several levels: (a) those directly affected by the infection, (b) groups at risk for the infection, including homosexual/bisexual men, i.v. drug users, people with multiple sex partners, and patients with serious psychiatric disorders, (c) partners, relatives, and friends of those with HIV infection, (d) health care providers, and (e) the general population. In this chapter, we will focus on the first of these levels. For information on the impact of the epidemic at the other levels, the reader is referred to the reviews by Maj (1991), Knox et al. (1994), and Coverdale (1996).

2 Mental Health Problems

HIV infection is often accompanied by significant mental health problems, for the following reasons:

- It produces a devastating physical illness, the psychological impact of which is comparable to that of cancer.
- In contrast to cancer, it elicits adverse social reactions (fear, stigma, rejection).
- It involves the brain, not only by inducing secondary infections or neoplasms, but also by producing a primary encephalopathy.
- In industrialized countries, it predominantly affects groups (homosexuals, i.v. drug users) who are already highly vulnerable to mental health problems.

Most people with HIV infection can be expected to experience mental health problems, although only a minority of them will suffer from proper psychiatric disorders.

The mental health problems most commonly reported by persons with HIV infection can be subdivided into the following groups: (a) distressing emotions, (b) cognitive disturbances, (c) relationship difficulties, (d) instrumental problems, and (e) suicidal ideation and/or attempts (Pakenham et al. 1996).

Distressing emotions most frequently experienced by those with HIV infection include the following: fear (of death, illness, pain, disability, disfigurement, abandonment, dependency, of infecting others, of others discovering their status, of the unknown), anger (directed at the partner or partners, family, friends, health staff, medical profession, institutions, fate, themselves), sadness, anxiety (frequently associated with overattentiveness to bodily changes), despair, guilt (about their lifestyle), uncertainty, and numbness. These emotions peak after the notification of serostatus, remain in the background or fluctuate during the asymptomatic stage of the infection, and then peak again after the appearance of physical symptoms, the start of antiretroviral treatment, the diagnosis of AIDS, and any worsening of physical status or exacerbation of HIV-related relationship and instrumental problems. For further information on emotional reactions to the notification of seropositivity and on pre- and post-test counseling, the reader is referred to Miller (1987).

Cognitive disturbances most frequently experienced by individuals with HIV infection include forgetfulness, poor concentration, slowness of thought, and speech difficulties. These disturbances are reported by 15%–25% of those with symptomatic HIV infection

(Maj et al. 1994b) and reflect, at least in a proportion of the cases, the initial involvement of the brain, as demonstrated by the concomitant impaired performance in some neuropsychological tests, in particular timed motor tasks with a cognitive component. Cognitive disturbances are relatively rare in the asymptomatic stage of the infection. For an overview of the data currently available on neuropsychological performance in physically asymptomatic HIV-seropositive subjects, the reader is referred to Newman et al. (1995).

Relationship difficulties most frequently experienced by individuals with HIV infection include the following: sexual problems (sexual frustration, detachment during sexual activity, difficulties in effecting appropriate changes in sexual behavior, problems in being open about their homosexuality), discrimination and/or rejection (by coworkers, neighbors, friends, family, or partner), dependence/independence conflicts and other problems related to caregiving, problems in establishing new relationships (in particular, dating problems), exacerbation of long-standing relationship problems (due to the underlying tension induced by having to cope with HIV infection), and isolation (as a consequence of rejection, self-inflicted, or due to physical disability). Problems in establishing and maintaining relationships predominate during the asymptomatic phase, whereas relationship difficulties reported by those with symptomatic infection tend to center around caregiving issues.

Instrumental problems most frequently encountered by those with HIV infection include employment, financial, and daily living difficulties and problems concerning access to health care services. These difficulties become prominent in the late stages of the infection, when physical health deteriorates.

Suicide is contemplated, at least transiently, by several individuals with HIV infection, especially in the symptomatic stages, although reliable prevalence rates of suicidal ideation are at present unavailable. In patients with AIDS, the suicide rate has been found to be seven to 36 times greater than in the age- and sex-matched population in the same geographic area (Marzuk et al. 1988; Wedler 1991; Coté et al. 1992; Pugh et al. 1993), whereas reliable estimates of the suicide risk in all individuals with HIV infection in a given area are currently lacking.

According to the currently available empirical research (for a review, see Catalàn et al. 1995), the factors which can affect the intensity of mental health problems in those with HIV infection and/or predispose them to or protect them from the development of proper psychiatric disorders include the following: (a) past psychiatric history (mental health problems and proper psychiatric disorders have been found to be more frequent in HIV-seropositive persons with a history of inpatient or outpatient psychiatric care), (b) social support (emotional, practical, and particularly informational support has been found to buffer HIV-related mental health problems and to affect the incidence of clinical depression), (c) concomitant life events (in particular, the AIDS-related death of one or more partners or friends has been found to affect the intensity of mental health problems and to influence the prevalence of clinical depression), (d) demographic factors (in particular age, since older persons are at higher risk for HIV encephalopathy), and (e) coping strategies (on which we will focus now briefly).

The coping strategies most frequently adopted by those with HIV infection can be classified as follows (Namir et al. 1990): (a) active and expressive/information seeking ("talked with others in the same situation," "tried to find out more about my illness"), (b) active and relying on others ("went to a friend or a professional to help me feel better"), (c) active and behavioral ("got involved in political activities related to my illness," "took more vitamins and ate healthy foods"), (d) active and cognitive ("prayed hard for a good outcome to the situation," "thought more about the meaning of life"), (e) passive and cognitive/ruminative ("daydreamed about better times," "thought about how I could have done things differently"), (f) passive and resigned ("tried to keep others from knowing how I was feeling"), (g) avoidant ("refused to think about it"), and (h) looking for distraction ("bought or did something special for myself"). Of these coping strategies, the "active and behavioral" and, to a lesser degree, the "active and expressive" and the "active and relying on others" have been found to be associated with lower anxiety and mood disturbance, whereas the avoidant strategy has been reported to be associated with higher levels of mood disturbance, including an increased prevalence of clinical depression (Namir et al. 1990; Folkman et al. 1993; Commerford et al. 1994).

3 Psychiatric Disorders

Virtually all psychiatric disorders have been described in individuals with HIV infection (for reviews, see Maj et al. 1993; Lyketsos and Federman 1995; Rabkin 1996). Some of them, such as psychotic disorders, mania, and anxiety disorders, do not seem to be more frequent in HIV-seropositive individuals than in matched seronegative control groups, while others, such as personality disorders and drug and alcohol abuse, are indeed more frequent, but only because they predispose subjects to the infection itself. The concept of adjustment disorder is of questionable value in those with HIV infection, in whom distressing

emotions are almost the rule, and looking for a boundary between normal and abnormal emotional reactions to the stress of the infection is merely a waste of time. We are therefore left with only two psychiatric disorders – dementia and delirium – whose prevalence is clearly increased in the symptomatic stages of the infection, and with another one – major depression – whose prevalence may be increased among symptomatic HIV-seropositive subjects in particular contexts. We will focus now on these three disorders.

3.1 Dementia

3.1.1 Clinical Picture

The onset of HIV-associated dementia is usually insidious. Early cognitive symptoms include forgetfulness, loss of concentration, mental slowing, and reduced performance in sequential mental activities of some complexity (the subject misses appointments or needs lists to recall ordinary duties; loses track of conversations or his or her own train of thought; and needs additional time and effort to organize thoughts and to complete daily tasks). Early behavioral symptoms include apathy, reduced spontaneity and emotional responsiveness, and social withdrawal (the subject becomes indifferent to his or her personal and professional responsibilities; his or her work production decreases, as does the frequency of his or her social interactions; and he or she complains of early fatigability, malaise, and loss of sexual drive). Depression, irritability, or emotional instability, agitation, and psychotic symptoms may also occur. Early motor symptoms include loss of balance and coordination, clumsiness, and leg weakness (the subject is less precise in normal hand activities, such as writing and eating; drops things more frequently than usual; trips and falls more frequently than usual; and feels the need to exercise more care in walking; Navia et al. 1986; Maj 1990).

Routine mental status tests, in this early stage, may be normal or show only slowing in verbal or motor responses and/or difficulty in recalling a series of objects after 5 or more minutes. Neurological examination may show tremor (best seen when the patient sustains a posture, such as holding the arms and fingers outstretched), hyperreflexia (particularly of the lower extremities), ataxia (usually seen only on rapid turns or tandem gait), slowing of rapid alternating movements (of the fingers, wrists, or feet), frontal release signs (snout reflex, palmar grasp), and dysarthria. Tests of ocular motility may show interruption of smooth pursuits and slowing or inaccuracy of saccades.

In the late stages of the disease, there is usually an overall deterioration of cognitive functions and severe psychomotor retardation. Speech is slow and monotonous, with word-finding difficulties and possible progression to mutism. Patients become unable to walk, due to paraparesis, and usually lie in bed, indifferent to their illness and their surroundings. Bladder and bowel incontinence are common. Myoclonus and seizures may occur. Pedal paresthesiae and hypersensitivity may appear, due to concurrent sensory neuropathy. The level of consciousness is usually preserved, except for occasional hypersomnolence.

It has been repeatedly maintained that, in HIV-associated dementia, higher cortical functions are spared, so that the picture is that of a "subcortical dementia" (Navia et al. 1986). This tenet, however, has been challenged by the report of patients showing such aspects as dyspraxia, dysgraphia, dyscalculia, mild paraphasia, and altered language comprehension (Poutiainen et al. 1991).

The course of HIV-associated dementia is variable, and no predictor of the pace of progression is currently available. The syndrome often progresses rapidly to severe deterioration and death, especially in patients with advanced systemic disease, but it may also have prolonged stable phases or may fluctuate, with reversible deterioration occurring in concomitance with opportunistic infections, such as *Pneumocystis carinii* pneumonia (World Health Organization 1990). Death usually occurs as a result of inanition, aspiration pneumonia, or systemic opportunistic infections.

3.1.2 Neuropsychological Picture

Formal neuropsychological examination of patients with HIV-associated dementia usually shows the most prominent impairment on tests of fine motor control (finger tapping, grooved pegboard), rapid sequential problem solving (trail making A and B, digit symbol), visuospatial problem solving (block design), spontaneity (verbal fluency), and visual memory (visual reproduction). In contrast, naming and vocabulary skills are largely preserved even in the most advanced cases. This pattern has been regarded as consistent with the clinical picture of a subcortical dementia.

3.1.3 Neuroradiological, Electroencephalographic and Laboratory Findings

The predominant neuroradiological finding in HIV-associated dementia is cerebral atrophy: both computed tomography (CT) and magnetic resonance imaging (MRI) demonstrate widened cortical sulci and, less

commonly, enlarged ventricles. Furthermore, MRI frequently shows high-intensity signal abnormalities on the T_2-weighted image (diffuse widespread involvement, patchy localized involvement, focal distinct areas of involvement, or punctuate white matter hyperdensities). These lesions are without mass effect and are most commonly located in the periventricular white matter and the semioval center (less frequently in the basal ganglia or in the thalamus).

The assessment of the regional cerebral metabolic rate for glucose by positron emission tomography (PET) has been reported to show a relative subcortical (thalamus and basal ganglia) hypermetabolism in the early stages and cortical and subcortical gray matter hypometabolism in the late stages of the disease (Rottenberg et al. 1987). The electroencephalogram (EEG) may be normal or show diffuse slowing, especially in late stages.

The most frequent cerebrospinal fluid (CSF) findings in HIV-associated dementia are an increase in total proteins (typically in the order of 50–100 mg/100 ml) and in the immunoglobulin (Ig) G fraction and index. Oligoclonal bands, either specific to HIV or not, may be present. A mononuclear pleocytosis (four to 50 cells/mm^3) may occur, and the ratio between the T-lymphocyte subsets (CD4 to CD8) may be reversed. The presence of the HIV core antigen p24 can be detected; although this finding is possible also in neurologically normal subjects, it does appear to reflect viral replication and may be useful in establishing the diagnosis and possibly in monitoring response to treatment (Price et al. 1988). A CSF to plasma ratio higher than 1 for neopterin and β_2-microglobulin has been reported and, in the absence of a correlation to the blood–brain barrier status, has been ascribed to an intrathecal synthesis of the two substances (Fuchs et al. 1989). CSF β_2-microglobulin levels have been found to correlate with the stage of dementia (Brew et al. 1989).

3.1.4 Neuropathology

Gross neuropathological examination of brains from patients with HIV-associated dementia usually reveals cerebral atrophy with sulcal widening and ventricular dilatation and occasionally meningeal fibrosis.

Microscopic abnormalities are most prominent in central white matter and deep gray structures. The two major lesion patterns are HIV encephalitis (multifocal and inflammatory) and HIV leukoencephalopathy (diffuse and noninflammatory), but it is recognized that overlapping and transition between these patterns may occur (Budka 1991).

The term *HIV encephalitis* corresponds to multifocal giant cell encephalitis, multinucleated cell encephalitis, and subacute encephalitis with multinucleated cells. This pattern is characterized by multiple disseminated microgranulomatous foci composed of microglia, macrophages, and macrophage-derived multinucleated giant cells, all or most of which show intense production of HIV antigens. Although the white matter, basal ganglia, and brain stem are preferentially involved, these foci are not rare in the cerebral cortex. In larger foci, central coagulative necrosis may develop. Damage to parenchyma in the foci is either inconspicuous or involves sponginess and myelin loss, whereas neurons and axons are usually perfectly preserved. Small vessels within a focus may show fibrinoid extravasation.

The term *HIV leukoencephalopathy* refers to progressive diffuse leukoencephalopathy. This picture consists of a diffuse damage to white matter, including myelin loss ("pallor"), astrogliosis, and infiltration by mono- and multinucleated microglia and macrophages. The white matter of the cerebral hemispheres is mainly affected, usually in a symmetric fashion, but cerebellar white matter may also be involved. Mono- or multinucleated macrophages, usually in a perivascular position, may incorporate myelin debris. Vacuolar swelling of myelin may be prominent in some cases.

A third histopathological pattern is that of *diffuse poliodystrophy*, defined as a diffuse reactive astrogliosis and microglial activation in the gray matter (basal ganglia, brain stem nuclei, and cerebral cortex). Neuronal dropout may be suspected, but remains to be verified by morphometry. This pattern has been proposed as a possible substrate of the cerebral atrophy which is frequently seen in neuroimaging studies and at autopsy. There are rare cases which exhibit prominent cortical diffuse poliodystrophy as the only correlate of severe progressive dementia.

The correlation between clinical and pathological aspects is reported to be poor in about one third of patients with HIV-associated dementia. In the majority of patients, however, more severe clinical manifestations tend to correspond to more severe neuropathology (Price et al. 1988).

3.1.5 Epidemiology

Data on the epidemiology of HIV-associated dementia are still preliminary. A recent prospective study, carried out on a cohort of homosexual men in the United States, found that the diagnosis was made concurrently with that of the initial AIDS-defining illness in 3% of patients and that the incidence of the dementia syndrome during the first 2 years after the diagnosis of AIDS was 7% per year (McArthur et al. 1994).

3.1.6 Pathogenesis

The pathogenesis of HIV-associated dementia is at present incompletely understood. Productive infection is almost exclusively restricted to macrophages and microglia, whereas neurons are not infected. Neuronal injury is currently believed to be produced by toxic products released directly by HIV-infected macrophages and microglia or by activated astrocytes. Some of these factors have been identified (Price and Perry 1994); they include platelet-activating factor, quinolinic acid, nitric oxide, and some metabolites of arachidonic acid, which are neurotoxic, and tumor necrosis factor-α, which is toxic for oligodendrocytes and can cause demyelination.

The strains of HIV which are isolated from the brain have in common the characteristic of infecting macrophages but not lymphocytes. This macrophage tropism corresponds to what was initially regarded as neurotropism. Macrophage tropism is related to a mutation in a specific region of gp120, the external glycoprotein of the virus. In the late stages of the infection, when active replication of the virus generates more mutants and the compromised immune system permits the escape of these mutants, the development of macrophage-trophic strains is more likely to occur, and this probably represents the limiting step for the occurrence of HIV encephalopathy and dementia.

3.1.7 Management

A beneficial effect of the antiretroviral agent zidovudine (AZT) on several cognitive measures in patients with HIV-associated dementia has been demonstrated in some studies (Yarchoan et al. 1987; Schmitt et al. 1988). One of the main disadvantages of this drug is the bone marrow suppression it causes.

An improvement of cognitive performance in AIDS patients has been also obtained using psychostimulants, such as methylphenidate and dextroamphetamine (Fernandez et al. 1988). Other drugs or nutritional factors that have been suggested for use in patients with HIV-associated dementia include peptide T, nimodipine, pyridoxin, and vitamin B_{12} (Maj et al. 1993).

Psychosocial interventions suggested for patients with HIV-associated dementia include maintenance of a structured daily schedule, titration of external stimuli, restriction to familiar environments, frequent orienting interactions with significant others, and monitoring of personal and financial affairs. Specific psychoeducation of HIV-infected subjects and their relatives before dementia develops has been recommended.

The care of patients with HIV-associated dementia will make increasing demands on health services as well as on volunteer and community support systems. It is uncertain, at present, whether such care is best provided in specialized units (e.g. inpatient AIDS units) or within general psychiatric or medical services. Special management problems may arise when the behavioral disturbance (e.g. poor impulse control, inappropriate sexual behavior) is such as to constitute a risk for other patients or staff members. Placement of patients in the terminal stage of the disease may also represent a problem; the lack of appropriate options in the community may obstruct their timely and humane discharge from the hospital.

3.2 Delirium

Delirium is a relatively frequent complication of the late stages of symptomatic HIV infection, although it often remains undetected or is misdiagnosed as psychosis or mania.

The factors which can lead to the development of a delirious state in symptomatic HIV infection include the following: central nervous system (CNS) opportunistic infections (in particular, cryptococcal meningitis, cerebral toxoplasmosis, and cytomegalovirus encephalitis), opportunistic infections of other organs (in particular, *Pneumocystis carinii* pneumonia), systemic opportunistic infections (such as staphylococcal bacteremia), opportunistic neoplasms of the brain (CNS lymphoma, Kaposi's sarcoma with CNS involvement), other space-occupying lesions of the brain (e.g. brain abscesses due to toxoplasmosis), metabolic derangements (disorders of fluid, electrolyte, or acid–base balance), nutritional deficits (e.g. of vitamin B_{12}), hepatic and renal dysfunction, surgical interventions, substance abuse or withdrawal, use of psychotropic drugs (in particular tricyclic antidepressants, due to their anticholinergic activity), and use of antiretroviral drugs (including AZT).

The clinical picture of HIV-associated delirium does not present specific features. No reliable estimate of the prevalence and incidence of this syndrome in patients with symptomatic HIV infection is currently available.

Medical management of delirium in HIV-infected subjects consists in treating the underlying cause (when possible), maintaining fluid and electrolyte balance and nutrition, and providing sedation and correction of disturbances of sleep–wake cycle. Low-dose neuroleptics (e.g. 1.75 mg haloperidol/day, 80 mg chlorpromazine/day) are frequently effective and safe in the treatment of agitation in AIDS patients with delirium. Haloperidol has been administered orally, intramuscularly, and even intravenously, alone or in

combination with lorazepam. The use of low-dose neuroleptics should prevent the development of the severe extrapyramidal reactions and the neuroleptic malignant syndrome reported in delirious AIDS patients treated with standard neuroleptic doses.

Environmental intervention and nursing care have been found to be very important. The former consists of the following: (a) provision of a well-lit, quiet room with a dimmed light at night and with a clock and a calendar to facilitate orientation, (b) restriction of the number of visitors and staff, while a trusted relative or friend may be allowed to stay with the patient outside of routine visiting hours (he or she will be instructed to talk calmly about matters of interest for the patient, stating frequently the date and the names of the hospital and of the doctors and nurses in attendance), and (c) provision of reassurance and explanation to the patient and significant others about the nature of delirium, its usual causes, and expected degree of reversibility. Nursing care includes careful observation, timely reporting of behavioral change, emotional support, and reorientation.

3.3 Major Depression

The ascertainment of major depression in patients with HIV infection is difficult for the following reasons: (a) the possible confounding effect of the physical symptoms of the infection (fatigue, diminished appetite and sleep, and loss of weight may be physical symptoms of HIV infection as well as depressive symptoms), (b) the possible confounding effect of the cognitive impairment related to HIV infection of the brain (psychomotor slowing, forgetfulness, and difficulties in concentration may be early symptoms of this impairment), and (c) the frequent occurrence of transient emotional and behavioral reactions in coincidence with the key points in the course of the infection (losing interest in human contact, feeling guilty about previous high-risk behaviors, and thinking of death may all be part of these reactions).

In order to diagnose major depression, the psychiatrist should ascertain that a depressive syndrome has persisted for at least 2 weeks, most of the day, nearly every day, including symptoms such as prominent depressive mood, markedly reduced interest or pleasure in all or almost all activities, true sense of worthlessness, and persistent suicidal ideation. Whether depression is understandable or not on the basis of the current situation of the subject should not influence the diagnostic decision. The clinical significance of symptoms such as psychomotor retardation or diminished ability to think or concentrate should be considered equivocal, unless detailed neuropsychological assessment demonstrates the absence of neurocognitive impairment. Similarly, the clinical significance of manifestations such as reduced weight or appetite, fatigue or loss of energy, and insomnia should be regarded as uncertain if the subject has reached the symptomatic stages of the infection.

The prevalence of major depression has been consistently found not to be increased in the typical samples of well-educated, middle class, mostly white homosexual/bisexual HIV-seropositive men studied in Western countries (Atkinson et al. 1988; Perry et al. 1990; Williams et al. 1991; Rosenberger et al. 1993), but may be increased in symptomatic HIV-seropositive subjects with different sociodemographic characteristics. In the WHO Neuropsychiatric AIDS Study, which we conducted in the early 1990s, we found a significant increase in the current prevalence of major depression in symptomatic seropositive i.v. drug users enrolled in Bangkok (Maj et al. 1994a). A similar finding has been subsequently reported in the United States in a sample predominantly composed of Afro-American i.v. drug users (Lipsitz et al. 1994). Current unemployment, history of prior depression, and the presence of AIDS-related symptoms have been found to predict the development of syndromal depression among subjects with symptomatic HIV infection in a recent longitudinal study (Lyketsos et al. 1996).

Major depression in individuals with HIV infection often remains undetected and untreated. In the WHO Neuropsychiatric AIDS Study, only 9% of those fulfilling the DSM III-R criteria for major depression (American Psychiatric Association 1987) were receiving antidepressant treatment.

The presence of depressive symptoms has been found to predict shorter longevity in AIDS patients after controlling for physical symptoms and $CD4^+$ cell count. The most likely interpretation of this finding is that depression may limit the motivation of individuals to continue treatment or to provide self-care in the presence of severe pain and disability (Patterson et al. 1996).

When major depression is present, drug treatment is, at the current state of knowledge, the first therapeutic option. Both tricyclic antidepressants and selective serotonin reuptake inhibitors (SSRIs) have been found to be effective and well tolerated in patients with major depression associated with HIV infection (Rabkin et al. 1994a,b). Some trials suggest that SSRIs may be better tolerated than tricyclics (except in subjects with chronic diarrhea), but systematic studies involving direct comparisons are presently lacking. Psychostimulants have been suggested to improve depressed mood and lethargy in AIDS patients by some case reports and studies conducted on small patients samples (Fernandez et al. 1995).

A few systematic studies have shown the efficacy of individual interpersonal psychotherapy (Markowitz et al. 1995) and of cognitive/behavioral group interventions (Kelly et al. 1993) in samples of mainly asymptomatic depressed HIV-seropositive subjects.

4 Conclusions

HIV infection has a clear, direct impact both on brain functioning and on psychological homeostasis and psychosocial adjustment. This impact occurs in most individuals at some point of the natural history of the infection, although only a minority of them actually develop a proper psychiatric disorder. The development of proper psychiatric disorders is conditioned by predisposing or protective factors, some of which operate at the level of the cerebral "hardware" (e.g. ageing), some at the level of the psychological/psychosocial "software" (e.g. coping strategies, social support, concomitant life events), and others at a level that remains unspecified (e.g. past psychiatric history).

Mental health teams should be equipped to deal not only with proper psychiatric disorders (most commonly dementia, delirium, and major depression in the symptomatic stages of the infection), but also with the much more common mental health problems of HIV-seropositive individuals (distressing emotions, cognitive disturbances, relationship and instrumental problems, and suicidal ideation and/or attempts). Moreover, they should be able to address the mental health problems of groups at risk, the loved ones and caregivers of affected individuals, and the general population. Of course, such an undertaking is only possible if adequate resources are made available in terms of personnel, training, and facilities, and the availability of these resources depends very much on the degree of recognition, by policy makers and public opinion, of the importance of the mental health component in the care of those with HIV infection. This recognition, however, requires opinion leaders in the field of mental health care themselves to become fully aware of the magnitude of the problem and motivated to face it.

5 References

American Psychiatric Association (1987) Diagnostic and statistical manual of mental disorders, 3rd edn revised (DSM-III-R). American Psychiatric Association, Washington, DC

*Atkinson JH, Grant I, Kennedy CJ et al (1988) Prevalence of psychiatric disorders among men infected with human immunodeficiency virus. A controlled study. Arch Gen Psychiatry 45: 859–864

Brew BJ, Bhalla R, Paul M et al (1989) CSF beta$_2$-microglobulin as a marker of the presence and severity of AIDS dementia complex. Abstracts of the Vth International Conference on AIDS, 4–9 June 1989, Montreal (abstr BP 233)

*Budka H (1991) Neuropathology of human immunodeficiency virus infection. Brain Pathol 1: 163–175

Catalàn J, Burgess A, Klimes A (1995) Psychological medicine of HIV infection. Oxford University Press, Oxford

Commerford MC, Gular E, Orr DA et al (1994) Coping and psychological distress in women with HIV/AIDS. J Community Psychol 22: 224–230

Coté T, Biggar R, Dannenberg A (1992) Risk of suicide among persons with AIDS. A national assessment. JAMA 268: 2066–2068

Coverdale JH (1996) HIV risk behavior in the chronically mentally ill. Int Rev Psychiatry 8: 149–156

Fernandez F, Adams F, Levy JK (1988) Cognitive impairment due to AIDS-related complex and its response to psychostimulants. Psychosomatics 29: 38–46

Fernandez F, Levy JK, Samley HR et al (1995) Effects of methylphenidate in HIV-related depression: a comparative trial with desipramine. Int J Psychiatr Med 25: 53–67

Folkman S, Chesney M, Pollack L et al (1993) Stress, control, coping, and depressive mood in human immunodeficiency virus-positive and -negative gay men in San Francisco. J Nerv Ment Dis 181: 409–416

Fuchs D, Chiodi F, Albert J et al (1989) Neopterin concentrations in cerebrospinal fluid and serum of individuals infected with HIV-1. AIDS 3: 285–288

Kelly JA, Murphy DA, Bahr GR et al (1993) Outcome of cognitive-behavioral and support group brief therapies for depressed, HIV-infected persons. Am J Psychiatry 150: 1679–1686

Knox MD, Davis M, Friedrich MA (1994) The HIV mental health spectrum. Community Ment Health J 30: 75–89

Lipsitz JD, Williams JBW, Rabkin JG et al (1994) Psychopathology in male and female intravenous drug users with and without HIV infection. Am J Psychiatry 151: 1662–1668

Lyketsos CG, Federman EB (1995) Psychiatric disorders and HIV infection: impact on one other. Epidemiol Rev 17: 152–164

Lyketsos CG, Hoover DR, Guccione M et al (1996) Changes in depressive symptoms as AIDS develops. Am J Psychiatry 153: 1430–1437

Maj M (1990) Organic mental disorders in HIV-1 infection. AIDS 4: 831–840

Maj M (1991) Psychological problems of families and health workers dealing with people infected with human immunodeficiency virus 1. Acta Psychiatr Scand 83: 161–168

Maj M, Starace F, Sartorius N (1993) Mental disorders in HIV-1 infection and AIDS. Hogrefe and Huber, Seattle

**Maj M, Janssen R, Starace F et al (1994a) WHO neuropsychiatric AIDS study, cross-sectional phase. I. Study design and psychiatric findings. Arch Gen Psychiatry 51: 199–212

**Maj M, Satz P, Janssen R et al (1994b) WHO neuropsychiatric AIDS study, cross-sectional phase. II. Neuropsychological and neurological findings. Arch Gen Psychiatry 51: 199–212

Markowitz JC, Klerman GL, Clougherty KF et al (1995) Individual psychotherapies for depressed HIV-positive patients. Am J Psychiatry 152: 1504–1509

*Marzuk P, Tierney H, Tardiff K et al (1988) Increased risk of suicide in persons with AIDS. JAMA 259: 1333–1337

McArthur JC, Selnes OA, Glass JD et al (1994) HIV dementia. Incidence and risk factors. In: Price RW, Perry SW (eds) HIV, AIDS, and the brain. Raven, New York, pp 251–272

**Miller D (1987) Living with AIDS and HIV. MacMillan, Houndmills

Namir S, Wolcott DL, Fawzy FI et al (1990) Implications of different strategies for coping with AIDS. In: Temoshok L, Baum A (eds) Psychosocial perspectives on AIDS. Erlbaum, Hillsdale, pp 173–190

**Navia BA, Jordan BD, Price RW (1986) The AIDS dementia complex. I. Clinical picture. Ann Neurol 19: 517–524

Newman SP, Lunn S, Harrison M (1995) Do asymptomatic HIV-seropositive individuals show cognitive deficit? AIDS 9: 1211–1220

Pakenham KI, Dadds MR, Terry DJ (1996) Adaptive demands along the HIV disease continuum. Soc Sci Med 42: 245–256

Patterson TL, Shaw BS, Semple SJ et al (1996) Relationship of psychosocial factors to HIV disease progression. Ann Behav Med 18: 30–39

Perry S, Jacobsberg LB, Fishman B et al (1990) Psychiatric diagnosis before serological testing for the human immunodeficiency virus. Am J Psychiatry 47: 89–93

Poutiainen E, Haltia M, Elobaara J et al (1991) Dementia associated with human immunodeficiency virus: subcortical or cortical? Acta Psychiatr Scand 83: 297–301

**Price RW, Brew B, Sidtis J et al (1988) The brain in AIDS: central nervous system HIV-1 infection and AIDS dementia complex. Science 239: 586–592

Price RW, Perry SW (1994) HIV, AIDS, and the brain. Raven, New York

Pugh K, O'Donnell I, Catalàn J (1993) Suicide in HIV disease. AIDS Care 4: 391–399

Rabkin JG, Rabkin R, Harrison W et al (1994a) Effect of imipramine on mood and enumerative measures of immune status in depressed patients with HIV illness. Am J Psychiatry 151: 516–523

Rabkin JG, Rabkin R, Wagner R et al (1994b) Fluoxetine effects on mood and immune status in depressed patients with HIV illness. J Clin Psychiatry 55: 92–97

Rabkin JG (1996) Prevalence of psychiatric disorders in HIV illness. Int Rev Psychiatry 8: 157–166

Rosenberger PH, Bornstein RA, Nasrallah HA et al (1993) Psychopathology in human immunodeficiency virus infection: lifetime and current assessment. Compr Psychiatry 34: 150–158

Rottenberg DA, Moeller JR, Strother SC et al (1987) The metabolic pathology of the AIDS dementia complex. Ann Neurol 22: 700–706

Schmitt FA, Bigley JW, McKinnis R et al (1988) Neuropsychological outcome of zidovudine (AZT) treatment of patients with AIDS and AIDS-related complex. N Engl J Med 319: 1573–1578

UNAIDS (1996) The HIV/AIDS situation in mid 1996. UNAIDS, Geneva

Wedler H (1991) Suicidal behaviour in the HIV-infected population: the actual situation in the FRG. In: Beskow JE, Bellini M et al (eds) HIV and AIDS-related suicidal behaviour. Monduzzi, Bologna

Williams JBW, Rabkin JG, Remien RH et al (1991) Multidisciplinary baseline assessment of homosexual men with and without human immunodeficiency virus infection. II. Standardized clinical assessment of current and lifetime psychopathology. Arch Gen Psychiatry 48: 124–130

**World Health Organization (1990) Report of the second consultation on the neuropsychiatric aspects of HIV-1 infection. World Health Organization, Geneva

Yarchoan R, Berg G, Brouwers P et al (1987) Response of human-immunodeficiency-virus-associated neurological disease to 3′-azido-3′-deoxythymidine. Lancet i: 132–135

CHAPTER

17

G. Fricchione, N. Cassem

Psychiatric Problems Related to Intensive Care and Organ Transplantation

1 Psychiatric Care in the Intensive Care Unit

1.1 Introduction

Affective, cognitive, and behavioral changes are all seen quite frequently in the intensive care unit (ICU) (Cassem 1995). All of these states may be associated with agitation, putting the patient at risk due to inability to maintain important intravenous and arterial lines, pacing wires, and other modern technologies such as the intra-aortic balloon pump.

1.2 Differential Diagnosis

In the ICU, the psychiatric differential diagnosis for abnormal mental states is a large one. Certain emergency conditions must be recognized rapidly. These include Wernicke's encephalopathy, hypoxia, hypoglycemia, hypertensive encephalopathy, intracerebral hemorrhage, meningitis/encephalitis (including human immunodeficiency virus), and poisoning (iatrogenic or exogenous). Although less urgent, several other conditions need rapid assessment, including subdural hematoma, septicemia, subacute bacterial endocarditis, hepatic and/or renal failure, thyrotoxicosis, myxedema, delirium tremens, anticholinergic psychosis, and complex partial seizures, particularly complex partial status epilepticus.

When reasonably assured that the above acute illnesses are not present, the psychiatric consultant will have time to eliminate other basic causes of altered mental status using a systematic approach such as that proposed by Ludwig (1980) (see Table 1).

1.3 Evaluation and Management

The treatment plan often involves diagnostic suggestions, including the organic workup, which is often comprised of thyroid-stimulating hormone level, vitamin B12 and folate levels, the venereal disease research laboratory (VDRL) or rapid plasma reagin (RPR) tests for syphilis, neuroimaging, and electroencephalogram. Cerebral spinal fluid analysis is also requested on occasion as part of this diagnostic evaluation. Recommended psychopharmacologic intervention might

Table 1. Differential diagnosis of brain dysfunction in critical care patients: Ludwig's differential diagnosis of the confusion–delirium–dementia–coma complex[a]

General etiology	Specific etiologies
Vascular	Hypertensive encephalopathy; cerebral arteriosclerosis; intracranial hemorrhage or thromboses; circulatory collapse (shock); systemic lupus erythematosus; polyarteritis nodosa; thrombotic thrombocytopenic purpura
Infectious	Encephalitis; meningitis, general paresis
Neoplastic	Space-occupying lesions such as gliomas, meningiomas, abscesses
Degenerative	Senile and presenile dementias such as Alzheimer's or Pick's dementia; Huntington's chorea
Intoxication	Chronic intoxication or withdrawal, effect of sedative-hypnotic drugs such as bromides, opiates, tranquilizers, anticholinergics, dissociative anesthetics, anticonvulsants
Congenital	Epilepsy; postictal states; aneurysm
Traumatic	Subdural and epidural hematomas, contusion; laceration; postoperative trauma; heat stroke
Intraventricular	Normal pressure hydrocephalus
Vitamin	Deficiencies of thiamine (Wernicke-Korsakoff syndrome), niacin (pellagra), B_{12} (pernicious anemia)
Endocrine/metabolic	Diabetic coma and shock; uremia; myxedema; hyperthyroidism, parathyroid dysfunction; hypoglycemia; hepatic failure; porphyria; severe electrolyte or acid base disturbances; remote side effects of carcinoma; Cushing's syndrome
Metals	Heavy metals (lead, manganese, mercury); carbon monoxide; toxins
Anoxia	Hypoxia and anoxia secondary to pulmonary or cardiac failure, anesthesia, anemia
Depression/other	Depressive pseudodementia; hysteria; catatonia

[a]From Ludwig (1980).

include antidepressants, antipsychotics, anxiolytics, and stimulants. Behavioral/psychosocial approaches may also be recommended, and family intervention is often required. Psychotherapy is most often within the realm of support and education. The psychiatry consultant guarantees regular follow-up for as long as indicated and helps in establishing goals for the patient in future treatment.

Review of current and past medications is very important. Many pharmacologic agents have central nervous system (CNS) effects and can therefore produce neuropsychiatric symptomatology. This may occur when these drugs have reached toxic level or when certain addictive medications are withdrawn (barbiturates, narcotics, benzodiazepines, and meprobamate). It should also be noted that steroids, anticonvulsants, methylphenidate and dextroamphetamine, beta blockers, and clonodine are also associated with withdrawal symptoms in some patients (Adler et al. 1982).

Of importance is the fact that acute reactive psychosis in the ICU is very rare, and thus the phrase "ICU psychosis," while popular with nonpsychiatric physicians, is a barren term. Most patients diagnosed as having ICU psychosis actually have delirium of unknown origin. Pharmacologic agents are probably the most frequent causes of delirium in the ICU. Table 2 lists some medications in clinical use that have been associated with delirium, although this list is by no means exhaustive. Drug–drug interaction merely complicates the issue. The usual management is to discontinue or reduce the dose of the drug; however, sometimes this is not feasible. One example involves patients with malignant ventricular arrhythmias who require lidocaine at a high infusion rate to maintain ventricular stability. In such patients, intravenous haloperidol is used to offset the agitating effects of the lidocaine delirium.

Another frequent occurrence is anticholinergic delirium. This delirium can be reversed by the use of intravenous physostigmine. Doses of 0.5–2 mg can be administered intravenously; however, the physician must be cautious given the potential for profound bradycardiac and hypotensive effects. Glycopyrrolate has been used prophylactically with an intravenous injection of 0.2 mg to protect against the cholinergic side effects of physostigmine, as glycopyrrolate does not pass the blood–brain barrier and therefore will not add to anticholinergicity.

If removal of the factors known to cause CNS impairment fails to lead to improvement in the delirium, the treatment of choice is treatment for nonspecific delirium. Psychosocial and behavioral approaches are usually not effective. The usual treatment includes the administration of antipsychotic

Table 2. Drugs associated with delirium and other psychiatric symptoms

Drug type	Drugs
Antiarrhythmics	Dilisopropamide Lidocaine Mexiletine Procainamide Tocainide
Antibiotics	Aminoglycosides Amphotericin Cephalosporins Chloramphenicol Chloroquine Ethambutol Gentamicin Isoniazid Rifampin Sulfonamides Tetracyclines Ticarcillin Vancomycin
Anticholinergics	Atropine Scopolamine Tricyclic antidepressants, amitriptyline, protriptyline, imipramine, desipramine, nortriptyline, trimipramine, maprotiline Trihexyphenidyl Benztropine Diphenhydramine Thioridazine Eye and nose drops
Anticonvulsants	Phenytoin
Antihypertensives	Captopril Clonidine Methyldopa Reserpine
Antineoplastic agents	Procarbazine L-Asparaginase Methotrexate (high dose) 5-Azacytidine Cytosine arabinoside (high dose) Vincristine Vinblastine 5-Fluorouracil Hexamethylmelamine Dacarbazine Aminoglutehthimide Tamoxifen
Antiviral agents	Acyclovir Gancyclovir Interferon
Barbituates	
Beta blockers	Propranolol Timolol
Cimetidine, rantidine	

Table 2 (Continued)

Drug type	Drugs
Digitalis preparations	
Disulfiram, metronidazole	
Dopamine agonists (central)	Amantadine Levodopa
Ergotamine	
GABA agonists	Benzodiazepines Baclofen
Immunosuppressives	ACTH Cyclosporine Steroids
Lithium	
Monomine oxidase inhibitors	Phenelzine Procarbazine
Narcotic analgesics	Meperidine (normeperidine) Pentazocine
Nonsteroidal anti-inflammatory drugs	Ibuprofen Indomethacin Naproxen Sulindac
Podophyllin (topical)	
Sympathomimetics	Amphetamine Cocaine Ephedrine Phenylephrine Phenylpropanolamine Aminophyline Theophyline

medications with or without benzodiazepines. If these medications are not sufficiently successful in controlling agitation and behavioral lack of control, even when physical restraints are in place, paralytics are sometimes used. Paralytic drugs such as pancuroninum bromide require the use of mechanical ventilation, which carries a clear risk in postoperative patients and also requires concurrent sedation.

Occasionally, the cause of agitation in a patient with delirium is pain, and in such situations morphine sulfate, Dilaudid (hydromorphone hydrochloride), or fentanyl are used to manage this aspect of agitation. Respiratory depression and hypotension are potential side effects, although the short-acting profile of fentanyl has led to increased safety and increased use in the ICU. Occasionally, patients whose agitation is being managed with fentanyl find it difficult to reach steady state tranquilization because of the short-acting nature of the drug. Intravenous haloperidol is the mainstay in the management of delirious agitation (Shapiro et al. 1995). In Europe, haloperidol is used intravenously to treat delirium tremens and acute psychosis and to premedicate electroconvulsive therapy patients. The effects of haloperidol on blood pressure, pulmonary artery pressure, heart rate, and respiration are milder than those of the benzodiazepines. This makes it a treatment of choice for severely ill patients with impaired cardiorespiratory status (Sos and Cassem 1980). Intravenous use of haloperidol is preferable to intramuscular administration for several reasons. Drug absorption may be poor in distal muscles if a patient's delirium is associated with hemodynamic instability. In addition, the agitated patient is usually quite paranoid, and intramuscular injections that are painful may worsen his or her paranoia. Interpretations of muscle enzyme reports may be erroneous if intramuscular injections are used frequently. Moreover, haloperidol used intravenously is less likely to produce extrapyramidal syndromes than intramuscular or even oral haloperidol (Menza et al. 1987). Intravenous haloperidol has a mean distribution time of 11 min; however, in critically ill patients and in the elderly, this may take even longer and thus observable tranquilizing effects may take 15–20 min. The mean half-life of intravenous haloperidol is approximately 14 h. In comparison, the mean half-life of intramuscular haloperidol (oral formula) is 24 h. Intravenous haloperidol is twice as potent as the oral form.

The intravenous line should be flushed with 2 ml normal saline prior to the use of haloperidol. Phenytoin precipitates haloperidol, and the two should therefore not be given in the same intravenous line. Haloperidol may also precipitate heparin, and since all lines in the ICU are usually heparinized, a 2-ml flush is advised. Haloperidol is usually given in 2- to 2.5-mg doses for mild agitation, 5-mg doses for moderate agitation, and 7.5- to 10-mg doses for severe agitation. Doses can be repeated every 30 min until the patient is calm. If the patient becomes agitated again after a calm state has been achieved, a repeat dose can be given. The dose of haloperidol on the second day should, if clinically feasible, be less than that given on the first day. If the sensorium and behavior improve, the patient will often only require haloperidol in the evening to protect against "sundowning" agitation. In this case, often only small, 1- to 5-mg doses of haloperidol will be required and can often be given by mouth. The psychiatric consultant tries to break the agitation rapidly as is done with delirium tremens. Partial control of agitation can sometime prolong the delirium behavior. Obviously, in the elderly delirious patient, doses should be approximately one third of what is usually used.

Maximum daily doses have not been established for intravenous haloperidol, although at the Massachusetts General Hospital over 1000 mg haloperidol per day has been used with safety (Tesar et al. 1985). Hypotension

may occasionally follow the use of intravenous haloperidol. It is often caused by a low-volume state; therefore, if patients have pulmonary artery catheters allowing for close monitoring, as is often the case in the ICU, volume replacement can be used to offset hypotension before administering further doses of haloperidol. Haloperidol may lower the seizure threshold, although it is one of the safest neuroleptics in this respect. It is a safe medication for patients with chronic obstructive pulmonary disease and is thus often used to reduce severe anxiety and agitation in patients who are having difficulty being weaned from the ventilator. Intravenous haloperidol does not block dopamine-related increase in renal blood flow. On occasion, it has been found to cause ventricular arrhythmia, sometimes in a torsades des pointes configuration, so electrocardiographic monitoring is important, especially in susceptible populations such as those with a prolonged corrected QT interval (Metzger and Freidman 1993).

Intravenous droperidol can also be used to manage delirious agitation; however, this agent is more potent at the α^1-adrenergic receptor as an antagonist and is thus more likely to cause problems with hypotension. It can also be more of a respiratory depressant. Intravenous benzodiazepines are often used alone or in combination with haloperidol for the treatment of agitation due to delirium, especially if patients have stable respiration while on mechanical ventilation. It may be beneficial to add intravenous lorazepam (1–2 mg IV every 2–4 h) to intravenous haloperidol in the management of severely agitated patients in the ICU setting, because intravenous lorazepam may reduce the dosage requirement for intravenous haloperidol as it adds a sedative effect. It may also reduce the extrapyramidal side effects of haloperidol, especially neuroleptic-induced catatonia and possibly the neuroleptic malignant syndrome. Lorazepam has also been found to successfully treat catatonic withdrawal states (Fricchione et al. 1997). Of course, it is also a good antiepileptic drug as well. Lorazepam can be used as an infusion at a rate of 1–3 mg per h, if necessary, to calm patients being ventilated. Intravenous lorazepam has a half-life of about 12–14 h and its sedating effect may therefore accumulate. This is an important factor when patients are about to be weaned from the ventilator. To avoid respiratory depression, lorazepam will occasionally need to be tapered before weaning can proceed.

Nevertheless, of the benzodiazepines, lorazepam is one of the least likely to cause respiratory depression. Along with oxazepam, it is safest for patients with hepatic insufficiency due to the fact that it does not require oxidative metabolism. Fentanyl can also be used to manage agitation in inpatients who are being mechanically ventilated, and propofol is another available sedating medication with which anesthesiologists are very familiar. Phenobarbital has been useful occasionally for severely burned patients who require prolonged sedation for burn care. Of course, phenobarbital and propofol are medications that are very depressing to respiratory drive, and therefore extreme caution must be applied, even in the ICU on the ventilator. Other parenteral neuroleptic drugs have also been used for treatment of agitation, including thiothixene, perphenazine (a medication often used intravenously for the nausea and vomiting associated with chemotherapeutic agents), and chlorpromazine, an extremely effective medication which can cause hypotension and a fall in cardiac output because of its very potent α_1-adrenergic antagonism. Chlorpromazine is also relatively anticholinergic and lowers the seizure threshold. In smaller doses, e.g. 10 mg, it can be both safe and effective.

1.4 Intensive Care Unit-Specific Problems and Their Management

1.4.1 Specific Unit Issues

In 1971, Cassem and Hackett reviewed the frequency of reasons for psychiatry consultation requests in various intensive care unit settings.

In the coronary care unit (CCU), in order of frequency, psychiatrists will be called to help with anxiety, depression, and behavioral problems such as signing out against medical advice and dependency/independency conflicts with the staff, hostility, and delirium psychosis.

In the surgical ICU, delirium, depression, and ventilator weaning anxiety are the most frequent reasons for consultation. In the respiratory ICU, depression, ventilator weaning anxiety, and dependency conflicts are the most frequent reasons for psychiatric consultation.

In the medical ICU, psychiatrists often see patients who have attempted suicide by overdose. They also see those with depression, character disorder, delirium, substance or alcohol abuse, and anxiety.

As an example of reactions to critical illness requiring intensive care, myocardial infarction serves as a good model. At the start of their illness, these patients exhibit fear and anxiety; then, as cardiac treatment stabilizes the patient and reduces symptomatology such as anginal pain, the patient may succumb to a resurgency of denial, which can often lead to maladaptive behavior, e.g. insisting on signing out of the hospital against medical advice. Through their presence and advocacy, the family may be able to encourage the patient to accept continued treatment. If

the ministrations of the physician and the family are successful and the patient remains in the hospital, he or she should be calmed by the prompt use of medication, usually haloperidol or lorazepam. The patient may go on to feel a sense of demoralization and vulnerability, culminating in depressed mood and affect. If the patient remains in the hospital for a prolonged period, certain personality problems may arise, such as passive/aggressive behavior or independency/dependency conflicts. Patients who, during the first days of adaptation to the crisis of critical illness, experience fear and anxiety should be treated with anxiolytic medications, with one priority being maintenance of sleep at night in the unit.

Short-acting benzodiazepines such as oxazepam, lorazepam, or alprazolam should be given as a standing dose to patients in the CCU unless there is some contraindication. Midazolam can be used intravenously if patients are not taking medications orally. It is short acting with a half-life of 1–12 h, whereas oxazepam and lorazepam have half-lives of 5–15 h and 10–20 h, respectively. Diazepam and chlordiazepoxide are moderately long acting benzodiazepines, and clonazepam is a longer-acting benzodiazepine. Clonazepam and alprazolam are often best for patients with panic-like anxiety, and clonazepam is often a useful agent in patients who are anxious and agitated with reactions to certain medications such as steroids and in managing manic symptomatology. Zolpidem is a sleeping agent with some non-benzodiazepine-like properties at the respiratory center level, thus making it perhaps a safer sleeping agent associated with slightly less respiratory depression. Buspirone is a non-benzodiazepine anxiolytic alternative. In the medically ill, it is occasionally a very worthwhile medication in the management of anxiety, having few side effects. This medication has been used with success particularly in patients in the respiratory ICU, as it is a respiratory stimulant and not a respiratory depressive. It can also be an adjuvant in patients with depression already on antidepressants and has been used to help manage agitation in patients with organic brain syndrome. A neuroleptic agent is often used if anxiety and fear become so severe that they approach psychotic agitation. Patients who are panicked to this degree are usually out of control. The preferred neuroleptic is haloperidol.

Psychiatrists can also aid patients with fear and anxiety through clarification, explanation, and reassurance. Behavioral relaxational therapies can also be used, although often relaxation breathing instructions may need to be altered. Correction of false myths is often fruitful, since these incorrect misconceptions can add to anxiety. Stressing the positive is appropriate. With a cardiac patient, this may mean telling the patient that normal activities can be resumed after an exercise program has been completed. In patients whose prognosis is poor, restatement of plans, goals, and treatments can go a long way in reducing fear and anxiety. Reassurance can be given by assuring the maintenance of comfort during the crisis. False reassurances regarding cure are not appropriate, as a physician needs to maintain credibility. The physician's equanimity may have the most therapeutic effect on patients, as long as the physician can express his or her empathy in the bargain.

The psychiatrist may become involved in consultation regarding capacity to make medical decisions such as signing out. Such a psychiatric evaluation will require a detailed mental status examination with a special focus on insight and judgment (Applebaum and Grisso 1988; see also Vol. 2, Chaps. 16 and 18).

While anxiety represents the fear of loss of health, of life, and of role as "spouse, parent, and citizen," depression represents a sense of having lost already. With any critical illness, the patient's mind may be said to have suffered an ego infarction. This can lead to psychological damage that can be referred to as despondency. This can be distinguished from major depressive disorder, which is a clinical illness. Recovery of psychological well-being often requires 2–3 months. Feeling more like one's self is the best treatment for illness in acute despondency. Before this occurs, patients should be encouraged to express concerns about what they feel the illness has done to their self-esteem and functioning, and it is important to let patients know that such concerns are normal emotional responses to physical illness and that they will disappear once the patient begins to feel healthy again.

If the patient has loss of energy and is unable to be mobilized, a psychostimulant such as methylphenidate in doses of 2.5–20 mg or dextroamphetamine in this same dose range may be safe choices for treatment. Patients should be checked in terms of their vital signs and response to the psychostimulant about 1 h after a dose. In patients with physiologically significant tachycardia, e.g. patients with atrial fibrillation with rapid ventricular rate, caution must be exercised with psychostimulants that can increase heart rate. These medications should not be given after 1–2 p.m. in the afternoon, as they can cause insomnia. On the other hand, they do not appear to cause anorexia in the medically ill, and in fact they may improve appetite. The serotonin reuptake-inhibiting medications (SSRI) sertraline, paroxetine, and fluoxetine are good alternatives to begin the treatment of major depression in the medically ill in the ICU. They are well tolerated in terms of respiration and cardiac functioning, but they may cause gastrointestinal dysfunction and restlessness. Because of their effects on certain cytochrome systems, drug–drug interactions may be a limiting factor in their use. Buproprion is also an excellent

choice in the management of depression in these patients and has been shown to be a very safe medication in the cardiac population. While it has been associated with seizures, when used in doses of less than 450 mg a day and with single doses of no more than 150 mg, it is safe in this regard.

The tricyclic antidepressants have four major problems associated with them:

1. Orthostatic hypotension (nortriptyline is the least likely to cause this)
2. Significant anticholinergic effects (desipramine has the least such effects of the tricyclics)
3. Conduction defects and increased risk of complete heart block especially in those patients with bifascicular block and second-degree heart block
4. Possible increase in the risk of ventricular arrhythmias (because they will increase a QTc interval)

1.4.2 Anxiety Caused by Weaning from the Ventilator

Patients with anxiety caused by weaning from the ventilator are often a great challenge for psychiatric consultants in the respiratory ICU. Patients will often be treated with a benzodiazepine such as lorazepam, which can be given prior to the weaning periods, or with a neuroleptic such as haloperidol if the patient's anxiety approaches the extremes of fear or agitation. Behavioral exercises can also be used, including hypnosis or relaxation techniques. It is always important to ask the patients themselves which medicine has been most effective. In some cases, for instance, the use of diphenhydramine has provided the calmness required for ventilator weaning.

1.4.3 Patients Exhibiting Maladaptive Behaviors

In the face of a crisis, some patients will regress to maladaptive behaviors, including noncompliance with the treatment regimen. These patients are usually thrust back into their own struggles with dependency and independency issues. They will sometimes become overdependent, refusing to get out of bed or to participate in physical therapy or refusing to eat unless spoon fed, etc. On the other hand, patients will occasionally become hyperindependent. These are patients who, instead of merely dangling their feet by the side of the bed at that stage of their recovery, will be found jogging on the spot or doing push-ups at the bedside. The physician should try to reestablish compliance in these individuals without becoming embroiled in battles over control. This will sometimes involve stressing to patients that they deserve the best treatment possible and that in order to get it they need to become partners in their own care.

Patients with narcissistic demanding traits must be informed repeatedly, in an appeal to their entitlement, that, when they comply, they are doing so in their own best interest. The use of humor, when indicated, can often diffuse some of the tension involved in these dependency/independency conflicts. When patients become hostile, angry, and aggressive, the physician can try a hearing-out process, avoiding the tactical error of saying "I understand." Reestablishing the common enemy – the disease itself – is often helpful. There are some patients who are not merely regressed in the context of a severe illness. Instead, they may have severe personality disorders that are premorbid in nature. Borderline personalities may be so difficult that the help of a psychiatrist is required. Occasionally, these patients are manipulative, help-rejecting complainers. They can stir up staff conflicts through the use of primitive defense mechanisms such as splitting, idealization–devaluation, projection, and pathological denial. While limit setting is often important, it is also important to acknowledge the real stressors in the patient's situation. Breaking down necessary defenses such as avoidance and idealization–devaluation should be avoided. The physician should also minimize overstimulation of the patient's wish for closeness while at the same time minimizing overstimulation of the patient's rage. Again, the physician can utilize patients' narcissistic entitlement by highlighting that they deserve the best treatment possible, which requires their cooperation.

1.4.4 Intensive Care Unit Transfer

Patients leaving the ICU environment experience both a loss and a gain. They may become increasingly anxious, because moving to a stepdown unit or to a ward environment is reflected in decreased staffing coverage and observation. On the other hand, this move is symbolic of the fact that their condition has improved to the point where they no longer require ICU care. Patients should be given educational information about the transfer at least 24 h in advance. They will then be able to anticipate less frequent checks and fewer nurses. At the same time, they can be assured that they are now healthy enough to no longer require intensive care and that, as a reward for their improvement, they are moving to a healthier environment.

1.5 Conclusions

The psychiatric consultant can be valuable in the recovery of critically ill patients by helping to relieve symptoms of behavioral, cognitive, and affective dysfunction that are related to the medical problems at hand. The psychiatrist accomplishes this through a careful examination and evaluation of the patient and the family, through an extensive differential diagnosis, and through the process of discovering etiologies and addressing them via a therapeutic multimodal approach that is closely integrated with the patient's critical care regimen.

Over the past 30 years, there has been an unprecedented increase in the pace of development of technology. The benefit of this technology in the majority of cases outweighs the risk. Nevertheless, significant psychiatric morbidity is associated with the use of medical technological innovations. Working at the interface between medicine and surgery, psychiatrists are perhaps most frequently confronted with the risk of psychiatric morbidity not only in the ICU setting but also in the area of organ transplantation.

2 Psychiatric Problems Related to Organ Transplantation

2.1 Introduction

The pioneering efforts in bone marrow and solid organ transplantation have begun to bear fruit during the last 20 years. At an early stage, surgical therapies and immunotherapy were relatively crude, with complicated postoperative courses and minimal psychosocial care available. Recently, in the setting of generally improved survival, more interest has been paid to the psychosocial and psychiatric care needs of transplant patients. This may stem from the fact that teams have lost patients after successful transplantations due to their severe psychiatric disorders in the form of suicide or drug abuse or noncompliance with postoperative immunosuppressive therapies.

2.2 Psychiatric Assessment of Organ Transplant Candidates

Transplantation assessment usually involves screening carried out by internists and surgeons who are specialized in the organ system being considered for transplantation. Nursing coordinators become involved in the formal assessment phase of potential transplant recipients and also take part in the education and support of patients and their families. There is usually also a social evaluation of the patient's home milieu and the quality of the support network. Considerations include the quality of the marital support available, support of others, including family and friends, financial resources, and geographic factors (Strouse et al. 1996).

By virtue of their education and training, transplantation psychiatrists are in a good position to direct the psychosocial evaluation of potential transplant donors and recipients (Levenson and Olbrisch 1996). They must be proficient in the examination of character and its effect on present and future compliance and will also have the expertise to evaluate and manage not only primary psychiatric disorders, but also secondary disorders pathophysiologically caused by the physical compromise organ failure and various treatments create. The psychiatrist will also need to be skilled at designing and co-managing a treatment regimen for each specific patient (Strouse et al. 1996).

Lowry and Martin (1992) have written about the ethical aspects of transplantation selection committees. Psychosocial evaluation must be done in a equitable manner, insuring equal access while avoiding the danger of harm to the patient. They urge that there should be informed consent for prospective transplant recipients, informing them of the psychosocial information used by the selection committee. Strouse and colleagues (1996) have summarized criteria for psychosocial screening. Absolute contraindications include:

1. Ongoing substance abuse
2. Psychosis which hinders informed consent or compliance in a significant way
3. Transplant refusal
4. Active suicidality
5. Physical symptoms without any pathophysiological basis (factitious disorder)

Relative contraindications include:

1. Dementia or static encephalopathy
   a) Inadequate psychosocial resources to maintain compliance
   b) Association of the brain disorder with a high incidence of post-transplantation neuropsychiatric morbidity, e.g. frontal lobe dementias
2. Psychiatric disorder such as a potentially lethal mood disorder, schizophrenia, or character disorder that has been refractory to treatment
3. A history of noncompliance with the transplant program, including refusal to take part in obligatory psychiatric or psychosocial treatments

Orentlicher (1996) has recently written about the use of psychosocial criteria to assess candidates for organ

transplantation in light of the American Disabilities Act (ADA). The ADA does not allow discrimination on the basis of any disability. However, it accepts that a person's mental health must be considered when organ allocation is being decided, as this involves a bonafide social interest, especially since a person's mental disability may reduce the benefit that he or she will receive from a transplant. Orentlicher therefore believes that the ADA will be interpreted by the courts to permit organ denial or lower waiting list priority for those with intractable psychiatric disorders as long as these diminished benefit forecasts are based on scientific criteria with good validity. They will also likely require that the assessment of candidates be individualized and not solely based on predictors that are generalizable and that transplant programs should involve a program of psychiatric care to help to improve the organ transplant candidate's mental disability.

2.3 Psychiatric Assessment of Living Donors

Psychiatrists specializing in organ transplantation also become involved in assessing living relatives and other living donors.

Strouse et al. (1996) recommend that potential organ donors be assessed for informed consent in essentially the same way as in any other presurgical condition. There are special considerations, however, including the following:

1. Assessment of family systems, focusing on coercion in the donation process and the concept of recipient obligation in response to the gift of an organ
2. Investigating the "black sheep" phenomenon in which a donor may have unrealistic beliefs that he or she will wipe clean any past indiscretions or problematic behavior by the organ donation
3. Investigating whether there are any individual, cultural, or religious apprehensions about the donation of an organ
4. Investigating the possible worsening of a donor's psychopathology secondary to biopsychosocial strain of organ donation

Studies in kidney donation appear to show that most kidney donors feel a sense of satisfaction and would repeat their donation if the opportunity arose again. A recent study of partial liver segment donors, which mostly involves parents of children requiring liver transplant, found that the donors tended to do well psychiatrically after the transplant, with increased self-esteem and satisfaction gains. This coincides with the findings on kidney donors (Fellner and Marshall 1968; Goldman 1993).

2.4 Preacceptance Phase

There are two rating scales which are used for the assessment of psychosocial variables in transplant candidates. These are the Psychosocial Assessment of Candidates for Transplant Scale (PACT) and the Transplantation Evaluation Rating Scale (TERS). The former was developed by Olbrisch et al. (1989), while the latter was developed by Twillman and colleagues (1993). They are both weighted scales with various psychiatric and behavioral categories. The TERS appears to have good interrater reliability, which is also true of the PACT, but the former may also have better predictive power in terms of post-transplant clinical outcomes. While reliability of psychosocial assessments has been established within some programs, Olbrisch and Levenson (1995) conclude that more validation studies are necessary with particular attention paid to patient and family outcomes in addition to patient survival.

In a recent study, the statistical relationship between psychiatric diagnosis, psychosocial adjustment, and health status was examined in organ transplant candidates. The authors interviewed and carried out psychometric testing on 311 heart, kidney, lung, and liver transplant candidates (Chacko et al. 1996). They found that more than 60% met criteria for DSM-IIIR axis I diagnoses and almost 32% met criteria for axis II diagnoses. The axis I disorders were associated with reduced psychosocial adjustment and health status. In patients with both axis I and axis II pathology, which comprised about 25% of the sample, the poorest premorbid coping and the worst degree of marital harmony was found. These patients were at particular risk for poor outcome. The most common axis I disorders found were adjustment disorders, with depression and anxiety. The authors conclude that psychiatric evaluation and management may be important in the maintenance of good health in transplant patients and believe that their findings underscore the importance of psychiatry consultation-liaison participation in the evaluation of transplant candidates.

Lucey and colleagues have developed an assessment of alcoholic transplant candidates (Lucey and Beresford 1992; Lucey et al. 1992). In what is called the Beresford Algorithm, alcoholic patients seeking transplantation require:

1. Accurate diagnosis
2. Understanding by both the family and the patient of the nature of alcoholism as a disease
3. Evaluation of social stability of the patient and family

4. Assessment of the patient's ability to maintain commonly accepted predictors of long-term abstinence

In a recent survey study of academic liver transplant team selection practices for alcohol-dependent applicants, Snyder and colleagues (1996) found certain consistencies between the centers. There was agreement that many alcoholic patients are suitable applicants for transplantation. None of the centers would reject out of hand an alcohol-dependent patient who had been abstinent for longer than 6 months, but some centers do reject applicants with less than 6 months of sobriety. The patients that most concern centers appear to be those with current alcoholism or with abstinence of less than 1 month. The centers are also most concerned about patients requiring a second transplantation after an alcoholic relapse and about those with comorbid antisocial personalities, schizophrenia, or mental retardation. Most centers do not proceed with outright rejection, with the majority preferring to require alcohol rehabilitation or observation. Of course, there are high degrees of mortality during the period in which alcoholics wait to complete an abstinence requirement.

In 1996, Kumar and colleagues presented an analysis of data concerning alcoholic liver transplant patients. They found drinking relapses in 11.5% of 73 alcoholic patients after liver transplantation. Most of these were said to be drinking socially. However, the majority of centers in the survey carried out by Snyder et al. (1996) (12 out of 14) had had experience with alcoholic patients who returned to alcohol intake after the transplantation. There were alcohol-related cases of graft failure in seven centers and of death in four centers. Nevertheless, Starzl and colleagues (1988) have found that survival of alcoholic patients after liver transplants is comparable to that of adult patients with other hepatic diseases.

Other substance abuse is highly correlated with noncompliance with medical and surgical postoperative management and with subsequent graft loss (Gastfriend et al. 1989). Another problem is that patients exposing themselves to drugs of abuse while they are in an immunosuppressed state as transplant recipients will be more susceptible to medical complications. Chang and colleagues (1997) recently found a significant reduction in survival for patients with substance abuse who had undergone bone marrow transplantation. Alcohol, marijuana, and opiates were identified as the principle forms of substance abuse.

The literature on the use of transplantation in patients with chronic psychotic disorder such as schizophrenia is scarce. While there have been reports of post-transplant outcomes such as suicide or psychotically induced noncompliance leading to transplant rejection, there is hope that more effective treatments for psychosis may improve the chances of these patients becoming transplant recipients. In terms of mental retardation and dementia, Strouse et al. (1996) recommend an approach which entails recognition of potential compliance difficulty, characterization of current psychosocial functioning and the patients' ability to care for themselves after transplantation, and organization of the network of support that will be necessary for post-transplantation care and compliance. These authors believe that exclusion out of hand of patients with dementia or mental retardation is an unfair policy given the fact that there may be few other risk factors for premature death other than organ failure.

In patients with severe treatment-refractory personality disorders, especially when comorbid mood or substance abuse disorder exists, the possibility of post-transplant noncompliance and acting-out behavior must be carefully scrutinized. Some researchers have found this group of patients to have postoperative surgical and psychiatric complications.

While self-destructive or lethal behaviors such as suicidality or refusal to accept transplant are grounds for not accepting a patient, some of these patients will have occult mood disorders related to their organ failure. Occasionally, management of the mood disorder will result in clearance of the suicidal ideation and in transplant acceptance. A situation in which organ failure is secondary to severe suicide attempts requires high-level ethical and clinical consideration (Strouse et al. 1996).

2.5 Pretransplant Phase

Once patients have been accepted onto a transplant waiting list, a new set of psychological challenges has to be faced, especially in areas such as heart transplantation, in which nonurgent transplants are becoming less rare. Potential transplant recipients must face mortality and increasing morbidity, which could eventually make transplantation impossible. It is not surprising that anxiety is therefore a common feature in patients awaiting transplantation. Physicians working with transplantation patients must recognize the anxiety and intervene promptly with referral for psychiatric evaluation and management. It is also important to realize that the risk of suicide is increased with anxiety and depression in the group awaiting a transplant.

Potential transplant recipients often have to wait for their procedure in the hospital because of the risk of increasing medical morbidity. Delirium is not uncom-

mon in these patients and will require psychiatric help with diagnosis and management. These patients also must deal with discouragement, demoralization, and depression, especially when comorbid medical conditions or complications put them at risk for coming off the transplantation list. These occurrences can be destabilizing not only for the patient but also for the family and loved ones. It is also important to take note of personality changes in the waiting period. Increased dependency or independency may be seen, and ambivalence toward transplantation may become pathological, leading to noncompliance. This requires periodic reexamination and support for compliance behavior.

2.5.1 Psychiatric Disorders

Patients Awaiting Renal Transplantation
Patients awaiting renal transplant usually do so while on dialysis. Major depressive disorder is often diagnosed in this population. There are also cognitive disorders, including dialysis dementia and organicity associated with end-stage renal disease. The transplant may improve these cognitive disorders.

Patients Awaiting Heart Transplantation
Patients who are potential heart transplant recipients often suffer from anxiety disorder. Cardiac diseases such as cardiomyopathy and congestive heart failure as well as valvular disease and arrhythmia are often associated with panic-like anxiety and depressive adjustment disorder. There are also patients with long-standing dysthymic disorder. This population will also be subject to cognitive disorder and other organicities with hypoxia and hyperfusion among the many etiologies. In addition, medications such as antiarrhythmics and digoxin used to stabilize these patients will often lead to delirium. In view of the shortage of organs available, patients awaiting a heart transplant will now sometimes require left ventricular assist devices (LVAD). Shapiro and his colleagues (1996) at Columbia Presbyterian have written about the psychosocial aspects of LVAD. While these mechanical marvels are successful in maintaining hemodynamic stability over a period of weeks to months, serving as a bridge to heart transplantation, they are associated with significant psychiatric and psychosocial sequelae. The same can be said for automatic implantable cardioverter defibrillators (AICD) used for ventricular arrhythmias (Fricchione et al. 1994). Panic/phobic and post-traumatic stress anxiety may be problematic in these patients, especially those with frequent device firings.

In the Shapiro group's LVAD experience, consultation was most often requested for family strain, major depressive disorder, organic mental syndromes, and significant adjustment problems. Cerebrovascular disease was a risk factor for depression and encephalopathy. Strokes were a risk of LVAD implantation, and the authors recommend treating major depression vigorously with antidepressants (serotonin reuptake inhibitors and stimulants) and psychotherapy. They feel that the treatment of both mood disorder and encephalopathy was a key factor in giving patients improved function and improved quality of life during the use of the LVAD.

Patients Awaiting Lung Transplantation
Patients awaiting single or double lung transplantation have a high incidence of anxiety disorder and depressed mood. Patients with chronic pulmonary disease are often dyspneic, and the level of their dyspnea seems to coincide with the degree of their mood dysfunction.

In addition to the anxiety problems including secondary panic-like episodes and major depression, dysthymia, and adjustment disorder depression, chronic lung disease patients often have a history of alcohol or other substance abuse and of encephalopathy.

Patients Awaiting Liver Transplantation
Because of the occurrence of hepatic encephalopathy in patients with liver failure, delirium is a frequent finding both in patients awaiting liver transplantation and in those who have undergone a transplantation. While patients may have improvement in neuropsychiatric functioning after a liver transplantation, postoperative delirium is frequent. In some cases, this delirium will lead to a dementing illness. In addition to these organic brain disorders, patients awaiting liver transplant will also suffer from adjustment reactions and major depression. Trzepacz and colleagues (1989) found that 9% of patients awaiting liver transplant have alcohol abuse or dependence, while 2% have abuse of other substances.

Patients Awaiting Bone Marrow Transplantation
Patients who have accepted bone marrow transplantation can become depressed during preparation and immunosuppression in the pretransplant stage. Anticipation of isolation while battling against cancer plays a role in mood changes. Total body irradiation which includes the brain can sometimes lead to acute radiation encephalopathy with cognitive changes (Jenkins et al. 1991).

2.5.2 Psychiatric Management

Surman and colleagues (1987) have discussed treatment in patients awaiting transplantation. They recommend symptom-directed behavioral intervention,

relaxation techniques, supportive psychotherapy, and psychopharmacologic interventions. Psychopharmacologic interventions include the SSRI and bupropion for depression in patients with end-stage organ diseases. Psychostimulants such as methylphenidate, dextroamphetamine, or pemoline can also be used, particularly for patients who are depressed with psychomotor retardation and fatigue. Anxiety reactions and panic-like states can be treated acutely with benzodiazepines such as lorazepam and oxazepam, which have no active metabolites and undergo simple glucuronidation in the liver. This is not to say that they cannot cause further confusion in patients who are encephalopathic, for example. Buspirone is a very well tolerated medication for anxiety disorders which avoids the benzodiazepine side effect profile, including respiratory depression. It has been used particularly in patients with lung disease. As noted, haloperidol can also be helpful in managing a high degree of anxiety and agitation states associated with delirium or psychosis.

2.6 Post-transplant Phase

2.6.1 Postoperative Considerations

Strouse et al. (1996) have outlined the factors predictive of psychiatric complications after transplantation surgery. They include the nature of the procedure performed, length of the surgery, intraoperative complications, metabolic derangements, physiologic stability, and medications given. Other issues must be attended to in the postoperative environment. These include the adaptation of those patients who require strict isolation to the sense of feeling alone and abandoned. They also include the questions the patient may raise about who the donor was. Some programs allow for very circumscribed information about the donor, such as a demographic description. In general, donors remain anonymous. Patients must also become acclimated to a new level of medical care. They must pay strict attention to their medication regimen and to the cataloguing of information such as weight and vital signs. The post-transplant patient should also be prepared to be vigilant in regard to the signs and symptoms of infection and rejection. Of course, the price to pay for this vigilance is often an elevation in anxiety level, leading to stress and adjustment reactions. If an organ is rejected, the patient will undergo a grief reaction, culminating in withdrawal and self-isolation and at times severe depression. These patients are often at elevated suicide risk. There are a variety of ways in which patients will deal with the potential for retransplantation once rejection takes hold. Since the decision for transplantation is often made in the context of a major depression or an anxiety reaction related to the graft rejection, it is often very important that psychiatry be an integral part of the treatment team.

2.6.2 Postoperative Neuropsychiatric Disorders

Delirium is not uncommon in patients who have had liver transplants. It is also a frequent occurrence in patients after lung transplantation. Patients undergoing heart and renal transplants also develop delirium, although not as often. Delirium in patients after transplantation will be caused by the added strain of major surgery and general anesthesia in patients who have end-stage organ disease. Metabolic derangements, intraoperative complications, postoperative use of CNS-active medications (including opiates, chemotherapeutic agents, and immunosuppressives such as cyclosporine, corticosteroids, and OKT3), early graft rejections, infections (including opportunistic ones), and residual effects of substance abuse (including withdrawal states) can all cause delirium. In bone marrow transplant recipients, brain irradiation may lead to acute, early-delayed, and late-delayed encephalopathies (Pollard and Young 1997). Encephalopathy is more likely if methotrexate is given during or after cranial irradiation. Cyclosporine is not only associated with delirium after transplantation, but also can result in neurotoxicity (Strouse et al. 1996). Cyclosporine neurotoxicity is associated with diffuse white matter lesions on magnetic residence imaging (MRI). It is well known that corticosteroids can also result in delirium as well as secondary affective syndromes. In terms of the management of post-transplant delirium, haloperidol has become a standard treatment to control agitation.

Seizures, both generalized tonic-clonic and complex partial ones, are not uncommon complications of transplantation. Those patients who have received liver transplants also can develop CNS lesions, including small vessel disease, leukoencephalopathy, and central pontine myelinolysis. Cyclosporine is a risk factor for the induction of seizures after transplantation. Strouse et al. (1996) point out that a cyclosporine holiday involving brief periods off the drug is sometimes required in order to manage seizures especially since anticonvulsants are occasionally unsuccessful in preventing them. There are frequent secondary organic mood disorders in patients after transplantation. Patients may become depressed and/or manic in this setting. Of course, medication side effects play a large role in the production of these affective changes. Even the antiviral agents acyclovir and gancyclovir can produce mood changes as well as delirium, panic, and isolated visual hallucinations.

Levenson and Olbrisch (1993) have reviewed the treatment of heart transplant recipients with psychopharmacological agents. They discourage the use of tricyclic antidepressants because of the hypotensive effects of these agents and the risk of precipitating ventricular arrhythmias and conduction defects. They sanction the use of newer antidepressants, including the SSRI and buproprion, although they acknowledge that very little has been published on this particular group of patients. They suggest using low starting doses with close monitoring, especially in regard to cytochrome effects in the liver. They also point out that fluoxetine, in addition to having a very long half-life, can elevate the serum levels of both cyclosporine and some antiarrhythmic medications. Our experience has been that SSRI and buproprion are well tolerated in transplant patients; as buproprion neither inhibits nor stimulates cytochrome, it is a particularly good choice in patients after liver transplantation, although vigilance for seizure activity needs to be maintained. Levenson and Olbrisch (1993) point out that electroconvulsive therapy (ECT) has been used safely in patients after transplantation.

Lithium can be hard to manage in patients with severe heart failure due to changes in cardiac output which affect the attainment of a steady state lithium level. Medications used in heart failure, including diuretics, salt restriction, angiotensin-converting enzyme inhibitors, and calcium channel blockers, may also affect lithium levels. Because of major fluid compartment shifts and electrolyte imbalances, lithium is occasionally even harder to manage in the immediate post-transplant period. When post-transplant stability returns, lithium can be used, but cyclosporine drug–drug interaction may pose its own difficulties, since cyclosporine has a tendency to elevate serum lithium levels. In addition, both agents cause tremor. Levenson and Olbrisch (1993) recommend the acute use of neuroleptics for mania in the post-transplant period. They do not feel that carbamazepine should be used because of its tendency to worsen congestive heart failure, potentiate hypotension, and produce arrhythmias and conduction defects. They do feel that valproic acid and clonazepam are reasonable agents in heart transplant recipients who are suffering from a manic episode. These agents can also be used with caution in patients who have had a renal or a lung transplant. Special caution must be used in liver transplant patients given the potential for hepatic toxicity. Cyclosporine, OKT3, and FK506 are all implicated in tremor. Attention must be paid to medications that increase cyclosporine levels, including methylprednisolone, ketoconazole, fluoconazole, cimetidine, verapamil, erythromycin, fluoxetine, and sertraline.

Strouse et al. (1996) report that central pontine myelinolysis in patients after liver transplantation is linked to cyclosporine use, perhaps due to cytochrome enzymes in the donor liver which increase cyclosporine levels. Patients may also complain of loss of taste, changes in hearing, incontinence, and sexual dysfunction after transplantation. Cognitive dysfunction may persist along with a fatigue syndrome after transplantation. It is difficult to distinguish the contribution of end-stage organ disease plus post-transplantation status and changes complicated by immunosuppressive and other medications. The use of psychostimulants and certain antidepressants with activating properties may be helpful in this condition. DiMartini (1996) has recently reviewed the case of a liver transplant recipient who had Wernicke-Korsakoff syndrome. She points out that transplantations may produce a propensity to Wernicke-Korsakoff syndrome secondary to factors such as gastrointestinal distress, chronic vomiting, and nutritional deficiencies brought on by conditions such as cytomegalovirus gastritis and nutritional depletion secondary to end-stage organ disease.

2.7 Long-Term Adjustment to Transplant

Craven (1990) outlined the coping challenges patients face after transplantation. These include readjusting to roles both in the family and in the workplace, adapting to a changed body image, coming to terms with rejection and infection fear, remaining compliant with immunosuppressants despite side effects, adapting to a new level of patienthood by meeting the expectations of the transplantation team, and realizing that the transplantation is not a cure and that even after the transplantation the physical hardships of chronic illness still have to be dealt with. In terms of quality of life following transplantation, a recent study looked at survivors of heart, liver, and lung transplantation procedures (Littlefield et al. 1996). The study included responses from 55 heart, 149 liver, and 59 lung transplant recipients and showed that those who had undergone successful lung transplants reported better functioning than heart or liver transplant patients in the three domains of physical, psychological, and social functioning. While lung patients reported levels of functioning equal to or better than published norms, heart and liver patients reported impairment in the areas of physical and social functioning. The heart patients felt that their new life style was intruded upon by their illness management, and they experienced more difficulty complying with it. Nevertheless, in all three groups, the majority of patients reported that life had improved in terms of health, energy, activity level,

and overall quality of life. Approximately 25% of the total subject group reported employment (full- or part-time). Others have found better employment and activity statistics, particularly in lung transplant survivors, although hepatic and renal transplant patients also have a high rehabilitation rate (Strouse et al. 1996). Altmaier and colleagues (1991) looked at the 2-year adjustment of bone marrow transplant survivors and found that they experienced greater vocational and sexual dysfunction than did a comparison sample of maintenance chemotherapy patients. However, the transplant survivors saw themselves as equally healthy and with low levels of psychological distress.

2.8 Conclusions

Stress and staff burnout are not uncommon in transplantation medicine. The organ transplant psychiatrist may be able to help members of the team through supportive relationships, referral to other psychiatrists or therapists when needed, and through group meetings of the staff, particularly at times of loss or disruption on the unit. Inviting the chaplain can sometimes help to reduce stress as well. The future of organ transplantation will involve certain built-in limits, particularly in the area of solid organ transplantation. These limits will include growing numbers of patients on waiting lists and longer waiting times due to the stagnant donor organ availability. This situation has led to more research on bridging technologies, including left and right ventricular assist devices and even a bile artificial liver (Coffman et al. 1996). However, as with all advances in medical technology, there will be secondary neurologic and psychiatric conditions. Psychiatrists working in cooperation with their transplantation team colleagues will need to employ the methods outlined above to insure that the cognitive and emotional quality of prolonged life is optimized.

3 References

Adler LE, Bell J, Kirch D et al (1982) Psychosis associated with clonodine withdrawal. Am J Psychiatry 139: 110–112

Altmaier EM, Gingrich RD, Fyfe MA (1991) Two year adjustment of bone marrow transplant survivors. Bone Marrow Transplant 7: 311–316

**Applebaum PS, Grisso T (1988) Assessing patients' capacities to consent to treatment. New Engl J Med 319: 1635–1638

**Cassem NH (1995) Psychiatric problems of the critically ill patient. In: Ayres SM, Grenvik A, Holbrook PR, Shoemaker WC (eds) Textbook of critical care, 3rd edn. Saunders, Philadelphia, pp 1589–1599

Cassem NH, Hackett TP (1971) Psychiatric consultation in the coronary unit. Ann Intern Med 75: 9

Chacko RC, Harper RG, Kunik N, Young J (1996) Relationship of psychiatric morbidity and psychosocial factors in organ transplant candidates. Psychosomatics 37: 100–107

Chang G, Antin JH, Orav EJ et al (1997) Substance abuse and bone marrow transplant. Am J Drug Alcohol Abuse 23: 301–308

Coffman KL, Hoffman A, Rosenthal P et al (1996) Neurological and psychological sequalae and transplant recipients after bridging with the bile artificial liver. Gen Hosp Psychiatry 18: 20S–24S

Craven JL (1990) Psychiatric aspects of lung transplant: the Toronto Lung Transplant Group. Can J Psychiatry 35: 759–764

DiMartini A (1996) Wernicke-Korsakoff syndrome in a liver transplant recipient. Psychosomatics 37: 564–567

Fellner CH, Marshall JR (1968) 12 kidney donors. JAMA 206: 2703–2707

Fricchione G, Olson L, Flay S (1994) Cardiac psychiatry and the management of malignant ventricular arrhythmias with the internal cardioverter defibrillator. Am Heart J 128: 1050–1054

Fricchione G, Bush G, Fozdar M, Francis A, Fink M (1997) Recognition and treatment of the catatonic syndrome. J Intensive Care Med 12: 135–147

Gastfriend DR, Surman OS, Gaffey G et al (1989) Substance abuse and compliance in organ transplantation. Substance Abuse 10: 149–153

Goldman LS (1993) Liver transplantation using living donors: preliminary donors psychiatric outcome. Psychosomatics 34: 235–240

House RM, Trzepacz PT, Thompson TL (1990) Psychiatric consultation to organ transplant services. In: Tasman A, Goldfinger SM, Kauffman CA (eds) Review of psychiatry, vol 9. American Psychiatric Press, Washington, DC, pp 515–536

Jenkins PL, Linington A, Whittaker JA (1991) A retrospective study of psychosocial morbidity in bone marrow transplant recipients. Psychosomatics 32: 65–71

Kumar S, Stauber RE, Gavaler JS et al (1996) Orthotopic liver transplantation for alcoholic liver disease. Hepatology 11: 159–164

*Levenson JL, Olbrisch ME (1993) Psychiatric aspects of heart transplantation. Psychosomatics 34: 114–122

Levenson JL, Olbrisch ME (1996) Psychiatric and psychosocial issues in organ transplantation. Gen Hosp Psychiatry 18: 25–45

Littlefield C, Abbey S, Fiducia D et al (1996) Quality of life following transplantation of heart liver and lungs. Gen Hosp Psychiatry 18: 36S–47S

Lowy F, Martin D (1992) Ethical considerations in transplantation, In: Rodin G (eds) Psychiatric aspects of organ transplantation. Craven J, Oxford University Press, Oxford, UK, pp 212–230.

Lucey MR, Beresford TP (1992) Alcoholic liver disease to transplant or not to transplant. Alcohol Alcoholism 27: 102–108

Lucey M, Marion R, Henley KS et al (1992) Selection for an outcome of liver transplantation in an alcoholic liver disease. Gastroenterology 102: 1736–1741

Ludwig AM (1980) Principles of clinical psychiatry. Free Press, New York, p 234

Menza MA, Murray GB, Holmes VG et al (1987) Decreased extrapyramidal symptoms with intravenous haloperidol. J Clin Psychiatry 48: 278–280

Metzger E, Freidman R (1993) Prolongation of the corrected QT and torsades des pointes cardiac arrhythmia associated with intravenous haloperidol in the medically ill. J Clin Psychopharmacol 13: 128–132

Olbrisch ME, Levenson JL (1995) Psychosocial assessment of organ transplant candidates current status of methodological and philosophical issues. Psychosomatics 36: 236–243

Olbrisch ME, Levenson J, Hamen R (1989) The PACT: a rating scale for the study of clinical decision making in psychosocial screening criteria for organ transplant candidates. Clin Transplant 3: 164–169

*Pollard JD, Young GAR (1997) Neurology and the bone marrow. J Neurol Neurosurg Psychiatry 63: 706–718

Orentlicher D (1996) Psychosocial Assessment of Organ Transplant Candidates in the Americans with Disabilities Act. General Hospital Psychiatry 18: 30S–35S

**Shapiro BA, Warren J, Egol AB, Greenbaum PM, Jacobi J, Nasraway SA, Schein R, Spevetz A, Stone JR (1995) Practice parameters for intravenous analgesia and sedation for adult patients in the intensive care unit: an executive summary. Crit Care Med 23: 1596–1600

Shapiro PA, Levin HR, Oz MC (1996) Left ventricular assistive devices. Psychosocial burden and implications for heart transplant programs. Gen Hosp Psychiatry 18: 30S–35S

Snyder SL, Drooker M, Strain JJ (1996) A survey estimate of academic liver transplant teams selection practices for alcohol dependent applicants. Psychosomatics 37: 432–437

Sos J, Cassem NH (1980) The intravenous use of haloperidol for acute delirium in intensive care settings. In: Speidel H, Rodewald G (eds) Psychic and neurological dysfunctions after open heart surgery. Thieme, Stuttgart

Starzl TE, Van Thiel D, Tzakis AG et al (1988) Orthotopic liver transplantation for alcoholic cirrhosis. JAMA 260: 2542–2547

Strouse TB, Wolcott DL, Skotzko CE (1996) Transplantation. In: Wise T, Rundell J (eds) Textbook of consultation-liaison psychiatry. American Psychiatric Press, Washington, DC, pp 640–670

Surman OS, Dienstag JL, Cosimi AB et al (1987) Psychosomatic aspects of liver transplantation. Psychother Psychosom 48: 26–31

Tesar GE, Murray GB, Cassem NH (1985) Use of high dose intravenous haloperidol in the treatment of agitated cardiac patients. J Clin Psychopharmacol 5: 344–347

Trzepacz PT, Brenner R, Van Thiel OH (1989) A psychiatric study of 247 liver transplantation candidates. Psychosomatics 30: 147–153

Twillman RK, Manetto C, Wolcott DL (1993) Transplant evaluation scale: a revision of psychosocial levels system for evaluating organ transplant candidates. Psychosomatics 34: 144–153

CHAPTER

18

M.F. Costantini-Ferrando, J.C. Holland

Psycho-oncology

1 Introduction

The field of psycho-oncology is a recently developed subspeciality within oncology which reflects the increased interest in the behavioral, psychological, and social factors in cancer prevention and in the quality of life of patients with cancer at all stages. It addresses the two psychological dimensions of cancer: (1) the impact of cancer on the psychological function of the patient, the patient's family, and staff (the psychosocial dimension) and (2) the role that psychological and behavioral variables may have in cancer risk and survival (the psychobiological dimension) (Holland 1998).

This subspeciality is comprised of health professionals from several disciplines and medical specialties who contribute to the wide breadth of clinical and research expertise in psycho-oncology. Psychiatry, psychology, social work, and nursing, in particular, have contributed significantly to clinical care, research, and training in psycho-oncology. Pastoral counselors increasingly bring in the spiritual and religious dimension, and cancer survivors, acting as advocates, also bring their own expertise based on their own unique personal experiences and insights.

The main objectives of the field of psycho-oncology are the following:

1. To foster concern for the integration of psychological care into the total care of cancer patients and their families by all health professionals in oncology
2. To foster the education and training of clinicians and investigators in the subspeciality of psycho-oncology
3. To develop a training curriculum in psycho-oncology which is comprised of a common body of knowledge relevant to professionals from all backgrounds, but with specific modules added to meet the educational needs of each particular oncologic discipline
4. To promote the study of psychological, social and behavioral factors in cancer prevention and detection and to study the effect of psychological, social, and behavioral factors on survival
5. To promote cross-cultural research that attempts to define the universal stresses of cancer and brings an understanding of the influences of cultural factors on adaptation
6. To encourage controlled trials of psychotherapeutic, behavioral, and psychopharmacological interventions to control the distress experienced by cancer patients
7. To increase the awareness of the psychological and social problems and their management that are relevant to cancer patients at all stages of illness
8. To include quality of life as an outcome variable in clinical trials research in order to aid patients in making treatment decisions
9. To incorporate quality of life and well-being as part of the treatment goals and to aim to obtain a disease-free state which includes patients regaining maximal functional outcome in the major domains of living
10. To promote policies at the local, community, state, national, and international levels, aimed at reducing cancer risk and fostering optimal rehabilitation and survival.

The areas reviewed in this chapter address the variables that influence the adjustment of patients to cancer, the range of psychological problems related to patients' adaptation to illness, and the behavioral and psychosocial variables in cancer morbidity and mortality.

2 Adaptation to Illness

Learning that one has cancer, or that a close relative has cancer, is always a frightening event. Thoughts of death, disability, disfigurement, dependence, and disruption of relationships are inevitably linked to a cancer diagnosis. While the initial response to such a diagnosis is fairly universal, individuals vary in how well they ultimately adapt to the disease. It is thus important to identify the factors that predict good or poor adjustment.

Three categories of factors contribute to adaptation to cancer: those contributed by the society, those derived from the patients, and those which are the consequence of the disease itself (Spencer et al. 1998).

1. *Society-based factors.* These are the social attitudes and beliefs held by a society about cancer. They deeply influence the patient's own beliefs and reactions and those of the people around them. The stigma that cancer carries and the long silence in telling patients their diagnoses both stem from long-held beliefs that "cancer equals death." While this is no longer true, attitudes change slowly.
2. *Patient-based factors.* The individual's developmental stage at the time of diagnosis and the life cycle events with which it interferes (e.g. marriage, fertility), the personality, the emotional level of maturity, and coping style, and the interpersonal resources of family and friends and social support

available to the individual are important resources that contribute to adaptation. Predictors associated with poor coping are social isolation, low socioeconomic status, alcohol or drug abuse, prior psychiatric history, prior experience with cancer (e.g. death of a relative), recent bereavement, rigidity of coping strategies, having a pessimistic philosophy of life, absence of a supportive belief/value system which allows the person to find meaning in the experience, and multiple stressors (such as financial or family obligations) (Spencer et al. 1998).

3. *Cancer-based factors.* The disease itself presents certain "givens" which determine the symptoms and disabilities with which the person has to cope, i.e. the stage and site of illness (e.g. colon, breast) and the symptoms (especially pain) and prognosis to which each patient must adapt. Important here as well is the critical role of the doctor–patient relationship and of the health care team, which can either be a strong positive or a troubling negative factor.

These three factors interact and determine the adaptation of each patient at any given point in time. This explains why each person's adaptation is unique to the particular situation and why universals about coping strategies are apt to be more helpful in the understanding of aggregate responses rather than in the application to a single person.

3 Normal Psychological Reactions

When individuals deal with serious illness, responses of fear, worry, and sadness are normal. These are the "normal" forms of distress which all patients manifest to a greater or lesser degree. The distress varies in relation to the challenges presented by the illness at a particular time and the issues to which the person has to adapt. Four areas can be identified that require different coping strategies and types of psychosocial interventions in order to achieve maximum quality of life. They are:

1. Adaptation to active treatment, which may be curative
2. Adaptation to palliative and terminal care
3. Adaptation to being a survivor
4. Adaptation to the knowledge of a high genetic risk and as yet asymptomatic cancer.

In individuals undergoing active and possibly curative treatment, the goal of psychosocial care is to support the patient's coping with the stresses of treatment and its side effects. Behavioral and psychopharmacological interventions, counseling, and self-help groups are all useful techniques to increase adherence to treatment by controlling symptoms of anxiety and depression and side effects of treatment such as nausea, vomiting, and alopecia.

For individuals at the time of transition from treatment with cure as a goal to care with palliation as the goal, there is usually a period of great distress as the new reality becomes clear. However, this is followed by an adaptation to the altered therapeutic goal of control of the disease. Psychosocial intervention should help patients and families with decisions about treatment and care and should be directed toward the control of anxiety and depression. Troublesome physical symptoms, particularly pain, commonly associated with advanced and terminal stages of illness, must also be controlled, and this often results in opiate-related encephalopathy (Breitbart and Payne 1998) and delirium, which require treatment.

In individuals who have successfully completed active treatment and who consider themselves survivors, there is often greater appreciation of life and a need to reexamine values and goals. On the negative side, however, there is a strong fear of recurrence, anxiety, and a sense of enhanced vulnerability to illness and death. Anxiety may dissipate over time, but it is exacerbated at the time of a medical checkup and when minor symptoms occur, which are rationally recognized not to be related to cancer, but are irrationally assumed to be due to cancer. Self-esteem is often lower, and problems with insurance and employment do occur (Cella et al. 1991). Reminders of treatment serve as stimuli for conditional responses of nausea, vomiting, and anxiety after chemotherapy (Cella et al. 1986). Studies have found that 15% of bone marrow transplant survivors experience the complete symptom cluster of a post-traumatic stress disorder, while 10% of them have nightmares, hypervigilence, and flashbacks. Up to 30% of mothers of children with transplants also have similar symptoms of distress (Pelcovitz et al. 1998).

In individuals who are at high genetic risk for cancers, there is an increasing awareness of their family history and a concern about whether genetic testing is wise to obtain. The anxiety about risk is handled differently; whereas some people use denial and refuse to come in, even for routine screening such as mammograms, others repeatedly check their bodies for signs of cancer and present to doctors frequently for evaluation and screening procedures. Regardless of their coping style, these individuals are vulnerable to high psychological distress, and support groups have proved valuable.

4
Psychiatric Disorders

When transient negative emotional reactions exceed normal limits, begin to pervade a person's life, and interfere with treatment and overall quality of life, these symptoms need to be evaluated for the presence of a psychiatric disorder (Roth et al. 1998). The most common psychiatric disorders seen in patients with cancer are adjustment disorders (with anxiety, depression, or both), anxiety disorders, depressive disorders, and delirium related to metabolic problems or medication side effects.

4.1
Anxiety Disorders

Symptoms of anxiety range from normal fears to incapacitating anxiety disorders. Three common sources of anxiety have been identified among cancer patients:

1. Situational anxiety may be related to the disease itself (the initial diagnosis, the anticipation of a new procedure, or, later on, the fear of recurrence). This is usually an adjustment disorder with anxious mood. Anxiety can be related to medical problems, such as poorly controlled pain, abnormal metabolic states, treatment side effects, drugs which produce anxiety or withdrawal symptoms (e.g. narcotic analgesics), hormone-secreting tumors, and paraneoplastic symptoms (Noyes et al. 1998).
2. Anxiety related to the treatment process, such as the fear of painful or frightening procedures (e.g. magnetic resonance imaging, MRI) and anticipatory anxiety prior to chemotherapy (Andrykowsky and Redd 1997; Redd et al. 1987).
3. A preexisting anxiety disorder that may be exacerbated in the context of cancer and its treatment. Examples occur with phobias stimulated by the need to deal with hospitals, needles, blood, or claustrophobia. Traumatic memories that are triggered by the medical setting may also resurface during the stress of cancer as a delayed posttraumatic stress disorder. All are symptoms that must be controlled to ensure patients' ability to tolerate optimal medical treatment.

In terms of treatment, simple situational anxiety can be easily managed by the reassuring and informative words of the treating physician reinforced by the oncology nurse or social worker. When more serious levels of anxiety are present, three means of treatment are appropriate: (1) counseling or formal psychotherapy using a supportive crisis intervention model; (2) behavioral interventions such as relaxation, systematic desensitization, and hypnosis; and (3) psychopharmacological agents (e.g. benzodiazepines, antidepressants, and neuroleptics).

4.2
Depression

Depression is a difficult disorder to diagnose among cancer patients due to the fact that the physical symptoms of depression, such as fatigue, weakness, loss of libido, anhedonia, and poor concentration, are also common symptoms caused by cancer, especially in advanced stages.

Most depressive disorders are reactions to the disease (adjustment disorders), and they occur in about one quarter of outpatients; higher percentages are reported among inpatients who are more physically ill (Bukberg et al. 1984). The symptoms of depression are several: depressed and/or dysphoric mood, restlessness or psychomotor slowing, insomnia, hopelessness and helplessness, guilt, and suicidal ideation. In diagnosing major depression, prior psychiatric and family history of depression and any substance abuse must also be taken into account.

Major risk factors for depression are prior history of depression or suicide attempts, substance abuse, poor social support, or recent bereavement. Risk factors related to the illness itself are advanced disease, poorly controlled pain, presence of another chronic disease or medical complication (such as metabolic, nutritional, endocrine, or neurological conditions), use of certain chemotherapeutic agents (e.g. interferon), and use of medications such as steroids and opiates.

The presence of a depressive disorder is particularly concerning because it is a risk factor for suicide. The incidence of suicide among cancer patients is increased compared to the general population (Bolund 1985, 1986). Recent studies have shown that requests for physician-assisted suicide occur primarily due to depression, not pain (Breitbart and Krivo 1998). Risk factors for suicide are prior history of suicide attempts, prior psychiatric or substance use disorders, depression and hopelessness, recent bereavement, and poorly controlled pain, delirium, advanced disease, debilitation, and exhaustion (Breitbart and Krivo 1998).

An evaluation of a suicide risk should gather information about the disease stage and prognosis, the patient's understanding of his or her symptoms and prognosis, the nature of the suicidal thoughts, the presence of other losses and the past history of psychiatric disorders or of poor adjustment to stressors, and family history of depression.

Depression is first treated by developing a good rapport with the patient and by then offering supportive psychotherapy, cognitive-behavioral treatment, and psychotropic agents.

4.3 Delirium

Medical causes for a change in the mood or behavior of an individual with advanced cancer need to be considered and evaluated for the presence of delirium: metabolic encephalopathy, electrolyte imbalance, treatment side effects (e.g. anticholinergics, chemotherapeutic agents, radiation), infection (e.g. septicemia), and paraneoplastic disorders (Posner 1978; Breitbart and Cohen 1998). Over 75% of hospitalized patients with advanced and terminal illness were found to develop a confusional state prior to death (delirium) (Massie et al. 1983).

Early signs of delirium are changes in sleep patterns, irritability, uncooperativeness, somnolence or agitation, and misinterpretation of sounds and objects. These early symptoms may later be followed by a refusal to cooperate with reasonable requests, anger, verbal abusiveness, illusions, delusions, and hallucinations as the symptoms escalate (Breitbart and Cohen 1998).

In terms of treatment, it is valuable to have a familiar person stay with the patient to provide a known, trusted, reassuring presence. Physical restraints should only be used with caution and checked regularly. Psychopharmacological intervention most commonly consists of administering low-dose haloperidol several times a day to reduce confusion. Lorazepam may be added to reduce agitation.

5 Behavioral and Psychosocial Variables in Cancer Morbidity and Mortality

Behavioral, psychological, and social variables have been found to have an impact on cancer morbidity and mortality. Specifically, lifestyle, poor socioeconomic status and education, early detection and treatment compliance, and the availability of social supports clearly reveal the impact of behavior and social factors on cancer morbidity and mortality (Cella et al. 1991; House et al. 1988). Personality type has not been found to be a significant factor. Bereavement and depression studies have failed to confirm a significant connection to cancer morbidity and mortality. While psychosocial interventions have proven useful in improving patients' well-being and quality of life, it is still unclear as to whether they can have a significant impact on survival (Classen et al. 1998).

6 Conclusion

Psycho-oncology is a new specialty within oncology, reflecting greater interest and concern about the psychological, behavioral, and social factors related to cancer prevention and treatment. The total care of cancer patients today must include concern for their psychological management as one of the basic elements. Research on prevention and detection depends on the social sciences for dealing with ways to change lifestyles and behavior. Theory and practice related to the effects of cancer on psychological function and its prevention depend heavily on the expertise provided by this new speciality. Psychiatry as a medical speciality and psycho-oncology as an oncologic subspeciality have a unique role to play in cancer treatment and prevention.

7 References

Andrykowski MA, Redd WH (1987) Longitudinal analysis of the development of anticipatory nausea. J Consult Clin Psychol 55: 36–41

Bolund C (1985) Suicide and cancer. I. Demographical and suicidological description of suicides among cancer patients in Sweden. J Psychosoc Oncol 3: 17–30

Bolund C (1986) Suicide and cancer. II. Medical and care factors in suicide by cancer patients in Sweden. J Psychosoc Oncol 3: 31–52

*Breitbart W, Cohen KR (1998) Delirium. In: Holland JC (ed) Psycho-oncology. Oxford University Press, New York, pp 564–575

Breitbart W, Krivo S (1998) Suicide. In: Holland JC (ed) Psycho-oncology. Oxford University Press, New York, pp 541–547

*Breitbart W, Payne DW (1998) Pain. In: Holland JC (ed) Psycho-oncology. Oxford University Press, New York

Bukberg J, Penman D, Holland JC (1984) Depression in hospitalized cancer patients. Psychosom Med 46: 199–212

Cella DF, Pratt A, Holland JC (1986) Persistent anticipatory nausea, vomiting, and anxiety in cured Hodgkin's disease patients after completion of chemotherapy. Am J Psychiatry 143: 641–643

Cella DF, Orav EJ, Kornblith AB, Holland JC, Silberfarb PM, Lee KW, Comis RL, Perry M, Cooper R, Maurer LH, Hoth DF, Perloff M, Bloomfield CD, McIntyre OR, Leone L, Lesnick G, Nissen N, Glicksman A, Henderson E, Barcos M, Crichlow R, Faulkner II CS, Eaton, W, North W, Schein PS, Chu F, King G, Chahinian AP (for the Cancer and Leukemia Group B) (1991)

Socioeconomic status and cancer survival. J Clin Oncol 9: 1500–1509

Classen C, Sephton SE, Diamond Susan, Spiegel David (1998) Studies of life-extending psychosocial interventions. In: Holland JC (ed) Psycho-oncology. Oxford University Press, New York, pp 730–742

*Holland JC (1998) Societal views of cancer and the emergence of psycho-oncology. In: Holland JC (ed) Psycho-oncology. Oxford University Press, New York, pp 3–15

House JS, Landis KR, Umberson D (1988) Social relationships and health. Science 241: 540–545

Massie MJ, Holland JC, Glass E (1983) Delirium in terminally ill cancer patients. Am J Psychiatry 140: 1048–1050

*Noyes R, Holt CS, Massie MJ (1998) Anxiety disorders. In: Holland JC (ed) Psycho-oncology. Oxford University Press, New York, pp 548–565

Pelcovitz D, Lebov B, Mandel F, Kaplan S, Weinblatt M, Septimins A (1998) Post-traumatic stress disorder and family functioning in adolescent cancer. J Trauma Stress 2: 205–221

Posner J (1978) Neurologic complications of systemic cancer. Dis Mon 25: 1–60

Redd WH, Jacobsen PB, Die-Trill M, Dermatis H, McEnvoy M, Holland JC (1987) Cognitive/attentional distraction in the control of conditioned nausea in pediatric oncology patients receiving chemotherapy. J Consult Clin Psychol 55: 391–395

*Roth AJ, Kornblith AB, Batel-Copel L, Peabody E, Scher HI, Holland JC (1998) Rapid screening for psychological distress in men with prostate cancer: a pilot study. Cancer 82(10): 1904–1908

Spencer SM, Carver CS, Price AA (1998) Psychological and social factors in adaptation. In: Holland JC (ed) Psycho-oncology. Oxford University Press, New York, pp 211–222

CHAPTER 19

H. Merskey

Pain and Pain Therapy

1 Definition of Pain

In practice, we often fail to find physical explanations for pain. It sometimes seems to arise from psychological processes, as with the headache of depression or the very rare pain which is an hallucination in schizophrenia. This raises a conceptual problem in the understanding of the word "pain". In normal conversation and in technical speech, pain describes an experience which one person relates by word of mouth to another. Normally, it connotes a type of experience which we associate with damage to the body. Thus the word for pain describes a subjective experience. It is a word for a psychological state and does not describe physical conditions. It tells us nothing about what is happening in the axon of the nerve or in nerve pathways. It has bodily reference so that the experience is tied to some portion of the body, but it is always a private experience and something which can only be inferred about a person, either from his or her report or by observation of behaviour which suggests that the individual has this condition. It requires confirmation by direct enquiry. Thus it cannot be defined by external criteria. Advances in neurophysiology and imaging may tell us one day that reports of pain are consistently associated with particular physiological changes, but to date those observations are limited and are not available for general use.

It is common for the word *pain* to be employed when an individual suffers injury or damage to the body and experiences a particular type of subjective state which is found to be unpleasant. The words used to describe pain commonly imply tissue damage (Devine and Merskey 1965). From this reasoning, it has seemed possible to form a definition of pain which would describe it as an unpleasant experience and which would say that it was one which we primarily associate with tissue damage or describe in terms of such damage. It is that characteristic experience which seems to be common to all mankind, with only very rare exceptions among those who do not withdraw from noxious stimuli and are thought to have a congenital absence of the experience of pain. The International Association for the Study of Pain (IASP) (1979) adopted the following definition: "An unpleasant sensory and emotional experience, associated with actual or potential tissue damage, or described in terms of such damage".

In a sense, this definition is only a semantic trick. It removes preoccupation with whether a noxious stimulus is pain and instead provides a psychological framework. It says that we will not use the word "pain" to describe a noxious stimulus, but we will use it to describe what a person feels. One of the advantages of this definition is that in those relatively few cases where pain seems to have a psychological cause and no other explanation, the patient can correctly use the same word as a patient with a broken limb.

This is a monistic view of pain. Often, we cannot distinguish in experience between pain of psychological origin and pain of physical origin. However, we should be very concerned to find the causes of pain. A given pain may have both physical and psychological causes, and in that sense it has a multiple aetiology.

2 Psychodynamic Hypotheses

"We feel a cut from the surgeon's scalpel more than ten blows of the sword in the heat of battle" (Montaigne 1580). This remark by Montaigne illustrates some aspects of the psychology of pain. Mild or moderate anxiety, which we often see in association with pain in clinical practice, seems to heighten pain. Being placed in a passive situation and unable to escape likewise increases pain. On the other hand, heightened arousal, as in battle, appears to minimize or abolish pain.

Besides anxiety, multiple symptoms, including pains in different parts of the body, have been associated historically with the diagnosis of hysteria. What was previously called hysteria has varied enormously in different periods of time, and it is best understood as a broad term used in the past for depression and for psychological complaints about the body and about mental function, e.g. memory. This approach has also led to the category of somatoform disorders in DSM-III and DSM-IV (American Psychiatric Association 1980, 1994) and its successors, separate from conversion and dissociative symptoms.

The historical tradition that pain could be a conversion symptom was popular in psychoanalytic literature and appeared in at least 31 separate psychoanalytic reports (Merskey and Spear 1967). This literature depended upon the use of the concept of repression, a concept which is now in considerable difficulties, particularly with reference to long-term effects (Holmes 1990). Most of the current difficulties with the concept flow from its over-use in the production of multiple personality disorder and from the development of false memories which have led to a re-appraisal of the theoretical and practical justification for the use and employment of the notion of repression (Merskey 1996). There may still be some place for repression as an acute mechanism in

conversion or dissociative disorders occurring with acute stress, but there is no good evidence to scientifically validate the notion of repression as a long-term mechanism.

Nevertheless, the idea that individual reactions may be determined by childhood experience persists. Schilder (1931) interpreted pains in one of his patients as a defence against sexuality and as a perverse sadomasochistic satisfaction. From this starting point, Engel (1951) described "atypical facial pain" as an hysterical conversion symptom associated with many varieties of self-punitive behaviour. He stressed the frequency in his cases of unnecessary surgery, the presence of other somatic symptoms, the tolerance of physical causes of pain – often with gusto – and remission of the pain at times of misfortune or when there was other cause for suffering, whether physical or mental. Later, he named this type of patient the "pain prone patient", whether the symptoms applied to facial pain or pain elsewhere in the body (Engel 1959).

Engel's work was anecdotal and has only been borne out in limited respects. These include the finding of increased resentment in patients with pain, increased numbers of operations, consultations, painful past illnesses and poor marital adjustment (Merskey 1965a,b). Alternative explanations are also available for many of the findings. Irritability and resentment are common phenomena in all patients with pain, whether of psychological origin or physical origin. In animals, aggression is one of the main responses to occur in the presence of pain from trauma (Ulrich et al. 1965; Ulrich 1966). Pain in humans similarly features hostile responses.

3 Personality Features

Traditionally, it was supposed that hysterical personality or histrionic personality, as it is now termed, would be liable to be associated with such pain as was not based upon organic disease. Work with the Minnesota Multiphasic Personality Inventory (MMPI) favoured the idea that there was a characteristic pattern of abnormality in patients with pain. Typically, the hypochondriasis scale is elevated above the mean, the depression scale is somewhat elevated and the hysteria scale is elevated more than depression but not as much as hypochondriasis. If the patient does not have organic illness, this may be accepted as evidence that the symptoms are hysterical in type. However, many patients with physical illness are liable to have a similar pattern of response on the MMPI. If a patient with physical illness gives positive responses to questions which appear on the hypochondriasis scale and which confirm that he or she has back pain, pain in other joints, headache or a liability to sleep badly, the result will often be interpreted as meaning that he or she is psychologically ill. In addition, half the items on the hysteria scale come from the hypochondriasis scale. Thus, if the hypochondriasis scale is elevated, the hysteria scale will be pulled upwards after it. Moreover, items such as sleeplessness, fatigue and difficulty in concentration, which can all be due to chronic physical causes of pain, are liable to be scored as positive, not only on the hypochondriasis scale, but also on the depression scale. Hence the pattern that has been described as a "conversion V" triad in the MMPI seems usually to be due to physical causes of pain (or other disruptive chronic physical illness). Causalgia (chronic regional pain syndrome type II) or other de-afferentation pain, is as likely a cause of this syndrome as the mind, if not more likely.

The problems of the MMPI limit its use in patients with chronic pain. The same applies to any other test, e.g. the Symptom Checklist-90, which relies upon a category called somatization and is based upon a count of physical symptoms. In patients with chronic pain, psychiatric measures must be free from questions with a somatic weighting. The Hospital Anxiety and Depression Scale (Zigmond and Snaith 1983) was designed specifically with this problem in mind and has particular merit in this field.

4 Selection

Selection influences which patients present with pain or in general medical care. Even very acute potentially serious conditions do not automatically produce an admission to hospital of a normal sample of the population. This is true even in general practice. Patients with migraine may say to themselves: "I have this dreadful headache again. I must see my family doctor about it. Mother always went to see him when she had headaches and vomiting and I should do the same." Or they may say: "Mother always went to see the doctor with her headaches and got no benefit. I will take the day off, take a pain killer and go to work tomorrow." Not surprisingly, patients in specialized migraine clinics tend to be more persistent in seeking relief than those with migraine in the community, and patients in other clinics are often anxious and hypochondriacal. Early studies held that migraine occurred in subjects who were hesitant over positive decisions, insecure, perfectionists and unduly worrying. They were of high intelligence; the women had not had a

successful heterosexual life, and the men had a low sexual drive. The migraine attack was seen as a conflict between the desire to escape from the mother's influence and the compulsion not to leave her. These sorts of descriptions were adopted by the late H.G. Wolff (1937, 1948), who particularly emphasized the presence of a rigid, ambitious, perfectionist personality with headache resulting from a variety of stresses.

Migraine is very common in the population as a whole. A survey by Crisp et al. (1977) in an actual general practice population showed that migraine patients were scarcely any different from a general practice population. There was a statistically significant increase in anxiety in migraine patients, but its quantitative significance was slight. Thus, while anxiety may provoke migraine, and migraine seems to be more common in patients who are going through a phase of anxiety and depression, it does not have an automatic relationship with emotional states. For all types of chronic pain, selection is a natural phenomenon, and the more specialized the clinic, the stronger the selection pattern.

Studies of facial pain lead to similar conclusions. Temporomandibular pain and dysfunction syndrome (TMPDS) is marked by pain in the face, particularly in the maxillary region, clicking of the joints in some patients, tenderness of the muscles of mastication and over the joint and limitation of mouth opening. Thirty-nine per cent of a normal population had clicking, 12% had pain in the face on wide opening of the jaw and 7% had limitation of movement (Agerberg and Carlsson 1972). In addition, facial pain affects about 5% of 25-year-old normal males and 11% of normal female (Heloe and Heloe 1979). In this latter study, 2% of the men and 5% of the women sought treatment. This resulted in a ratio of men to women of 1:6 in the clinic compared with a ratio of 1:2 in the community. In clinical practice, women typically present far more often with a complaint of TMPDS than do men. TMPDS used to be held to result from psychophysiological dysfunction. We found that only 48% of a population with TMPDS could be classified as likely to have psychiatric illness even when we used very liberal cut-off points on the General Health Questionnaire-28, a screening measure for psychological illness (Salter et al. 1983). Speculand et al. (1983) likewise observed that the abnormal illness behaviour patterns of patients with this condition were more like those of general practice patients than of a chronic pain population.

Strong evidence of selection effects has been produced by Crook et al. (1989) with respect to patients with chronic pain. Patients in a pain clinic were more likely to have been injured than those in a community sample and also reported a greater intensity and constancy of pain and had more difficulties with the activities of daily living. They were more depressed and withdrawn socially and showed more long-term consequences due to unemployment, litigation and alcohol and drug abuse. Whereas only 2% of patients in the family practice group were disabled or unemployed, 38% in a pain clinic were in that situation.

5 Clinical Contributions

Many specific painful syndromes have long been recognized. Bonica (1953, 1990) established that both acute and chronic pain were damaging pathophysiologically in and of themselves and provided a method for diagnosis and treatment of the different chronic pain syndromes. His major work, *The Management of Pain*, put together a wealth of knowledge about the neurology of chronic painful syndromes and how to treat them by regional blocks, many of which he himself had pioneered. He established the importance of multi-disciplinary consultation, including the significance of psychiatric and psychological advice, which would be incorporated in a comprehensive pain clinic. Bonica always acknowledged that the development of the comprehensive pain clinic was not only his idea but was also promulgated by the late F.A.D. Alexander (1978). By 1990, Bonica was able to identify 336 pain control facilities worldwide, two thirds of which were in the United States. Forty-nine per cent of clinics were run by anaesthetists, 13% by rehabilitation specialists and 12% by neurosurgeons. Psychiatry and psychology specialists together ran 7%. In general, the primary responsibility for organizing comprehensive pain clinics is in the hands of other disciplines, but psychiatry and psychology play a vital role in many cases. Bonica also described a system of classification for chronic pain syndromes which was adopted by the Task Force on Taxonomy of the IASP and forms the basis for a classification of chronic pain syndromes (Merskey and Bogduk 1994).

It is useful to know the rates of depression in patients with pain and the rates of pain in patients with depression, even if selection is taking place. The incidence of depression in clinics ranges from 10% to 100% (Romano and Turner 1985) in patients with chronic pain. By depression, one means a state which is of the order of severity of a major depressive disorder rather than only a symptom. In surveys of patients in different clinics, we found that the lowest probable incidence of the likely presence of psychiatric disorder appeared in an oral medicine clinic where patients were referred by dentists for consultations on facial pain. The highest rate was in a psychiatric clinic (Merskey et al. 1987). Patients with headache were between these two extremes (Merskey et al. 1985), as

were patients with pain in various regions attending anaesthetist nerve block clinics. In these latter studies, the measure used was the General Health Questionnaire with 28 items (GHQ-28; Goldberg 1978). The values presented were obtained with a cut-off score of 4/5, which tended to maximize the presence of supposed psychiatric conditions.

When a fairly stringent instrument for the diagnosis of depression such as the Levine-Pilowsky Depression Questionnaire was used, the findings tended to be more conservative and produced the low figure of 10% mentioned above. The latter questionnaire was designed to diagnose well-established cases of depression and to show also whether they appeared to have a reactive or an endogenous pattern.

6 Survey Data

Community figures for chronic pain have generally been based upon pain which may only have lasted as long as 1 month. Most survey figures for chronic pain (e.g. Crook et al. 1984) find that about 11% of the population overall has pain present for most of the time over a period of 2 weeks, and another 5% has had some incidental pain in a previous period of 2 weeks. Magni et al. (1990) examined data collected by the United States National Center for Health Statistics. In this data, 14.4% of a representative sample of the United States population between the ages of 25 and 74 suffered from chronic pain related to the joints and musculoskeletal system. The definition of chronic pain in this case was pain present for most of the time during a period of at least 1 month in the preceding 12 months. In addition, 7.4% had some pain of uncertain duration. On the basis of scores on the Depression Scale of the Center for Epidemiologic Studies (CESD), it appeared that 18% of the population with chronic pain were found to have depression, whereas only 8% of the population who did not have chronic pain reached similar scores for depression. There were significantly more women with pain (ratio approximately 6:4) and older people as well as people with a lower income.

These data imply that nearly one in five patients with joint pain present for most of 1 month had clinical depression. In the study, 83% of patients with chronic pain had received some formal medical treatment for their condition. Of those who received treatment, whether from family practitioners or others, the numbers who were depressed at any one time appear to be about two out of nine. Of those who had pain, 7.9% lost between 1 and 30 days of work, while 8.5% lost more than 30 days of work. This figure of 8.5% is approximately 1.2% of the whole population studied. In this group, one would expect to find more depression and of course more frequent attendance for medical care.

In follow-up studies, patients in the community with "chronic pain", as defined above, had a spontaneous remission rate of 32.5% after an interval of 8 years. The frequency of subjects with chronic pain was slightly more than twice that found in the previous survey. In this follow-up using logistic regression analysis, the odds ratio for patients with depression to develop pain 8 years later was 2.14, while the odds ratio for patients with pain to develop depression 8 years later was 2.85. Thus pain preceded depression and depression preceded pain in different cases. These results were highly significant statistically ($p < 0.001$), although the effect was relatively small. The high significance depended upon the large numbers of subjects in the study (Magni et al. 1994). A necessary connection has not been established between the predictive and outcome variables, but the evidence supports both the view that depression promotes pain and that pain promotes depression, and the latter effect is a little more powerful than the former. In Magni's study, the low quantitative effect may be related to the long interval of follow-up. At first we thought that a long interval would be an excellent opportunity to test the interaction between these two variables in a prospective fashion. On further consideration, it seems that a short follow-up interval – say, 18 months to 3 years – might be more effective in telling us about the closeness of the relationship between the two variables in either direction.

In most pain clinics, the frequency of depression or psychological problems appears to be about 30% (Pilowsky et al. 1977; Large 1980; France et al. 1985; Merskey et al. 1987).

7 Behavioural Concepts

Another approach towards pain treats it as a behavioural phenomenon (Fordyce et al. 1968; Fordyce 1976). According to this approach, pain behaviour is augmented by favourable contingencies, i.e. causes, inducements, stimuli or events which favour its presentation and diminished by events which reduce the benefits to the organism of complaining of pain or acting as if in pain. It has long been recognized that exercise helps to reduce pain, and behavioural theory invokes this link. Stiffness produced by increased activity may be reduced by more activity after a day or two if deconditioning is overcome. However, some patients get worse because their pain is increased each time they are active and does not diminish on

subsequent occasions. In this instance, we might think that there is an obdurate cause for increased pain with activity and that this cause is not responsive to improved muscle conditioning. It can be difficult to know who will benefit from exercise and who will be made worse. Patients with acute inflammation and acute injuries may still do better to rest. In chronic pain, exercise programmes appear to have great success with relatively early cases, but give rise to a good deal of discontent in patients who are pushed to exercise and in whom the pain is exacerbated.

Current books or articles on pain frequently indicate a behavioural element in programmes for the treatment of pain (e.g. Gatchel and Turk 1996). The critics of operant behavioural treatment particularly question the view that behavioural methods in pain treatment programmes claiming to treat excess disability and expressions of suffering are also treating pain. They doubt this on logical grounds and also suspect that the clinician will not always be as skillful as is necessary in defining "excess disability and expressions of suffering". Anyone who gets this wrong is going to be pushing patients repeatedly to do things which are increasingly difficult and painful. This follows from the fundamental position which Fordyce (1976) has stated, namely, that the subjective state of the patient is not a matter of concern to him or her provided behaviour can change. Schmidt (1988) demonstrated that the operant approach confused pain and ratings of pain by the sufferer. No matter how much denial there is on this topic, it seems that the notion of treating pain behaviour involves some rejection of the patient's experience. A more moderate approach recognizes that behavioural management of patients can help patients to overcome difficulties in activity which they thought were insuperable and can discourage maladaptive behaviour.

Another grave problem with the behavioural approach is that it has become extremely popular in circumstances where the practitioner appears to have a profound conflict of interest. In the United States especially, the treatment of a considerable proportion of patients in pain clinics is paid for by insurance companies. The interest of an insurance company is to get the insured person back to health and strength, but also to establish that he or she will receive treatment directed to his employability. These aims are in the interest of the insured as well, but patients are more interested than insurance companies in relieving pain. In order to get paid, clinics may have to provide programmes which offer to remove *disability* rather than pain. One aim may be sacrificed for the sake of another. This is rarely discussed in the United States in the technical literature, but behaviour therapists claim that they are "rescuing" patients from disability which would otherwise be interminable. Such an approach is particularly noteworthy in a recent report by a Task Force of the IASP (Fordyce 1995), which takes the view that back pain is so much subject to psychosocial influences that, after the exclusion of surgically correctable lesions and investigation over a period of 6 weeks, individuals with chronic "non-specific back pain" should not receive further medical benefits. Both medical care and disability benefits should be denied and the individuals declared "activity intolerant". Such individuals may be candidates for rehabilitation and further benefits on those grounds, but not otherwise medically, and would need to be compensated through unemployment insurance programmes. This approach has given rise to considerable opposition.

Cognitive measures do not share these objections of principle. Cognitive therapy for depression is now well established, at least for mild and moderate cases, and does not necessarily compete with pharmacotherapy. Flor and Turk (1988) showed that cognitive variables in rheumatoid arthritis accounted for between 32% and 60% of the variance in pain and disability and did so more effectively than physical measures. The patient's awareness of disability and distress in other situations also tends to predict outcome better than the physician's estimate of physical status in a number of instances. This does not prove that the disability is of psychological origin. Patients' awareness effectively takes into account both their physical state and their psychological condition, the latter being based at least in part upon the former. Hence the subjective measure is likely to be the more comprehensive one.

Anderson et al. (1988) have shown that psychological variables do not necessarily independently predict pain behaviour in rheumatoid arthritis. Pain behaviour such as guarding, bracing, grimacing, sighing, rigidity, passive rubbing and active rubbing most closely related to physical illness. This is in accordance with expectation, since the model of pain behaviour was developed out of the model of physical disease. Pain behaviour is thought to be "psychological" or anomalous only when it is not matched by physical disease. In many cases of pain for which X-ray measurements and even computed tomographic scan (CT) and magnetic resonance imaging (MRI) do not contribute much extra information, it is often thought that pain behaviour which is not supported by these measures, or by neurological or orthopaedic findings, must be psychological in origin. This is an important fallacy, and *psychiatric illness in patients with pain should only be diagnosed on its own grounds.* Pain cannot be attributed to psychological illness if an adequate psychiatric diagnosis cannot be made, and a psychiatric diagnosis should not be made simply because the patient complains a lot and organic evidence is insufficient.

Provided one takes a broad view, the identification of pain behaviour and treatment by exercise may be

part of a useful comprehensive treatment of chronic pain. It is generally agreed that it is important to have contact with family members and to work through them to discourage undue dependence at home. With chronic low back pain, in particular, it can help to discuss the patient's abilities and difficulties with family members and to see that they are not unduly solicitous but give patients every opportunity to do things for themselves to the extent possible. However, in dealing with patients' relatives, caution still applies with respect to behavioural interpretations.

8 Psychological and Physiological Mechanisms of Pain

When pain occurs as a result of a psychological process, several mechanisms may apply. Hallucinations can produce pain but do so very rarely. It is hardly ever a clinical problem with schizophrenia. A chronic mono-symptomatic complaint of pain without any other evidence of psychiatric illness or physical cause may be the precursor of schizophrenia, but the diagnosis can only be made when the illness becomes apparent on other grounds. Depression can cause pain as a delusional symptom, but as a rule that only happens in severe endogenous depression and usually it responds very quickly to suitable treatment. Interestingly, somatic hallucinations in schizophrenia are much more often related to passivity experiences and are not painful, although various forms of change may be described in the body (Watson et al. 1981).

Pain associated with depression and anxiety could be due to muscle contraction. This is not nearly as common as has been supposed, or at least it is hard to prove that it is common. Excess muscle contraction in the presence of ischaemia will lead to pain (Lewis et al. 1931). Repeated contraction of the muscles of a limb while blood inflow is restricted by a tourniquet quickly leads to pain, which is presumed to be due to a failure to carry away the waste products of metabolism. Chronic headache is likewise presumed to be due to over-contraction of muscles and a similar possible effect. Some evidence exists in favour of this point of view. The difficulty is that attempts to measure muscle contraction and to quantify it have not generally been in proportion either to the anxiety from which the patient suffers or the extent of the pain. The International Headache Society only describes "tension-type headache", recognizing the poor relationship between demonstrated muscle contraction and chronic headache. It seems that at the present time there is no firm explanation of headache of psychological origin, except that tension-type headache and episodes of migraine increase with anxiety or with depression.

The view that some pain is hysterical is perhaps best reinforced by the example of the Couvade syndrome, in which the husbands of pregnant women develop abdominal pains and even abdominal "contractions". There is no reason to expect physical disorder in such cases and no plausible pathophysiological explanation, even if air swallowing may cause some gastrointestinal distention. While the diagnosis of hysterical pain should be accepted on the basis of this and cognate evidence cited above, its frequency is also rare – or at least it is extremely hard to demonstrate – as we shall discuss subsequently.

One more mechanism of the production of pain deserves consideration. Anxiety may, in some fashion, alter activity in the spinal cord and in the mid-brain in a way which enhances input from noxious stimulation or inhibits suppressor mechanisms. Thus an effect might be transmitted from the limbic system through descending pathways to the spinal cord, which would increase a current abnormal activity. Ascending and descending activity in cerebro-spinal pathways has been repeatedly demonstrated experimentally and provides a potential mechanism, linked in principle to the pattern theory approach of Melzack and Wall (1965). This evolution of the "gate theory" of pain allows for the increase and decrease of pain from a physical illness in accordance with the presence or absence of emotional arousal and also takes account of various types of input and mechanisms of control within the nervous system. While it is not feasible to apply these theories to individual cases, they give a credible reason for believing in the occurrence of some chronic pain as a consequence of pathophysiological dysfunction within the nervous system.

The origins of this pathophysiological dysfunction may well be more organic than psychological. Important changes take place in the spinal cord in response to noxious stimulation and have ramifications at several levels of functioning of the central nervous system. This knowledge arises from developments in appreciation of the plasticity of nervous activity within the spinal cord and brain stem. The most strong focus has to do with regional pain syndromes.

It has been common to date to suppose that regional effects were signs of an emotional state in the individual. Loss of use or spread of pain in one region such as an upper limb or a lower limb fits ideally with a thought in the mind of a patient. Hence regional symptoms have often been categorized as hysterical or likely to be hysterical and as produced by thought processes, but plasticity offers an alternative explanation.

Wall (1984) described abundant evidence that the receptive fields of afferent neurons in the dorsal horn

of the spinal cord can change and extend. In the rat, 3 or 4 days after de-afferentation, cells that formerly responded to stimulation within the usual anatomical area begin to respond to stimuli from other areas. We can compare this in humans to cells in the area of the spinal cord usually related to the little finger only responding to local stimulation at first, but beginning after some time to respond to stimulation in the thumb or the radial side of the arm. Such effects can result from causing a punctate burn in one part of a limb (McMahon and Wall 1984). The presence of the burn in one area permits excitation of cells throughout the whole region. Cook et al. (1987) showed that electrical conditioning stimuli at 1 Hz for 20 s to C-fibre afferents from the gastrocnemius muscle of the rat would more than triple the receptive field of a cutaneous afferent neuron, whether it responded originally to a firm mechanical stimulus or to pinch. Moreover, neurons that originally responded only to pinch would begin to respond to touch about 5 min after the stimulation. Thus mechanisms exist in the spinal cord that allow regional pain to develop from a localized disturbance, including subcutaneous changes.

Woolf and Wall (1986) made the following observation:

> Because brief afferent inputs from deep tissue have even more pronounced effects than cutaneous inputs, this may explain the more widespread sensory disturbances that accompany deep injury. The varied pattern of post injury pain hypersensitivity resulting from injury to different tissues may be the consequence therefore of the activation of different afferents with different central actions. The presence of widespread tenderness (allodynia) with this sort of movement is frequently the most disturbing symptom in patients in chronic pain.

This work demonstrates that mechanisms exist in the spinal cord that allow regional pain to develop from a localized disturbance, including subcutaneous changes. In these circumstances, non-anatomical pain cannot automatically reflect the presence of hysteria. Hyposensitivity and patchy sensory loss in the presence of pain are also unreliable. Traditional explanations of cases of pain of this type as hysterical are unsound.

Other features which have also traditionally been taken to reflect the presence of hysteria or psychological motivation in chronic pain are likewise unreliable. Give-way weakness may be due to the patient not wishing to use an arm which hurts. Signs that have been thought to be typically hysterical – namely, a history of hypochondriasis, potential secondary gain, belle indifférence, non-anatomical patchy sensory loss, change in boundaries or hypoalgesia, sensory loss that splits at the mid-line on pin-prick or vibratory stimulation, and give-way weakness – have all been examined by Gould et al. (1986). In 30 patients with acute central nervous system damage, these authors showed that 29 had at least one feature of a supposedly non-physiologic nature, and the mean number of supposedly hysterical items per patient was 3.4. These authors conclude that hysteria is easily misdiagnosed if the above signs or items of history are accepted as pathognomonic and that the tests which are said to provide absolute evidence of hysteria lack validity.

Other reasons for not diagnosing chronic pain as psychological in origin include the finding that many patients whose pain was supposed to be due to accidents or injuries and dependent upon their search for compensation continue to have pain long after the legal issues have ceased to figure in their circumstances (Mendelson 1982). The psychological implications of accidents and injury, however, deserve some consideration.

9 Compensation and Motive

As just mentioned, many injured patients lack physical signs, but appear to have so-called "soft-tissue pain" and continue to suffer from pain long after their claims for compensation have been judged and either accepted or rejected. Nevertheless, controversy still exists in many countries about the part played by accidents in promoting chronic pain. Two particular sites of injury, the neck and the back, are most involved in this argument. Cervical sprain injuries are a common consequence of rear-end collisions by motor vehicles. Typical cervical sprain injuries ("whiplash") occur at speeds varying from 8 to 30 kmh. A passenger wearing a seat belt will be held to the seat while the vehicle is propelled forward. When this takes place, the head moves backwards and may be restrained somewhat by a head-rest, but nevertheless can undergo an excessive degree of excursion. The organs that take the strain of hyperextension of the head are the muscles of the neck, which may be torn and bleed. This has been demonstrated experimentally in monkeys by MacNab (1973). Information has accumulated which suggests that a variety of physical lesions in the neck may account for these phenomena. While not all individuals suffer from this disorder after a rear-end motor vehicle accident, perhaps as many as one in seven develop some pain.

Follow-up studies and retrospective examinations of material lead to the conclusion that two factors determine the likelihood of chronic pain after rear-end collisions and somewhat similar sideways collisions with respect to the neck. The first is the prior

history of the individual, i.e. whether he or she has had previous headaches or neck pain or injury. The second is the severity of the pain at the time of impact. Psychosocial circumstances and personality factors have no bearing on the outcome at 6 months (Radanov et al. 1991), but the amount of initial pain influences the occurrence of continuing pain at 6 months after the injury as well as difficulties in concentration and the occurrence of secondary effects such as depression. Of course, anyone may develop a painful syndrome after a relatively minor injury on the basis of hypochondriasis, but the majority of those who suffer chronic pain after cervical sprain injury are likely to have a valid physical explanation for their pain.

This view has been reinforced by the work of Bogduk and his colleagues (Lord et al. 1993), who have demonstrated that anaesthetization of the small medial ramus of the dorsal branch of the cervical afferent nerves as well as the descending branch of the first cervical nerve root will anaesthetize very specifically the zygapophysial joint which it supplies (actually, upper and lower rami have to be anaesthetized for each joint). The zygapophysial (facet) joints are highly likely to be injured in a rear-end collision in which they are strained as the head is thrust backwards on the neck. Bogduk has demonstrated in double-blind controlled trials that patients with chronic pain are highly likely to be able to distinguish long-acting injections from short-acting ones, thus demonstrating that they are not having a placebo response. (The investigators had doubts about the ethics of saline injection alone as a control). Not less than 56% and up to 75% of cervical sprain victims were thus found in their series to have pain from those joints and to be rid of their pain temporarily by this investigation.

There is pathological evidence that pain from cervical sprain injuries may be due to lesions in which the posterior ligament in particular is torn and separated from the disk, or there may be partial lesions of disks with clefts where the ingrowth of small aberrant vessels and nerves will give rise to chronic pain in the neck. This brief sample of the available information may serve to indicate that there is a strong potential physical basis for an illness which has frequently been described as being due to economic or psychological motives.

One study has emerged to challenge these findings scientifically. Schrader et al. (1996) followed up 202 individuals who had suffered rear-end collisions in Lithuania and matched them with a very similar number of controls who had not had such collisions. At the end of between 1 and 3 years of follow-up, they found no difference between the two groups and concluded that the cervical sprain injury "loses its validity". Unfortunately, 80% of the subjects were men, which is unusual in the West, where 60% of individuals with cervical sprain injuries are women. Secondly, the study was retrospective and relied upon asking individuals at least 1 year, and sometimes 3 years later, whether they had had pain 1–3 years previously. Thirdly, and most importantly, it appears that only 31 individuals out of 202 (i.e. 15%) acknowledged any pain at the time of the first injury. About one in seven of individuals who are involved in rear-end collisions actually experience pain from the accident (Bjørgen 1996), while many individuals who are involved in such collisions do not and thus never come to medical notice. Follow-up studies in Australia, Britain, Canada and the United States tend to put the frequency of persistent symptoms at between 10% and 45% in patients seen at hospital emergency departments following motor vehicle accidents. Only a selected sample of individuals in developed countries experience pain in the neck sufficiently badly to attend hospital. The main weakness of the Norwegian/Lithuanian study is that it lacked the power to make the necessary discrimination and it handled its statistics incorrectly (Merskey 1997).

Similar problems of interpretation beset low back pain, which may also follow motor vehicle accidents but is more common after industrial injuries. An extensive study (Bigos et al. 1991) of Boeing factory workers demonstrated that, in a follow-up of a sample of industrial workers, not liking one's job and having a previous accident both contributed towards time off work. The findings were based upon acute absence. It is commonplace that if one does not like one's boss or one's employment, a minor illness will be treated with more concern than if one is keen to get back to work. This has little bearing on the frequency of pain from physical injuries which may cause chronic disability. While there is evidence according to the IASP Task Force (Fordyce 1995) that cases of disability from back pain are rising and this may present a financial burden to a number of countries, the direct evidence that the pain is due to psychosocial factors is weak (Teasell and Merskey 1997). None of the arguments that have been brought forward about the increase in back pain have faced the possibility that this may be due to significant organic processes of the type discussed here. Likewise, they have not taken into account the weakness of the diagnosis of psychological causes in most current clinical cases with chronic pain.

10 Patients with Pain

Who are the patients who complain of pain? A few major groups of patients with chronic pain can be delineated from the experience of pain clinics. The

easiest to identify are those with clear nerve lesions, such as causalgia, a nerve entrapment syndrome, post-herpetic neuralgia, trigeminal neuralgia, etc. These forms of chronic pain attract attention and receive treatment readily from most practitioners. The largest group of patients with chronic pain are those with non-malignant causes of musculoskeletal pain, usually cervical sprain injury or a low back sprain. Patients with chronic pain from disk lesions are rare, except for the few who post-operatively develop complications such as arachnoiditis or the rather larger numbers who have additional musculoskeletal discomforts associated with their disk lesions. These have often been categorized as having "mechanical back pain" or myofascial pain, the latter being identified by the occurrence of taut muscle bands (which are not always reliably found), a tender location over the band or in the muscle and radiation of pain from the site of pressure.

Some patients develop generalized pain affecting both the upper and lower parts of the body and the right and left sides as well as the spinal region. Many such patients, who have no other particular cause for widespread pain, appear to develop it as a result of the existence of chronic pain in one or more areas previously, such pain often following from trauma. In such instances, it is quite common to find that the patient has developed fibromyalgia (FM), defined as diffuse pain with the above characteristics, accompanied by at least 11 out of 18 specific tender points. The significance of this illness has been much disputed, but is becoming clearer. First, diffuse pain and specific tender points may be associated with several types of physical disorder that precipitate the condition. These include rheumatoid arthritis, auto-immune disease and prior chronic pain. When it is present, a sleep disorder consisting of the appearance of alpha-type rhythms in deep sleep (alpha/delta pattern) can be found in a significant proportion of patients (Moldofsky et al. 1975). This sleep pattern is not specific, but is found in many patients with FM and is associated with waking and not feeling refreshed. In addition to the specific tender points, there is usually a widespread lowering of the threshold for pain on noxious stimulation such as pressure. Thus the illness appears to be one of diffuse pain accompanied by disruptions of sleep and the presence of widespread aching in the musculature. In addition, it appears to be associated more often than by chance with auto-immune disease, as already mentioned, migraine, Raynaud's syndrome and the irritable bowel syndrome.

In FM, increases in substance P in the cerebro-spinal fluid (CSF) to a level of two or three times the normal level have been demonstrated in patients and controls in three independent studies (Vaerøy et al. 1988; Russell et al. 1994; Bradley et al. 1996). FM is also associated with a somewhat increased rate of psychological disturbance, but not sufficient to establish that psychological factors are always a cause. Thus, in some studies, the amount of psychological illness has been found to be comparable to that in control patients with rheumatoid arthritis, and in others it has been rather more, but not affecting 100% of patients. It would not be surprising if emotional stress were to exacerbate a physical condition of this type, presumably by way of an effect upon spinal cord dysfunction.

Other musculoskeletal conditions such as rheumatoid arthritis have to be taken into account as causes of chronic pain. Patients with cancer pain comprise a large separate group who are commonly treated by a number of methods, both psychological and physical.

There are four types or groups of patients of psychiatric interest:

1. Minor lesion: emotional exacerbation (very common)
2. Hysterical emulation of pain complaints or factitious simulation (rare)
3. Primary psychiatric illness, e.g. depressive disorder, anxiety state (common in psychiatric practice, less common in pain clinics)
4. Major physical illness: secondary depression or other psychological response (very common)

In the first group, there is a definite physical lesion, e.g. mild post-herpetic neuralgia, which becomes worse if the patient is under stress or in emotional difficulties, depressed or perhaps gaining benefit from the symptom. Next, it is always possible, despite the case made above for organic explanations, that individuals who do not have physical illness may mimic, either deliberately or unintentionally, the experience and complaints of pain of others who do have a physical basis for them. This group does not appear to be large. In the third group of patients who have chronic pain, some will have other independent psychological illness, such as chronic depression, or sub-acute depression, anxiety states, hypochondriasis or "somatization disorder", etc.

The worst problems probably comprise the fourth group, i.e. patients with physical illness that is sufficient to cause disability who develop psychological complications. These patients with severe pain develop depression at about 6 months after the onset of their pain, when several measures have been tried and relief is inadequate. They see themselves losing their income, perhaps losing their houses, unable to support their families, unable to maintain normal sexual and other relationships with their spouse and suffering from broken sleep, difficulty in concentration and a general handicap in running their life. This handicap is poorly recognized by others and often rejected by insurance

companies, workers' compensation boards, etc. The recent tendency is clearly negative in North America towards these patients. In the province of Quebec, in Canada, for example, individuals who have cervical sprain injuries usually cease to receive benefits after they have been paid for 27 days.

The most common association of pain with psychiatric conditions is between a physical illness which has disabled and distressed the patient and the onset of depression. Patients with independent psychiatric illness such as depression from bereavement who receive appropriate psychiatric treatment commonly recover. This is not so for the patients with chronic pain causing psychiatric illness and, indeed, is another reason for supposing that their illnesses are not independent psychiatric conditions and may have a more intractable peripheral somatic basis.

11 Management and Treatment

11.1 Evaluation

Psychiatric management in chronic pain begins with the review of the physical state. Whether a psychiatrist works alone or within the team of a comprehensive pain clinic, he or she should be ready to contribute actively to the evaluation of all types of case. This includes diagnosis, although his or her colleagues in a pain clinic are more likely to be looking particularly for advice on issues of management such as how to respond to the presence of anxiety or depression, what techniques of psychological care should be used and whether supportive interviews, cognitive formulations, assistance with job issues or family matters are required, and of course they may seek information on the use of psychotropic drugs.

It is important to note the pain of which the patient complains and to discuss it as well as the patient's ideas and beliefs about its physical origin. The psychiatrist must also determine whether the physical evaluation is adequate. The fact that there is "nothing neurological" or "no bones broken, no nerve injury, no reflex impairment or nothing more than a so-called soft tissue injury" does not prove psychological causation. "Soft tissue disorder" might include physical damage to small joints, latent damage to larger joints or disks which has not been seen and perhaps also some fibrosis in muscles and around nerves subsequent to the rupture of muscles and bleeding into them. If the psychiatrist accepts the patient as "psychiatric", he or she will have some obligation to produce a cure – which may be much more available for some potential psychiatric causes than for the organic ones at this stage. Psychiatrists must therefore be very cautious in accepting these types of cases for psychiatric care, and if they recognize a physical state on which they can make some useful comments, they should be prepared to do so, without claiming authority in the diagnostic field of another discipline. Further, one should be able to provide a psychiatric perspective which will assist in encouraging treatment of complicating psychological factors in an "organic" case or in minimizing their relevance where psychological factors are perhaps being over-emphasized.

Among physical conditions which I have seen, and which were being misdiagnosed as "neurosis", there are many cases of cervical sprain syndrome, a number of cases of the thoracic outlet syndrome, paroxysmal hemicrania, cluster headache, peripheral neuropathy and the syndrome of painful legs and moving toes. Whether or not there is a physical state which can be recognized, the psychiatrist must be careful and thorough in deciding whether the evidence for psychiatric change in terms of a recognizable psychiatric diagnosis matches the symptoms. There may be some evidence of separate disturbances in a person's life, stress at work, marital problems or failure to achieve adequate relationships at work or in other interpersonal relationships. That may not be enough to match or account for the symptoms for which the individual complains. Whatever the psychiatric diagnosis, it should be reached through standard procedures and treated without prejudice to the possibility of a physical illness being present as well.

Patients who have had chronic pain prior to the onset of depression are likely to have a primary physical problem, and patients who have depression first and then chronic pain may have some or much of their pain as a result of the depression. This is most likely in the case of headache and least likely in the case of unilateral or focal symptoms.

In investigating patients with chronic pain, it is important to take the complaint of pain seriously, to suspend judgement on aetiology until adequate grounds exist for giving the patient information and, if there are no grounds for a firm diagnosis, to say so.

Psychiatrists may also be involved in the management of patients with cancer pain and palliative care. In these settings, psychiatrists often function in three ways. First, there may be patients with formal psychological illnesses which, although recognized by family practitioners and others, seem to need extra help, particularly with the treatment of depression. This will most often be achieved by extending the amount of psychological support and assistance to the patient to adapt psychologically to a changing physical state. Help may also be needed sometimes from the psychiatrist in regard to particular applications of

psychotropic medications, such as phenothiazines for analgesia or a choice of a second or third alternative among anti-depressants. Thirdly, the help of a psychiatrist may be sought in some clinics (perhaps many) in providing support for the professional staff in their continuing exposure to patients with great distress and tragic conditions. This help from professional staff is likely to be in terms of psychological discussion as well as help with particular patients.

11.2 General Management

In a comprehensive pain clinic, a psychiatrist will join his or her colleagues in determining the application of the different measures of treatment that may be employed. Some of these will clearly be specific, and specific diagnoses will lead to appropriate specific treatments. Few of these are psychiatric in nature, except for some types of major depressive illness, but psychotropic drugs or other drugs acting on the central nervous system may often be used. Trigeminal neuralgia can be treated with carbamazepine or other anti-convulsants which have an effect upon it. Cluster headache responds to lithium carbonate among other prophylactic measures, but most perhaps to the withdrawal of nicotine. Chronic paroxysmal hemicrania, a rather rare variant, almost always found in women, can be controlled specifically by indomethacin. Anaesthetists may provide nerve blocks for defined syndromes in particular regions, and repeated nerve blocks, with or without steroids in the injection, can help to produce abatement of pain and enable the patient to progress in physiotherapy. Transcutaneous electrical nerve stimulation (TENS) may be useful and is essentially harmless.

Patients with pain, especially when it is chronic, need support and tolerance. They should be encouraged where possible to undertake specific exercises related to any musculoskeletal area that causes trouble and to improve their general physical state. Management by a physiotherapist for these purposes is often desirable in severe cases. A psychiatrist will be likely to offer additional guidance within pain clinics in the choice of techniques of psychological value.

11.3 Medication

Several tricyclic anti-depressants, including clomipramine, doxepin, imipramine, maprotiline and nortriptyline, have analgesic effects. However, amitriptyline is the tricyclic drug which has the most evidence of proven analgesic actions that are independent of the presence of depression. It is particularly useful in restoring sleep impaired by pain. Its main disadvantage is the tendency it has to promote weight gain. This may be counteracted by skilled psychopharmacotherapy, either with the use of fluoxetine (adjusting the dose because of the interaction of fluoxetine with tricyclics via the P450 cytochrome system) or by the administration of fluvoxamine, citalopram or sertraline to assist in night sedation. Each of these four specific serotonin re-uptake inhibitors, and also venlafaxine and bupropion can be helpful in promoting weight loss in some patients, but none of them has any particular effect on pain, so far as is known, unless the pain itself is due to a depressive illness. Amitriptyline, or a similar tricyclic, is therefore still necessary for many patients despite the side-effects and problems. Neuroleptics, e.g. methotrimeprazine, also appear to have direct analgesic effects in some patients.

Treatment should be comprehensive. The patient's life situation should be looked at as well as the prescription of pharmacotherapy. Some non-psychotropic analgesics may be prescribed by psychiatrists and others. Non-steroidal anti-inflammatory drugs such as naproxen or indomethacin or paracetamol (acetaminophen) should be avoided if possible. All these drugs, even including acetaminophen (paracetamol), are liable to give rise to chronic gastrointestinal problems, most often dyspepsia or ulceration, but not infrequently diarrhoea. The best available non-steroidal anti-inflammatory drugs in North America are the Cox-2 inhibitors, because they have less effect on the stomach than other drugs as they do not inhibit prostaglandins in the stomach wall, but inhibit prostaglandins elsewhere. Drugs which produce muscle relaxation, including baclofen (a γ-aminobutyric acid antagonist), cyclobenzaprine (closely similar in structure to amitriptyline), carisoprodol or hydroxyzine, are used in many countries as adjuvant treatments in patients with chronic pain. Baclofen can cause depression or confusion, while carisoprodol breaks down to meprobamate, and those using these drugs should be aware of these problems.

Most patients coming to the psychiatrist for the treatment of pain have already had a number of these measures, but it pays to review them and to see how carefully they have been employed and whether some small benefits may be obtained by giving the amitriptyline earlier in the evening so that the patient falls asleep when intended and so that the amitriptyline, with its long half-life, does not last too long when he or she is ready to wake up.

Narcotics, including sustained-release preparations of oxycodone or of morphine, are also recommended for chronic non-malignant pain. Acceptance of this is not universal. Patients receiving these drugs should

have had extensive alternative treatment and, ideally, no history of substance abuse.

11.4 Relaxation

Among general measures, relaxation is the foremost. Progressive relaxation was introduced originally by Jacobson (1929). Biofeedback has been widely used, but it seems to be no more effective than relaxation (Jessup et al. 1979; Turk et al. 1979; Nuechterlein and Holroyd 1980). Specific efforts with biofeedback occasionally produce reports of extra benefit, but none seems to have established wide reliability.

Likewise, hypnosis, which used to be a focus of attention and hope in the treatment of chronic pain, has long been found to be extremely sporadic in its effects and generally not worthwhile. Hypnosis also produces little more benefit, if any, than systematic procedures for relaxation. There are those who claim occasional dramatic responses, but results with hypnosis remain unpredictable and unreliable, and occasional cases of apparent great success never seem to be followed through or maintained with other workers. Some current opinion holds that hypnosis amounts only to cooperative role-playing on the part of both patient and physician (Merskey 1995). Used in a non-committal way, hypnosis may be encouraging and helpful to some patients, provided it is not employed for interpretative purposes, regression, memory recall or some of the other activities which have lately made it notorious (Merskey 1996).

11.5 Cognitive Treatments

Cognitive therapy is often linked currently with encouragement of coping mechanisms and behavioural treatment, as already indicated. More pure cognitive therapy without emphasis on deconditioning focuses upon the patient's ideas about himself or herself being insufficiently able to master pain. Controlled studies have given validation to the idea that cognitive treatment is an effective method of assisting patients with pain. It seems unlikely that it can help patients with severe pain as much as those with mild or moderate chronic pain. It is as appropriate to try it in patients who have a physical cause for pain as in those who do not. Misconceptions about pain deserve attention, and fears about pain may be resolved during cognitive therapy which becomes, in that case, virtually identical with the practice of supportive therapy. The aim of cognitive therapy includes encouragement to patients to believe that they are not helpless and that they have some resources and ability to function regardless of pain. This clearly overlaps with support for coping mechanisms and the standard type of sympathetic tolerance, encouragement, guidance and re-orientation which many psychiatrists have long employed. The methods used in these approaches also include emphasis on self-efficacy in the sense of control over aversive stimulation. It is to be expected that they will be helpful and also that they will only assist in adjustment within limits. Support for the patient should also be provided through interviews with other members of the family and consistent recognition that the patient is affected by a disabling and difficult problem.

11.6 Psychological Techniques Overall

Psychological treatment or psychotherapy will attend to the various problems of patients with depression and anxiety on conventional lines. It may be repetitive to say that this will include supportive psychotherapy of the cognitive and interpretive type with attention to social issues and marital counselling and other forms of psychological treatment, whether these approaches are needed because of a physical illness or because the pain appears to be giving rise to significant emotional change. Some influence may be exerted on behaviour by the selection of topics and goals – more often than by providing a formal cognitive or behavioural programme. In my experience, a behavioural technique such as deconditioning is often more appropriate to patients who have associated fear of motor vehicles (where they suffered the injuries which have given rise to their pain) and of driving than in the direct reduction of "pain behaviour", which is, in any case, a concept that is too readily misapplied.

The psychiatrist would obviously participate here in the assessment and provision of rehabilitation plans and in their encouragement. He or she may also look at other issues which occur in patients with chronic pain. For example, many individuals in mid-adult life – particularly those below the age of 55 – form a very gloomy view about their future because they anticipate that their chronic pain will only get worse as they get older. It often helps with patients with chronic musculoskeletal pain related to the spinal axis to explain that these types of pain may not decline dramatically, but that they do decline somewhat over decades. The patient may be more stiff and less mobile, a little more bent in stature, but will not have quite as much pain as currently. This advice is frequently received with a good deal of relief on the part of patients who have anticipated that because most

physical conditions seem to get worse with age, so their chronic pain will inevitably be worse likewise. Other aspects of adaptation to alterations in life will also be needed for patients with chronic pain, and psychiatrists as well as family practitioners and other specialists will be able to add their particular quota to the general support and psychological assistance which should always be offered with medical care of patients with chronic disorders.

12 References

Agerberg G, Carlsson GE (1972) Functional disorder of the masticatory system. I. Distribution of symptoms according to age and sex judged from investigation by questionnaire. Acta Odont Scand 30: 579–613

Alexander FAD (1978) The genesis of the pain clinic. In: Pain abstracts, vol 1. Second World Congress on Pain, 27 August–1 September 1978, IASP, Seattle, WA, p 250

American Psychiatric Association (1980) Diagnostic and statistical manual of mental disorders, 3rd edn. American Psychiatric Association, Washington, DC

American Psychiatric Association Task Force on DSM-IV (1994) Diagnostic and statistical manual of mental disorders, 4th edn. American Psychiatric Association, Washington, DC

Anderson KO, Keefe RR, Bradley LA et al (1988) Prediction of pain behavior and functional status of rheumatoid arthritis patients using medical states and psychological variables. Pain 33: 25–32

Bigos SJ, Battie MC, Spengler DM et al (1991) A prospective study of work perceptions and psychosocial factors affecting the report of back injury. Spine 16: 1–6

Bjørgen IA (1996) Late whiplash syndrome. Lancet 348: 124

Bonica JJ (1953) The management of pain. Lea and Febiger, Philadelphia

Bonica JJ (1990) The management of pain, vol 1, 2nd edn. Lea and Febiger, Philadelphia, pp 197–208

Bradley LA, Mountz JW, Blalock JE et al (1996) Regional cerebral blood flow (RcBF) in the caudate nucleus and thalamus of fibromyalgia (FM) patients is associated with cerebrospinal fluid (CSF) levels of substance P (SP). Abstracts of the 8th World Congress on Pain. International Association for the Study of Pain. IASP, Seattle, WA

Cook AJ, Woolf CJ, Wall PD et al (1987) Dynamic receptive field plasticity in rat spinal cord dorsal horn following C-primary afferent input. Nature 325: 151–153

Crisp AH, Gaynor Jones M, Slater P (1977) The Middlesex Hospital Questionnaire. Br J Med Psychol 51: 269–280

Crook J, Rideout E, Browne G (1984) The prevalence of pain complaints in a general population. Pain 18: 299–314

Crook J, Weir R, Tunks E (1989) An epidemiological follow-up survey of persistent pain sufferers in a group family practice and specialty pain clinic. Pain 36: 49–61

Devine R, Merskey H (1965) The description of pain in psychiatric and general medical patients. J Psychosom Res 9: 311–316

Engel GL (1951) Primary atypical facial neuralgia. An hysterical conversion symptom. Psychosom Med 13: 375–396

Engel GL (1959) "Psychogenic" pain and the pain prone patient. Am J Med 26: 899–918

Flor H, Turk DC (1988) Chronic back pain and rheumatoid arthritis: predicting pain and disability from cognitive variables. J Behav Med 11: 251–265

Fordyce WE (1976) Behavioral methods in chronic pain and illness. Mosby, St Louis, p 236

Fordyce WE (1995) Pain in the workplace. International Association for the Study of Pain Press, Seattle

Fordyce WE, Fowler RS, Lehmann JE, Delateur BJ (1968) Some implications of learning in problems of chronic pain. J Chron Dis 21: 179–190

France RD, Krishnan KRR (1985) The dexamethasone suppression test as a biologic marker of depression in chronic pain. Pain 21: 49–55

Gatchel RJ, Turk DC (1996) Psychological approaches to pain management: a practitioner's handbook. Guilford, New York

Goldberg DP (1978) Manual of the General Health Questionnaire. NFER, Windsor

Gould R, Miller BL, Goldberg MA et al (1986) The validity of hysterical signs and symptoms. J Nerv Ment Dis 174: 593–598

Heloe B, Heloe LA (1979) Frequency and distribution of myofascial pain dysfunction syndrome in a population of 25-year olds. Commun Dent Oral Epidemiol 7: 357–360

Holmes DS (1990) The evidence for repression: an examination of 60 years of research. In: Singer JL (ed) Repression and dissociation – implications for personality theory, psychopathology and health. Univ Chicago Press, Chicago, pp 85–102

International Association for the Study of Pain (1979) Pain terms: a list with definitions and notes on usage. Pain 6: 249–252

Jacobson E (1929) Progressive relaxation. Univ Chicago Press, Chicago

Jessup BA, Neufeld RWJ, Merskey H (1979) Biofeedback therapy for headache and other pain: an evaluative review. Pain 7: 255–270

Large RG (1980) The psychiatrist and the chronic pain patient: 172 anecdotes. Pain 9: 253–263

Lewis T, Pickering GW, Rothschild P (1931) Observations upon muscular pain in intermittent claudication. Heart 15: 359–383

Lord S, Barnsley L, Bogduk N (1993) Cervical zygapophysial joint pain in whiplash. In: Teasell RW, Shapiro AP (eds) Cervical flexion/exterior injuries. Spine: state of the art reviews, vol 7. Hanley and Belfus, Philadelphia, pp 355–372

MacNab I (1973) The whiplash syndrome. Clin Neurosurgery 20: 232–241

Magni G, Caldieron C, Rigatti-Luchini S, Merskey H (1990) Chronic musculoskeletal pain and depressive symptoms in the general population. An analysis of the 1st National Health and Nutrition Examination Survey data. Pain 43: 299–307

Magni G, Moreschi C, Rigatti-Luchini S, Merskey H (1994) Prospective study of the relationship between depressive symptoms and chronic musculoskeletal pain. Pain 56: 289–297

McMahon SB, Wall PD (1984) Receptive fields of rat lamina 1 projection cells move to incorporate a nearby region of injury. Pain 19: 235–247

Melzack R, Wall PD (1965) Pain mechanisms: a new theory. Science 150: 971

Mendelson G (1982) Not "cured by a verdict". Effect of legal settlement on compensation claimants. Med J Aust 2: 219–230

Merskey H (1965a) The characteristics of persistent pain in psychiatric illness. J Psychosom Res 9: 291–298

Merskey H (1965b) Psychiatric patients with persistent pain. J Psychosom Res 9: 299–309

Merskey H (1995) The analysis of hysteria: understanding conversion and dissociation, 2nd edn. Gaskell, London

Merskey H (1996) Ethical issues in the search for repressed memories. Am J Psychother 50: 322–335

Merskey H (1997) Whiplash in Lithuania. Pain Res Management 2: 13–14

Merskey H, Bogduk N (1994) Classification of chronic pain: descriptions of chronic pain syndromes and definitions of pain terms, 2nd edn. International Association for the Study of Pain, Seattle

Merskey H, Spear FG (1967) Pain: psychological and psychiatric aspects. Baillière Tindall, London

Merskey H, Brown A, Brown J et al (1985) Psychological normality and abnormality in persistent headache patients. Pain 23: 35–47

Merskey H, Lau CL, Russell ES et al (1987) Screening for psychiatric morbidity. The pattern of psychological illness and premorbid characteristics in four chronic pain populations. Pain 30: 141–157

Moldofsky H, Scarisbrick P et al (1975) Musculoskeletal symptoms and non-REM sleep disturbance in patients with "fibrositis syndrome" and healthy subjects. Psychosom Med 37: 341–351

Montaigne ME (1580) Book 1. In: LeClerk J-V (ed) Essais. Garnier, Paris, pp 374–375

Nuechterlein KM, Holroyd JC (1980) Biofeedback in the treatment of tension headache – current status. Arch Gen Psychiatry 37: 866–873

Pilowsky I, Chapman CR, Bonica JJ (1977) Pain and depression. Br J Psychiatry 141: 30–36

Radanov BP, Stefano GD, Schnidrig A et al (1991) Role of psychosocial stress in recovery from common whiplash. Lancet 338: 712–715

Romano JM, Turner JA (1985) Chronic pain and depression: does the evidence support a relationship? Psychol Bull 97: 18–34

Russell IJ, Orr MD, Littman B et al (1994) Elevated cerebrospinal fluid levels of substance P in patients with the fibromyalgia syndrome. Arthritis Rheum 11: 1593–1601

Salter M, Brooke RI, Merskey H et al (1983) Is the temporomandibular pain and dysfunction syndrome a disorder of the mind? Pain 17: 151–166

Schilder P (1931) Notes on the psychopathology of pain in neuroses and psychoses. Psychoanal Rev 18: 1–22

Schmidt AJM (1988) Reply to letter. Pain 33: 388–389

Schrader H, Obelieniene D, Bovim G et al (1996) Evolution of late whiplash syndrome outside the medico-legal context. Lancet 347: 1207–1211

Speculand B, Goss AN, Hughes A et al (1983) Temporo-mandibular joint dysfunction: pain and illness behaviour. Pain 17: 139–150

Teasell RW, Merskey H (1997) Chronic pain disability in the work place. Pain Res Management 2: 197–205

Turk DC, Meichenbaum DH, Herman WH (1979) Application of biofeedback for the regulation of pain: a critical review. Psychol Bull 86: 1322–1338

Ulrich RE (1966) Pain as a cause of aggression. Am Zool 6: 643

Ulrich RE, Hutchison PR, Azrin NH (1965) Pain-elicited aggression. Psychol Rec 15: 11

Vaerøy H, Helle RE, Førre Ø et al (1988) Elevated levels of substance P and high incidence of Raynaud phenomenon in patients with fibromyalgia: new features for diagnosis. Pain 32: 21–26

Wall PD (1984) Introduction: textbook of pain. In: Wall PD, Melzack R (eds) Churchill-Livingstone, Edinburgh

Watson GD, Chandarana PC, Merskey H (1981) Relationships between pain and schizophrenia. Br J Psychiatry 138: 33–36

Wolff HG (1937) Personality factors and reactions of subjects with migraine. Arch Neurol Psychiatry 37: 895

Wolff HG (1948) Headache and other head pain. Oxford University Press, London

Woolf CJ, Wall PD (1986) Relative effectiveness of C primary afferent fibers of different origins in evoking a prolonged facilitation of the flexor reflex in the rat. J Neurosci 6: 1433–1442

Zigmond AS, Snaith RP (1983) The hospital anxiety and depression scale. Acta Psychiatr Scand 67: 361–370

Part 1
Schizophrenic, Affective and Related Disorders

CHAPTER 1

A. Jablensky

Symptoms of Schizophrenia

1
Definition of Schizophrenia by Its Symptoms

1.1
Diagnosis

1.1.1 Clinical Diagnosis

A century since the delineation of the diagnostic entity of *dementia praecox*, schizophrenia is still an "epigenetic puzzle" (Gottesman and Shields 1982). It can only be diagnosed by a painstaking analysis of the subjective experience reported by patients, the history and course of symptoms, the observation of behaviour and (to a lesser extent) by evaluation of the pre-morbid development, personality traits and family background. Recent research has highlighted "candidate" biological markers, such as neurocognitive deficits, peculiarities in brain morphology and neurochemical abnormalities, yet none of these variables possesses at present the sensitivity and specificity required of a diagnostic test. Schizophrenia remains, essentially, a clinical entity defined by its symptoms and their course. The existence of a specific disease (or diseases) underlying the clinical entity is a working hypothesis for which definitive proof or refutation is yet to be produced, although the link of schizophrenia to brain disorder is no longer disputed.

1.1.2 Polymorphism of Schizophrenic Symptoms

The symptoms of schizophrenia span the entire range of psychopathology and display an extraordinary amount of inter-individual variation. The diagnostic concept of schizophrenia is an omnibus term for clinical pictures as disparate as the insidious development of the "clinical poverty syndrome" (Wing and Brown 1970) and the expansive delusional preoccupation with themes of cosmic significance. The polymorphism of the manifestations of schizophrenia (incorrectly referred to as "heterogeneity") is an essential feature of the disorder and a persisting irritant to any theories about its unitary nature.

1.1.3 Symptoms as Diagnostic Criteria

The introduction of explicit diagnostic criteria in psychiatry, with their latest versions incorporated in ICD-10 (WHO 1992) and DSM-IV (American Psychiatric Association 1994), was a major advance which reduced semantic variation and the effect of bias due to selective emphasis on particular manifestations of schizophrenia by different "schools of thought".

The diagnostic criteria for schizophrenia proposed since Feighner et al. (1972) have strived to achieve at least three different, not fully compatible goals: (1) to identify groups of patients with similar prognosis (e.g. poor outcome), (2) to define a "heritable" diagnostic category and (3) to enable early diagnosis and treatment. "Narrow" definitions, such as those in the Research Diagnostic Criteria (RDC; Spitzer et al. 1978), DSM-III (American Psychiatric Association 1980) and DSM-IIIR (American Psychiatric Association 1987), were successful in achieving the first goal, simply because they included previous chronicity as one of the diagnostic criteria. As regards the second goal, the evidence is conflicting: the "narrow" DSM-III definition has been associated with a higher heritability in twin studies but not in adoptive, family or linkage studies where higher heritability estimates are obtained with a broader definition. As regards the third goal, the evidence available does not allow us to determine whether the choice of "narrow" versus "broad" criteria, and of the corresponding ratio of false-positive to false-negative diagnoses, is associated with better or worse outcome of early intervention in incipient psychotic illness.

It is now generally accepted that no single symptom is absolutely necessary for diagnosing schizophrenia, and that any subset of symptoms from an agreed list can be sufficient for the diagnosis (this method of classifying is called "polythetic", as opposed to the "monothetic" method, which requires the presence of at least one "strong" indicant; Sokal 1974). In practice, the polythetic definition of the symptomatology of schizophrenia means that patients can be allocated to the diagnostic category without having a single symptom in common. The symptom checklists contained in current diagnostic criteria for schizophrenia (Table 1) treat symptoms as interchangeable, so that the presence of any one, or any two, would suffice for the criterion to be met. As a consequence, the agreement between different operational definitions of schizophrenia in current use is less than perfect. The proportion of patients that are concordantly identified as schizophrenic by the different diagnostic systems can be disappointingly low, not only between different diagnostic systems but also within the same system because different subsets of symptoms can be used for the diagnosis (McGorry et al. 1992). The "polydiagnostic" approach proposed by Berner and Katschnig (1984), in which alternative sets of diagnostic criteria are applied to the same patients in order to identify a "core" group meeting all criteria, is still of uncertain practical value.

Table 1. Symptoms as part of diagnostic criteria for schizophrenia

Diagnostic criteria	Delusions	Hallucinations	First-rank symptoms	Thought and speech disorders	Negative symptoms	Other symptoms
Research Diagnostic Criteria, RDC (Spitzer et al. 1978)	Delusions of being controlled or influenced; bizarre or multiple delusions Somatic, grandiose, religious, nihilistic, or other, lasting for longer than 1 week Any type, if accompanied by hallucinations for longer than 1 week	Non-affective verbal, spoken to the subject Any type throughout the day for several days or intermittently for longer than 1 month	Thought broadcasting, insertion or withdrawal A voice keeps up a running commentary on the subject's behaviours or thoughts Two or more voices converse	Marked formal thought disorder accompanied by either blunted or inappropriate affect, delusions or hallucinations of any type, or grossly disorganised behaviour		
DSM-III-R (American Psychiatric Association 1987)	Delusions Bizarre (totally implausible) delusions for longer than 1 week	Prominent (throughout the day for several days or several times a week for several weeks) Voice having no apparent relation to depression or elation	Voice keeping up a running commentary on the person's behaviour or thoughts Two or more voices conversing	Incoherence or marked loosening of associations	Flat or grossly inappropriate affect	Catatonic behaviour
DSM-IV (American Psychiatric Association 1994)	Delusions present for a significant portion of time during a 1-month period (or less if treated) One delusional symptom required, if bizarre	Hallucinations	Voice keeping up a running commentary on the person's behaviour or thoughts Two or more voices conversing	Disorganised speech Frequent derailment or incoherence	Affective flattening Alogia Avolition	Grossly disorganised or catatonic behaviour
ICD-10 (WHO 1993)	Delusions of control, influence or passivity, referred to body or limb movements or specific thoughts, actions or sensations Persistent delusions of other kind	Persistent hallucinations, any modality, occurring every day for more than 1 month, accompanied by delusions without affective content or by persistent over-valued ideas	Thought echo, insertion, withdrawal or broadcasting Delusional perception	Neologisms, breaks or interpolations in the train of thought Incoherence Irrelevant speech	Apathy Paucity of speech Blunting or incongruity of emotional response	Catatonic behaviour: excitement, posturing or waxy flexibility, negativism, mutism, stupor

1.2 Is Schizophrenia a Disease Entity?

1.2.1 Critique of the Concept

The proposition that clinical pictures as disparate as hebephrenia, catatonia and paranoid psychosis could have anything in common has provoked vigorous debate since its inception. Hoche (1912) compared Kraepelin's theory to the chasing of a phantom. Towards the end of his career, Kraepelin himself changed his point of view and admitted that, rather than being a disease, schizophrenia could be seen as "a common reaction of mankind to the most varied forms of noxious events" (Kraepelin 1920). The protean nature of the symptoms and behaviours embraced by the diagnosis of schizophrenia continues to be the focus of critique.

Modern critics of the concept of schizophrenia advance three main arguments.

1. The concept lacks coherence because of the polymorphism of its attributes – a point which no one at present seems to dispute.
2. A diagnosis of schizophrenia does not predict other attributes (behavioural or biological) of the individuals given that label.
3. No genetic base specific to schizophrenia has been demonstrated to exist, and the findings of family studies (including twin and adoptive data), as well as those of genetic linkage studies, are fully compatible with alternative models.

The solutions proposed by critics are (a) to split schizophrenia into smaller, narrowly defined syndromes which may have discrete genetic and pathophysiological bases, (b) to replace the current nosology with a rehashed "unitary psychosis" concept, or (c) to discard the concept altogether since it is no more than a "social construction" (Sarbin 1990) or a "scientific delusion" (Boyle 1990).

1.2.2 Relevance of the Concept

Whereas each of the critical arguments above has a point, it is hard to see demonstrable advantages stemming from any one of the proposed alternatives. Both "splitting" and "lumping" strategies have been tried before, e.g. by the Kleist–Leonhard school (Leonhard 1995) and the psychobiology school tradition (Meyer 1958), respectively, without substantial advances in the scientific understanding of schizophrenia. Radical calls to dismantle the concept have so far failed to propose a plausible alternative model that would account for the subjective experiences and behaviours we designate as schizophrenic, nor for the consistent observation that about 1% of the population develop such disorders in their lifetime.

Perhaps the strongest evidence that schizophrenia is not an artifactual construct comes from the relative invariance of its clinical presentation and incidence across different populations and over time. Field research carried out by the World Health Organization (WHO) has shown that patients matching the diagnostic concept can be reliably identified in over 20 different populations and cultures and that the incidence of such disorders is within the range of 1.8–4.2 per 10,000 population at risk per year (WHO 1973, 1979; Jablensky et al. 1992). Comparative studies based on hospital archives and on a re-analysis of Kraepelin's own case records (Jablensky et al. 1993; Jablensky and Woodbury 1995) have led to the conclusion that the psychopathological pictures classified as dementia praecox at the turn of the century, and as schizophrenia today, are essentially similar. The ubiquitous occurrence and striking constancy of the manifestations of schizophrenia amid general human diversity has led to the conjecture that they are related in some direct fashion to the "speciation characteristics" of *homo sapiens* (Crow 1997).

1.2.3 Is There a "Basic" Disturbance?

There have been many attempts to define a common denominator for the bewildering variety of the manifestations of schizophrenia. The fundamental disturbance has been sought in "a weakening of the mainsprings of volition" and the "loss of inner unity of mental activities" (Kraepelin 1919), a "structural loosening of associations" (Wernicke 1900; E. Bleuler 1911), a "weakness in consciousness" (Berze 1914), "intrapsychic ataxia" (Stransky 1904) and a "loss of vital contact with reality" (Minkowski 1927). While intuitively appealing, such formulations simply replace one level of description with another.

More recent attempts at reduction involve theoretical models which allow testable predictions to be made. Such models assume that schizophrenia is essentially a cognitive disorder resulting from a "neurointegrative defect" (Meehl 1990), a neurocognitive "vulnerability" (Zubin and Spring 1977; Nuechterlein et al. 1990) or a "misconnection disorder" involving the multimodal association cortices and the basal ganglia (Frith 1992; Hemsley 1994). The predictions that can be derived from such theoretical models are currently generating a substantial body of empirical research.

It remains an open question whether a single basic disturbance or a *causa prima* of the symptoms of

schizophrenia will ever be found. It is possible that the primary genetic, physiological and cognitive deficits are multiple and only loosely related, although they manifest a "common final pathway". It is also possible that the primary deficits are not intrinsically pathological but represent extreme normal variants of structure and function. Their interaction, either additive or non-linear, could produce the diagnostic symptoms, but subclinical manifestations would be detectable in otherwise healthy individuals. There is a fairly convincing evidence of a preliminary nature to support the emerging research agenda along such lines, with "candidate" markers of pathogenetic processes in attention and memory dysfunction, brain morphology and physiology.

To sum up the argument, we do not know whether schizophrenia is a single process with pleiotropic manifestations at the level of cerebral organisation or a collection of aetiologically unrelated but interacting processes. Its manifestations in toto do not fit neatly into any of the current disease models, including the multifactorial ones usually associated with cancer, ischaemic heart disease or diabetes. For the time being, the validity of the clinical construct of schizophrenia is based on a polythetic logic in the sense that we designate as schizophrenia any co-occurring subset of manifestations in the following domains:

- Characteristic symptoms
- Characteristic patterns of course and outcome
- Characteristic brain morphology and neurochemistry
- Characteristic neurocognitive performance
- Characteristic genetic associations

The key word "characteristic" refers to internal consistency and discriminant power vis-à-vis abnormalities associated with other disorders or clinical constructs.

1.2.4 Why Study Symptoms?

In the absence of an identified "basic disturbance" and validated biological markers, the clinical concept of schizophrenia hangs together on the strength of the empirical evidence that its many facets form a broad syndrome with some internal cohesion and a characteristic evolution over time. The evidence is neither final nor static and needs to be re-examined as new concepts and technologies coming from molecular genetics, cognitive science or brain imaging bring forth new perspectives on disease causation and brain function. The construction of "top-down" models based on cognitive theory must be complemented by a "bottom-up" approach involving a reliable and valid description of its symptoms and signs which are the building blocks of the whole edifice. However, the study of symptoms should no longer be purely descriptive. The mapping of clinical phenomenology on specific brain dysfunction (and vice versa) is becoming feasible, and the resulting "functional psychopathology" (van Praag 1993) may substantially recast the present nosology. The dissection of schizophrenia into clinical phenotypes with specific neurocognitive or neurophysiological underpinnings is beginning to be perceived as a promising approach in psychiatric genetics amid growing doubts that the diagnostic categories of DSM-IV or ICD-10 are coded in the human genome.

1.2.5 How to Capture the "Schizophrenic" in the Symptoms

The symptoms of schizophrenia relate to virtually all domains of psychological function. However, the characteristic that sets them apart from other psychiatric disorders is the pervasive way in which they affect the core perceptions of the self and the external world. Scharfetter (1983) coined the term "ego psychopathology" and proposed to group the key symptoms of schizophrenia into five psychopathological domains of ego dysfunction: vitality, activity, consistency, demarcation and identity.

Generally, the manifestations of schizophrenia fall into two large categories: abnormal subjective experiences and objectively observable disorders of behaviour and performance. The characteristic "schizophrenic" quality of the subjectively experienced symptoms can only be grasped by phenomenological analysis (Jaspers 1948) which relies on patients' self-reports, whether spontaneously produced or elicited by means of clinical interviews such as the Present State Examination (PSE; J.K. Wing et al. 1974, 1990). A common denominator for many of the characteristic subjective symptoms is that they represent "disorders of significance" in which the normal categories of cognitive apperception, i.e. time, space and causality, break down or are replaced by a pervasive delusional awareness of connectedness between objectively unrelated events, persons and intentions ("the patient knows there is a meaning but not what that meaning is"; Gruhle 1915). In one of those symptoms, delusional perception, the individual perceptual components are split from their natural context and acquire different properties from those which they have when the normal context prevails (Matussek 1952). In this example, phenomenological analysis suggests the hypothesis of an underlying neurocognitive dysfunction (inability to hold context in working memory) which is empirically testable.

2 Overview of Symptoms

2.1 Classification and Frequency

2.1.1 Ways of Classifying the Symptoms

Different ways have been proposed of classifying the manifold symptoms and signs in schizophrenia (Cutting 1995). E. Bleuler (1911) introduced the distinction between fundamental and accessory symptoms. The fundamental symptoms (loosening of associations, ambivalence and autism) are present, at least to a minimally discernible degree, in every case of schizophrenia and are therefore necessary for the diagnosis. The accessory symptoms, which include hallucinations, delusions and catatonic signs, are much more conspicuous than the fundamental symptoms but are not necessary for the diagnosis, though in the absence of clearly identifiable fundamental symptoms they may be sufficient for the diagnosis.

Other ways of grouping symptoms involve the "positive/negative" distinction (see below) or their temporal sequence (prodromal, acute and residual). However, neither Bleuler's distinction between fundamental and accessory symptoms, nor the positive/negative or temporal sequence schemes are strict classifications based on explicit principles and rules. In the following overview, symptoms are therefore grouped according to domains of psychopathology, beginning with those that are subjective in the phenomenological sense and then proceeding to those accessible to objective observation and measurement.

2.1.2 Most Frequent Symptoms

Figure 1 provides frequency profiles for 44 psychotic and affective symptoms in 1288 patients with schizophrenia (the majority of them experiencing a first psychotic episode) assessed in the WHO ten-country study (Jablensky et al. 1992). The group profiles of patients in developing countries (India, Nigeria, Colombia) are remarkably similar to those of patients in developed countries (Czech Republic, Denmark, Ireland, Japan, Russia, UK and USA). Table 2 lists the 15 most common symptoms along with their frequency at the index psychotic episode and 2 years later in 811 schizophrenic patients evaluated in the WHO International Pilot Study of Schizophrenia (WHO 1973). In most clinical samples, the relatively non-specific psychotic symptoms such as hallucinations and delusions (i.e. accessory symptoms in Bleuler's sense) and negative symptoms such as flat affect, apathy and slowness tend to predominate. However, they are intermingled with "nuclear" symptoms such as delusional mood and thought alienation for which claims of relative specificity have been made.

2.2 First-Rank Psychotic Symptoms

2.2.1 Definition

Kurt Schneider (1950, 1957) proposed that nine groups of symptoms had a "decisive weight" in the differential diagnosis between schizophrenia and affective psychoses and designated them as "first-rank symptoms" (FRS). These symptom groups are as follows:

- Audible thoughts
- Voices arguing about or discussing the patient
- Voices commenting on the patient's actions
- Experiences of influences on the body (somatic passivity experiences)
- Thought withdrawal and other interference with thought
- Thought broadcast (diffusion of thought)
- Delusional perception
- Any other experiences involving "made" volition, impulses and feelings that are experienced as caused by an outside agency

In acknowledging the empirical derivation of the FRS, Schneider noted that they lacked a common structure (admitting nevertheless that some of them could be described as "loss of the contours of the ego") and had no theoretical implications for the understanding of schizophrenia.

The FRS concept, which had been familiar to German-speaking psychiatrists since the 1940s, became internationally accessible largely through the English translation published in 1959 of the 5th edition of *Klinische Psychopathologie* (Schneider 1950), the work of Fish (1967) and the WHO International Pilot Study of Schizophrenia (WHO 1973). At present, FRS are incorporated, explicitly or implicitly, in the RDC (Spitzer et al. 1978), DSM-III and its successors (American Psychiatric Association 1980, 1987, 1994) and ICD-10 (WHO 1992, 1993). Their international promotion since the 1960s was probably due to the sharpness of their definition and the belief that they could be ascertained reliably.

2.2.2 Versions

There are several modifications in the presentation of FRS (Table 3). Mellor's list (Mellor 1970) includes definitions which correspond closely to those of Schneider and are accompanied by illustrative vignettes. Data are provided on their frequency and

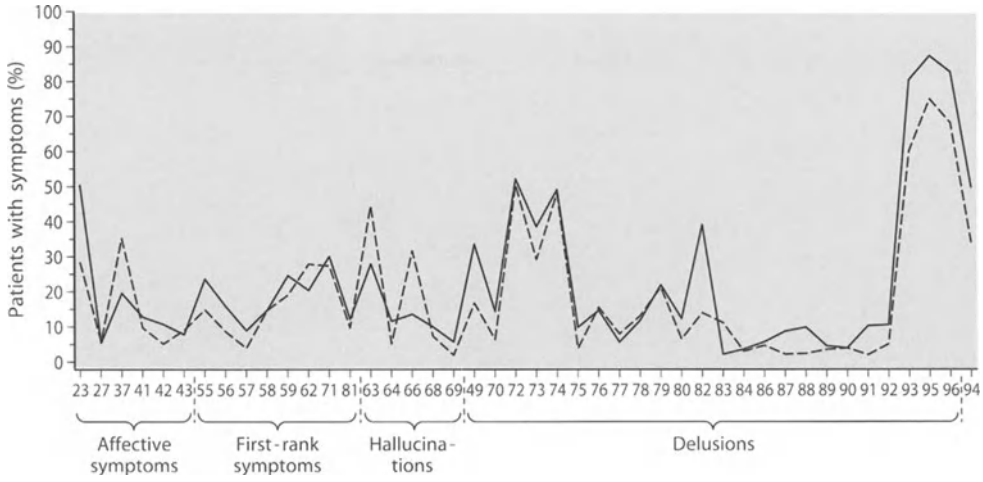


Fig. 1. Frequency of 44 selected symptoms in 551 patients in developing countries (*dashed line*) and 737 patients in developed countries (*solid line*), all meeting ICD/Catego criteria for schizophrenia. *Affective symptoms:* 23, depressed mood; 27, morning depression; 37, early waking; 41, expansive mood; 42, ideomotor pressure; 43, grandiose ideas and actions. *First-rank symptoms:* 55, thought insertion; 56, thought broadcast; 57, thought echo; 58, thought withdrawal; 59, thoughts being read; 49, delusional mood. *Hallucinations:* 62, voices in third person; 63, voices speaking to subject; 64, dissociative hallucinations; 66, visual hallucinations; 68, olfactory hallucinations; 69, delusion of smell; 70, other hallucinations. *Delusions:* 71, control; 72, reference; 73, delusional misinterpretation; 74, persecution; 75, assistance; 76, grandiose abilities; 77, grandiose identity; 78, religious; 79, paranormal; 80, physical forces; 81, alien forces; 82, primary delusions; 83, subcultural beliefs; 84, morbid jealousy; 86, sexual; 87, fantastic; 88, guilt; 89, appearance; 90, depersonalisation; 91, hypochondriacal; 92, catastrophe; 93, systematisation of delusions; 94, evasiveness; 95, preoccupation with delusions or hallucinations; 96, acting out of delusions. Data from the WHO ten-country study on schizophrenia (Jablensky et al. 1992)

Table 2. The 15 most frequent symptoms in a cohort of 811 patients with diagnosis of schizophrenia assessed in the WHO International Pilot Study of Schizophrenia (WHO 1979)

Symptom	Frequency (%) at index psychotic episode	Frequency (%) at 2-year follow-up
Ideas of reference	55.1	18.0
Suspiciousness	60.0	25.2
Delusions of reference	50.3	14.2
Delusions of persecution	48.1	12.7
Auditory hallucinations	43.8	11.6
Verbal auditory hallucinations	37.9	10.7
Voices speak to patient	36.3	9.4
Delusional mood	47.5	10.5
Thought alienation	33.5	7.4
Restricted quantity of speech	17.5	12.9
Flat affect	51.0	27.1
Apathy	30.4	18.8
Lack of insight	82.7	42.5
Inadequate description of symptoms	67.2	25.2

associations in a sample of 173 patients with schizophrenia. The tenth edition of the PSE, incorporated in the WHO Schedules for Clinical Assessment in Neuropsychiatry (SCAN; J.K. Wing et al. 1990) provides a semi-structured clinical interview, a glossary of definitions and rating rules for 11 FRS, but neither the composition of the list nor the glossary definitions are fully identical with those of Mellor.

Table 3. Comparison between three sets of definitions of Schneider's first-rank symptoms (FRS)

Mellor (1970)	Koehler (1979)	SCAN (Wing et al. 1990)
	Delusional continuum	
	Delusional mood	Delusional mood
	Delusional notion linked to, or provoked by, a perception	
Delusional perception	Delusional perception	Delusional perception
Influence on body (somatic passivity)	*Passivity continuum*	
"Made" impulses (drives)	Passivity mood	"Made" impulses
"Made" volitional acts	General experience of influence	Replaced control of actions
Thought insertion	Specific experience of influence	Thought insertion
"Made" feelings	Experience of influenced depersonalisation	Replaced control of affect
	Positive experience of alienation	
Thought withdrawal	Negative-active experience of alienation (includes thought withdrawal)	Thought withdrawal
Thought broadcasting (diffusion)	Negative-passive experience of alienation (includes thought broadcast)	Thought broadcast
	Sense deception continuum	
Voices commenting	Pseudohallucinatory voices (commenting, arguing, audible thoughts)	Voices commenting; loud thoughts; thought echo
Voices arguing	Hallucinatory voices (commenting, arguing,	
Audible thoughts (*écho de la pensée*)	audible thoughts)	Hallucinations of other senses

SCAN, Schedules for Clinical Assessment in Neuropsychiatry.

The PSE Catego (diagnostic categories) algorithm (J.K. Wing et al. 1974), used in the WHO cross-cultural studies, defines a "nuclear" class of schizophrenia (S+) characterised by at least three out of six of Schneider's FRS or, in their absence, by the concurrent presence of other, non-affective hallucinations and delusions.

In reviewing the "boundary criteria" of Schneider's symptoms, Koehler (1979) proposed re-formatting the FRS into three phenomenological continua (delusional, passivity and sense deception). Each symptom can be dichotomised into a "narrow" and a "wide" form. A clinical questionnaire compiled by O'Grady (1990) combines Mellor's set of FRS and the narrow/wide dichotomies.

2.2.3 Critique of the Concept

How well do the FRS perform in clinical practice and research? What is their frequency among schizophrenic patients, and how specific are they for schizophrenia? Critique of the diagnostic utility of FRS has been prompted by exaggerated claims about their "pathognomonicity" (made mainly in the British and American literature; see Andreasen and Flaum 1994) or the lack of adequate evidence (by present-day conventions) in Schneider's own writings to support their reliability and validity (Boyle 1990; Crichton 1996).

2.2.4 Are First-Rank Psychotic Symptoms Pathognomonic?

As regards "pathognomonicity", Schneider's own pronouncements are somewhat inconsistent: within the same chapter they are ascribed both a "strong" role in clinical decision-making ("undisputed precedence when it comes to the allocation of the individual case") and a "weaker" one (as simply having a "special value in helping us to determine the diagnosis"). However, he never referred to them as "pathognomonic" and stated explicitly that FRS may also occur in psychotic states associated with organic brain disease. Their original purpose was pragmatic: "when I find thought withdrawal in a psychosis which has no demonstrable somatic basis, then according to convention I call that psychosis schizophrenia" (Schneider 1957). As regards evidence for their diagnostic utility, the FRS have been more extensively investigated than any other group of symptoms in schizophrenia.

2.2.5 Frequency

The frequency of FRS is not as low as assumed. Allowing for some variation in the inclusion criteria, the prevalence of FRS in clinical populations with schizophrenia is in the range of 35%–68%, i.e. a frequency that is optimal for the use of a likelihood ratio in clinical diagnostic decisions. As Geddes et al.

(1996) point out, the published data suggest a likelihood ratio of about 30. With an a priori probability of schizophrenia in a clinical population of 30%–50%, the ascertainment of FRS would increase the *posterior* probability of a patient meeting RDC criteria for schizophrenia to 85%–95%. With a lower a priori probability of 5%–10%, the posterior probability would still be of the order of 65%–75%.

2.2.6 Specificity

The data available tend to confirm the relative specificity of FRS (Table 4). The positive predictive value (PPV), i.e. the ratio of FRS-positive cases that are "true" schizophrenia (diagnosed according to criteria which do not contain FRS) to all FRS-positive cases is in the range of 0.68–1.00. Allowing for sample size, selection and diagnostic bias, it would still be reasonable to accept that the PPV is no less than 0.80. In other words, the FRS indicate the likely presence of other psychotic symptoms diagnostic of schizophrenia. On the other hand, the absence of FRS (i.e. their negative predictive value, NPV) does not rule out schizophrenia.

2.2.7 Cross-Cultural "Robustness"

FRS are cross-culturally "robust". In clinical studies totalling over 450 cases, FRS have been described in 20%–56% of non-European schizophrenic patients in India (Radhakrishnan et al. 1983), Pakistan (Malik et al. 1990), South Africa (Teggin et al. 1985), Saudi Arabia (Zarrouk 1978) and among immigrants to the United Kingdom (Ndetei and Vadher 1984). In the WHO studies (WHO 1973; Jablensky et al. 1992), the prevalence of FRS (Catego class S+) among patients with ICD schizophrenia ranged from 38% (a rural area in India) to 84% (Nigeria). The subjective reports of patients on such symptoms were strikingly similar.

Moreover, the morbid risk (age, 15–54 years) for Catego S+ schizophrenia in the WHO ten-country study was in the narrow range of 0.27%–0.54%, in contrast to the risk for the "broad" clinical diagnosis, which varied between 0.50% and 1.72%. Thus FRS delineate a subset of cases that would meet criteria for "positive" schizophrenic symptomatology in different populations and cultural settings.

2.2.8 Common Pathophysiology

The cross-cultural "robustness" of the FRS suggests that they may have a common pathophysiological basis, e.g. in a temporo-limbic dysfunction. FRS occur in psychoses associated with temporal lobe epilepsy (Trimble 1990). Catego S+ patients in the WHO study (Jablensky et al. 1992) had an increased probability (relative risk, RR $\geqslant$ 2) of a history of seizures when compared to non-S+ patients. Another possible lead is provided by evidence that FRS may be markers of postsynaptic dopaminergic supersensitivity, as suggested by the significantly increased growth hormone response to apomorphine in Catego S+ patients compared to non-S+ patients (Whalley et al. 1984).

2.3 Prodromal and Residual "Basic" Symptoms

2.3.1 Definition

The basic deficits or basic symptoms (BS; Huber 1983; Süllwold and Huber 1986; Gross and Huber 1995) are conceptually related to Schneider's FRS. They are described as subjective experiences which precede or follow episodes of florid psychotic symptoms in schizophrenia and schizoaffective disorder. As "sentinel" symptoms or prodromes, they may be the earliest precursors of psychosis; and as postpsychotic "pure defects", they are presumed to express aspects of the biological vulnerability to schizophrenia, e.g. subtle impairments in information processing.

2.3.2 Precursors of Psychosis

The BS are not negative symptoms; they have been described as "microproductive" symptoms or as "positive symptoms *in statu nascendi*". As such, they are claimed to be "the real primary symptoms of schizophrenia and the basis of the fluctuating productive psychotic symptomatology", especially the FRS (Huber and Gross 1989). Patients have at least partial insight into them and may develop coping strategies. They can remain compensated in a favourable environment.

2.3.3 Assessment and Measurement

The BS are rated on a continuum ranging from the least specific (level 1) to the most specific or "typical schizophrenic phenomena", which include FRS (level 3). The Bonn Scale for Assessment of Basic Symptoms (BSABS; Gross et al. 1987) contains 98 BS items divided into five groups:

1. Direct minus symptoms (fatiguability, diminished "psychic tolerance", reduced energy)

Table 4. Frequency, positive predictive value (PPV) and negative predictive value (NPV) of Schneider's first-rank symptoms (FRS) calculated from data reported in six studies

Author(s)	Total study population (*n*)	Patients with schizophrenia (*n*)	FRS definition	Proportion of schizophrenia patients with one or more FRS	PPV	NPV
WHO IPSS (1973)	1131[a]	809[g]	CATEGO S (PSE-8)	0.68	0.91	0.51
Carpenter et al. (1973)	165[b]	103[h]	8 FRS (PSE-8)	0.51	0.85	0.50
Radhakrishnan et al. (1983)	266[c]	88[i]	PSE-8	0.35	0.68	0.74
Tandon and Greden (1987)	294[d]	81[j]	SADS	0.51	0.82	0.83
O'Grady (1990)	99[e]	21[k]	Mellor's "narrow" definition	0.62	1.00	0.90
Jablensky et al. (1992)	1288[f]	1087[l]	CATEGO S+ (PSE-9)	0.58	0.87	0.19

IPSS, International Pilot Study of Schizophrenia; PSE, Present State Examination; SADS, Schedules for Affective Disorders and Schizophrenia.
[a]Hospital-based sample of patients with schizophrenia, depression and non-psychotic disorders.
[b]Subset of the WHO IPSS population.
[c]Unselected consecutive admissions.
[d]Patients admitted to a research unit, drug-free for 2 weeks.
[e]Acute hospital admissions.
[f]Epidemiological sample of first-contact, non-affective psychoses.
[g]ICD-8.
[h]DSM-II.
[i]Feighner's criteria.
[j]Research Diagnostic Criteria (RDC).
[k]Carpenter's Flexible System.
[l]ICD-9.

2. Indirect minus symptoms (heightened impressionability, obsessions, depersonalisation phenomena, phobic avoidance)
3. Subjective cognitive disorders, involving (a) thought, speech and memory (lowering of "thought energy", diminished ability for "re-visualisation", impaired ability to comprehend symbols, impaired, false or overinclusive recall of biographical memories, poverty of content of speech), (b) sensory distortions (e.g. acoasms, micro- and macropsies, visual perception of "pseudomovement" or distorted outlines of objects, "aroused state of perceptual awareness", captivation by detail, seeing own face in mirror as changed), (c) motor phenomena (transient motor automatisms, freezing or blockage, loss of normal automated routines)
4. Coenesthesias or alterations of general sensibility (paroxysmal numbing or pains, migrating sensations, "electric" or thermal sensations, sensations of body enlargement or shrinkage, levitation, "dysaesthetic crises" similar to panic attacks, occurring mainly at night)
5. Autonomic symptoms (pupillary abnormalities, hyperhidrosis, vasomotor disturbances, nycturia and polyuria, paroxysmal tachycardia, systolic hypertension)

The description of BS suggests a substantial overlap with symptomatology that is usually classified under the rubrics of anxiety, obsessive-compulsive, somatoform or other neurotic manifestations.

2.3.4 Predictive Value

The clinical epidemiology of BS has been less well studied than that of FRS. The hypothesis that early presence of BS is predictive of later schizophrenia has been tested in an 8-year follow-up of 96 patients with provisional diagnoses of neurotic, mood or personality disorders, of whom 78 had BS on initial examination (Klosterkötter et al. 1997). On follow-up, 58% had either developed Schneider's FRS or otherwise met the DSM-III-R criteria for schizophrenia. All of the patients who made a transition to schizophrenia had had BS at the initial examination, and none of the patients who remained non-psychotic on follow-up had had BS at the initial examination. As screening criteria, the BS performed at the perfect level of 1.0 sensitivity but with low specificity (0.45), which resulted in 23% of the cases being misclassified (false positive for schizophrenia). The predictive value of BS was explained mainly by the subjective cognitive symptoms, such as thought interference and sensory distortions. One cannot, therefore, exclude the possibility that in this study psychotic symptoms had been present but masked by other symptomatology at index examination. Further evaluation of the clinical utility of BS is required.

2.4 Thought, Language and Communication Disorders

2.4.1 Earlier Studies

Kraepelin (1919) paid close attention to thought and speech disorders. Based on his previous research in Wilhelm Wundt's psychological laboratory, he introduced the word association test in the study of dementia praecox. Under the rubrics of disorders of self-expression, internal speech and the train of thought, he described poverty of thought (the "cessation of the need to express oneself"), incoherence, rhyming and word play, stereotypies, derailments, paraphasias and neologisms in ways anticipating present-day concepts. E. Bleuler (1911) regarded the "loosening of associations" as pathognomonic and detectable "in every case and at every period of the illness". Various other facets of thought disorder were described as "alogia" and "paralogia" (Kleist 1930), "overinclusiveness" (Cameron 1939), "concretism" (K. Goldstein 1944) and "cognitive slippage" (Meehl 1990).

2.4.2 Terminology and Classification

Despite the wealth of clinical description, the terminology and classification in this area of psychopathology have been inconsistent. Andreasen (1982) proposed designating this group of disturbances as "thought, language and communication disorders" and, using linguistic principles, divided them into disorders of the following:

- Morphology (word approximations, neologisms, clanging, paraphasias)
- Syntaxis (incoherence)
- Textual discourse (derailment, loss of goal)
- Pragmatics (poverty of speech, poverty of content, pressure of speech, distractible speech)

2.4.3 Measures and Scales

This classification underlies the Thought, Language and Communication (TLC) scales (Andreasen 1979) and the Communication Disturbance Index (CDI; Docherty et al. 1996). The latter is focused specifically on communication pragmatics and includes items such as vague, confused or ambiguous references, missing information and lack of structural clarity.

The Thought Disorder Index (TDI; Johnston and Holzman 1979) is derived from transcripts of subjects' verbatim responses to Rorschach cards, which are rated according to 20 categories.

2.4.4 Linguistic Perspective

It is hard to see whether the different scales measure a common construct. In this respect, psychopathological and neurocognitive research can benefit from a reference standard provided by linguistics. Chaika (1974, 1990) performed linguistic analysis of speech transcripts from patients diagnosed with schizophrenia. She concluded that, from a linguist's point of view, the speech of schizophrenic patients was as rule based and analysable as "normal" discourse. However, it manifested several regularly appearing peculiarities, such as a sporadic disruption in the ability to match semantic features with actual lexical items in the language, production of sentences according to phonological or semantic features of previously uttered words (rather than according to topic) and a failure to self-monitor. None of these features would qualify a "schizophrenic" speaker as linguistically incompetent. Together, the abnormalities seem to point to memory and information-processing deficits rather than to a primary defect in linguistic competence.

2.4.5 Neurocognitive Studies

Neuropsychological research using error detection and semantic priming paradigms provides some insight into the nature of these underlying deficits. Leudar et al. (1994) reported experimental support for the notion of deficient self-monitoring (the internal error detection rate in a sample of patients with schizophrenia was about 50% lower than in the control subjects). Vinogradov et al. (1992) investigated semantic memory in schizophrenic patients using a semantic priming paradigm. Their results suggested an intact semantic memory network and a normal priming effect (reflecting the spread of activation in the network) when "automatic" processing was involved, e.g. in a word pronunciation task. However, no significant priming effects were observed in lexical decision tasks (deciding whether a string of letters is a word or a non-word) which involve "controlled" or post-lexical information processing. Studies of acutely thought-disordered schizophrenic patients suggest the presence of other abnormalities which may be related to their state. "Hyperpriming" was described in such patients by Manschreck et al. (1988), and Spitzer (1997) found an increased indirect semantic priming effect (i.e. an excess of low-probability, unusual or remote responses to primed stimulus words). These findings are consistent with an inhibitory deficit causing hyperactivation in semantic association networks and a consequent lowering of the signal-to-noise ratio.

Drawing on analogies from primate research, Goldman-Rakic (1994) proposed that both types of findings may be related to a primary working memory defect manifesting as an impaired ability to maintain relevant semantic contexts "on line".

2.4.6 Are Thought, Language and Communication Disorders Specific to Schizophrenia?

Such impairments may not be specific to schizophrenia. Studies comparing the speech of patients with different psychotic disorders have failed to produce a clear separation of the schizophrenic patients from the rest, and some of the linguistic features displayed by "schizophrenic" speakers are shared by patients with manic disorder (Docherty et al. 1996). It seems, therefore, that contrary to E. Bleuler (1911), thought and language disorders may be neither fundamental nor pathognomonic for schizophrenia but a secondary expression of a more widespread neurophysiological deficit involving the mechanisms of memory.

2.5 Movement Disorders: Catatonic Phenomena

2.5.1 Definition

Catatonic phenomena, first identified as a clinical cluster by Kahlbaum (1874), comprise disorders of movement, speech and autonomic function. The motor disturbances consist of hyperkinesias (excitement), hypokinesias (inhibition and slowness) and dys- or parakinesias (abnormal postures, loss of the "fluidity" of spontaneous movements, mannerisms, grimacing, staring, stereotypies, iterations). These disturbances are usually intertwined with phenomena described as disorders of volition (negativism, excessive or automatic compliance, echo-phenomena, impulsivity). Catatonic speech disorders include perseveration, verbigeration, echo- or palilalia, aprosody and mutism. Characteristic autonomic signs are dilatation of pupils or anisoconia, seborrhoea, sweating, oedema and acrocyanosis and alterations in muscle tone (rigidity, hypotonia).

2.5.2 Clinical Rating Scales

Two rating scales, the 17-item Modified Rogers Scale (Lund et al. 1991) and the 23-item Bush-Francis

Catatonia Rating Scale (Bush et al. 1996), have been shown to be reliable. A checklist of catatonic signs and glossary definitions is included in the WHO SCAN (J.K. Wing et al. 1990).

2.5.3 Leonhard's Typology of Catatonia

Catatonic signs tend to be underdiagnosed and underreported in clinical populations. In this context, the differentiated typology of catatonic phenomena proposed by Leonhard (1957, 1995) merits a wider clinical use. Within the group of systematic schizophrenias, Leonhard distinguished six types of catatonia:

1. Parakinetic (grotesque distortions or fragmentation of movements and speech which are reactive to external stimuli)
2. Manneristic (predominance of odd stereotyped or stylised movements suggestive of special meaning or purpose)
3. Proskinetic (automatic or impulsive motions triggered by external stimuli)
4. Negativistic (predominance of "contrariness" or resistance to interference)
5. Speech-prompt (disordered speech production provoked by questions)
6. Speech-sluggish (slowing and poverty of speech to the extent of mutism)

It is doubtful that these types constitute independent entities, but their description should help to differentiate catatonic phenomena from other movement disorders such as parkinsonism, tardive dyskinesia or akinetic mutism.

2.5.4 Frequency of Catatonic Phenomena

The evidence suggests that, on proper examination, catatonic signs can be detected in at least 7%–14% of consecutive admissions with diagnosis of schizophrenia (Bush et al. 1996). Such signs were present in as many of 23% of the patients of the pre-neuroleptic era admitted during the 1950s (Fenton et al. 1997). This is comparable with the frequency of 19.5% referred to by Kraepelin (1919). In the WHO ten-country study (Jablensky et al. 1992), the catatonic subtype of schizophrenia was diagnosed in 5.2% among 1151 first-contact patients, but the proportion varied markedly from 10.3% in the developing countries to 1.2% in the developed countries.

2.5.5 Specificity in Schizophrenia

There is a widespread belief that catatonic phenomena are diagnostically non-specific and that schizophrenia is one of the many conditions in which such disorders may occur. Catatonic signs have been described in depression (20% of elderly depressives, according to Starkstein et al. 1996), mania and organic brain disease (a catatonic type was listed as one of the "exogenous reaction types" by Bonhoeffer 1912). However, the published evidence is not fully consistent with this view. In a retrospective cohort of 273 neuroleptic-naïve patients rated for movement disorders, Fenton et al. (1997) found that the only statistically significant association of "spontaneous dyskinesias" was with a diagnosis of schizophrenia, although dyskinesias were also present in 14.3% of the bipolar affective cases. It is possible that the catatonic syndrome includes components which occur predominantly in the context of schizophrenia and autistic spectrum disorders (L. Wing 1996), while other components are shared with motor disturbances in affective disorder (e.g. melancholic stupor) or organic brain disorder such as parkinsonism.

2.5.6 Pathophysiological Basis

The pathophysiological basis of catatonia remains quite elusive. There is a tendency to subsume catatonic phenomena under the extrapyramidal disorders (Rogers 1991), but surprisingly little evidence has been produced that they are "organic" in the neurological sense or that they are strictly involuntary. Catatonic disorders are peculiarly responsive to the environment, as evidenced by patients' subjective reports (Strömgren 1992). Indirect evidence for an environmental effect is provided by the decline in the incidence of the catatonic subtype in "modern" sociocultural settings and its relative persistence in traditional rural communities. Despite their "neurological" appearance, catatonic phenomena may in fact be more closely related to the dissociative disorders. Their designation by E. Bleuler (1911) as "disorders of the will" may therefore be quite appropriate.

2.6 Disorders of Affect and Mood

2.6.1 Fundamental Affective Disturbances

Disturbances of affect are one of Bleuler's "fundamental" disorders in schizophrenia (E. Bleuler 1911). Certainly, abnormalities of affect, emotional response and mood are also among its most visible manifestations, since they are readily elicited in any social situation and permeate the patients' psychomotor behaviour and expression. The profound disturbance

of emotional rapport or of "affective exchange", perceived by a clinically trained interviewer interacting with a patient with schizophrenia, was designated by Rümke (1941) as "praecox feeling" and was considered to be pathognomonic for the disorder. The fundamental affective disturbances that can be described as deficits or negative symptoms include the following:

- Blunted, or flat, affect, including a general diminution of emotional response (flatness) and indifference to events or topics that normally evoke such response
- Incongruity of affect (affect not in keeping with, or not understandable in, its ideational context), also referred to as parathymia when it is a marked and persistent characteristic
- Restricted affect or affective rigidity (lack of modulation of affect)
- Anhedonia (a pervasive and refractory reduction in the capacity to experience pleasure)

2.6.2 Perplexity

Another characteristic symptom, perplexity, or "the oppressive awareness of being unable to maintain a consistent grasp of reality" (Störring 1939) is a mood state with a strong cognitive component.

2.6.3 Dissociation Between Experience and Expression

Although affective symptoms are common and persistent (see Table 2), they have attracted less research than the other negative or positive symptoms in schizophrenia. Some recent research findings on emotions in schizophrenia seem to challenge the conventional clinical view on affective disturbances as a manifestation of inner emotional impoverishment. Thus, while patients exposed to emotion-arousing video segments were rated as significantly less facially expressive of both positive and negative emotions than control subjects, they reported experiencing as much positive and negative emotions as the controls (Kring et al. 1993). Further, the "anhedonia" hypothesis that schizophrenic subjects may be selectively impaired in experiencing positive emotions while not differing from normal individuals in the experience of negative emotions was not supported. This suggests a dissociation between the inner emotional experience, which may be relatively intact, and the outward display of emotion, which is grossly disturbed in schizophrenia.

2.6.4 Frequency of Depressive Mood Disorder

Depressive symptoms are extremely common in schizophrenia and may appear at any stage of the disorder. In the prodromal phase, or in the first psychotic episode, depressed mood can be found in over 40% of the patients (Leff et al. 1988; Bustamante et al. 1994). In the WHO ten-country study (Jablensky et al. 1992), depressed mood was present on initial examination in 50% of the patients in the developed countries and in 30% of the patients in developing countries. Patients with Schneider's FRS tended to have a higher prevalence of depressive symptoms (over 60%) than patients without FRS. During a 5-year follow-up, 15% of the schizophrenia patients developed purely affective episodes (Leff et al. 1992). In the stabilisation phase of schizophrenia, the point prevalence of depressive symptoms is of the order of 30% (Birchwood et al. 1993). The suicide risk is correspondingly high (13%–15% of schizophrenic patients eventually kill themselves; Caldwell and Gottesman 1990) and represents a central problem in the long-term management of patients with schizophrenia.

2.6.5 Heterogeneity of Depressive States

Depression in schizophrenia is heterogeneous. Firstly, depressed mood needs to be distinguished from anhedonia, which is a primary deficit, and from akinetic parkinsonism, which may be a side-effect of neuroleptic treatment. Secondly, when correctly diagnosed, depression may be one of the following:

- A feature of the disease process itself
- A psychological response to the schizophrenic illness as a traumatic life situation (Birchwood et al. 1993)
- A true co-morbid disorder

2.6.6 Schizoaffective Disorder

The high frequency of depressive symptoms in schizophrenia and their temporal relationship to the psychotic phenomena raise doubts about the validity of establishing a separate clinical entity for schizoaffective disorder. The same reservation applies to the diagnostic rules in ICD-10 and DSM-IV, which give precedence to a diagnosis of schizoaffective or mood disorder if characteristic symptoms of schizophrenia or other mood-incongruent psychotic symptoms follow, or co-occur with, a major depressive syndrome. Considering that genes predisposing to mood disor-

ders must be very common, and that they may interact with other common genes associated with the risk of schizophrenia, a phenotypic co-occurrence of schizophrenic and affective symptoms could be expected to be the rule rather than an exception.

2.7 Negative Symptoms and Deficits

2.7.1 Origin of the Concept

A general "weakening" of mental processes resulting in a "defect" was the cornerstone of the construction of dementia praecox out of the previously unrelated syndromes of hebephrenia, catatonia and paranoid dementia. Kraepelin (1919) suggested that precursors of the "defect" could be detected at the earliest stages of the process, co-existing with "productive" or "florid" symptoms. Since the 1970s, the terms "negative symptoms" and "positive symptoms" have virtually replaced "defect" and "productive" symptoms in the literature on schizophrenia. According to Berrios (1985), the terms "negative" and "positive" symptoms were first mentioned by Reynolds in 1858. However, their current usage derives from Hughling Jackson's doctrine of the dissolution of higher nervous functions (Jackson 1887). In the 1960s, J.K. Wing (1961) proposed a symptom profile of schizophrenia consisting of four measures: flatness of affect, poverty of quantity or content of speech, incoherence of speech and coherently expressed delusions and hallucinations. Snezhnevskij (1975) in Russia developed a syndromal scale, based on Jackson's evolutionary concepts, with nine levels of positive syndromes and ten levels of negative syndromes of schizophrenia.

2.7.2 "Type I" and "Type II" Schizophrenia

The present popularity of the terms owes much to an influential paper by Crow (1980), who proposed to simplify the description of schizophrenia by grouping its symptoms and signs as either positive or negative. "Type I" (positive) schizophrenia was characterised by hallucinations, delusions and formal thought disorder, with a presumed underlying dopaminergic dysfunction, as evidenced by the good response of these symptoms to neuroleptics. "Type II" (negative) schizophrenia displayed the "clinical poverty syndrome" (Wing and Brown 1970) with social withdrawal, loss of volition, affective flattening and poverty of speech and thought content. Its response to neuroleptics was poor, and the presumed pathophysiology was associated with structural brain abnormalities, as evidenced in enlarged ventricles and an attenuated inter-hemispheric asymmetry. This typology can be clinically misleading, since it implies that the two clusters or "types" are mutually exclusive, which is certainly not the case (see also Eaton et al. 1995, p. 28).

2.7.3 Assessment

The scales for assessment of negative and positive symptoms (SANS and SAPS) were developed by Andreasen and Olsen (1982) on the basis of factor analysis indicating that the two groups of symptoms loaded on a single bipolar factor and that their extreme forms, similar to Crow's type I and type II, occupied the two ends of a continuum. The claim that the symptoms of schizophrenia could be reduced to a single bipolar factor was subsequently retracted, and the Type I/Type II dichotomy was replaced by three- or four-factor models. However, the SANS and SAPS continue to be widely used in schizophrenia research, providing a "shortcut" to the evaluation of mental state. That such "shortcut" methods are not without a price is illustrated by the findings of Bell et al. (1994), who applied a latent trait logistic model (item-response analysis) to the SANS/SAPS ratings obtained in 149 patients. They demonstrated a high proportion (85%) of threshold order errors inherent in the scales themselves. It is likely that similar scaling problems are common to many of the other instruments offering simple, aggregate measures of psychotic symptomatology.

2.7.4 Are the Terms "Negative" and "Positive" Misnomers?

The negative/positive typology may be contributing to a loosening of the concepts in the clinical description of schizophrenia. In their current usage, the terms are misnomers insofar as they imply a reference to Jackson's concept of hierarchical dissolution of functions (Jackson 1887). According to the latter, negative signs indicate a loss or "suspension" of a higher brain function due to a lesion. A positive sign cannot be caused by a lesion but results from the release of lower-level functions through the removal of inhibitory processes. Negative symptoms, therefore, precede, and could be said to cause, the positive symptoms. Although Jackson's dissolution theory, which was explicitly espoused by Kraepelin (1920), may still have some relevance for the understanding of the symptoms of schizophrenia, the actual usage of the terms negative and positive symptoms is "atheoretical" and purely descriptive. Since no theory-based rule exists for classifying a symptom of schizophrenia as either negative or positive, the placement of symptoms in these categories is by and large arbitrary. In reviewing

the literature, Walker and Lewine (1988) came to the conclusion that only six symptoms were both consistently classified as either negative or positive and included in the widely used clinical scales; a further 19 symptoms were consistently classified but not included, and seven symptoms (including Bleuler's "loosening of associations") were neither consistently classified nor included in rating scales. In many instances, it is hard to see what considerations might help to decide whether a symptom is negative or positive. Thought and speech disorders, as well as catatonic phenomena, which are difficult to place in either group, illustrate the ambiguity of the negative/positive classification. The expectation that the statistical processing of symptoms by factor analysis or other techniques would result in a "natural" classification of schizophrenic symptoms is unwarranted unless the raw data used as input to such an analysis are exhaustive or representative, unbiased by selection and phenomenologically sound.

2.7.5 Primary Deficits

The proposal by Carpenter et al. (1988) to distinguish between primary and secondary negative symptoms is an attempt to refine the classification of negative symptoms. The primary deficit symptoms are "enduring traits" which may precede the initial psychotic episode, are present both during and between episodes, are not accounted for by depression, anxiety or drug side-effects and do not respond to conventional anti-psychotic medication or its withdrawal. They include anhedonia, flat or restricted affect, poverty of content of speech, lack of a sense of purpose and diminished social drive. In the Schedule for Deficit Syndrome (SDS; Kirkpatrick et al. 1989), the presence of any two such symptoms is required for the diagnosis of a "deficit syndrome". In contrast, non-deficit negative symptoms, such as psychomotor slowness, anergia or social withdrawal, lack persistence, fluctuate in severity and show a temporal association with mood states or side-effects of medication. Samples of patients identified by deficit syndrome criteria have been shown to have higher rates of poor pre-morbid adjustment and neurological impairment, as well as lower depression scores and lower rates of alcohol abuse (Buchanan et al. 1990; Kirkpatrick et al. 1996a,b).

2.7.6 A Discrete Syndrome or a Dimension?

Notwithstanding the advantage of defining a core primary deficit, the validity of designating a discrete deficit syndrome, as opposed to a continuous dimension of primary deficits, is open to question. It is conceivable that the primary deficits seen in schizophrenic patients represent the severe end of a continuum which extends to borderline states and to normal temperament and character trait variation. The designing of neurobehavioural probes and psychometric measures capable of capturing such continuous variation could be a promising avenue for research.

2.8 Other Symptoms

2.8.1 Neurological Signs

Neurological abnormalities (both "hard" and "soft" signs) have repeatedly been shown to occur more frequently in schizophrenic patients than in normal control subjects. The difference between schizophrenic patients and controls persists after controlling for effects of medication. The deviations are widespread, but those most commonly ascertained include the following:

- Motor coordination deficits (e.g. dysdiadochokinesia, ataxia)
- Involuntary movements (e.g. choreiform movements)
- Integrative sensory functions (e.g. astereognosis, right–left discrimination)
- Primitive reflexes (e.g. snout reflex)

Patients with neurological abnormalities tend to perform poorly on neurocognitive tasks when compared to patients without such abnormalities and control subjects (Flashman et al. 1996). Perhaps more importantly, neurological deficits appear to be familial, with non-psychotic siblings of patients scoring intermediate between probands and healthy controls (Ismail et al. 1998). It is possible, therefore, that the diffuse neurological abnormalities represent the mildest expression of the genetic vulnerability underlying schizophrenia.

2.8.2 Minor Physical Anomalies

Dysmorphic or dysplastic signs have been described in patients with psychoses since the nineteenth century. Lombroso (1887) took systematically anthropometric measures in institutional populations in Italy and Switzerland and reported an excess of anomalies (especially craniofacial) among the inmates of psychiatric asylums, of whom many were undoubtedly schizophrenic. Using the Waldrop scale (Waldrop

and Halverson 1971), Gualtieri et al. (1982) compared schizophrenic patients with other psychiatric patients and with controls and found higher scores for minor physical anomalies in the schizophrenia group. The majority of the subsequent studies have confirmed the higher incidence among schizophrenic patients, compared to controls, of minor abnormalities including the following:

- High palate, abnormal palate ridges
- Bifid tongue
- Epicanthus
- Protruding ears
- Transverse palmar crease and a decreased total a–b (palmar) ridge count

These abnormalities are not specific to schizophrenia and do not appear to be strongly associated with any of the clinical features or known risk factors (McGrath et al. 1995). Using a refined measurement scale, Lane et al. (1997) identified 12 craniofacial anomalies which reliably distinguished schizophrenic patients from controls. These signs were normally distributed in both patients and controls, with a distribution around a higher mean score for the patient group. It is generally accepted that minor physical anomalies are signs of a "developmental instability", but the exact contribution of genetic and environmental factors, such as obstetric events between 8 and 22 weeks of gestation, is not known (see also Chaps. 4 and 6, Vol. 3, Part 1).

3
Reducing Variation: Grouping and Ordering

3.1
Grouping on the Basis of Clinical Concepts

3.1.1 Bleuler's Fundamental and Accessory Symptoms

Great clinicians are able to grasp "meaningful connections" (Jaspers 1948) – a performance which computers and statistical analysis can only imperfectly emulate but are able to test. There is no scarcity of attempts at reducing the complexity of schizophrenia to simple "laws" on the basis of clinical observation and analysis. One simplification which has not only survived but profoundly influenced present-day thinking about schizophrenia is Bleuler's distinction between fundamental and accessory symptoms (E. Bleuler 1911).

Bleuler's fundamental "four A's" include the following:

1. Loosening of associations (formal thought disorder)
2. Affective disturbances (blunting, flattening, incongruity)
3. Ambivalence (disturbance of volition and behaviour)
4. Autism (withdrawal from reality)

Any of these disorders can vary from minimum intensity (as in "latent" schizophrenia) to maximum intensity (as in schizophrenic psychosis), but they are persistent, never absent and impart on the clinical presentation the specific quality that allows the experienced clinician to diagnose schizophrenia. They take precedence over any other symptoms (e.g. mood disorder) that may be present concurrently and are decisive for the diagnosis. The accessory symptoms, on the other hand, are far more conspicuous and usually dominate the clinical picture without being either specific or necessary for the diagnosis:

- Perceptual disorders (hallucinations and illusions)
- Delusions
- Disorders of memory
- Changes in verbal expression (speech and writing)
- Personality change
- Catatonic signs
- Somatic signs
- Acute states (depressive, manic, catatonic)

The distinction between fundamental and accessory symptoms is purely clinical and should not be confused with Bleuler's less successful attempt to identify aetiologically "primary" and "secondary" symptoms.

3.1.2 Kraepelin's "Registers"

In one of his last journal articles, Kraepelin (1920) questioned and virtually rejected the existence of discrete disease entities in psychiatry (which he himself had promoted) and proposed instead a model based on Jackson's evolutionary concepts. In this model, the major psychopathological syndromes formed an hierarchical continuum ranging from "encephalopathic" mental states associated with organic brain damage to syndromes expressing pre-formed, heritable forms of reaction to noxious events. The psychopathological syndromes were compared with "the different registers of an organ, any of which may be brought into play according to the severity and extent of the pathological changes involved". Kraepelin distinguished three main "registers" of mental disorder:

1. Affective, paranoid and hysterical forms
2. The schizophrenic form
3. Encephalopathic and paroxysmal forms

Pathological processes operating at the encephalopathic or paroxysmal level of disorder usually recruit functions belonging to the higher levels and can therefore manifest schizophrenic, affective and hysterical features. The reverse is not the case: a primary affective or hysterical mechanism of reaction has no means of involving the lower levels. The intermediate position of the schizophrenic "register", on the other hand, implies that its manifestations can combine with either, or both, of the other two levels.

This model is, of course, an over-simplification, but a productive one. It has modern repercussions in the "continuum of psychosis" hypothesis (Crow 1995) and can provide an organising framework for many current genetic and neurophysiological concepts.

3.2 Grouping by Statistical Methods

3.2.1 Need for Data Reduction

Considering the variation in schizophrenic symptomatology, there is a need for data reduction to simplify the description of patient samples and increase the statistical power of clinical and biological studies. Factor analysis has been applied to psychiatric rating scales since the 1960s (Lorr et al. 1963; WHO 1973). A renewed interest in its application to the symptoms of schizophrenia has emerged recently with the availability of desktop computing statistical packages.

3.2.2 Factor Analysis

Essentially, factor analysis and related methods reduce the co-variation of the primary data matrix to a small number of latent variables or "factors" which account for the interrelationships among the primary variables and explain a proportion of their observed variance. By explicating relationships that might remain hidden from inspection or univariate analysis, exploratory factor analysis (EFA) generates hypotheses about dimensions which require replication by confirmatory factor analysis (CFA) and validation against external criteria (which should not be of the original data matrix).

Results of factor-analytical studies of symptoms of schizophrenia are strongly dependent on input. For example, studies using the SANS and SAPS generate factors that are different from those produced by studies using the Positive and Negative Syndrome Scale for Schizophrenia (PANSS) or other instruments.

3.2.3 Three-Factor Model

Based on a relatively small number of input variables (such as SANS/SAPS scores), a three-factor structure has been described by Liddle (1987) and replicated by other investigators (Brown and White 1992; Johnstone and Frith 1996). In this model, the negative symptoms of schizophrenia load on a single factor or dimension ("psychomotor poverty"), while the positive symptoms tend to split into a delusions and hallucinations factor ("reality distortion") and a thought and speech disorder factor ("disorganisation"). However, in comparing Liddle's and several other factorial models based on SANS and SAPS in terms of goodness-of-fit, Peralta et al. (1994) concluded that a four-factor model accorded best with the original data. The fourth factor proposed has been labelled "relational impairment" and involves the SANS items "anhedonia-asociality" (Peralta et al. 1994; Toomey et al. 1997).

3.2.4 How Many Factors?

Generally, the number of factors required to explain the variance is greater when a wider range of symptoms is used as input. For example, the use of the 50 SANS/SAPS items rather than the ten SANS/SAPS global ratings, or of alternative instruments such as the 30-item PANSS (Kay 1990) or the 27-item Comprehensive Psychopathological Rating Scale (CPRS; Åsberg et al. 1978), results in different factor solutions, which include a depressive factor (Kay and Sevy 1990; White et al. 1997; Arora et al. 1997; Salokangas 1997), a cognitive factor (Lindenmayer et al. 1995), a general neurotic factor (Rey et al. 1994) and a pre-morbid social impairment factor (Lenzenweger and Dworkin 1996). Using 21 items of the OPCRIT (Operational Criteria for Diagnosis) symptom checklist, rated on 102 patients with schizophrenia, Cardno et al. (1996) obtained negative, disorganisation and three positive symptom factors (paranoid delusions, FRS subjective thought disorder and FRS hallucinations). Factor solutions therefore are not unique, and the question of how many factors can parsimoniously describe the symptomatology of schizophrenia can only be answered in the context of a particular selection of symptoms and measurement method.

3.2.5 Stability of Factors

Most factor analytical studies have been cross-sectional, based on clinical populations of varying length of illness. One study (Arndt et al. 1995) of recent-onset patients examined the three-factor model based on

SANS/SAPS scores longitudinally over 2 years and found the three factors to vary independently of one another. The negative factor was identifiable at initial examination and was stable over 2 years. In another longitudinal study of recent-onset patients (Salokangas 1997), both the negative symptoms factor and a delusional factor were stable over 5 years. However, factor stability may be eroded over longer follow-up periods.

A prospective follow-up of 90 patients in Madras, India (Eaton et al. 1995), who were rated monthly for PSE symptoms over a 10-year period indicated that the initial two-factor structure of negative and positive symptoms faded in the course of time, evolving into a single factor loaded by both negative and positive symptoms.

3.2.6 Limitations of Factor Analysis

Results of factor-analytical studies of schizophrenia should be interpreted with caution, considering the diversity of clinical populations investigated (in terms of sample size, age, length of illness and diagnostic criteria employed) and the limitations of the instruments used to generate the input data. Measurement errors are bound to result in inconsistencies or in method-related co-variation in the primary correlation matrix of factor analysis which will ultimately affect its output. It is premature, therefore, to refer to the factor output of studies based on measures that may be psychometrically and phenomenologically flawed as established "dimensions" or "syndromes" of schizophrenia.

3.2.7 Cluster Analysis

Whereas factor analysis groups variables, cluster analysis groups individuals on the basis of maximum shared characteristics. Farmer et al. (1983) identified four clusters into which patients with schizophrenia could be grouped according to symptoms. A combination of cluster analysis and multidimensional scaling using SANS/SAPS ratings (Minas et al. 1992) produced four groups of symptoms (negative, thought disorder, hallucinations and delusions and persecutory delusions). In a more recent study (Dollfus et al. 1996), a sample of 138 patients with schizophrenia assessed with the PANSS also yielded four clusters, including one characterised by negative symptoms, one mainly psychotic, one mixed and one oligosymptomatic. The psychotic cluster could be split further into one group with conceptual disorganisation and another with delusions and hallucinations.

Cluster analysis is as dependent on the selection of input variables as factor analysis. Its advantage of grouping individuals, rather than variables, makes it suitable for epidemiological purposes but less so for clinical research focused on disease constructs.

3.2.8 Grade of Membership

A statistical approach which combines the features of factor analysis and cluster analysis is the grade of membership model (GoM), in which psychiatric disorders are represented as "fuzzy sets" (Woodbury et al. 1994; Manton et al. 1994a), allowing individuals to be members of more than one set. The method is computationally complex and was designed to analyse large data sets and to handle categorical variables and missing values. The GoM model simultaneously extracts from the data matrix a number of "pure types" (corresponding to factors in a general sense) and assigns scores describing the degree to which an individual belongs to a pure type. GoM analysis was applied to 1065 patients from the WHO International Pilot Study of Schizophrenia, with ratings on 170 symptoms used as input (Manton et al. 1994a). A statistically optimal solution (maximum likelihood) for schizophrenia was obtained with five pure types, each characterised by a unique set of symptoms. There were two clearly affective pure types (one depressive and one manic) and a "negative" pure type characterised by apathy and restricted speech. The positive symptoms of schizophrenia formed two pure types: one with marked FRS and another characterised by multiple perceptual disorders, depersonalisation–derealisation phenomena and bizarre delusions. When the pure types were correlated with the 5-year course and outcome, the pure type characterised by perceptual disorders, depersonalisation–derealisation and bizarre delusions had the worst prognosis. The potential of GoM to handle complexity in data sets makes it well suited to the study of schizophrenic symptomatology, but its actual applications are still in an initial stage.

3.2.9 General Remarks on Data Reduction

By applying statistical modelling and data analytic techniques to large, clinically well characterised samples, it should be possible to tease out specific dimensions or syndromes within the broad spectrum of schizophrenic symptomatology. This would represent a considerable advance for clinical and biological research, which at present relies mainly on a global diagnosis of the phenotype. However, this goal may not be easy to achieve in schizophrenia. A serious constraint which reduces the power of this strategy to

discover a "natural classification" within schizophrenia is the narrow information base on symptoms due to small or biased samples, selective measurement and the lack of validation criteria.

3.3 Grouping on the Basis of Familial Clustering

3.3.1 Schizophrenia Spectrum

The application of the methods of molecular genetics to the study of the aetiology of schizophrenia in groups of related individuals (entire families, siblings, twins) presupposes an ability to distinguish between "affected" and "unaffected" or "undiagnosed" subjects. Since the results of genetic linkage methods are very sensitive to false-positive identification of individuals as "affected", it would be safe to designate as "affected" only probands who meet restrictive ("narrow") diagnostic criteria, such as DSM-III-R schizophrenia. However, the rarity of families with multiple cases of schizophrenia among first-degree relatives means that samples would be small and the statistical power to detect linkage low. To overcome this limitation, use can be made of epidemiological and family study data which suggest that the genetic influence on the liability to schizophrenia in families is shared with liability to other disorders or syndromes, i.e. that schizophrenia is part of a spectrum of conditions which may have a common genetic basis (Kendler and Diehl 1993). The main criterion for deciding whether a given disorder can be admitted to membership in the schizophrenia spectrum is its relative risk ratio, which compares its prevalence among the first-degree relatives of schizophrenic probands with its prevalence among the relatives of probands with other psychiatric diagnoses or with its population rate.

A number of disorders and syndromes meet this criterion to a varying degree and have been included in the schizophrenia spectrum:

- "Typical" schizophrenia (core member of the spectrum)
- Schizotypal and paranoid personality disorders
- Schizoaffective disorder (depressed type)
- Other non-affective psychotic disorders (schizophreniform and delusional disorders, atypical psychosis)
- Affective disorders with psychotic features

The shared liability to these five disorders, in terms of a correlation coefficient, has been estimated at 0.36 on the basis of an epidemiologically based family study (Kendler et al. 1995). However, while the evidence is strong that "typical" schizophrenia and affective disorders with psychotic features are the two ends of the spectrum, the relative positions of the schizophrenia-like personality disorders, schizoaffective disorder and the mixed group of "other non-affective" psychotic disorders within the spectrum remain ambiguous. Some investigators treat schizophrenia, together with schizotypal personality disorder (SPD), as "hard" spectrum, in contrast to a "soft" spectrum which includes the remaining three and sometimes even further conditions. In all its variations, however, the spectrum concept depends critically on the validity of the concept of SPD.

3.3.2 Schizotypal Personality Disorder

The origin of the concept of SPD, which is the cornerstone of the schizophrenia spectrum theory, is in Bleuler's belief that "there also exist, beyond doubt, latent schizophrenias which never become manifest" (E. Bleuler 1920). "Latent schizophrenia" was listed as a subtype of schizophrenia in ICD-8 and ICD-9. The term "schizotypy" was introduced by Rado (1956) and subsequently adopted by Meehl (1990) to describe a personality type characterised by anhedonia, ambivalence, "interpersonal aversiveness", body image distortion, perceptual anomalies and sensory, kinaesthetic and vestibular aberrations. Along similar lines, Chapman et al. (1982) developed scales to measure perceptual aberrations and "magical ideation" as traits predicting "psychosis proneness". These two constructs were amalgamated with clinical descriptions from the Danish–U.S. adoptive study into a diagnostic category (Kety et al. 1978), which was used to explain the excess prevalence of oddities of personality and behaviour among the biological relatives of adopted children who developed schizophrenia as adults. The diagnostic category was incorporated, with some modifications, into DSM-III and its successors and into ICD-10. The criteria emphasise the following:

- Pervasive social and interpersonal deficits
- Constricted affect
- Odd speech
- Cognitive or perceptual distortions
- Odd beliefs or magical thinking
- Occasional, brief quasi-psychotic episodes

The relatively frequent occurrence of SPD among the first-degree relatives of probands with schizophrenia has been replicated by the majority (though not by all) of the family studies (Kendler and Walsh 1995). There is also some evidence that individuals sampled for presence of SPD features are more likely to have relatives with clinical schizophrenia than control subjects without SPD (Thaker et al. 1993). Overall, however, the evidence that SPD is a well-defined and coherent cluster of traits is not overwhelming. No

population prevalence data are yet available, and the reliability of the diagnosis based on self-report measures has been questioned (Kendler et al. 1996). Twin data suggest that the current SPD criteria may be heterogeneous and that its manifestations fall into a "negative" cluster (odd speech and behaviour, inappropriate affect and social anxiety) that is more common among the relatives of schizophrenic probands and a "positive" cluster (magical ideation, brief quasi-psychotic episodes) that may be related to major affective disorder (Torgersen et al. 1993).

4 Course of Symptoms

The systematic investigation of the course of schizophrenic symptoms was initiated by Kraepelin, who believed that, in the absence of demonstrable brain pathology and identifiable causes, careful observation of the natural history could establish the validity of the disease entity: "The complexity of the conditions which we observe in the domain of dementia praecox is very great, so that their inner connection is at first recognizable only by their occurring one after the other in the course of the same disease" (Kraepelin 1919).

4.1 Adult Schizophrenia

Towards the end of his career, Kraepelin revised his earlier claim that the prognosis of dementia praecox was invariably poor and acknowledged that "permanent cures" had occurred in about 15% of his patients. Many subsequent longitudinal studies have highlighted the wide variability of the course of schizophrenia, which seems to be the most salient characteristic of its natural history. Manfred Bleuler's observations on 208 patients followed up intensively for 22 years or until death (M. Bleuler 1972; M. Bleuler et al. 1976) are worth retaining as an important record of the pre-neuroleptic prognosis of schizophrenia (although some of the patients were medicated at later stages of the follow-up):

- Lasting recovery ("complete cure") in 20%–26%, severe chronic states in 14%–24%
- No further deterioration after the fifth year since onset and development of a clinically stable state in 50%–75% of the patients
- Remitting course characterised by multiple episodes and full remissions in 22%
- Catastrophic course (rapid onset of chronic deterioration) in 4%
- 20-year suicide rate of 14%–22%

4.1.1 Recent Course and Outcome Studies

More recent studies tend to corroborate the pattern of outcomes outlined by the earlier studies. The reported rates of improvement without relapse range from 21% (Bland and Orn 1978) to 30% (Scottish Schizophrenia Research Group 1992), and poor outcome in terms of continuous psychotic symptoms and/or increasing social disability from 24% to 43% (Shepherd et al. 1989).

The prospective WHO studies (WHO 1973, 1979; Leff et al. 1992; Jablensky et al. 1992) provided a cross-cultural database on the course and outcome of schizophrenia which comprises initial and follow-up information on a total of 2736 patients diagnosed as schizophrenic according to strict criteria. The general conclusions of the WHO studies are as follows:

- A wide continuum of outcomes can be observed in patients with similar initial symptom characteristics, ranging from stable recovery after a single psychotic episode to a chronic unremitting psychosis
- The relapse rate for psychotic symptoms tends to increase over time, but their average duration tends to remain stable or to decrease
- Only a minority of patients (22% in the WHO cohorts) experience unremitting psychotic symptoms
- The level of social impairment associated with negative symptoms at 2 years changes little during a 5-year follow-up; most of the clinical change occurs between the 2-year follow-up and the 5-year follow-up and is in the direction of improvement

The recent resurgence of interest in the early detection and treatment of first episodes of psychosis, driven by theoretical considerations and clinical concerns, is supported by empirical evidence suggesting that the pre-onset period is marked by a characteristic sequence of sub-stages (Häfner et al. 1995; Hambrecht et al. 1994) that can be described in finer detail. Plausible clinical considerations have been proposed in support of the view that the first episode of psychosis represents a critical developmental transition that may impact the subsequent course of schizophrenia. Thus clinical research bridging the gap between statistical investigations of risk factors or antecedents of disease and individual pathways to psychotic illness may have an important role to play in understanding and, ultimately, influencing the development and course of schizophrenia.

4.1.2 Stability of Clinical Syndromes

Longitudinal studies suggest that the symptomatology of schizophrenia "breeds true" in the sense that few patients are eventually reclassified into other disease categories following a firm initial diagnosis of a schizophrenic illness. In the 2-year follow-up of the International Pilot Study of Schizophrenia (WHO 1979), 75% of the patients with an initial diagnosis of schizophrenia who experienced relapses of any type had schizophrenic symptoms only, and another 3% had both schizophrenic and other types of episodes. On 5-year follow-up, these proportions were 59% and 17%, respectively, i.e. there was some increase in the number of patients who in the course of time developed other symptoms (mainly affective) in addition to persisting or episodic schizophrenic symptoms. Patients with FRS on initial examination had a relative risk of 2.7 of experiencing the same symptoms at later stages of the disorder compared to patients with no initial FRS (Jablensky et al. 1992).

The most common non-psychotic symptoms developing in the course of schizophrenia are depressive. In the course of 2 years, 17% of the patients in the International Pilot Study of Schizophrenia with a diagnosis of schizophrenia who remitted and relapsed had clear-cut depressive episodes (Sheldrick et al. 1977). This proportion remained unchanged at 15% at the end of the 5-year follow-up (Leff et al. 1992). Similarly, the frequency of major depressive episodes was 24% during the 2–12 years of follow-up in the National Institute of Mental Health (NIMH) study (Breier et al. 1991). These data suggest that depression is part of the clinical spectrum of schizophrenia.

4.1.3 Predictors of Course and Outcome

The predictors of course and outcome in schizophrenia fall into six classes:

1. Socio-demographic and family background characteristics
2. Characteristics of the pre-morbid personality and pre-index functioning
3. History of past psychotic episodes and treatments
4. Characteristics of the onset
5. Symptoms of the initial clinical state
6. Mixed findings related to brain morphology, treatment response and habit behaviour

Many of these predictors have been established independently by different investigators, and there is agreement on the general direction of their effects. Their explanatory power varies depending on the setting, sample size, homogeneity of patient groups, number of predictors and dependent variables and measurement error, but generally tends to be low. Generally, male sex, single marital status, pre-morbid social withdrawal, insidious onset and pre-index chronicity emerge as robust predictors of a poor outcome in the short to medium term (2–5 years), while female sex, being married, having social contacts outside the home and acute onset predict a relatively good outcome (Childers and Harding 1990; Angermeyer et al. 1990; Munk-Jørgensen et al. 1991; Jablensky et al. 1992). In the short term, the best predictor of relapse is anti-psychotic drug withdrawal (Dencker et al. 1986). With the exception of negative symptoms, the initial symptoms of schizophrenia have less predictive power than the variables listed above.

4.2 Pathoplastic Influences on Symptom Expression

4.2.1 Age

Of all pathoplastic influences on the manifestations of schizophrenia, age has the most significant impact. Most studies on the psychopathology of schizophrenia have been based on clinical samples of young adults. However, schizophrenic symptoms may have their onset at any age. Recent research has highlighted more systematically than previous studies the clinical profile of the disorder at a very young age and also after the age of 45. The frequency and severity of the different types of schizophrenic symptoms are age dependent and seem to reflect particular domains of vulnerability associated with the developmental stages. However, the precise nature of the relationship is not yet clear. Notably, this age-dependent expression appears to be much more marked in positive symptoms than in negative symptoms. The latter may occur at any age and show little variation across developmental stages.

4.2.2 Childhood Schizophrenia

Childhood schizophrenia is extremely rare (prevalence of less than 0.9% of all psychiatric inpatient admissions of children under 15; Thomsen 1996). Nevertheless, its clinical manifestations can be distinguished from those of other early-onset disorders such as autism or Asperger syndrome. They show similarity to, and continuity with, the symptoms of schizophrenia manifesting at a later age. Several studies of children with schizophrenia beginning before the age of 12 (Russel 1994; Remschmidt et al. 1994; Alaghband-Rad et al. 1997) indicate that hallucinations in all sensory modalities, delusions (often bizarre or involving childhood themes and beliefs), catatonic signs and

negative symptoms are common. Command auditory hallucinations and visual hallucinations are more frequent in childhood than in schizophrenia of later onset. They may occur in the context of an oneiroid (dream-like) state with a strong affective colouring and are difficult to distinguish from dissociative states and vivid imagery. Thought disorder, however, is rare in schizophrenic illnesses beginning before puberty. Its frequency relative to other symptoms increases sharply in adolescence, mainly in the context of the hebephrenic syndrome, which is also characterised by inappropriate or shallow affect and severe conduct disorder.

4.2.3 Childhood Precursors of Symptoms in Adult Schizophrenia

Both the cognitive and psychomotor symptoms of schizophrenia have early developmental precursors. In the Copenhagen prospective follow-up of 207 children born to schizophrenic mothers, early measures of thought disorder ("vague and drifting thinking" and poverty of speech) were found to correlate positively with adult thought disorder, a finding suggesting that "schizophrenic symptomatology develops by gradual accretion" (Parnas and Schulsinger 1986).

4.2.4 Ageing Process and Symptoms

The effect of ageing on the symptoms of schizophrenia concerns two different issues: the clinical manifestations in elderly patients with an onset of the disorder at a younger age, and the symptom picture of schizophrenia with first onset in middle or older age. As regards the ageing schizophrenic patient, the results of follow-up studies suggest that the florid psychotic symptoms tend to recede and that residual, mainly negative symptoms occupy the foreground (Harvey et al. 1996). Cognitive deficits are correlated with the negative symptoms and may be conspicuous, although the negative symptoms themselves show little or no progression with age. Such symptoms may not be totally irreversible, and remarkable symptomatic improvement (in 24% of patients, according to M. Bleuler et al. 1976) has been described in long-term follow-up studies.

4.2.5 Late-Onset Schizophrenia

First onset of schizophrenia after the age of 65 years has been documented in 4% of males and 18% of females in a community prevalence sample of non-affective functional psychoses (Castle and Murray 1993). The terms late-onset schizophrenia (onset after age 45) and late paraphrenia (onset after age 60) are used somewhat inconsistently to denote syndromes which overlap but may be associated with different causal and risk factors (Almeida et al. 1995). Generally, the late-onset schizophrenia-like illnesses may display any of the symptoms of schizophrenia occurring at an earlier age, including FRS, but usually have several characteristic features (Henderson and Kay 1997):

- Rare occurrence of formal thought disorder and catatonic signs
- Preserved affect (absence of restricted, flat or inappropriate affect)
- Predominance of paranoid delusions and auditory hallucinations
- Schizoid or paranoid pre-morbid personality and social isolation
- Frequent finding of visual impairment or hearing loss

The paranoid ideation often takes the form of the so-called permeable walls or partition delusions (people, substances or forces are entering through the walls). Fantastic delusions and delusional memories are also characteristic and may have a basis in deficits in episodic memory and in recognition memory for faces which have been found in elderly psychotic patients (David and Howard 1994).

4.2.6 Sex Differences

Both brain maturation differentials and environmental influences in pre-adolescence could be expected to result in male–female differences in the presentation of psychotic disorders with an onset in early adulthood. The epidemological evidence suggests sex differences in the population incidence of schizophrenia (higher in males) and affective disorders (higher in females) in 16- to 25-year-olds and a tendency to an earlier onset of schizophrenia in males which may be attenuated by environmental factors (Jablensky and Cole 1997). However, the evidence for a strong pathoplastic effect of gender on the symptoms of schizophrenia and their course is far less clear-cut than might be predicted. There is a tendency for women to have a better premorbid social adjustment (Shtasel et al. 1992), a greater admixture of affective symptoms (Copolov et al. 1990) and better long-term outcomes (J.M. Goldstein 1988), but studies with a focus on specific negative, positive and cognitive symptoms have failed to demonstrate sex differences in their frequency and treatment response (Lindström and von Knorring 1994; Fennig et al. 1995; Pinals et al. 1996). Generally, the data do not support the view (Lewine 1981) that fundamental male–female differences exist in symptom expression in schizophrenia.

4.2.7 Effects of Culture and Ethnicity

The WHO cross-cultural studies of schizophrenia (WHO 1973, 1979; Leff et al. 1992; Jablensky et al. 1992) provide consistent and comprehensive evidence that the symptoms and signs associated with the construct of schizophrenia are basically the same in all populations studied and occur with a similar frequency among people presenting for treatment at mental health facilities. Social anthropologists who have commented critically on some of the conclusions of the WHO studies (Hopper 1991; Edgerton and Cohen 1994) have pointed out (correctly) that the methodology of these studies was "etic", i.e. proceeding from a diagnostic construct which has its historic origin in one culture and seeking to replicate and validate the construct in other cultures. It is theoretically possible that other manifestations of "reality distortion" exist in various traditional cultures but remain undetected or unaccounted for in studies using instruments and symptom definitions that are "ethnocentric". This argument is irrefutable, since it is impossible to design a study that would be exhaustive in investigating, without an a priori hypothesis, all possible mental states and behaviours. However, no evidence has been produced up to date of any culture or ethnic group in which schizophrenic disorders meeting the WHO criteria and description simply do not exist. This is not to say that ethnic variation in schizophrenic symptomatology is non-existent. It has not been systematically explored and its extent is practically unknown.

5 Symptoms and Brain Function

5.1 Neurocognitive and Neurobiological Correlates

5.1.1 Differentiated Symptom Measurement

Some of the symptoms of schizophrenia may be windows to neurocognitive and neurobiological abnormalities involved in the pathogenesis of the disorder. Unfortunately, we do not know which symptoms are best suited for this role, because the greater part of biological and cognitive research in schizophrenia has been conducted on clinical samples defined by diagnosis, without a differentiated subtyping by symptoms or syndromes. Potentially relevant findings from those less numerous studies which have targeted neurocognitive and biological correlates of individual symptoms or areas of psychopathology are reviewed below.

5.1.2 Neuropsychological and Neurocognitive Studies (see Chap. 8, Vol. 3, Part 1)

Certain consistent patterns of abnormalities have been linked to negative or deficit symptoms. Impairment of "executive" functions (categorical reasoning, sequencing and use of representations to guide behaviour) and attention deficits (impairments of selective and sustained attention, reduced processing capacity) have been replicated across studies of patients with "psychomotor poverty" (Gold and Weinberger 1995; Buchanan et al. 1997).

Less consistent findings have been elicited with regard to the neurocognitive mechanisms underlying positive symptoms or the "reality distortion" and "disorganisation" factors. Patients with speech and thought disorder tend to exhibit source memory difficulties and to perform poorly on word generation tasks and on tasks demanding suppression of irrelevant responses. Patients with hallucinatory and delusional symptoms have been shown to be deficient in auditory learning and recognition memory and to perform poorly on eye movement tasks (Strauss 1993). The neurocognitive impairments associated with positive symptoms may be part of an overarching dysfunctional organisation of cerebral activity in schizophrenia which involves functional disconnections between frontal lobe structures (initiation of action) and posterior areas (monitoring and perception of the effects of action). Such disconnection may result in an inability to distinguish between external events and self-generated mental activity (Frith 1992, 1996).

5.1.3 Event-Related Potentials (see Chap. 9, Vol. 3, Part 1)

Electrophysiological measures such as event-related brain potentials (ERP) are sensitive to dysfunctions involving the medial temporal lobe and superior temporal gyrus which are thought to underlie positive symptoms. Thus a well-replicated abnormality in the P50 component (which may reflect a failure in the inhibitory mechanism "gating" sensory input) has been found to be strongly associated with positive symptoms but not with negative symptoms (Waldo et al. 1991). Abnormalities in the mismatch negativity (MMN) wave, linked to an auditory sensory memory deficit, have been found to be negatively correlated with SANS scores (Catts et al. 1995). In contrast, the finding of a prolonged latency of the P300 wave (for which a more widely distributed topography involving the prefrontal cortex is assumed) may be an early sign of cognitive impairment and negative symptoms (McCarley et al. 1997).

5.1.4 Structural Brain Imaging (see Chap. 11, Vol. 1, Part 1; Chap. 5, Vol. 3, Part 1)

Anatomical brain imaging studies involving comparisons between schizophrenic and healthy control subjects indicate grey matter and volume reductions in schizophrenia (Johnstone et al. 1989; Frazier et al. 1996). Most consistently, such reductions have been demonstrated to affect the left hemisphere temporal grey matter and the hippocampus (Mozley et al. 1994; Marsh et al. 1997). However, few studies so far have attempted to relate magnetic resonance imaging (MRI) volumetric changes to particular symptom clusters in schizophrenia. The results are not entirely consistent. Thus at least two recent studies (Turetsky et al. 1995; Wibble et al. 1995) report left temporal lobe abnormalities, rather than the predicted frontal lobe abnormalities, in patients with severe negative symptoms. In subjects mainly characterised by marked positive symptoms, Barta et al. (1990) found a reduced temporal gyral volume in patients with prominent auditory hallucinations, and Rossi et al. (1994) described reduced temporal plane asymmetry in patients with thought disorder. In a recent study of 12 schizophrenic patients, Liddle's three factors were used as co-variates for regional grey matter volume differences. Psychomotor poverty was found to be negatively correlated with volume in a large region of the left ventro-medial prefrontal cortex. Disorganisation, however, was significantly associated with relative increases in regional grey matter volume in the medial temporal cortex bilaterally, which included the hippocampus and the parahippocampal gyrus. No significant correlations, positive or negative, were found for reality distortion (Chua et al. 1997). The results of this study suggest that, although schizophrenic patients as a group tend to have volume decreases in several areas compared to healthy subjects, there may be differences within the schizophrenic group that are related to the predominant clinical symptomatology. In this study, the temporal cortex volume reduction characterising the schizophrenic patients as a group was relatively less pronounced in those patients who had marked thought and speech disorders than in the patients without such symptoms.

5.1.5 Functional Imaging Using Positron Emission Tomography and Single Photon Emission Tomography

Several regional cerebral blood flow (rCBF) studies using positron emission tomography (PET) or single photon emission tomography (SPET) have demonstrated in patients with predominantly negative symptoms, as compared to control subjects, a bilaterally decreased regional activation of the prefrontal cortex during performance of executive tasks. The rCBF reduction in the dorsolateral prefrontal cortex has been found to be particularly significant in patients with marked poverty of speech. Convergent evidence has been contributed by ^{31}P magnetic resonance spectroscopy (MRS) showing a decrease in phosphomonoesthers in the left frontal lobe, correlated with poor executive task performance (Deicken et al. 1995). Such evidence has been interpreted as supporting the "hypofrontality" model of the pathophysiology of negative symptoms in schizophrenia (Weinberger et al. 1986). However, the findings are not consistent and several studies have failed to replicate the hypofrontality effect. More recent interpretations of the PET evidence suggest a major role for an abnormal pattern of functional connectivity between frontal and temporal lobes rather than a circumscribed regional abnormality as a possible correlate of negative symptoms (Liddle 1995, 1997).

PET studies of patients characterised by active positive symptoms report a widely distributed pattern of rCBF decreases and increases which involve the left temporal lobe, left parietal lobe, right medial prefrontal cortex, anterior and posterior cingulate, left ventral striatum and thalamus (Kaplan et al. 1993; Sabri et al. 1997). Experiments attempting to detect a PET "signature" of auditory hallucinations in real time have highlighted a pattern of activation which includes the subcortical nuclei (thalamus and striatum), limbic and paralimbic structures and the orbitofrontal complex (involved in the generation of speech). Group experiments combined with a single case study of a patient with both auditory and visual hallucinations (Silbersweig et al. 1995) led to the tentative conclusion that the generation of hallucinations is associated with activity in the deep brain structures, while their specific perceptual content is determined by activity in the neocortical regions. Failure to activate areas suspected to be related to the monitoring of inner speech (rostral supplementary motor area and left middle temporal gyrus) distinguished patients "prone to hallucinations" from other patients with schizophrenia and controls (McGuire et al. 1996).

5.1.6 Functional Magnetic Resonance Imaging

Functional MRI (fMRI) is a new research tool which provides excellent temporal resolution, uses no radioactive tracers and therefore offers a capacity for practically unlimited serial measurements. Although the number of published fMRI studies on schizophrenic patients (reviewed by Kindermann et al. 1997) is small, the potential of the method has been demonstrated by findings of attenuated exogenous

auditory activation in patients experiencing auditory hallucinations. This is interpreted as an indication of a physiological competition for a common neural substrate between exogenous and endogenous (hallucinogenic) activation (David et al. 1996).

5.2 Current Limitations and Future Prospects of Correlational Studies

5.2.1 Current Constraints

The modest progress in the elucidation of specific pathophysiological mechanisms underlying schizophrenic symptoms is due, in part, to the general constraints of sample size, symptom variability and capacity of the research technology. In addition, there are specific constraints related to clinical assessment, limitations of the neurocognitive and neuropsychological paradigms and limitations of the analysis and interpretation.

5.2.2 "Visualising" Studies

With regard to clinical assessment, the majority of imaging studies use aggregate symptom measures or factors which may be too crude and do not provide an adequate differentiation between relevant clinical states. Few studies have attempted to probe systematically individual symptoms. The few, highly imaginative studies that aim to "visualise" cerebral activation differences linked to the subjective experience of auditory hallucinations tend to rely on the subjects' ability to report such experiences "on line". Such research subjects must be selected for an unusual degree of insight into their own "reality testing" dysfunction. Consequently, it is possible that either the subjective experience reported and "visualised" on functional imaging in such studies is phenomenologically different from hallucinations in the usual sense or that what the method is measuring is activation associated with a high level of self-monitoring.

5.2.3 Need for Specific Neurocognitive Probes

As regards neuropsychological measurement, there is a paucity of validated neurocognitive probes specifically designed for the study of schizophrenic symptoms. The majority of the standard neuropsychological batteries and tests used in schizophrenia research were originally calibrated on brain-damaged clinical populations. The sensitivity of the present neuropsychological tools to detect "differential deficits" in brain functioning related to individual symptoms (as opposed to diffuse impairment) may be too low (Frith 1996; Gur et al. 1997).

5.2.4 Possible Fallacies of Interpretation

Finally, both neurocognitive and dynamic neuroimaging studies tend to reveal a complex picture of distributed activity (Schröder et al. 1996; Andreasen et al. 1997) that has no unique interpretation. The variation in brain activation that is due to processes specific to schizophrenia is difficult to disentangle from confounding variation originating in many other sources, both intra- and extracerebral. At the level of interpretation, predictions based on current cognitive models of schizophrenia often lack the specificity that would match the phenomenological description of symptoms. In the absence of a strong predictive hypothesis and of robust symptom measurement, varying ad hoc interpretations may be fitted to the data.

5.2.5 Possibility and Promise of Correlational Research

Despite such limitations, there is a forming consensus that the identification of neurobehavioural phenotypes correlated with imaging data is possible. For example, one of the few neuroimaging studies comparing rCBF profiles across diagnostic categories (Dolan et al. 1993) led to the conclusion that the dysfunction in the dorsolateral prefrontal cortex, associated with poverty of speech, is related to symptoms rather than to the disease. Such findings are of particular relevance to genetic studies where tentative causal associations are more likely to be found with intermediate "endophenotypes" rather than with the diagnostic category.

6 References

Alaghband-Rad J, Hamburger SD, Giedd JN, Frazier JA, Rapoport JL (1997) Childhood-onset schizophrenia: biological markers in relation to clinical characteristics. Am J Psychiatry 154: 64–68

Almeida OP, Howard RJ, Levy R, David AS, Morris RG, Sahakian BJ (1995) Clinical and cognitive diversity of psychotic states arising in late life (late paraphrenia). Psychol Med 25: 699–714

American Psychiatric Association (1980) Diagnostic and statistical manual of mental disorders, 3rd edn. American Psychiatric Association, Washington DC

American Psychiatric Association (1987) Diagnostic and statistical manual of mental disorders, 3rd edn, revised. American Psychiatric Association, Washington DC

American Psychiatric Association (1994) Diagnostic and statistical manual of mental disorders, 4th edn American Psychiatric Association, Washington DC

Andreasen NC (1979) Thought, language, and communication disorders. Arch Gen Psychiatry 36: 1315–1330

Andreasen NC (1982) Should the term "thought disorder" be revised? Compr Psychiatr 23: 291–299

Andreasen N, Flaum M (1994) Characteristic symptoms of schizophrenia. In: Widiger TA, Frances AJ, Pincus HA, First MB, Ross R, Davis W (eds) DSM-IV Sourcebook, vol 1. American Psychiatric Association, Washington DC, pp 351–380

Andreasen NC, Olsen S (1982) Negative vs positive schizophrenia: definition and validation. Arch Gen Psychiatry 39: 789–794

Andreasen NC, O'Leary DS, Flaum M, Nopoulos P, Watkins GL, Ponto LLB, Hichwa RD (1997) Hypofrontality in schizophrenia: distributed dysfunctional circuits in neuroleptic-naive patients. Lancet 349: 1730–1734

Angermeyer MC, Kuhn L, Goldstein JM (1990) Gender and the course of schizophrenia: differences in treated outcomes. Schizophr Bull 16: 293–308

Arndt S, Andreasen NC, Flaum M, Miller D, Nopoulos P (1995) A longitudinal study of symptom dimensions in schizophrenia. Arch Gen Psychiatry 52: 352–360

Arora A, Avasthi A, Kulhara P (1997) Subsyndromes of chronic schizophrenia: a phenomenological study. Acta Psychiatr Scand 96: 225–229

Åsberg M, Montgomery SA, Perris C, Schalling D, Sedvall G (1978) A comprehensive psychopathological rating scale. Acta Psychiatr Scand Suppl 271: 5–28

Barta PE, Pearlson GD, Powers RE, Richards SS, Tune LE (1990) Auditory hallucinations and smaller superior temporal gyral volume in schizophrenia. Am J Psychiatry 147: 1457–1462

Bell RC, Low LH, Jackson HJ, Dudgeon PL, Copolov DL, Singh BS (1994) Latent trait modelling of symptoms of schizophrenia. Psychol Med 24: 335–345

Berner P, Katschnig H (1984) Approche polydiagnostique en recherche psychiatrique. Ann Méd Psychol 142: 825–831

*Berrios GE (1985) Positive and negative symptoms and Jackson: a conceptual history. Arch Gen Psychiatry 42: 95–97

Berze J (1914) Die primäre Insuffizienz der psychischen Aktivität. Deuticke, Leipzig

Birchwood M, Mason R, MacMillan F, Healy J (1993) Depression, demoralization and control over psychotic illness: a comparison of depressed and non-depressed patients with a chronic psychosis. Psychol Med 23: 387–395

Bland G, Orn H (1978) 14-year outcome in early schizophrenia. Acta Psychiatr Scand 58: 327–338

*Bleuler E (1911) Dementia praecox oder die Gruppe der Schizophrenien. Deuticke, Leipzig

Bleuer E (1920) Lehrbuch der Psychiatrie, 3rd edn. Springer, Berlin

**Bleuler M (1972) Die schizophrenen Geistesstörungen im Lichte langjähriger Kranken- und Familiengeschichten. Thieme, Stuttgart

Bleuler M, Huber G, Gross G, Schüttler R (1976) Der langfristige Verlauf schizophrener Psychosen. Nervenarzt 47: 477–481

Bonhoeffer K (1912) Die Psychosen im Gefolge von akuten Infektionen, Allgemeinerkrankungen und inneren Erkrankungen. In: Aschaffenburg W (ed) Handbuch der Psychiatrie B. Deuticke, Leipzig, pp 1–110

Boyle M (1990) Schizophrenia: a scientific delusion? Routledge, London

Breier A, Schreiber JL, Dyer J, Pickar D (1991) National Institute of Mental Health longitudinal study of chronic schizophrenia. Arch Gen Psychiatry 48: 239–246

Brown KW, White T (1992) Syndromes of chronic schizophrenia and some clinical correlates. Br J Psychiatry 161: 317–322

Buchanan RW, Kirkpatrick B, Heinrichs DW, Carpenter WT (1990) Clinical correlates of the deficit syndrome of schizophrenia. Am J Psychiatry 147: 290–294

Buchanan RW, Strauss ME, Breier A, Kirkpatrick B, Carpenter WT (1997) Attentional impairments in deficit and nondeficit forms of schizophrenia. Am J Psychiatry 154: 363–370

Bustamante S, Maurer K, Löffler W, Häfner H (1994) Depression im Frühverlauf der Schizophrenie. Fortschr Neurol Psychiatr 62: 317–329

Bush G, Fink M, Petrides G, Dowling F, Francis A (1996) Catatonia. I. Rating scale and standardized examination. Acta Psychiatr Scand 93: 129–136

Caldwell CB, Gottesman II (1990) Schizophrenics kill themselves too: a review of risk factors for suicide. Schizophr Bull 16: 571–589

Cameron N (1939) Deterioration and regression in schizophrenic thinking. J Abnorm Soc Psychol 34: 265–270

Cardno AG, Jones LA, Murphy KC, Asherson P, Scott LC, Williams J, Owen MJ, McGuffin P (1996) Factor analysis of schizophrenic symptoms using the OPCRIT checklist. Schizophr Res 22: 233–239

Carpenter WT, Strauss JS, Muleh S (1973) Are there pathognomonic symptoms in schizophrenia? Arch Gen Psychiatry 28: 847–852

Carpenter WT, Heinrichs DW, Wagman AMI (1988) Deficit and nondeficit forms of schizophrenia: the concept. Am J Psychiatry 145: 578–583

Castle DJ, Murray RM (1993) The epidemiology of late-onset schizophrenia. Schizophr Bull 19: 691–700

Catts SV, Shelley AM, Ward PB, Liebert B, McConaghy N, Andrews S, Michie PT (1995) Brain potential evidence for an auditory sensory memory deficit in schizophrenia. Am J Psychiatry 152: 213–219

Chaika E (1974) A linguist looks at "schizophrenic" language. Brain Language 1: 257–276

*Chaika E (1990) Understanding psychotic speech. Thomas, Springfield

Chapman LJ, Chapman JP, Miller EN (1982) Reliabilities and intercorrelations of eight measures of proneness to psychosis. J Consult Clin Psychology 50: 187–195

Childers SE, Harding CM (1990) Gender, premorbid social functioning, and long-term outcome in DSM-III schizophrenia. Schizophr Bull 16: 309–318

Chua SE, Wright IC, Poline JB, Liddle PF, Murray RM, Frackowiak RSJ, Friston KJ, McGuire PK (1997) Grey matter correlates of syndromes in schizophrenia. Br J Psychiatry 170: 406–410

Copolov DL, McGorry PD, Singh BS, Proeve M, Van Riel R (1990) The influence of gender on the classification of psychotic disorders – a multidiagnostic approach. Acta Psychiatr Scand 82: 8–13

Crichton P (1996) First-rank symptoms or rank-and-file symptoms? Br J Psychiatry 169: 537–540

Crow TJ (1980) The molecular pathology of schizophrenia: more than one disease process? Br Med J 280: 66–68

Crow TJ (1995)A continuum of psychosis, one human gene, and not much else – the case for homogeneity. Schizophr Res 17: 135–145

*Crow TJ (1997) Is schizophrenia the price that Homo sapiens pays for language? Schizophr Res 28: 127–141
Cutting J (1995) Descriptive psychopathology. In: Hirsch SR, Weinberger DR (eds) Schizophrenia. Blackwell, Oxford
David AS, Howard R (1994) An experimental phenomenological approach to delusional memory in schizophrenia and late paraphrenia. Psychol Med 24: 515–524
David AS, Woodruff PWR, Howard R, Mellers JDC, Brammer M, Bullmore E, Wright I, Andrew C, Williams SCR (1996) Auditory hallucinations inhibit exogenous activation of auditory association cortex. NeuroReport 7: 932–936
Deicken RF, Merrin EL, Floyd TC, Weiner MW (1995) Correlation between left frontal phospholipids and Wisconsin Card Sort Test performance in schizophrenia. Schizophr Res 14: 177–181
Dencker SJ, Malm U, Lepp M (1986) Schizophrenic relapse after drug withdrawal is predictable. Acta Psychiatr Scand 73: 181–185
Docherty NM, DeRosa M, Andreasen NC (1996) Communication disturbances in schizophrenia and mania. Arch Gen Psychiatry 53: 358–364
Dolan RJ, Bench CJ, Liddle PF, Friston KJ, Frith CD, Grasby PM, Frackowiak RSJ (1993) Dorsolateral prefrontal cortex dysfunction in the major psychoses: symptom or disease specificity? J Neurol Neurosurg Psychiatry 56: 1290–1294
Dollfus S, Everitt B, Ribeyre JM, Assouly-Besse F, Sharp C, Petit M (1996) Identifying subtypes of schizophrenia by cluster analysis. Schizophr Bull 22: 545–555
Eaton WW, Thara R, Federman B, Melton B, Liang K (1995) Structure and course of positive and negative symptoms in schizophrenia. Arch Gen Psychiatry 52: 127–134
Edgerton RB, Cohen A (1994) Culture and schizophrenia: the DOSMD challenge. Br J Psychiatry 164: 222–231
Farmer AE, McGuffin P, Spitznagel EL (1983) Heterogeneity in schizophrenia: a cluster-analytic approach. Psychiatry Res 8: 1–12
*Feighner JP, Robins E, Guze SB, Woodruff RA, Winokur G, Munoz R (1972) Diagnostic criteria for use in psychiatric research. Arch Gen Psychiatry 26: 57–63
Fennig S, Putnam K, Bromet EJ, Galambos N (1995) Gender, premorbid characteristics and negative symptoms in schizophrenia. Acta Psychiatr Scand 92: 173–177
Fenton WS, Blyler CR, Wyatt RJ, McGlashan TH (1997) Prevalence of spontaneous dyskinesia in schizophrenic and nonschizophrenic psychiatric patients. Br J Psychiatry 171: 265–268
Fish F (1967) Clinical psychopathology. Wright, Bristol
Flashman LA, Flaum M, Gupta S, Andreasen NC (1996) Soft signs and neuropsychological performance in schizophrenia. Am J Psychiatry 153: 526–532
Frazier JA, Giedd JN, Hamburger SD, Albus KE, Kaysen D, Vaituzis AC, Rajapakse JC, Lenane MC, McKenna K, Jacobsen LK, Gordon CT, Breier A, Rapoport JL (1996) Brain anatomic magnetic resonance imaging in childhood-onset schizophrenia. Arch Gen Psychiatry 53: 617–624
*Frith CD (1992) The cognitive neuropsychology of schizophrenia. Erlbaum, Hove
Frith C (1996) Neuropsychology of schizophrenia. Br Med Bull 52: 618–626
Geddes JR, Christofi G, Sackett DL (1996) Commentaries on "First-rank symptoms or rank-and-file symptoms?". Br J Psychiatry 169: 544–545
Gold JM, Weinberger DR (1995) Cognitive deficits and the neurobiology of schizophrenia. Curr Opin Neurobiol 5: 225–230
*Goldman-Rakic PS (1994) Working memory dysfunction in schizophrenia. J Neuropsychiatry Clin Neurosci 6: 348–357
Goldstein JM (1988) Gender differences in the course of schizophrenia. Am J Psychiatry 145: 684–689
Goldstein K (1944) Methodological approach to the study of schizophrenic thought disorder. In: Kasanin JS (ed) Language and thought in schizophrenia. Norton, New York, pp 17–40
Gottesman II, Shields J (1982) Schizophrenia: the epigenetic puzzle. Cambridge University Press, New York
Gross G, Huber G (1995) Psychopathology and biological-psychiatric research. Neurol Psychiatry Brain Res 3: 161–166
Gross G, Huber G, Klosterkötter J, Linz M (1987) BSABS. Bonner Skala für die Beurteilung von Basissymptomen (Bonn Scale for the Assessment of Basic Symptoms). Springer, Berlin Heidelberg New York
*Gruhle HW (1915) Selbstschilderung und Einfühlung. Z Ges Neurol Psychiatr 28: 148–231
Gualtieri CT, Adams A, Shen CD, Loiselle D (1982) Minor physical anomalies in alcoholic and schizophrenic adults and hyperactive and autistic children. Am J Psychiatry 139: 640–643
Gur RC, Ragland JD, Gur RE (1997) Cognitive changes in schizophrenia – a critical look. Int Rev Psychiatry 9: 449–457
*Häfner H, Maurer K, Löffler W, Bustamante S, an der Heiden W, Riecher-Rössler A, Nowotny B (1995) Onset and early course of schizophrenia. In: Häfner H, Gattaz WG (eds) Search for the causes of schizophrenia, vol III. Springer, Berlin Heidelberg New York, pp 43–66
Hambrecht M, Häfner H, Löffler W (1994) Beginning schizophrenia observed by significant others. Soc Psychiatry Psychiatr Epidemiol 29: 53–60
Harvey PD, Lombardi J, Leibman M, White L, Parrella M, Powchik P, Davidson M (1996) Cognitive impairment and negative symptoms in geriatric chronic schizophrenic patients: a follow-up study. Schizophr Res 22: 223–231
Hemsley DR (1994) Cognitive disturbances as the link between schizophrenic symptoms and their biological bases. Neurol Psychiatry Brain Res 2: 163–170
*Henderson AS, Kay DWK (1997) The epidemiology of functional psychoses of late onset. Eur Arch Psychiatry Clin Neurosci 247: 176–189
*Hoche A (1912) Die Bedeutung der Symptomkomplexe in der Psychiatrie. Z Ges Neurol Psychiatr 12: 540–551
Hopper K (1991) Some old questions for the new cross-cultural psychiatry. Med Anthropol Q 5: 299–330
*Huber G (1983) Das Konzept substratnaher Basissymptome und seine Bedeutung für Theorie und Therapie schizophrener Erkrankungen. Nervenarzt 54: 23–32
Huber G, Gross G (1989) The concept of basic symptoms in schizophrenia and schizoaffective psychoses. Recent Prog Med
Ismail B, Cantor-Graae, McNeil TF (1998) Neurological abnormalities in schizophrenic patients and their siblings. Am J Psychiatry 155: 84–89
Jablensky A, Cole SW (1997) Is the earlier age at onset of schizophrenia in males a confounded finding? Results from a cross-cultural investigation. Br J Psychiatry 170: 234–240
Jablensky A, Woodbury MA (1995) Dementia praecox and manic-depressive insanity in 1908: a grade of membership

analysis of the Kraepelinian dichotomy. Eur Arch Psychiatry Clin Neurosci 245: 202–209

*Jablensky A, Sartorius N, Ernberg G, Anker M, Korten A, Cooper JE, Day R, Bertelsen A (1992) Schizophrenia: manifestations, incidence and course in different cultures. A World Health Organization ten-country study. Psychol Med Monogr Suppl 20: 1–97

Jablensky A, Hugler H, von Cranach M, Kalinov K (1993) Kraepelin revisited: a reassessment and statistical analysis of dementia praecox and manic-depressive insanity in 1908. Psychol Med 23: 843–858

Jackson JH (1887) Remarks on the evolution and dissolution of the nervous system. J Ment Sci 33: 25–48

**Jaspers K (1948) Allgemeine Psychopathologie, 5th edn Springer, Berlin [English translation: Hoenig J, Hamilton MW (1963) General psychopathology. Manchester University Press, Manchester/University of Chicago Press, Chicago]

Johnston MH, Holzman PS (1979) Assessing schizophrenic thinking. Jossey-Bass, San Francisco

Johnstone EC, Frith CD (1996) Validation of three dimensions of schizophrenic symptoms in a large unselected sample of patients. Psychol Med 26: 669–679

Johnstone EC, Owens DGC, Bydder GM, Colter N, Crow TJ, Frith CD (1989) The spectrum of structural brain changes in schizophrenia: age of onset as a predictor of cognitive and clinical impairments and their cerebral correlates. Psychol Med 19: 91–103

Kahlbaum KL (1874) Die Katatonie oder das Spannungsirresein. Hirschwald, Berlin [English translation: Catatonia. Johns Hopkins University Press, Baltimore, 1973]

Kaplan RD, Szechtman H, Franco S, Szechtman B, Nahmias C, Garnett ES, List S, Cleghorn JM (1993) Three clinical syndromes of schizophrenia in untreated subjects: relation to brain glucose activity measured by positron emission tomography (PET). Schizophr Res 11: 47–54

Kay SR (1990) Positive-negative symptom assessment in schizophrenia: psychometric issues and scale comparison. Psychiatr Q 61: 163–178

Kay SR, Sevy S (1990) Pyramidical model of schizophrenia. Schizophr Bull 16: 537–545

*Kendler KS, Diehl SR (1993) The genetics of schizophrenia: a current, genetic-epidemiological perspective. Schizophr Bull 19: 261–285

Kendler KS, Walsh D (1995) Schizotypal personality disorder in parents and the risk for schizophrenia in siblings. Schizophr Bull 21: 47–52

Kendler KS, Neale MC, Walsh D (1995) Evaluating the spectrum concept of schizophrenia in the Roscommon family study. Am J Psychiatry 152: 749–754

Kendler KS, Thacker L, Walsh D (1996) Self-report measures of schizotypy as indices of familial vulnerability to schizophrenia. Schizophr Bull 22: 511–520

*Kety SS, Rosenthal D, Wender PH, Schulsinger F, Jacobsen B (1978) The biologic and adoptive families of adoptive individuals who became schizophrenic: prevalence of mental illness and other characteristics. In: Wynne LC, Cromwell RL, Matthysse S (eds) The nature of schizophrenia. New approaches to research and treatment. Wiley, New York, pp 25–37

Kindermann SS, Karimi A, Symonds L, Brown GG, Jeste DV (1997) Review of functional magnetic resonance imaging in schizophrenia. Schizophr Res 27: 143–156

Kirkpatrick B, Buchanan RW, McKenney PD, Alphs LD, Carpenter WT (1989) The schedule for the deficit syndrome: an instrument for research in schizophrenia. Psychiatry Res 30: 119–123

Kirkpatrick B, Amador XF, Flaum M, Yale SA, Gorman JM, Carpenter WT, Tohen M, McGlashan T (1996a) The deficit syndrome in the DSM-IV field trial. I. Alcohol and other drug abuse. Schizophr Res 20: 69–77

Kirkpatrick B, Amador XF, Yale SA, Bustillo JR, Buchanan RW, Tohen M (1996b) The deficit syndrome in the DSM-IV field trial. II. Depressive episodes and persecutory beliefs. Schizophr Res 20: 79–90

Kleist K (1930) Zur hirnpathologischen Auffassung der schizophrenen Grundstörungen. Die alogische Grundstörung. Schweizer Arch Neurol Psychiatr 26: 99–102 [English translation in: Cutting J, Shepherd M (eds) (1987) The clinical roots of the schizophrenia concept. Cambridge University Press, Cambridge, pp 75–78]

Klosterkötter J, Schultze-Lutter F, Gross G, Huber G, Steinmeyer EM (1997) Early self-experienced neuropsychological deficits and subsequent schizophrenic diseases: an 8-year average follow-up prospective study. Acta Psychiatr Scand 95: 396–404

Koehler K (1979) First rank symptoms of schizophrenia: questions concerning clinical boundaries. Br J Psychiatry 134: 236–248

**Kraepelin E (1919) Dementia praecox and paraphrenia. Livingstone, Edinburgh

*Kraepelin E (1920) Die Erscheinungsformen des Irreseins. Z Ges Neurol Psychiatr 62: 1–29 [English translation in: Hirsch SR, Shepherd M (eds) (1974) Themes and variations in European psychiatry. Wright, Bristol, pp 7–30]

Kring AM, Kerr SL, Smith DA, Neale JM (1993) Flat affect in schizophrenia does not reflect diminished subjective experience of emotion. J Abnorm Psychol 102: 507–517

Lane A, Kinsella A, Murphy P, Byrne M, Keenan J, Colgan K, Cassidy B, Sheppard N, Horgan R, Waddington JL, Larkin C, O'Callaghan E (1997) The anthropometric assessment of dysmorphic features in schizophrenia as an index of its developmental origins. Psychol Med 27: 1155–1164

Leff J, Tress K, Edwards B (1988) The clinical course of depression in schizophrenia. Schizophr Res 1: 25–30

Leff J, Sartorius N, Jablensky A, Korten A, Ernberg G (1992) The International Pilot Study of Schizophrenia: five-year follow-up findings. Psychol Med 22: 131–145

Lenzenweger MF, Dworkin RH (1996) The dimensions of schizophrenia phenomenology. Not one or two, at least three, perhaps four. Br J Psychiatry 168: 432–440

Leonhard K (1957) Aufteilung der endogenen Psychosen. Akademie-Verlag, Berlin [English translation: The classification of endogenous psychoses, 5th edn Irvington and Wiley, New York, 1979

*Leonhard K (1995) Aufteilung der endogenen Psychosen und ihre differenzierte Ätiologie, 7th revised edn. Thieme Stuttgart

Leudar I, Thomas P, Johnston M (1994) Self-monitoring in speech production: effects of verbal hallucinations and negative symptoms. Psychol Med 24: 749–761

Lewine RJ (1981) Sex differences in schizophrenia – timing or subtypes? Psychol Bull 90: 432–444

Liddle PF (1987) The symptoms of chronic schizophrenia: a re-examination of the positive-negative dichotomy. Br J Psychiatry 151: 145–151

Liddle PF (1995) Inner connections within domain of dementia praecox: role of supervisory mental processes in schizophrenia. Eur Arch Psychiatry Clin Neurisci 245: 210–215

Liddle PF (1997) Dynamic neuroimaging with PET, SPET or fMRI. Int Rev Psychiatry 9: 331–337

Lindenmayer JP, Grochwski S, Hyman RB (1995) Five factor model of schizophrenia: replication across samples. Schizophr Res 14: 229–234

Lindström E, von Knorring L (1994) Symptoms in schizophrenic syndromes in relation to age, sex, duration of illness and number of previous hospitalizations. Acta Psychiatr Scand 89: 274–278

Lombroso C (1887) Genie und Irrsinn. Reclam, Leipzig

Lorr M, Klett CJ, McNair DM (1963) Syndromes of psychosis. Pergamon, New York

Lund CE, Mortimer AM, Rogers D, McKenna P (1991) Motor, volitional and behavioural disorders in schizophrenia. 1. Assessment using the modified Rogers scale. Br J Psychiatry 158: 323–327

Malik SB, Ahmed M, Bashir A, Choudhry TM (1990) Schneider's first-rank symptoms of schizophrenia: prevalence and diagnostic use. A study from Pakistan. Br J Psychiatry 156: 109–111

Manschreck TC, Maher BA, Milavetz JJ, Ames D, Weisstein CC, Schneyer ML (1988) Semantic priming in thought disordered schizophrenic patients. Schizophr Res 1: 61–66

Manton KG, Woodbury MA, Tolley HD (1994a) Statistical applications using fuzzy sets. Wiley, New York

Manton KG, Korten A, Woodbury MA, Anker M, Jablensky A (1994b) Symptom profiles of psychiatric disorders based on graded disease classes: an illustration using data from the WHO International Pilot Study of Schizophrenia. Psychol Med 24: 133–144

Marsh L, Harris D, Lim KO, Beal M, Hoff AL, Minn K, Csernansky JG, DeMent S, Faustman WO, Sullivan E, Pfefferbaum A (1997) Structural magnetic resonance imaging abnormalities in men with severe chronic schizophrenia and an early age at clinical onset. Arch Gen Psychiatry 54: 1104–1112

Matussek P (1952) Untersuchungen über die Wahnwahrnehmung. Arch Psychiatr Z Neurol 189: 279–318 [English translation in: Cutting J, Shepherd M (eds) (1987) The clinical roots of the schizophrenia concept. Cambridge University Press, Cambridge, pp 89–103]

McCarley RW, O'Donnell BF, Niznikiewicz MA, Salisbury DF, Potts GF, Hirayasu Y, Nestor PG, Shenton ME (1997) Update on electrophysiology in schizophrenia. Int Rev Psychiatry 9: 373–386

McGorry PD, Singh BS, Connell S, McKenzie D, van Riel RJ, Copolov DL (1992) Diagnostic concordance in functional psychosis revisited: a study of inter-relationships between alternative concepts of psychotic disorder. Psychol Med 22: 367–378

McGrath JJ, van Os J, Hoyos C, Jones PB, Harvey I, Murray RM (1995) Minor physical anomalies in psychoses: associations with clinical and putative aetiological variables. Schizophr Res 18: 9–20

McGuire PK, Silbersweig DA, Wright I, Murray RM, Frackowiak RSJ, Frith CD (1996) The neural correlates of inner speech and auditory verbal imagery in schizophrenia: relationship to auditory verbal hallucinations. Br J Psychiatry 169: 148–159

**Meehl PE (1990) Toward an integrated theory of schizotaxia, schizotypy, and schizophrenia. J Pers Dis 4: 1–99

Mellor CS (1970) First rank symptoms of schizophrenia. Br J Psychiatry 117: 15–23

Meyer A (1958) Psychobiology: a science of man. Thomas, Springfield

Minas IH, Stuart GW, Klimidis S, Jackson HJ, Singh BS, Copolov DL (1992) Positive and negative symptoms in the psychoses: multidimensional scaling of SAPS and SANS items. Schizophr Res 8: 143–156

Minkowski E (1927) La schizophrénie. Payot, Paris

Mozley PD, Gur RE, Resnick SM, Shtasel DL, Richards J, Kohn M, Grossman R, Herman G, Gur RC (1994) Magnetic resonance imaging in schizophrenia: relationship with clinical measures. Schizophr Res 12: 195–203

Munk-Jørgensen P, Mortensen PB, Machón RA (1991) Hospitalization patterns in schizophrenia. A 13-year follow-up. Schizophr Res 4: 1–9

Ndetei DM, Vadher A (1984) A cross-cultural study of the frequencies of Schneider's first rank symptoms of schizophrenia. Acta Psychiatr Scand 70: 540–544

Nuechterlein KH, Dawson ME, Ventura J, Fogelson D, Gitlin M, Mintz J (1990) Testing vulnerability models: stability of potential vulnerability indicators across clinical states. In: Häfner H, Gattaz WF (eds) Search for the causes of schizophrenia, vol II. Springer, Berlin Heidelberg New York, pp 178–191

O'Grady JC (1990) The prevalence and diagnostic significance of Schneiderian first-rank symptoms in a random sample of acute psychiatric in-patients. Br J Psychiatry 156: 496–500

Parnas J, Schulsinger H (1986) Continuity of formal thought disorder from childhood to adulthood in a high-risk sample. Acta Psychiatr Scand 74: 246–251

Peralta V, Cuesta MJ, de Leon J (1994) An empirical analysis of latent structures underlying schizophrenic symptoms: a four-syndrome model. Biol Psychiatry 36: 726–736

Pinals DA, Malhotra AK, Missar CD, Pickar D, Breier A (1996) Lack of gender differences in neuroleptic response in patients with schizophrenia. Schizophr Res 22: 215–222

Radhakrishnan J, Mathew K, Richard J, Verghese A (1983) Schneider's first rank symptoms – prevalence, diagnostic use and prognostic implications. Br J Psychiatry 142: 557–559

Rado S (1956) Psychoanalysis of behavior. Grune and Stratton, New York

Remschmidt HE, Schulz E, Martin M, Warnke A, Trott GE (1994) Childhood-onset schizophrenia: history of the concept and studies. Schizophr Bull 20: 727–745

Rey ER, Bailer J, Bräuer W, Händel M, Laubenstein D, Stein A (1994) Stability trends and longitudinal correlations of negative and positive symptoms within a three-year follow-up of initially hospitalized schizophrenics. Acta Psychiatr Scand 90: 405–412

Rogers D (1991) Catatonia: a contemporary approach. J Neuropsychiatry 3: 334–340

Rossi A, Serio A, Stratta P, Petruzzi C, Schiazza G, Mancini F, Casacchia M (1994) Planum temporale asymmetry and thought disorder in schizophrenia. Schizophr Res 12: 1–7

Rümke HC (1941) Het Kernsymptoom der Schizophrenie en het Praecoxgevoel. Studies en Voordrachten over Psychiatrie. Scheltema & Holkema, Amsterdam, pp 53–58

Russel AT (1994) The clinical presentation of childhood-onset schizophrenia. Schizophr Bull 20: 634–646

Sabri O, Erkwoh R, Schreckenberger M, Owega A, Sass H, Buell U (1997) Correlation of positive symptoms exclusively to

hyperperfusion or hypoperfusion of cerebral cortex in never-treated schizophrenics. Lancet 349: 1735–1739

Sarbin TR (1990) Toward the obsolescence of the schizophrenia hypothesis. J Mind Behav 11: 259–284

Salokangas RKR (1997) Structure of schizophrenic symptomatology and its changes over time: prospective factor-analytical study. Acta Psychiatr Scand 95: 32–39

Scharfetter C (1983) Schizophrene Menschen. Urban and Schwarzenberg, Munich

*Schneider K (1950) Klinische Psychopathologie, 8th edn. Thieme, Stuttgart [English translation by Hamilton MW, Anderson EW (1959) Clinical psychopathology. Grune and Stratton, New York]

Schneider K (1957) Primäre und sekundäre Symptome bei der Schizophrenie. Fortschr Neurol Psychiatr 25: 487–490

Schröder J, Buchsbaum MS, Siegel BV, Geider FJ, Lohr J, Tang C, Wu J, Potkin SG (1996) Cerebral metabolic activity correlates of subsyndromes in chronic schizophrenia. Schizophr Res 19: 41–53

Scottisch Schizophrenia Research Group (1992) The Scottish first episode schizophrenia study. VIII. Five-year follow-up: clinical and psychosocial findings. Br J Psychiatry 161: 496–500

Sheldrick C, Jablensky A, Sartorius N, Shepherd M (1977) Schizophrenia succeeded by affective illness: catamnestic study and statistical enquiry. Psychol Med 7: 619–624

Shepherd M, Watt D, Falloon I, Smeeton N (1989) The natural history of schizophrenia: a five-year follow-up study of outcome and prediction in a representative sample of schizophrenics. Cambridge University Press, Cambridge

Shtasel DL, Gur RE, Gallacher F, Heimberg C, Gur RC (1992) Gender differences in the clinical expression of schizophrenia. Schizophr Res 7: 225–231

Silbersweig DA, Stern E, Frith C, Cahill C, Holmes A, Grootoonk S, Seaward J, McKenna P, Chua SE, Schnorr L, Jones T, Frackowiak RSJ (1995) A functional neuroanatomy of hallucinations in schizophrenia. Nature 378: 176–179

Snezhnevskij AV (1975) Mesto kliniki v issledovanii prirodi shizofrenii [The role of clinical investigation in the study of the nature of schizophrenia]. Zh nevropatol psihiatr 75: 1340–1347

Sokal RR (1974) Classification: purposes, principles, progress, prospects. Science 185: 115–123

Spitzer M (1997) A cognitive neuroscience view of schizophrenic thought disorder. Schizophr Bull 23: 29–50

Spitzer RL, Endicott J, Robins E (1978) Research diagnostic criteria. Rationale and reliability. Arch Gen Psychiatry 35: 773–782

Starkstein SE, Petracca G, Teson A, Chemerinski E, Merello M, Migliorelli R, Leiguarda R (1996) Catatonia in depression: prevalence, clinical correlates, and validation of a scale. J Neurol Neurosurg Psychiatr 60: 326–332

Störring G (1939) Wesen und Bedeutung des Symptoms der Ratlosigkeit bei psychischen Erkrankungen. Thieme, Leipzig, pp 65–69 [English translation in: Cutting J, Shepherd M (eds) (1987) The clinical roots of the schizophrenia concept. Cambridge University Press, Cambridge, pp 79–82]

Stransky E (1904) Zur Auffassung gewisser Symptome der Dementia praecox. Neurol Centralblatt 23: 1137–1143 [English translation in: Cutting J, Shepherd M (eds) (1987) The clinical roots of the schizophrenia concept. Cambridge University Press, Cambridge, pp 37–41]

Strauss ME (1993) Relations of symptoms to cognitive deficits in schizophrenia. Schizophr Bull 19: 215–231

Strömgren E (1992) The concept of schizophrenia: the conflict between nosological and symptomatological aspects. J Psychiatr Res 26: 237–246

Süllwold L, Huber G (1986) Schizophrene Basisstörungen. Springer, Berlin Heidelberg New York

Tandon R, Greden JF (1987) Schneiderian first rank symptoms: reconfirmation of high specificity for schizophrenia. Acta Psychiatr Scand 75: 392–396

Teggin AF, Elk R, Ben-Arie O, Gillis LS (1985) A comparison of Catego class "S" schizophrenia in three ethnic groups: psychiatric manifestations. Br J Psychiatry 147: 683–687

Thaker G, Adami H, Moran M, Lahti A, Cassady S (1993) Psychiatric illnesses in families of subjects with schizophrenia-spectrum personality disorders: high morbidity risks for unspecified functional psychoses and schizophrenia. Am J Psychiatry 150: 66–71

Thomsen PH (1996) Schiziohrenia with childhood and adolescent onset – a nationwide register-based study. Acta Psychiatr Scand 94: 187–193

Toomey R, Kremen WS, Simpson JC, Samson JA, Seidman LJ, Lyons MJ, Faraone SV, Tsuang MT (1997) Revisiting the factor structure for positive and negative symptoms: evidence from a large heterogeneous group of psychiatric patients. Am J Psychiatry 154: 371–377

Torgersen S, Onstad S, Skre I, Edvardsen J, Kringlen E (1993) "True" schizotypal personality disorder: a study of co-twins and relatives of schizophrenic probands. Am J Psychiatry 150: 1661–1667

Trimble MR (1990) First-rank symptoms of Schneider. A new perspective? Br J Psychiatry 156: 195–200

Turetsky B, Cowell PE, Gur RC, Grossman RI, Shtasel DL, Gur RE (1995) Frontal and temporal lobe brain volumes in schizophrenia. Arch Gen Psychiatry 52: 1061–1070

van Praag HM (1993) "Make-believes" in psychiatry or the perils of progress. Brunner/Mazel, New York

Vinogradov S, Ober BA, Shenaut GK (1992) Semantic priming of word pronunciation and lexical decision in schizophrenia. Schizophr Res 8: 171–181

Waldrop MF, Halverson CF (1971) Minor physical anomalies and hyperactive behaviour in young children. In: Helmmuth J (ed) Exceptional infant. Studies in abnormalities. Brunner/Mazel, New York

Waldo MC, Carey G, Myles-Worsley M, Cawthra E, Adler LE, Nagamoto HT, Wenedr P, Byerley W, Plaetke R, Freedman R (1991) Co-distribution of sensory gating deficit and schizophrenia in multi-affected families. Psychiatry Res 39: 257–268

Walker E, Lewine RJ (1988) The positive/negative symptom distinction in schizophrenia. Validity and etiological relevance. Schizophr Res 1: 315–328

*Weinberger DR, Berman KF, Zec RF (1986) Physiologic dysfunction of dorsolateral prefrontal cortex in schizophrenia. I. Regional cerebral blood flow evidence. Arch Gen Psychiatry 43: 114–124

Wernicke C (1900) Grundriss der Psychiatrie in klinischen Vorlesungen. Thieme, Leipzig

Whalley LJ, Christie JE, Brown S, Arbuthnott GW (1984) Schneider's first-rank symptoms of schizophrenia. An association with increased growth hormone response to apomorphine. Arch Gen Psychiatry 41: 1040–1043

White L, Harvey PD, Opler L, Lindenmayer JP and the PANSS Study Group (1997) Empirical assessment of the factorial structure of clinical symptoms in schizophrenia. Psychopathology 30: 263–274

WHO (1973) The international pilot study of schizophrenia, vol 1. World Health Organization, Geneva

*WHO (1979) Schizophrenia. An international follow-up study. Wiley, Chichester

WHO (1992) The ICD-10 classification of mental and behavioural disorders. Clinical descriptions and diagnostic guidelines. World Health Organization, Geneva

WHO (1993) The ICD-10 classification of mental and behavioural disorders. Diagnostic criteria for research. World Health Organization, Geneva

Wible CG, Shenton ME, Hokama H, Kikinis R, Jolesz FA, Metcalf D, McCarley RW (1995) Prefrontal cortex and schizophrenia. Arch Gen Psychiatry 52: 279–288

Wing JK (1961) A simple and reliable subclassification of chronic schizophrenia. J Ment Sci 107: 862–875

*Wing JK, Brown GW (1970) Institutionalism and schizophrenia. Cambridge University Press, London

*Wing JK, Cooper JE, Sartorius N (1974) Measurement and classification of psychiatric symptoms. Cambridge University Press, London

Wing JK, Babor T, Brugha T, Burke J, Cooper JE, Giel R, Jablensky A, Regier D, Sartorius N (1990) SCAN Schedules for clinical assessment in neuropsychiatry. Arch Gen Psychiatry 47: 589–593

Wing L (1996) The autistic spectrum. Constable, London

Woodbury MA, Manton KG, Tolley HD (1994) A general model for statistical analysis using fuzzy sets: sufficient conditions for identifiability and statistical properties. Information Sci 1: 149–180

Zarrouk EA (1978) The usefulness of first rank symptoms in the diagnosis of schizophrenia in a Saudi Arabian population. Br J Psychiatry 132: 571–573

Zubin J, Spring B (1977) Vulnerability – a new view of schizophrenia. J Abnorm Psychol 86: 103–126

CHAPTER

2

J. Leff

Epidemiology of Schizophrenic Disorders

1 Studies of Incidence – Benefits and Problems

The two main measures of rates used in epidemiological studies are incidence and prevalence. The latter is determined both by the incidence of a disorder and by its course, which itself is a product of natural history and available treatments. Hence the interpretation of differences in prevalence rates is far from simple. Consequently, in this chapter we shall concentrate on what has been learned from incidence rates.

The principle value of comparisons of incidence rates is the possible identification of aetiological factors. The aetiology of schizophrenia remains obscure despite decades of research focused on the problem, so that studies of the incidence of the disorder continue to be a prime strategy. We have deliberately referred to schizophrenia as a *disorder* because its status as a disease has yet to be firmly established. There are boundary problems on at least two frontiers: on the one hand with the affective psychoses and on the other with schizophrenia spectrum disorders. Some psychiatrists point to the considerable overlap in symptoms and inheritance of schizophrenia and the affective psychoses and support the unitary psychosis concept, abolishing the Kraepelinian split between dementia praecox and manic-depressive psychoses. Among those who accept schizophrenia as a discrete entity, there are differing views about how narrow or broad the definition should be. There is also increasing support for the heterogeneity of the clinical concept, echoing Bleuler's formulation of a group of disorders under the title of the schizophrenias. The lack of agreement on the definition of schizophrenia has undermined much of the epidemiological research in the past. It has only been the advent of accepted international diagnostic systems, particularly the Diagnostic and Statistical Manual (DSM) and the International Classification of Diseases (ICD), that has enabled the field to progress. In fact, these systems have established a dialogue between epidemiology and classification. Internationally agreed definitions allow the comparison of incidence rates across countries, which in turn reflects back on the definitions, indicating modifications and adjustments.

2 International Studies of Schizophrenia

The first international study of psychoses managed by the World Health Organisation (WHO) was the International Pilot Study of Schizophrenia (IPSS). This pioneering research broke new ground in equipping a group of psychiatrists from nine different countries with the ability to use the same semi-structured clinical interview, the Present State Examination (PSE) (Wing et al. 1974). Data from the PSE were processed by a computerised program, CATEGO, incorporating an algorithm which assigned each patient to a diagnostic class. The application of this reference classification showed that there was close agreement between psychiatrists from seven of the centres in their diagnosis of schizophrenia, with a heavy reliance on Schneider's (1957) first-rank symptoms. The two outliers were Moscow and Washington (World Health Organisation 1973; Leff 1977). Subsequently, diagnostic practices in the United States swung dramatically from a broad concept of schizophrenia to a very narrow one, which is embodied in DSM-III.

The demonstration by the IPSS of reasonable international agreement on the diagnosis of schizophrenia and the value of a computerised reference classification enabled the WHO to proceed with the next epidemiological study, the Determinants of Outcome of Severe Mental Disorders (DOSMD). Whereas the IPSS was designed to compare series of consecutive admissions or contacts with the services in the participating centres, DOSMD had the much more ambitious aim of collecting all psychotic patients in the catchment areas of the centres making contact with psychiatric services during the study period. Thus the samples were intended to be truly epidemiological, allowing a comparison of first contact rates for schizophrenia across centres. One of the main points of interest was to compare rates between developing and developed countries in view of major social and cultural differences between them which might be implicated in aetiology (Sartorius et al. 1986). Consequently, two of the participating centres were in developing countries, Ibadan in Nigeria and Chandigarh in India. Chandigarh was divided into a city, which is relatively developed, and the surrounding countryside, in which a traditional agricultural lifestyle still thrives. The problems of establishing a comprehensive case-finding network in a developing centre are formidable and need to include traditional healers. In the event, it was considered that too much leakage of cases occurred in Ibadan for the data to be included. The results from the other centres are displayed in Table 1.

It is evident that the rates for narrowly defined, Schneiderian schizophrenia (S+ in the CATEGO classification) do not differ across centres (Jablensky et al. 1992). This remarkable finding has attracted much interest, since no other disease has an invariant incidence across the world. This finding suggests that the aetiological factors for Schneiderian schizophrenia

Table 1. First contact rates of schizophrenia per 100,000 population according to different diagnostic definitions (adapted from Jablensky et al. 1992)

Country	Broad schizophrenia (n)*	S+ (n)**	Non-S+ (n)***
Aarhus	15	7	8
Chandigarh rural	42	11	31
Chandigarh urban	35	9	26
Dublin	22	9	13
Honolulu	16	9	7
Moscow	28	12	16
Nagasaki	20	10	10
Nottingham	22	14	8
χ^2	39.8	7.7	61.8

S+, patients with at least one of Schneiders' first-rank symptoms; non S+, patients without Schneiders' symptoms but with other non-affective delusions and/or hallucinations.
*$p < 0.00001$; **not significant; ***$p < 0.000001$; $df = 7$.

are likely to be independent of the social and cultural environment. By contrast, the rates for broadly defined schizophrenia, which include S+ and non-S+ classes, differ significantly. It is curious that in the publications from this study there has been no analysis reported for the non-S+ rates separately from the S+ rates. This analysis is shown in Table 1, from which it can be seen that the greatest variability across centres occurs with this category. There is a fourfold difference between the lowest rate in Aarhus, Denmark and the highest rate in the rural area of Chandigarh. It is likely that variation in the operation of environmental factors between centres accounts for this striking difference in rates. Hence this study illustrates the way in which an epidemiological study can contribute to the diagnostics of schizophrenia. The findings suggest that the broad rubric of schizophrenia encompasses two entities with different aetiologies, Schneiderian schizophrenia and another group including paranoid schizophrenia without first-rank symptoms and catatonia.

3 Environmental Factors as Aetiological Agents and Gender Effects

The only intrinsic aetiological factor that has been identified as important in schizophrenia is inheritance (see Chap. 3, this volume). A number of environmental factors have been suggested as possibly aetiological, but the evidence for each of them remains inconclusive or controversial. Most of the evidence derives from epidemiological studies, which will be reviewed here. Hypotheses concerning physical factors have focused on oestrogens, viral infections in utero, anoxic brain damage at birth and maternal starvation. A protective effect of oestrogens has been postulated as the explanation of the gender difference in the age of onset of schizophrenia. It has generally emerged from epidemiological studies of the incidence of schizophrenia that affected men make contact with the psychiatric services at an earlier age than women, on average about 5 years earlier, although some carefully conducted studies have failed to find this difference (e.g. Iacono and Beiser 1992; Salokangas 1993; Jablensky and Cole 1997). Hambrecht et al. (1992) list the methodological problems of obtaining a true as opposed to a treated incidence rate and suggest that some of the contradictory findings are due to differing proportions of affected people in the community receiving treatment by the psychiatric services. They identify a gender bias in the diagnosis of early cases of schizophrenia as another confounder. For example, Löffler et al. (1994), using the Danish national psychiatric case register, found that, in comparison with PSE-CATEGO diagnoses, females were less often diagnosed as schizophrenic than males.

Häfner et al. (1991a) conducted a large-scale epidemiological study in a catchment area in Germany of 1.5 million people. They collected a sample of 392 people over 2 years who were between the ages of 12 and 59 and who were admitted for the first time with a diagnosis of schizophrenia. Applying a broad definition based on ICD categories, the cumulative incidence rates are displayed in Fig. 1.

The graphs for men and women show clearly the earlier onset in men. However, there is a second peak of onsets in women between the ages of 45 and 59, resulting in an equal lifetime incidence for both genders. Hambrecht et al. (1992) point out that, because women catch up with men only after the age of 45, studies that are confined to subjects younger than this will inevitably find a higher incidence in men. Iacono and Beiser (1992) found no new cases of first-onset schizophrenia after the age of 45, which they attribute to the exclusion of delusional psychosis from their study. Generally, European researchers tend to include late-onset psychoses with predominantly para-

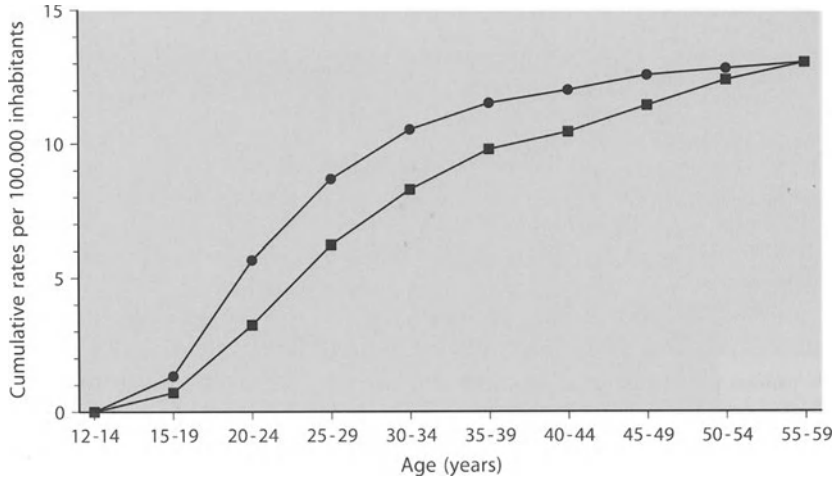


Fig. 1. Cumulative incidence rates for schizophrenia (broad definition; ICD 295, 297, 298.3 and 298.4). *Solid line*, males (patients in 2 years, 187; total population, 707,905; total rate, 13.21%); *dashed line*, females (patients in 2 years, 205; total population, 780,300; total rate, 13.14%). The source of data was a representative first-admission sample (1987/1989). The catchment area was Mannheim, Heidelberg, the Rhine-Neckar district and the Eastern Palatinate (Häfner et al. 1991b)

noid delusions in their series, which increases the incidence rate in women.

Hambrecht et al. (1992) have postulated that the delayed early onset in women is due to a neuromodulatory effect of oestrogens on the sensitivity of dopamine-2 receptors in the brain. This would also account for the late rise in incidence for women after the menopause. Häfner et al. (1991b) tested this hypothesis in animal experiments and found, as predicted, that long-term exposure to 17-β-oestradiol-2 reduced dopamine induced behaviour. Post-mortem striatal binding experiments with [^{3}H]sulpiride showed reduction of D2 receptor sensitivity in the brain as a result of the exposure to oestradiol. These findings are of great interest, but clearly need to be replicated.

A recent study by Jablensky and Cole (1997) merits presentation in some detail because it involves analysis of one of the largest data sets available. This comprises 1431 individuals with a first onset of schizophrenia collected for the WHO ten-country study of the incidence of severe mental disorders (Jablensky et al. 1992). A sequence of log linear models was fitted to the data to determine the contribution of gender, marital status, family history of mental illness, premorbid personality and setting (developed versus developing countries) on the age of onset. The results showed that marriage had a much larger effect on age at onset than gender. On average, marriage delayed the onset by 5 years for women and more than 8 years for men. In one of the log linear models, the effect of gender was replaced by the effects of the interaction between gender and marital status, and gender and centre. This apparently protective effect of marriage on the onset of schizophrenia is liable to a variety of interpretations, but it does throw into question biological explanations for the link between gender and age of onset.

4 Season of Birth Effect and Viruses as Aetiological Agents

A possible pathogenic environmental factor that has attracted interest as a result of epidemiological studies is the influenza virus. The hypothesis that this might be involved in the aetiology of schizophrenia was engendered by the uneven distribution throughout the year of the births of people who later developed schizophrenia. There has been an interest in the months of birth of schizophrenic patients since the 1930s, partly because these were readily available from the records of psychiatric hospitals. One of the earliest statistical analyses of a number of data sets was conducted by Barry and Barry (1961). They emphasised the importance of multiple independent studies with geographically separated samples because of the possible effect of local factors on the prevalence of mental illness, the policy on hospital admissions and the monthly distribution of births in the general population. Since their pioneering study, a substantial body of research has been built up which satisfies their recommendations. Studies have been conducted in many different countries, including some in the southern as well as the northern hemisphere, with the same excess appearing in the winter months. The most recent review has been that by Bradbury and Miller (1985), which demonstrates that this phenomenon is one of the most widely replicated findings in the epidemiological research on schizophrenia. However, the explanation for it remains elusive. Lewis

(1989) suggested that it might be an artefact resulting from the fact that more patients born in the early part of the year are likely to be included in a sample because the risk for schizophrenia increases with age. However, Rodrigo et al. (1992) carried out a study of season of birth controlling for the age–incidence effect and still found an excess for the months of January and December. The link between season of birth and influenza epidemics is at best associational, showing that women who were pregnant at the time of a major epidemic are more likely to produce offspring who later develop schizophrenia than those who were not exposed.

Few studies present unequivocal evidence that the women in question actually suffered an influenzal infection. One of the most solid is the analysis by Crow and Done (1992) of a prospective sample of children born to mothers in the United Kingdom in March 1958. Fortuitously, these children would have been in the second trimester of pregnancy during the 1957 influenza epidemic. Crow and Done failed to find an excess of schizophrenia in the offspring of those mothers who reported having had influenza at the time. McGrath and Castle (1995) point out the unreliability of mothers' recall of whether they had had influenza some months previously and recommend that future studies should document aetiological evidence of an antibody response to the influenza virus. In the absence of this vital evidence, Mednick et al. (1988) caution that the association could be mediated by over-the-counter drugs taken by the mothers to relieve symptoms of influenza or the high fever developed.

Two factors have been identified which modify the season of birth effect: gender and place of birth. McGrath and Castle (1995) draw attention to the fact that, in five out of the six studies in which the effect of gender was examined, the positive association found was mainly or exclusively in females. This finding has been repeated in a study from Japan (Kunugi et al. 1995) which was not included in the review by McGrath and Castle. These results suggest that female foetuses are more vulnerable than males to whatever pathogen is operating during pregnancy. The other factor concerns urban as opposed to rural births. Machón et al. (1983) were among the first investigators to link the season of birth effect to place of birth. They studied a sample of children at high risk for schizophrenia because they were born to mothers with severe chronic schizophrenia. The proportions of the high-risk children who developed schizophrenia were 23.3% for winter urban births, 8.4% for non-winter urban births, 0% for winter rural births and 6.3% for non-winter rural births. The association between winter births and urban place of birth was significant. This finding has been replicated by Takei et al. (1995) for England and Wales and O'Callaghan et al. (1995) for Ireland. One interpretation is that infectious diseases are transmitted more easily in the dense populations of towns than in the countryside.

In order to investigate this possibility directly, Verdoux et al. (1997) used data from public departments of psychiatry in metropolitan France. They related the month of birth of 4139 schizophrenic patients to the density of population in the administrative region in which they were born. Patients born in the five most densely populated areas were 1.21 times more likely to be born during the winter months than those born in other areas. This effect was not found to differ between males and females. The authors acknowledge that the specific environmental risk factor or factors underlying the winter birth excess in urban areas have not yet been identified, but consider that sociological factors are implausible because they would not preferentially affect subjects born in the winter. The association between the urban excess of schizophrenic patients making a first contact with the psychiatric services and sociological factors was one of the first observations to interest psychiatric epidemiologists. It has been neglected for many years, since the explanation was generally believed to have been established, but there has been a recent revival of interest in this topic.

5 Distribution of Schizophrenia in Cities

The classic work on the ecology of schizophrenia was conducted by Faris and Dunham (1939) in the city of Chicago. They found that the distribution of mental hospital admissions for schizophrenia was not random, but was concentrated in the "zone in transition", an inner city ring immediately adjacent to the centre in which there was a high level of social deprivation. Two contrasting explanations have been advanced for this finding, which has been extensively replicated in developed countries. The social breeder hypothesis states that poor social conditions lead to a higher incidence of severe mental illness, while the social drift hypothesis, first proposed by Myerson (1940), attributes the accumulation of severely mentally ill in inner cities to a descent down the social scale. After decades of controversy about the evidence for and against each of these hypotheses, the argument appeared to be clinched by two studies with unequivocal findings. Dunham (1965) conducted a study in the city of Detroit in which he documented the length of residence of patients making a first contact with the psychiatric services. He focused on two areas, one with a high rate of schizophrenia, the other with a low rate.

He found that the excess rate was entirely contributed to by patients who had moved into the area within the past 5 years. These mobile people were mostly born in places other than large cities and were of a lower social class than the residentially stable patients. In fact, the mobile patients were responsible for the excess in the lowest social class found by Dunham in his sample. Dunham's study of geographical drift was complemented by a study of social drift by Golberg and Morrison (1963). They compared the social class of young men admitted to a mental hospital for the first time with the social class of their fathers at a comparable age. They found the usual excess in social class V for the young men, whereas the social class distribution of their fathers did not differ from that of the general population. They concluded that the young men had started to drift down the social scale before they developed overt signs of schizophrenia, usually in early adolescence.

These two studies were generally accepted to have resolved the controversy, and little work was done in this area for many years. However, it is rash to assume in psychiatry that one or two studies can settle the dust on an ideological argument, and contradictory evidence has been appearing lately. Lewis et al. (1992) used data from a cohort of over 49,000 male Swedish conscripts which was linked to the Swedish National Register of Psychiatric Care. Men who were brought up in cities had an incidence of schizophrenia which was 1.65 times that of men who were reared in the countryside. After adjusting for possible confounding factors, including family finances, a family history of mental illness and parental divorce, the odds ratio was reduced slightly to 1.57, which was still highly significant. An urban excess of this magnitude is common in the literature, but in this sample could not be explained by geographical drift, since all the urban patients had been brought up in cities since their childhood. Castle et al. (1993) used a case-control design to investigate the same issue. They utilised the Camberwell Case Register to identify schizophrenic patients making their first contact with psychiatric services and selected the next non-psychotic patient on the register matched for age and sex as a control. Cases were more likely than controls to have been born in an inner city area and to have fathers whose occupations were manual rather than non-manual. These observations contradict the hypotheses of social and geographical drift and suggest that socio-economic deprivation during early life predisposes to the later development of schizophrenia. These findings reopen an area of research and demand further work, but if correct, would implicate some social or physical factor or factors in the urban environment in the aetiology of schizophrenia. Proponents of the viral hypothesis view them as supportive evidence for their theory.

6 Food Intake During Pregnancy

Another environmental factor that has been invoked is maternal nutrition during pregnancy. The opportunity to explore this arose out of a tragic occurrence during the Second World War when a Nazi blockade of Holland was set up. Between 1944 and 1945, famine conditions existed in the west of Holland. Susser and Lin (1992) studied the birth cohorts from these years and compared cohorts exposed to severe food deprivation with those not so exposed. The authors found that birth cohorts exposed to severe food deprivation during the first trimester of pregnancy showed a greater than twofold increase in admissions for schizophrenia, although this was found to apply only to women and not to men. The apparently specific vulnerability of the female foetus to poor maternal nutrition echoes the finding that females rather than males show the season of birth effect. The possible pathogenic effect of poor maternal nutrition requires confirmation, but it is difficult to imagine another situation in which famine was so clearly delimited in time.

7 Is the Incidence of Schizophrenia Falling?

So far we have considered variations in the incidence of schizophrenia which point to possible aetiological factors. However, there have been recent claims that the incidence of schizophrenia generally is falling. If this were correct, it would have enormous implications for the provision of psychiatric services as well as stimulating a search for the possible reasons for the decline in incidence, although no explanation has been proposed by those who have documented the decline other than a suggestion that schizophrenia is becoming a milder disease and hence those affected are less likely to contact psychiatric services. Because of problems of changing methods of ascertaining cases and of assigning diagnoses over time, the most reliable technology for investigating temporal variations in incidence is the case register. Der et al. (1990) reviewed six studies documenting a decreasing first admission or first contact rate for schizophrenia, three from Scotland and one each from Denmark, Australia and New Zealand. They mounted their own investigation using data on admissions obtained from the Mental Health Enquiry for England and Wales. They identified a problem in distinguishing first admissions from readmissions, a common source of error with this type of data (Ni Nullain et al. 1987). Der et al. grouped together schizophrenia, schizoaffective psychosis and

paranoia and found a considerable decrease in first admissions between 1970 and 1986. They concluded that possible artefacts did not entirely account for this decline. However, the Mental Health Enquiry data were derived from psychiatric departments throughout the country and the diagnoses were recorded by hundreds of different people, not all of whom were psychiatrists.

Changing fashions in diagnosis could constitute the major explanation for this finding. Kendell et al. (1993) analysed first-admission rates for schizophrenia and other psychoses in Edinburgh between 1971 and 1989. The inception rate for schizophrenia fell significantly over this period, but they found evidence that the diagnostic criteria had narrowed during this time and also drew attention to the fact that a substantial and changing proportion of recorded first admissions had had previous admissions. For these reasons, they considered that it was not possible to conclude that the incidence of schizophrenia had fallen. Case register diagnoses are the responsibility of a small number of people and hence tend to be much more consistent than hospital admission diagnoses. Since the review by Der et al. (1990), analyses of data from three case registers have been published. Data from the case registers in Nottingham (Harrison et al. 1991) and Camberwell (Castle et al. 1991) in the United Kingdom showed no decline in the incidence of schizophrenia over time, whereas a decline was indicated by data from the nation-wide case register in Denmark (Munk-Jorgensen and Mortensen 1992). An intriguing possibility raised by Castle et al. (1991) is that a decline in the incidence of schizophrenia affecting the majority white population may be obscured by a high incidence in Afro-Caribbeans who migrated to the United Kingdom in the 1950s.

8
High Rate of Schizophrenia in Afro-Caribbeans

A number of studies showed a high incidence of schizophrenia in the Afro-Caribbean population in the United Kingdom (Cochrane 1977; Rwegellera 1977, Dean et al. 1981). These early studies used hospital admission diagnoses or unstructured interviews, so the results were suspect. In recent years, a series of studies has been published which have employed standardised interviewing and diagnostic techniques and have been rigorously conducted. The first of these was sited in the WHO centre in Nottingham (Harrison et al. 1988) and found incidence rates in the Afro-Caribbean population that were remarkably higher than those for native whites. Using ICD-9 or DSM-III criteria, the rates for the age-groups 16–29 and 30–44 were 12 to 13 times higher than among whites. The calculation of the denominator for the two ethnic groups was based on data derived from a survey of heads of households, which may have led to some inaccuracies. In addition, the numerator for Afro-Caribbeans in the age-group 30–44 was only seven. Consequently, doubt has been cast on these extraordinarily high rates. Two subsequent studies used data from the 1991 census to calculate the denominators, which are likely to have been more accurate as this was the first census in which respondents were asked to ascribe their ethnicity. King et al. (1994) found that the incidence of schizophrenia for the Black population in a district in north London was four times that of the whites. They included Africans and Afro-Caribbeans in their Black sample, which numbered 24 in all. They also calculated the incidence for the Asian population, which was even higher than that of the Black group, but was based on only eight patients. Bhugra et al. (1997) also determined incidence rates for a variety of ethnic groups in London, in two districts in the south and east. They found that the rate for Afro-Caribbeans was twice that of the whites, being based on 38 patients. The rate for Asians was based on 24 cases and was slightly but not significantly higher than that for whites. It is noteworthy that most of the Afro-Caribbean patients in the two recent studies were born in the United Kingdom.

These findings are supported by one study outside the United Kingdom, in the Netherlands. Selten and Sijben (1994) used data from the Dutch national register to determine first-admission rates for schizophrenia for four immigrant groups, those from Surinam, the Dutch Antilles, Turkey and Morocco. The age-stratified rates for people from Surinam and the Antilles were from twice to five times the rates for the native-born. Young males from Morocco also had very high rates, but those for the other ethnic groups did not exceed the rates for the native-born. This study is not as rigorous as the two recent British studies, as the researchers relied on hospital diagnoses rather than standardised interviews. Nevertheless, this body of work taken as a whole indicates an exceptionally high incidence of schizophrenia in Afro-Caribbeans who have migrated to Europe or who are born of migrant parents. While other migrant groups have elevated rates, they do not appear to have the same level of susceptibility as Afro-Caribbeans. The fact that these high rates persist in the second generation rules out the stress of the migratory experience as an explanation. Another possibility is a high rate of schizophrenia in the West Indian islands, but two epidemiological studies, in Jamaica (Hickling et al. 1996) and Trinidad (Bhugra et al. 1996) have shown that the rates there are not elevated. Selective migration remains a possible explanation, but is less plausible for the second generation than for the first, and moreover the incidence of mania is also several times higher in the

Afro-Caribbean population in the United Kingdom than among the whites (Leff et al. 1976; Harrison et al. 1988). Hence it would have to be postulated that people of schizoid and of cyclothymic personality were more likely to migrate than the rest of the West Indian population.

The most likely explanation is that factors in the social or physical environment are responsible for the excess rate of schizophrenia, although it will be necessary to identify factors that operate differentially on Afro-Caribbeans as opposed to other ethnic minority groups that do not show such an excess. Substance abuse, particularly cannabis, has been suggested as a cause of the high incidence rate, but there is no evidence that more U.K. Afro-Caribbeans smoke cannabis than U.K. whites. These findings direct attention again to the possibly pathogenic environment of deprived inner cities, as these are where the recent migrants settle. This area is deserving of further research, as it holds out the hope of illuminating the aetiology of schizophrenia.

9 Is Schizoid Personality Related to Schizophrenia?

The hypothesis of selective migration raises the question of the relationship between schizoid personality and schizophrenia. Kraepelin (1913) first introduced the idea that the personality changes which can be detected well before the onset of frank schizophrenic symptoms are likely to be a manifestation of the same causes as the psychosis. In a historical review of the field, Kendler (1985) noted that there have been two conceptual trajectories, one concerned with schizoid traits in relatives of schizophrenic patients, the other with borderline personalities. The latter term has been applied to people who are not necessarily related to schizophrenic patients, but who are deemed to show the fundamental features of chronic schizophrenia without any of the characteristic signs or severe deterioration. Kendler pointed out that there has often been a failure to distinguish between these different concepts, leading to confused thinking and equivocal findings.

Kety et al. (1971), in the Danish adoption study, stated that they were operating on the hypothesis that schizophrenia need not be transmitted as such, but could appear as a broader spectrum of disorders. They coined the term "schizophrenia spectrum", which has been used in subsequent family studies. Spitzer et al. (1979) used data from the Danish study to develop personality inventories for schizotypal personality and borderline personality, which were to be incorporated in the Diagnostic and Statistical Manual. Kendler (1985), in reviewing the criteria employed by Kety and colleagues for "spectrum disorder", identifies some as defining schizoid traits to be expected in relatives of schizophrenic patients, such as social isolation, aloof/cold affect and suspiciousness, and others which characterise borderline personality, such as magical thinking, recurrent illusions and ideas of reference.

Kendler et al. (1995) have attempted to clarify this area by determining the ability of a broad range of schizotypal factors to discriminate between the relatives of patients with various psychiatric conditions and relatives of normal controls. The data they used for the analysis were derived from the Roscommon Family Study, which is being conducted in a relatively poor, rural county in the West of Ireland. Roscommon is one of the counties covered by the Three County Case Register, established in 1973. Patients with psychosis and those with major affective disorder were ascertained from the case register, and age- and sex-matched controls were identified from the electoral register. Interviews were conducted with all probands and their first-degree relatives over the age of 16. The interviewers, who were blind to the diagnosis or control status of the respondents, used the Structured Interview for Schizotypy (SIS), which comprises 14 symptoms and 11 signs.

Factor analysis of the data on 1272 relatives revealed seven distinct factors, indicating the complexity of the construct. The relatives of patients with schizophrenia were most clearly differentiated from the relatives of controls by four factors: odd speech, social dysfunction, avoidant symptoms and negative symptoms (poor rapport, aloofness/coldness, guardedness and odd behaviour). Borderline symptoms failed to discriminate usefully, confirming the distinction made by Kendler (1985) in his historical review. The SIS factors showed some ability to discriminate relatives of schizophrenic patients from relatives of patients with affective disorders, but the distinction was not as clear as from relatives of controls, weakening the specificity with which these factors reflect a familial vulnerability to schizophrenia. Kendler et al. (1995) point out that a major component of the SIS factors were signs as opposed to symptoms, which would inevitably be omitted from self-report questionnaires. This study represents a major advance in illuminating the complex nature of the concept of schizotypy and in identifying the key signs and symptoms to be ascertained in relatives of patients with schizophrenia.

10 Conclusions

This review has highlighted an increasing research interest in using epidemiological techniques to search

for the causes of schizophrenia. This is certainly due in part to the ground-breaking international studies of schizophrenia initiated by WHO. While epidemiology can do no more than indicate potential candidates for aetiological factors, the studies reviewed have thrown up a number of innovative hypotheses which will undoubtedly be explored experimentally in the near future. They have also opened up old controversies about the role of social deprivation, which deserve renewed interest alongside the burgeoning of biological investigations.

11 References

Barry H, Barry H (1961) Season of birth. Arch Gen Psychiatry 5: 292–300

Bhugra D, Hilwig M, Hossein B, Marceau H, Neehall J, Leff J, Mallett R, Der G (1996) First contact incidence rates of schizophrenia in Trinidad and one-year follow-up. Br J Psychiatry 169: 587–592

**Bhugra D, Leff J, Mallett R, Der G, Corridon B, Rudge (1997) Incidence and outcome of schizophrenia in whites, African Caribbeans and Asians in London. Psychol Med 27: 791–798

Bradbury TN, Miller GA (1985) Season of birth in schizophrenia: a review of evidence, methodology, and etiology. Psychol Bull 98: 569–594

Castle D, Wessely S, Der G, Murray RM (1991) The incidence of operationally defined schizophrenia in Camberwell, 1965–84. Br J Psychiatry 159: 790–794

Castle D, Scott K, Wessely S, Murray RM (1993) Does social deprivation during gestation and early life predispose to later schizophrenia? Soc Psychiatry Psychiatr Epidemiol 28: 1–4

Cochrane R (1977) Mental illness in immigrants to England and Wales. Soc Psychiatry 12: 25–35

Crow TJ, Done DJ (1992) Prenatal exposure to influenza does not cause schizophrenia. Br J Psychiatry 161: 390–393

Dean G, Walsh D, Downing H, Shelley E (1981) First admissions of native-born and immigrants to psychiatric hospitals in South-East England 1976. Br J Psychiatry 139: 506–512

*Der G, Gupta S, Murray RM (1990) Is schizophrenia disappearing? Lancet 1: 513–516

Dunham HW (1965) Community and schizophrenia: an epidemiological analysis. Wayne State University Press, Detroit

Faris REL, Dunham HW (1939) Mental disorders in urban areas. Chicago University Press, Chicago

Goldberg EM, Morrison SL (1963) Schizophrenia and social class. Br J Psychiatry 109: 785–802

Häfner H, Maurer K, Löffler W, Reicher-Rössler A (1991a) Schizophrenie und Lebensalter. Nervenarzt 62: 536–548

Häfner H, Behrens S, De Vry J et al (1991b) Warum erkranken Frauen später an Schizophrenie? Erhöhung der Vulnerabilitätsschwelle durch Östrogen. Nervenheilkunde 4: 153–163

Hambrecht M, Maurer K, Häfner H (1992) Evidence for a gender bias in epidemiological studies of schizophrenia. Schizophren Res 8: 223–231

Harrison G, Owens D, Holton A, Nedson D, Boot D (1988) A prospective study of severe mental disorder in Afro-Caribbean patients. Psychol Med 18: 643–657

Harrison G, Cooper JE, Gancarczyk R (1991) Changes in the administrative incidence of schizophrenia. Br J Psychiatry 159: 811–816

Hickling F, Rodgers-Johnson P (1995) The incidence of first contact schizophrenia in Jamaica. Br J Psychiatry 167: 193–196

Iacono WG, Beiser M (1992) Where are the women in first-episode studies of schizophrenia? Schiz Bull 18: 471–480

Jablensky A, Cole SW (1997) Is the earlier age at onset of schizophrenia in males a confounded finding? Results from a cross-cultural investigation. Br J Psychiatry 170: 234–240

**Jablensky A, Sartorius N, Emberg G et al (1992) Schizophrenia: manifestations, incidence, and course in different cultures, a World Health Organization Ten-Country-Study. Psychol Med Monogr Suppl 20: 1–97

Kendell RE, Malcolm DE, Adams W (1993) The problem of detecting changes in the incidence of schizophrenia. Br J Psychiatry 162: 212–218

Kendler KS (1985) Diagnostic approaches to schizotypal personality disorder: a historical perspective. Schiz Bull 11: 538–553

Kendler KS, McGuire M, Gruenberg AM, Walsh D (1995) Schizotypal symptoms and signs in the Roscommon Family Study. Their factor structure and familial relationships with psychotic and affective disorders. Arch Gen Psychiatry 52: 296–303

Kety SS, Rosenthal D, Wender PH, Shulsinger F (1971) Mental illness in the biological and adoptive families of adopted schizophrenics. Am J Psychiatry 128: 302–306

King M, Coker E, Leavey G, Hoare A, Johnson-Sabine E (1994) Incidence of psychotic illness in London: comparison of ethnic groups. Br Med J 39: 1115–1119

Kraepelin E (1913) Psychiatrie, 8th edn, vol 3, part 2. Barth, Leipzig, p 922

Kunugi H, Nanko S, Takei N (1992) Influenza and schizophrenia in Japan. Br J Psychiatry 161: 274–275

Kunugi H, Nanko S, Takei N, Saito K, Hayashi N, Kazamatsuri H (1995) Schizophrenia following in utero exposure to the 1957 influenza epidemics in Japan. Am J Psychiatry 152: 450–452

Leff J (1977) International variations in the diagnosis of psychiatric illness. Br J Psychiatry 131: 329–338

Leff J, Fisher M, Bertelsen A (1976) A cross-national epidemiological study of mania. Br J Psychiatry 129: 428–437

Leff J (1977) International variations in the diagnosis of psychiatric illness. Br J Psychiatry 131: 329–338

*Lewis G, David A, Andreasson S, Allebek P (1992) Schizophrenia and city life. Lancet 340: 137

Lewis MS (1989) Age incidence and schizophrenia. 1. The season of birth controversy. Schizophr Bull 15: 59–73

Löffler W, Fätkenheurer B, Maurer K et al (1994) Validation of Danish case register diagnosis for schizophrenia. Acta Psychiatr Scand 90: 196–203

Machón RA, Mednick SA, Schulsinger F (1983) The interaction of seasonality, place of birth, genetic risk and subsequent risk in a high risk sample. Br J Psychiatry 143: 383–388

*McGrath J, Castle D (1995) Does influenza cause schizophrenia? A five year review. Aust NZ J Psychiatry 29: 23–31

Mednick SA, Machón RA, Huttunen MO, Bonett D (1988) Adult schizophrenia following prenatal exposure to an influenza epidemic. Arch Gen Psychiatry 45: 189–192

Myerson A (1940) Review, mental disorders in urban areas. Am J Psychiatry 96: 995–997

Munk-Jorgensen P, Mortensen PB (1992) Incidence and other aspects of the epidemiology of schizophrenia in Denmark, 1971–87. Br J Psychiatry 161: 489–495

Ni Nullain M, O'Hare A, Walsh D (1987) Incidence of schizophrenia in Ireland. Psychol Med 17: 943–948

O'Callaghan E, Cotter D, Colgan K, Larkin C, Walsh D, Waddington JL (1995) Confinement of winter birth excess in schizophrenia to the urban-born and its gender specificity. Br J Psychiatry 166: 51–54

Rodrigo G, Lusiardo M, Briggs G, Ulmer A (1992) Season of birth of schizophrenics in Mississippi, USA. Acta Psychiatr Scand 86: 327–331

Rwegellera GGC (1977) Psychiatric morbidity among West Africans and West Indians living in London. Psychol Med 7: 317–329

Salokangas RK (1993) First-contact rate for schizophrenia in community psychiatric care. Consideration of the oestrogen hypothesis. Eur Arch Psychiatry Clin Neurosci 242: 337–346

Sartorius N, Jablensky A, Korten A et al (1986) Early manifestations and first-contact incidence of schizophrenia in different cultures. Psychol Med 16: 909–928

Schneider K (1957) Primäre und sekundäre Symptome bei der Schizophrenie. Fortschr Neurol Psychiatrie 25: 487–490

Selten JP, Sijben N (1994) First admission rates for schizophrenia in immigrants to the Netherlands. Soc Psychiatry Psychiatr Epidemiol 29: 71–77

Spitzer RL, Endicott J, Gibbon M (1979) Crossing the border into borderline personality and borderline schizophrenia. Arch Gen Psychiatry 36: 17–24

Susser ES, Lin SP (1992) Schizophrenia after prenatal exposure to the Dutch Hunger Winter of 1944–1945. Arch Gen Psychiatry 49: 983–988

Takei N, Sham P, O'Callaghan E, Glover G, Murray RM (1995) Early risk factors in schizophrenia: place and season of birth. Eur Psychiatry 10: 165–170

Verdoux H, Takei N, de Saint-Mathurin RC, Murray RM, Bourgeois ML (1997) Seasonality of birth in schizophrenia: the effect of regional population density. Schizophren Res 23: 175–180

Wing JK, Cooper JE, Sartorius N (1974) The measurement and classification of psychiatric symptoms. Cambridge University Press, Cambridge

World Health Organisation (1973) The international pilot study of schizophrenia. World Health Organisation, Geneva

CHAPTER

3

H.W. Moises, I.I. Gottesman

Genetics, Risk Factors, and Personality Factors

1 Introduction

It is no longer heresy to state that schizophrenia has a neurobiological basis (Henn 1995) and that the most convincing risk factor is genetic loading for the disorder (Eaton et al. 1995). A recent workshop on schizophrenia recommended that a major focus should be on the search for predisposing genes and that there should be parallel research in many other areas (Barondes et al. 1997). Theorizing about schizophrenia now seems to have abandoned fiction and to have rediscovered facts. The fact is that, of all the currently known risk factors for the disease, genes are the most important: winter birth, stressful life events, unmarried, and low economic status increase the relative risk for schizophrenia by 1.1, 2.7, 2, and 4 respectively, whereas inferred genetic factors lead to relative risks ranging from 10 to 50 (Häfner 1987; Eaton et al. 1995; Jablensky 1995). Today, no serious scientist denies the importance of genetic factors in schizophrenia, while the magnitude of the genetic contribution is still a matter of debate with estimates ranging from low (Torrey 1992) to completely genetic (McGuffin et al. 1994). Heritability was estimated to lie between 70% and 89% (Rao et al. 1981; Risch and Baron 1984; Farmer et al. 1987).

We shall briefly summarize the current state of knowledge about the effect and likely chromosomal locations of schizophrenia risk genes (for reviews, see Propping 1989; Gottesman 1991; Kendler and Diehl 1993; Moises 1995; McGuffin and Owen 1996; Vogel and Motulsky 1996; Moldin and Gottesman 1997; Plomin et al. 1997).

2 Formal Genetics

2.1 Involvement of Genes in the Etiology of Schizophrenia

Whether genes are involved in the etiology of schizophrenia can be answered by controlling for gene dosage as the independent variable and observing the risk for schizophrenia and related disorders as the dependent variable. Genetic similarity and hence the number of risk genes for schizophrenia (gene dosage) increases from biologically unrelated (adoption studies) to biologically related individuals of a schizophrenic proband (family studies) and reaches its maximum in monozygotic twins (twin studies).

2.1.1 Adoption Studies

Adoption studies (Heston 1966; Tienari 1991; Kendler et al. 1994; Kety et al. 1994) are especially valuable to examine separately genetic and environmental factors. Extensive adoption studies in Denmark observed a significantly greater prevalence of chronic schizophrenia and spectrum disorders in biological relatives of adoptees with chronic schizophrenia compared to controls. Interestingly, the Finnish study found that most of the psychotics lived in disturbed adoptive families, while similar family situations in the control group did not lead to psychoses (Tienari et al. 1994). The findings are consistent with genetic control of sensitivity to the environment (Wahlberg et al. 1997).

2.1.2 Family Studies

The risk for developing schizophrenia increases with the degree of genetic similarity to a schizophrenic relative (see Fig. 1), severity of illness (Schulz 1932; Kallmann 1938), and age of onset (Kay 1963). The sex difference in age of onset is smaller in cases with a high genetic load (Häfner and an der Heiden 1997). All these factors can be interpreted as the result of a gene-dosage effect. A larger number of schizophrenia risk genes result in earlier onset, reduced sex differences, a more severe course of the disease, and a higher risk for family members.

2.1.3 Twin Studies

The similarities in gene dosage between members of monozygotic (100% identical) and dizygotic twin pairs (50% on average) can be used to estimate the "relative powers of nature and nurture" (Galton 1875), assuming that there is no systematic difference in environmental factors between these two twin types. A review of the six most recent twin studies reveals a median concordance rate for monozygotic twins of 48%, approximately three times the corresponding rate for dizygotic twins (17%), suggesting that genes play a significant role in the pathogenesis of schizophrenia (Gottesman 1991). In summary, adoption, family, and twin studies demonstrate that genes play an important role in the pathogenesis of schizophrenia.

2.2 Environmental Risk Factors

A monozygotic concordance rate of 48% strongly implicates a role for nongenetic factors. Nearly one

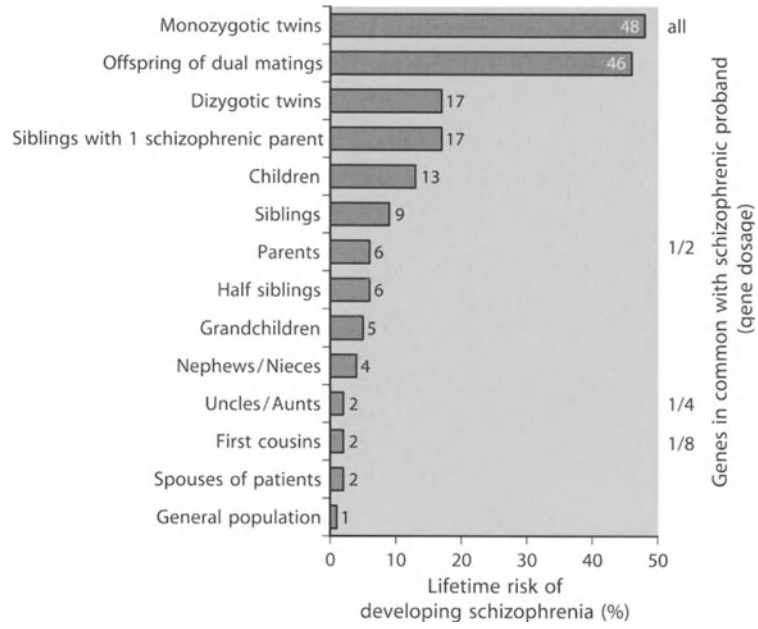


Fig. 1. Average risks for developing schizophrenia compiled from the family and twin studies conducted in European populations between 1920 and 1987. (Adapted from Gottesman 1991)

quarter of the total liability to schizophrenia has been found to be due to chance environmental factors, where as the importance of shared familial environmental factors seems to be marginal (McGue et al. 1983; Heath et al. 1989). It is possible that "nongenetic" factors consist entirely of stochastic events affecting gene expression or structure (McGuffin et al. 1994; Woolf 1997). The facts do not support any purely psychological or environmental theory of schizophrenia.

Environmental risk factors possibly contributing towards the liability to schizophrenia, or to its release, include stress in the form of life events (Norman and Malla 1993) or disruptive patterns of emotional expression within families (Bebbington and Kuipers 1994; Tienari et al. 1994; Wahlberg et al. 1997), winter birth (Machon et al. 1983; Pulver et al. 1990; Beckmann and Franzek 1992), maternal virus infections, pregnancy, and birth complications (Cardno and McGuffin 1996). The winter birth effect might be explained by an intrauterine viral etiology of schizophrenia during the second trimester of gestation, a critical period for brain development due to extensive neuronal migration. Stress has never been shown to cause schizophrenia but to trigger a new acute psychotic episode in individuals already suffering from the disease.

In the future, the identification of risk genes by molecular genetic methods (see below) will provide the means for searching for environmental risk factors in genetically sensitive individuals. Since it is easier to change the environment than to change genes, this knowledge might be the key to the prevention of schizophrenia.

2.3 Number of Genes Involved

Family data are in agreement with models involving two loci (oligogenic model) or numerous genes of small effect (polygenic model) (Risch 1990; Gottesman 1991), whereas the results of linkage analyses suggest, depending on the criteria employed, at least two (Moldin and Gottesman 1997), eight (see Fig. 2) or more vulnerability-increasing genes for schizophrenia.

2.4 Transmission of Schizophrenia Genes

The multifactorial-polygenic model (MFP) of schizophrenia developed by Gottesman and Shields (1967) has survived three decades of testing by segregation and linkage analyses and is therefore most likely to be correct (Moises 1995). Although monogenic or single major locus models (SML) of schizophrenia cannot be ruled out for some families or postulated subforms such as periodic catatonia (Leonhard 1979), they seem unlikely to be correct for the large majority of schizophrenia patients (Risch 1990). Additional factors, such as heterogeneity and anticipation, have been postulated, leading to more complex models.

Etiological or genetic heterogeneity of schizophrenia has often been hypothesized (e.g. Bleuler 1911; Tsuang et al. 1990) and has especially been emphasized by the

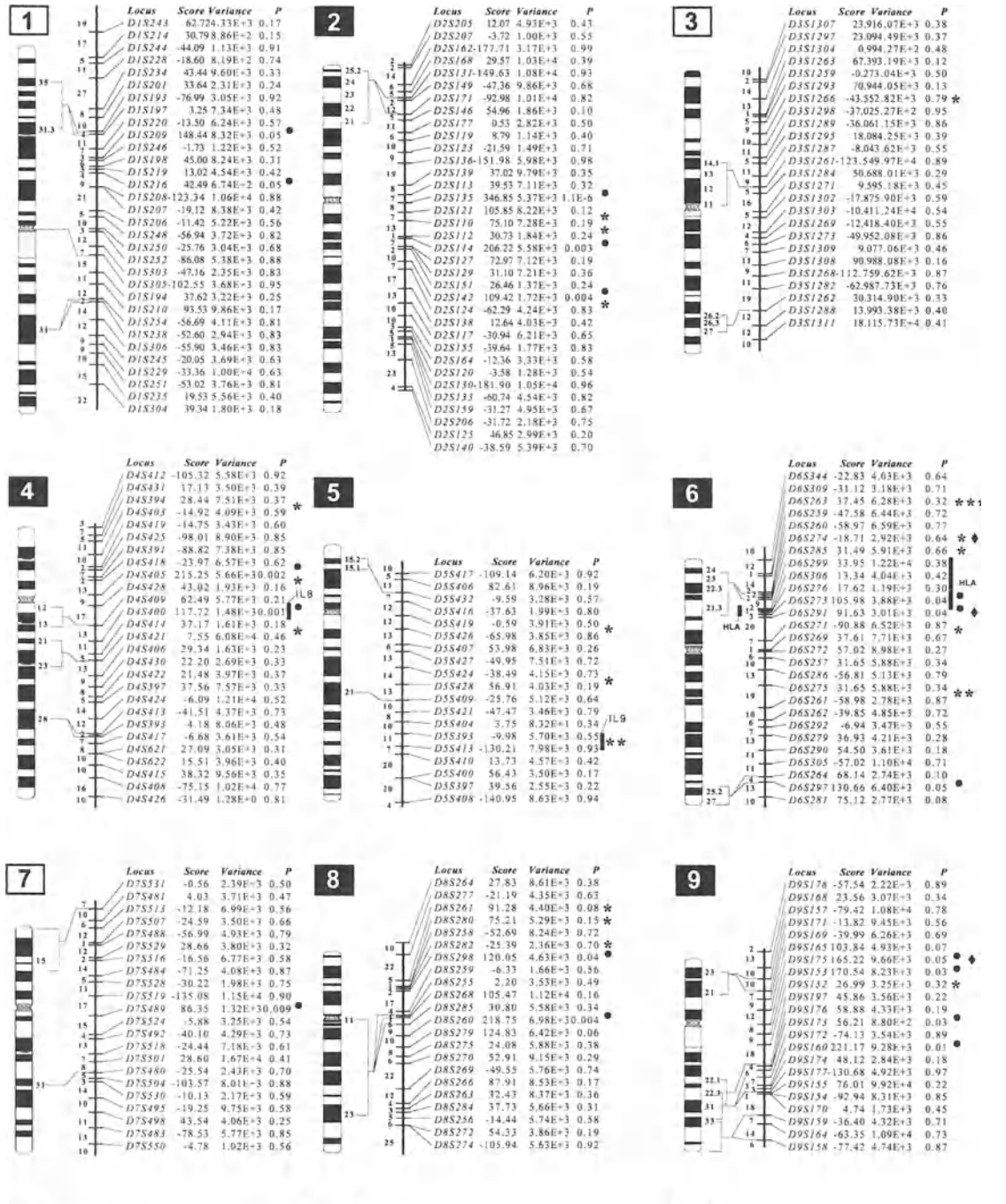

1
Locus Score Variance P
D1S243 62.72 4.33E+3 0.17
D1S214 30.79 8.86E+2 0.15
D1S244 -44.09 1.13E+3 0.91
D1S228 -18.60 8.19E+2 0.74
D1S234 43.44 9.60E+3 0.33
D1S201 33.64 2.31E+3 0.24
D1S193 -76.99 3.05E+3 0.92
D1S197 3.25 7.34E+3 0.48
D1S220 -13.50 6.24E+3 0.57
D1S209 148.44 8.32E+3 0.05
D1S246 -1.73 1.22E+3 0.52
D1S198 45.00 8.24E+3 0.31
D1S219 13.02 4.54E+3 0.42
D1S216 42.49 6.74E+2 0.05
D1S208 -123.34 1.06E+4 0.88
D1S207 -19.12 8.38E+3 0.42
D1S206 -11.42 5.22E+3 0.56
D1S248 -56.94 3.72E+3 0.82
D1S250 -25.76 3.04E+3 0.68
D1S252 86.08 5.38E+3 0.88
D1S303 -47.16 2.35E+3 0.83
D1S305 -102.55 3.68E+3 0.95
D1S194 37.62 3.22E+3 0.25
D1S210 93.53 9.86E+3 0.17
D1S254 -56.69 4.11E+3 0.81
D1S238 -52.60 2.94E+3 0.83
D1S306 -55.90 3.46E+3 0.83
D1S245 -20.05 3.69E+3 0.63
D1S229 -33.36 1.00E+4 0.63
D1S251 -53.02 3.76E+3 0.81
D1S235 19.53 5.56E+3 0.40
D1S304 39.34 1.80E+3 0.18
2
Locus Score Variance P
D2S205 12.07 4.93E+3 0.43
D2S207 -3.72 1.00E+3 0.55
D2S162 -177.71 3.17E+3 0.99
D2S168 29.57 1.03E+4 0.39
D2S133 -149.63 1.08E+4 0.93
D2S149 -47.36 9.86E+3 0.68
D2S171 -92.98 1.01E+4 0.82
D2S146 54.96 1.86E+3 0.10
D2S177 0.53 2.82E+3 0.50
D2S119 8.79 1.14E+3 0.40
D2S123 -21.59 1.49E+3 0.71
D2S136 -151.98 5.98E+3 0.98
D2S139 37.02 9.79E+3 0.35
D2S113 39.53 7.11E+3 0.32
D2S135 346.85 5.37E+3 1.1E-6
D2S121 105.85 8.22E+3 0.12
D2S110 75.10 7.28E+3 0.19
D2S112 30.73 1.84E+3 0.24
D2S114 206.22 5.58E+3 0.003
D2S127 72.97 7.12E+3 0.19
D2S129 31.10 7.21E+3 0.36
D2S151 26.46 1.37E+3 0.24
D2S142 109.42 1.72E+3 0.004
D2S124 -62.29 4.24E+3 0.83
D2S138 12.64 4.03E+3 0.42
D2S117 -30.94 6.21E+3 0.65
D2S155 -39.64 1.77E+3 0.83
D2S164 -12.36 3.33E+3 0.58
D2S120 -3.58 1.28E+3 0.54
D2S130 -181.90 1.05E+4 0.96
D2S133 -80.74 4.54E+3 0.82
D2S159 -31.27 4.95E+3 0.67
D2S206 -31.72 2.18E+3 0.75
D2S125 46.85 2.99E+3 0.20
D2S140 -38.59 5.39E+3 0.70
3
Locus Score Variance P
D3S1307 23.916.07E+3 0.38
D3S1297 23.094.49E+3 0.37
D3S1304 0.994.27E+2 0.48
D3S1263 67.393.19E+3 0.12
D3S1259 -0.273.04E+3 0.50
D3S1293 70.944.05E+3 0.13
D3S1266 -43.552.82E+3 0.79
D3S1298 -37.025.27E+2 0.95
D3S1289 -36.061.15E+3 0.86
D3S1295 18.084.25E+3 0.39
D3S1287 -8.043.62E+3 0.55
D3S1261 -123.549.97E+4 0.89
D3S1284 50.688.01E+3 0.29
D3S1271 9.595.18E+3 0.45
D3S1302 -17.875.90E+3 0.59
D3S1303 -10.411.24E+4 0.54
D3S1269 -12.418.40E+3 0.55
D3S1273 -49.952.08E+3 0.86
D3S1309 9.077.06E+3 0.46
D3S1308 90.988.08E+3 0.16
D3S1268 -112.759.62E+3 0.87
D3S1282 -62.987.73E+3 0.76
D3S1262 30.314.90E+3 0.33
D3S1288 13.993.38E+3 0.40
D3S1311 18.115.73E+4 0.41
4
Locus Score Variance P
D4S412 -105.32 5.58E+3 0.92
D4S431 17.13 3.50E+3 0.39
D4S394 28.44 7.51E+3 0.37
D4S403 -14.92 4.09E+3 0.59
D4S419 -14.75 3.43E+3 0.60
D4S425 -98.01 8.90E+3 0.85
D4S391 -88.82 7.38E+3 0.85
D4S418 -23.97 6.57E+3 0.62
D4S405 215.25 5.66E+3 0.002
D4S428 43.02 1.93E+3 0.16
D4S409 62.49 5.77E+3 0.21
D4S400 117.72 1.48E+3 0.001
D4S414 37.17 1.61E+3 0.18
D4S421 7.55 6.08E+4 0.46
D4S406 29.34 1.63E+3 0.23
D4S430 22.20 2.69E+3 0.33
D4S422 21.48 3.97E+3 0.37
D4S397 37.56 7.57E+3 0.33
D4S424 -6.09 1.21E+4 0.52
D4S413 -41.51 4.37E+3 0.73
D4S393 4.18 8.06E+3 0.48
D4S417 -6.68 3.61E+3 0.54
D4S621 27.09 3.05E+3 0.31
D4S622 15.51 3.96E+3 0.40
D4S415 38.32 9.56E+3 0.35
D4S408 -75.15 1.02E+4 0.77
D4S426 -31.49 1.28E+0 0.81
5
Locus Score Variance P
D5S417 -109.14 6.20E+3 0.92
D5S406 82.61 8.96E+3 0.19
D5S432 -9.59 3.28E+3 0.57
D5S416 -37.63 1.99E+3 0.80
D5S419 -0.59 3.91E+3 0.50
D5S426 -65.98 3.85E+3 0.86
D5S407 53.98 6.83E+3 0.26
D5S427 -49.95 7.51E+3 0.72
D5S424 -38.49 4.15E+3 0.73
D5S428 56.91 4.03E+3 0.19
D5S409 -25.76 5.12E+3 0.64
D5S421 -47.47 3.46E+3 0.79
D5S404 3.75 8.32E+1 0.34
D5S393 -9.98 5.70E+3 0.55
D5S413 -130.21 7.98E+3 0.93
D5S410 13.73 4.57E+3 0.42
D5S400 56.43 3.50E+3 0.17
D5S397 39.56 2.55E+3 0.22
D5S408 -140.95 8.63E+3 0.94
6
Locus Score Variance P
D6S344 -22.83 4.03E+3 0.64
D6S309 -31.12 3.18E+3 0.71
D6S263 37.45 6.28E+3 0.32
D6S239 -47.58 6.44E+3 0.72
D6S260 -58.97 6.59E+3 0.77
D6S274 -18.71 2.92E+3 0.64
D6S285 31.49 5.91E+3 0.66
D6S299 33.95 1.22E+4 0.38
D6S306 13.34 4.04E+3 0.42
D6S276 17.62 1.19E+3 0.30
D6S273 105.98 3.88E+3 0.04
D6S291 91.63 3.01E+3 0.04
D6S271 -90.88 6.52E+3 0.87
D6S269 37.61 7.71E+3 0.67
D6S272 57.02 8.98E+3 0.27
D6S257 31.65 5.88E+3 0.34
D6S286 -56.81 5.13E+3 0.79
D6S275 31.65 5.88E+3 0.34
D6S261 -58.98 2.78E+3 0.87
D6S262 -39.85 4.85E+3 0.72
D6S292 -6.94 3.47E+3 0.55
D6S279 36.93 4.21E+3 0.28
D6S290 54.50 3.61E+3 0.18
D6S305 -57.02 1.10E+4 0.71
D6S264 68.14 2.74E+3 0.10
D6S297 130.66 6.40E+3 0.05
D6S281 75.12 2.77E+3 0.08
7
Locus Score Variance P
D7S531 -0.56 2.39E+3 0.50
D7S481 4.03 3.71E+3 0.47
D7S513 -12.18 6.99E+3 0.56
D7S507 -24.59 3.50E+3 0.66
D7S488 -56.99 4.93E+3 0.79
D7S529 28.66 3.80E+3 0.32
D7S516 -16.56 6.77E+3 0.58
D7S484 -71.25 4.08E+3 0.87
D7S528 -30.22 1.98E+3 0.75
D7S519 -135.08 1.15E+4 0.90
D7S489 86.35 1.32E+3 0.009
D7S524 -5.88 3.25E+3 0.54
D7S492 -40.10 4.29E+3 0.73
D7S518 -24.44 7.18E+3 0.61
D7S501 28.60 1.67E+4 0.41
D7S480 -25.54 2.43E+3 0.70
D7S504 -103.57 8.01E+3 0.88
D7S530 -10.13 2.17E+3 0.59
D7S495 -19.25 9.75E+3 0.58
D7S498 43.54 4.06E+3 0.25
D7S483 -78.53 5.77E+3 0.85
D7S550 -4.78 1.02E+3 0.56
8
Locus Score Variance P
D8S264 27.83 8.61E+3 0.38
D8S277 -21.19 4.35E+3 0.63
D8S261 91.28 4.40E+3 0.08
D8S280 75.21 5.29E+3 0.15
D8S258 -52.69 8.24E+3 0.72
D8S282 -25.39 2.36E+3 0.70
D8S298 120.05 4.63E+3 0.04
D8S259 -6.33 1.66E+3 0.56
D8S255 2.20 3.53E+3 0.49
D8S268 105.47 1.12E+4 0.16
D8S285 30.80 5.58E+3 0.34
D8S260 218.75 6.98E+3 0.004
D8S279 124.83 6.42E+3 0.06
D8S275 24.08 5.88E+3 0.38
D8S270 52.91 9.15E+3 0.29
D8S269 -49.55 5.76E+3 0.74
D8S266 87.91 8.53E+3 0.17
D8S263 32.43 8.37E+3 0.36
D8S284 37.73 5.66E+3 0.31
D8S256 -14.44 5.74E+3 0.58
D8S272 54.33 3.86E+3 0.19
D8S274 -105.94 5.63E+3 0.92
9
Locus Score Variance P
D9S178 -57.54 2.22E+3 0.89
D9S168 23.56 3.07E+3 0.34
D9S157 -79.42 1.08E+4 0.78
D9S171 -13.82 9.45E+3 0.56
D9S169 -39.99 6.26E+3 0.69
D9S165 103.84 4.93E+3 0.07
D9S175 165.22 9.66E+3 0.05
D9S153 170.54 8.23E+3 0.03
D9S152 26.99 3.25E+3 0.32
D9S197 45.86 3.56E+3 0.22
D9S176 58.88 4.33E+3 0.19
D9S175 56.21 8.80E+2 0.03
D9S172 -74.13 3.54E+3 0.89
D9S160 221.17 9.28E+3 0.01
D9S174 48.12 2.84E+3 0.18
D9S177 -130.68 4.92E+3 0.97
D9S155 76.01 9.92E+4 0.22
D9S154 -92.94 8.31E+3 0.85
D9S170 4.74 1.73E+3 0.45
D9S159 -36.40 4.32E+3 0.71
D9S164 -63.35 1.09E+4 0.73
D9S158 -77.42 4.74E+3 0.87
10
Locus Score Variance P
D10S249 0.65 1.29E+1 0.43
D10S189 -3.05 2.62E+3 0.52
D10S226 25.59 3.30E+3 0.33
D10S191 -133.67 4.42E+3 0.98
D10S203 -32.69 3.68E+3 0.71
D10S199 -94.13 6.40E+3 0.88
D10S207 19.64 3.70E+3 0.37
D10S188 -50.20 3.06E+3 0.82
D10S201 -5.09 2.73E+3 0.54
D10S185 -11.87 1.05E+3 0.64
D10S192 -74.79 2.71E+3 0.92
D10S187 -9.54 4.00E+2 0.68
D10S216 12.96 4.33E+2 0.27
D10S212 18.55 4.37E+3 0.79
11
Locus Score Variance P
D11S922 -40.26 3.73E+3 0.75
D11S909 -10.73 2.14E+3 0.59
D11S928 -54.81 2.98E+3 0.84
D11S907 12.78 2.03E+3 0.39
D11S935 -29.90 4.41E+3 0.67
D11S905 -38.63 7.03E+3 0.68
D11S903 -84.73 4.05E+3 0.91
D11S913 2.28 1.37E+3 0.48
D11S916 -102.12 5.10E+3 0.92
D11S901 0.12 1.64E+3 0.50
D11S920 2.64 1.90E+3 0.48
D11S925 45.37 1.43E+3 0.12
D11S934 20.78 3.11E+3 0.35
D11S910 10.01 6.89E+3 0.45
D11S968 74.15 4.59E+3 0.14
12
Locus Score Variance P
D12S99 43.41 5.33E+3 0.28
D12S98 -5.29 1.53E+2 0.67
D12S87 40.14 3.83E+3 0.26
D12S83 -51.65 3.55E+3 0.81
D12S92 -5.39 1.58E+3 0.55
D12S81 -10.83 1.96E+3 0.60
D12S82 27.35 2.57E+3 0.29
D12S101 3.09 6.00E+3 0.48
D12S78 25.75 2.28E+3 0.29
D12S79 113.66 2.50E+3 0.01
D12S76 24.59 3.47E+3 0.34
D12S97 28.35 4.02E+2 0.08

Fig. 2

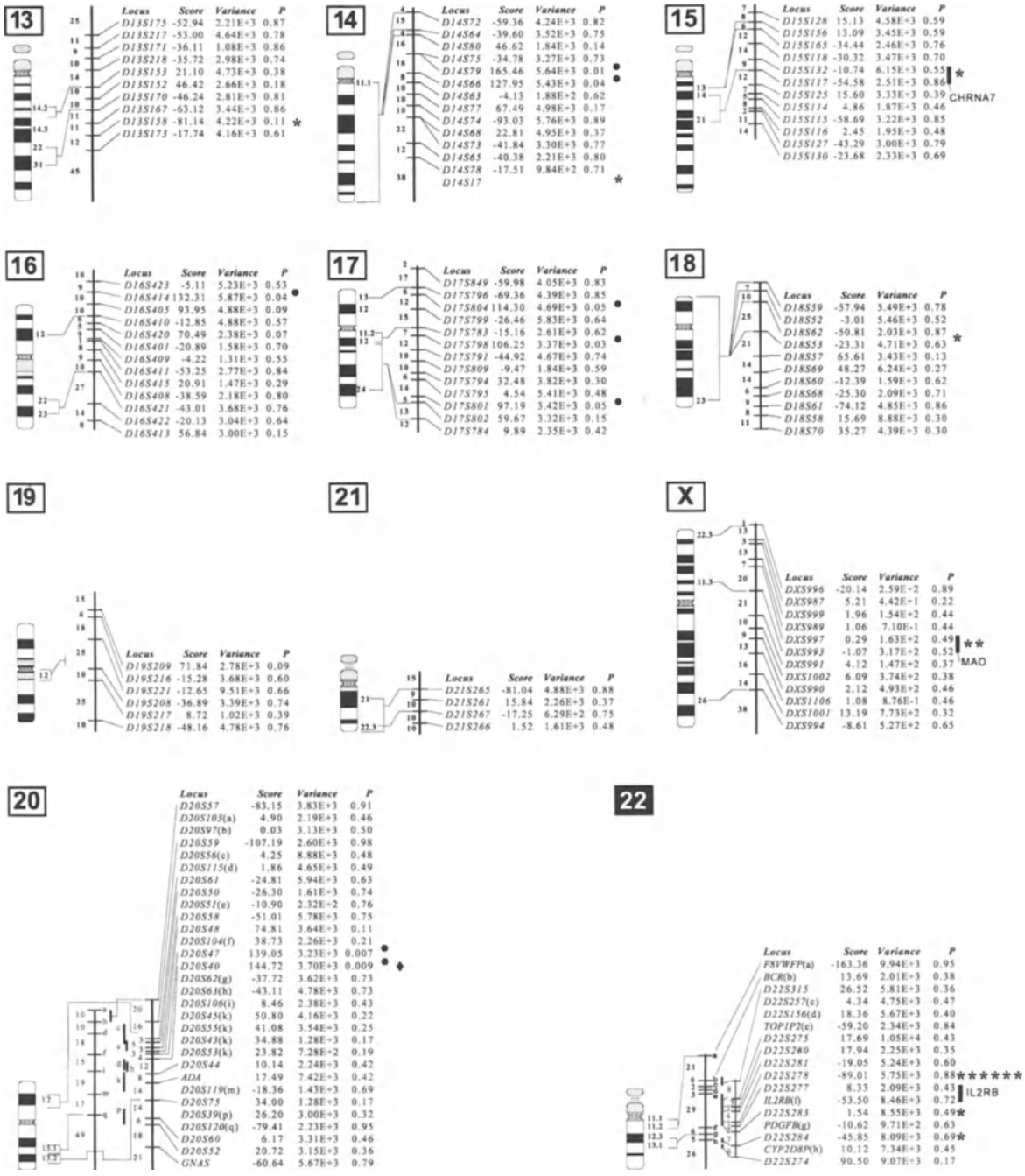


Fig. 2. Genome-wide probability map of schizophrenia susceptibility genes. Each *dot*, *diamond*, or *asterisk* represents evidence for linkage from independent samples. *Dots* (as well as score, variance, and *p* value) and *diamonds* display the results of stage I and II, respectively, of the two-stage genome scan by Moises et al. 1995a, *asterisks* the results from other studies (see Table 1) with lod scores ranging from 1.0 to 6.49 or *p* values from 0.05 to 0.000001 (chromosome 2). Eight chromosomes, highlighted in *black*, show clustering of positive linkage results (defined as $p < 0.05$ or lod score > 1.0) in the same region from at least two independent groups of investigators and therefore have a high probability of harboring susceptibility genes for schizophrenia. (Adapted from Moises et al. 1995a with the kind permission of *Nature Genetics*)

Wernicke-Kleist-Leonhard-Beckmann school (Leonhard 1979; Beckmann et al. 1996), while others found evidence in favor of continuity between affective disorders and schizophrenia (Crow 1986; Maier et al. 1993). The cloning of risk-increasing genes will provide powerful tools for testing these hypotheses.

Anticipation, defined as an increase in severity and a decrease in age of onset across generations, has been observed in all clinical studies (e.g. Thibaut et al. 1995; Bassett and Husted 1997; Johnson et al. 1997). In several monogenic neuropsychiatric disorders, anticipation is caused by unstable DNA, an expansion of triplet repeats such as CAG or CTG through consecutive generations leading to dysfunctional proteins (Petronis and Kennedy 1995). The molecular search for triplet expansions in schizophrenia has so far yielded negative results with two exceptions (Morris et al. 1995; O'Donovan et al. 1996).

2.5
Phenotype Corresponding to the Schizophrenia Genotype

In families and twins, Eugen Bleuler and Essen-Möller observed an extension of the schizophrenia genotype to schizophrenia-like personality disorders, the so-called schizoid disorders (Essen-Möller 1946). Adoption studies and blindfolded family studies later confirmed this concept in the form of schizophrenia spectrum disorders, including schizophrenia, schizoaffective psychosis, and paranoid, schizoid, and schizotypal personality disorders.

Several phenotypes serving as intermediators between genes and clinical disease, termed "endophenotypes" (Gottesman and Shields 1972) or "vulnerability markers" (Nuechterlein et al. 1990), have been postulated in schizophrenia, e.g. schizothymic constitution (Kretschmer 1925), Minnesota Multiphasic Personality Inventory (MMPI) psychometric index (Moldin et al. 1990), reduced reaction time for modality shift (Maier et al. 1994), structural brain abnormalities (Cannon and Marco 1994), reduced span of apprehension (Suslow and Arolt 1996), auditory event-related potentials (Frangou et al. 1997), backward masking performance (Green et al. 1997), P50 auditory-evoked potential gating deficit (Freedman et al. 1997), and deviant smooth-pursuit eye movements (SPEM) (Holzman 1992). For the latter two phenotypes, chromosomal linkages have been reported (Arolt et al. 1996; Freedman et al. 1997).

3
Molecular Genetics

3.1
Genes Predisposing to Schizophrenia

If risk genes play a significant role in the etiology of schizophrenia, as numerous formal genetic studies have demonstrated beyond doubt, one of the most important steps to follow is the identification of risk genes and their function. The development of recombinant DNA technology (S.N. Cohen et al. 1973) and genetic maps (Botstein et al. 1980) laid the foundations for identifying genes involved in schizophrenia. The positional cloning strategy (Collins 1992; Ghosh and Collins 1996) offers a straightforward approach for identifying risk genes. It has been successfully employed in nearly 100 diseases, among them such important disorders as Alzheimer's disease, breast cancer, colon cancer, Crohn's disease, diabetes, and Huntington's disease (see the NIH webpage http://www.nhgri.nih.gov/dir/gtb/clone/) and consists mainly of four steps:

1. *Chromosomal localization* of disease genes using families, genetic markers, and genetic maps
2. *Gene identification* by allelic association (linkage disequilibrium), mutation search, and phenotypic covariations
3. Investigations into the *pathogenesis of the disorder* using expression studies and transgenic and gene knockout animal models
4. *Therapeutic and preventional studies*

3.1.1 Chromosomal Localizations

In principle, chromosomal localizations of risk genes can be found by using as clues chromosomal aberrations producing a schizophrenia-like phenotype (Propping 1983; DeLisi et al. 1994b; Moises 1995) or by screening the entire human genome using linkage analysis and a dense map of genetic markers (Weissenbach et al. 1992). To date, the results of three complete genome scans have been reported (Coon et al. 1994; Moises et al. 1995a; Levinson et al. 1998). One of these employed a two-stage strategy (Moises et al. 1995a), a procedure that in principal has a built-in validation facility (Peltonen 1995). Figure 2 gives a genome-wide overview of positive linkage findings (for references, see Table 1).

Some of the reported locations are probably true, while others are false-positive findings. Several groups found evidence for linkage to schizophrenia on chromosomes 2, 4–6, 8, 9, 11, and 22, indicating that these chromosomal regions are likely harboring risk genes for the disease. The linkage findings on chromosome 2 and 6 fulfill the criteria for significant linkage ($p < 0.00002$) according to Lander and Kruglyak (1995). The situation on chromosome 6p24-p21 is complex (see Fig. 3), involving approximately 55 cM (for reviews, see Peltonen 1995; Straub et al. 1996).

3.1.2 Methodological Problems

Linkage Analysis

Linkage analysis, primarily developed for monogenic disorders, is employed to find chromosomal localiza-

Table 1. Possible chromosomal locations of risk genes for schizophrenia

Chromosomes	Reports (loci)
1	Moises et al. 1995a (stage I)
2	Moises et al. 1995a (stage I); Levinson et al. 1998
3	Pulver et al. 1995
4	Barr et al. 1994; Coon et al. 1994; Moises et al. 1995a (stage I); Levinson et al. 1998
5	Sherrington et al. 1988; Silverman et al. 996; Schwab et al. 1997; Straub et al. 1997
6	Straub et al. 1995; Moises et al. 1995a stage I & II); Schwab et al. 1995b; Antonarakis et al. 1995; SLCG 1996; Arolt et al. 1996; Maziade et al. 1997; Cao et al. 1997
7	Moises et al. 1995a (stage I)
8	Moises et al. 1995a (stage I); Pulver et al. 1995; Kendler et al. 1996; SLCG 1996
9	Moises et al. 1995a (stages I and II); Levinson et al. 1998
10	Levinson et al. 1998
11	St. Clair et al. 1990; Maziade et al. 1995; Levinson et al. 1998
12	Moises et al. 1995a (stage I)
13	Lin et al. 1995, 1997
14	Coon et al. 1994; Moises et al. 1995a (stage I)
15	Coon et al. 1994; Freedman et al. 1997
16	Moises et al. 1995a (stage I)
17	Moises et al. 1995a (stage I)
18	Wildenauer et al. 1996
20	Moises et al. 1995a (stages I and II)
22	Pulver et al. 1994a,b; Coon et al. 1994; Lasseter et al. 1995; Vallada et al. 1995; Schwab et al. 1995a; Moises et al. 1995b; Gill et al. 1996
X and Y	Collinge et al. 1991; Crow et al. 1994; DeLisi et al. 1994a; Dann et al. 1997

tion of genes in schizophrenia and other such complex genetic disorders. However, unlike monogenic diseases, in the complex disorders a certain gene under investigation is neither necessary nor sufficient, resulting in variations of risk genes between individuals and incomplete cosegregation within families (Greenberg 1993; Sing et al. 1996). As a consequence, significant linkage results are much more difficult to obtain in complex disorders. For genes with a small effect, required sample sizes are well over 2000 families (Risch and Merikangas 1996), and the original sample size has to be multiplied by approximately the number of genes involved in the disorder in order to obtain sufficient power for replication (Suarez et al. 1994). Not surprisingly, confirmation of initial linkage results is difficult and a matter of debate in schizophrenia (Crow 1997; Moldin 1997). The debate in schizophrenia is not dissimilar to the situation in other complex disorders after reports of inconsistent linkage findings and prior to the identification of the relevant gene (e.g. in atopy see, Cookson et al. 1989; Morton 1992; Shirakawa et al. 1994; Fölster-Holst et al. 1998). To avoid a flood of false-positive claims, strict guidelines for interpreting linkage results in complex disorders have been proposed by Lander and Kruglyak (1995). While strict linkage criteria are necessary to decrease the probability of false-positive results (type II error), they will also increase the probability of false-negative results (type I error), an especially undesirable effect in complex disorders considering the difficulties in obtaining evidence for linkage, even if the risk genes are known, as in some HLA-associated disorders, breast cancer, atopy, and ApoE in Alzheimer's disease (Thorlacius et al. 1995; Fölster-Holst et al. 1998; Cai et al. 1997). The only way to reduce type II and I errors is to have further relevant evidence available, such as several studies reporting evidence for linkage in the same region (see Fig. 2) and the cloning of the gene. The latter has never been successful in complex disorders by using a pure positional cloning approach (Collins et al. 1997) except when one of two strategies has been applied: (1) haplotype studies in isolated human populations or (2) identification of a subphenotype in pedigrees in which the disease behaves in a near-mendelian fashion. In other words, cloning of the first schizophrenia gene is most likely to be successful by using isolated populations from Iceland, Northern Sweden, Finland, and Micronesia or near-mendelian endophenotypes such as P50 auditory-evoked potential gating deficit or deviant SPEM. In any case, we should not forget that linkage statistics cannot prove the existence of a risk-increasing gene; the inference only becomes more or less probable. The final proof has to come from the demonstration of concomitant variations between the gene and the phenotype. Problems and strategies in complex disorders are discussed by Weeks and Lathrop (1995), Ott (1996), Risch and Merikangas (1996), Sing et al. (1996), and Pawlowtzki et al. (1997).

Association Studies

Association studies in the general population of unrelated cases using a case-control design are a rather ambiguous approach (Kidd 1993) because of the high rate of false-positive findings of up to 99.5% (Crowe 1993). In other words, unreplicated positive findings from case-control studies and the associated media hype about gene "discoveries" are probably incorrect.

3.1.3 Candidate Genes

Using mostly association studies and a case-control design, a large number of candidate genes for schizo-

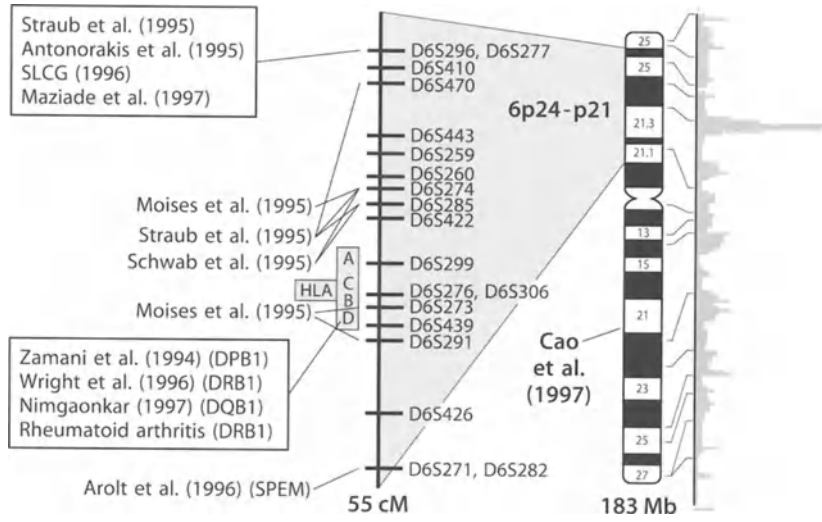


Fig. 3. Positive linkage and association studies on chromosome 6. The density of genes mapped to chromosome 6 is shown to the *right* (data from Schuler et al. 1996). *SPEM*, smooth-pursuit eye movement

phrenia have been investigated (for a review, see Moises 1995). More recently, positive results have been reported for genes of the androgen receptor (Crow et al. 1993), D_2–D_5 dopamine receptors (Arinami et al. 1997; Ebstein et al. 1997; Catalano et al. 1993; N.M. Williams et al. 1997), dopamine transporter (Persico and Macciardi 1997), dopamine-β-hydroxylase (Wei et al. 1997), serotonin 2A receptor (HTR2A) (Inayama et al. 1996; J. Williams et al. 1996), SCA1 (Wang et al. 1996), in the HLA-region B35, DRB1, DPB1, CD4, DQB1 (Blackwood et al. 1996; Wright et al. 1996; Zamani et al. 1994; Nimgaonkar et al. 1997), and the α_7-nicotinic receptor (Leonard et al 1996; Freedman et al. 1997).

More importantly, the α_7-nicotinic receptor has been found by linkage analysis and genome scanning using as endophenotype a measure of the known deficit of inhibitory mechanism in schizophrenics, the P50 auditory-evoked potential gating deficit. In schizophrenia families, significant linkage was obtained between an abnormal P50 ratio and a marker on chromosome 15q13-q14, D15S1360, the localization of the α_7-nicotinic receptor gene (Leonard et al. 1996; Freedman et al. 1997). Different subforms of epilepsy also show evidence of linkage to this gene (Elmslie et al. 1997; Neubauer et al. 1998). Hints for candidate genes can also come from an association between schizophrenia and other diseases with a known genetic component, such as rheumatoid arthritis (RA) and cancer.

Negative Associations with Rheumatoid Arthritis and Cancer

In schizophrenic patients, a decreased risk in cancer mortality and in RA has been documented (Dupont et al. 1986; Vinogradov et al. 1991). Despite the fact that schizophrenics are heavy smokers, the rate for lung cancer is only 38% of normal subjects. In turn, lung cancer and RA are positively associated (Mellemkjaer et al. 1996). For the latter, an association with loci of the HLA region on chromosome 6p is well known (Weyand and Goronzy 1995). Chromosome 6p has also been implicated in schizophrenia by linkage studies.

4 Past, Future, and Ethical Problems

In the nineteenth century, Charles Darwin's (1809–1882) cousin Francis Galton (1822–1911) developed a plan to "improve" the human race and societies by excluding individuals with "bad" genes from reproduction; he named his strategy "eugenics." In the twentieth century, the eugenics movement – also termed "race hygiene" – became influential in many countries, leading in the USA to preferential immigration from Northern Europe and racial segregation, to compulsory sterilization of nearly 500,000 individuals in 11 American states, Germany, and Sweden, and finally under the Nazi regime to the killing of five million Europeans of Jewish ancestry and nearly 70,000 patients, many of them schizophrenic patients (Propping 1989, 1992; Gottesman and Bertelsen 1996; Weber 1996; Gejman 1997; Moldin and Gottesman 1997; Watson 1997). Not only is mass, government-mandated eugenics inhuman but also unfeasible, because the genes for many common disorders such as schizophrenia are so widely distributed in populations that preventing the reproduction of risk genes seems only be possible by preventing the reproduction of the human species. Furthermore, even if we could reduce the rate of vulnerability genes for schizophrenia, it is

rather doubtful whether we should, since they may also be involved in disease resistance or some positive, even creative, human behaviors.

Summing up the current state of knowledge, the majority of people suffering from schizophrenia most probably have a polygenic disorder with a strong genetic component involving unknown environmental factors and probably more than eight risk genes of unknown function, some at least localized on chromosomes 2, 4–6, 8, 9, 11, and 22. In the near future, one or more of the genes will be cloned, thereby revealing through expression studies the network of genes involved in the pathophysiological pathways of schizophrenia. As in depression research (e.g. Edwards et al. 1992), genetically modified animals will provide models to study schizophrenia's pathogenesis, course, rational treatment, and prevention. The discovery process will be considerably accelerated by the Human Genome Project.

By 2005, the Human Genome Project, a three-step, three billion dollar program initiated by the U.S. government and arguably the single most important organized international research project in the history of medicine (Guyer and Collins 1993), is expected to reveal to mankind the chromosomal localization and structure of the estimated 100,000 human genes (Cohen 1993; Guyer and Collins 1995; Rowen et al. 1997, webpage: http://www.ornl.gov/TechResources/Human_Genome/home.html). The first two steps of the program have already been completed. Comparable only to the nineteenth century's discovery of the building blocks of chemistry in the form of the periodic table, the twentieth century's discovery of the building blocks of life will give us biology's "periodic table," leading to more predictability in the life sciences (Lander 1996). Likewise, the impact on biology, industry, medicine, and society will be profound.

New medical disciplines will emerge, such as molecular and predictive medicine based on genetic testing for disease susceptibilities. Medicine's emphasis will probably shift from treatment to prevention and counseling of individuals at risk (Dausset 1986; Gottesman and Collins 1994). Ethical, legal, and social issues (ELSI) are anticipated, and solutions are being discussed within the framework of the Human Genome Project (see webpage http://lbl.gov/education/ELSI/ELSI.html). A central ethical problem of predictive medicine is confidentiality. There is agreement among geneticists that results of risk tests should not be made available to employers or insurance agencies. However, according to surveys, most health professionals are not prepared to integrate genetics into clinical practice (Collins 1997).

The appreciation of genetic nonidentity as a simple fact of life will be another important consequence of the "genetic revolution." It is impossible to fight against existing genetic inequality by denying its existence. What is needed are medical and social strategies to counterbalance existing suffering whatever their proximate or distant origins may be. The famous Nobel laureate James D. Watson (1997) put it this way:

> Genetics as a discipline must strive to be the servant of the people, as opposed to governments, working to mitigate the genetic inequalities arising from the random mutations that generate our genetic diseases.

In human history, lesser crimes against humanity than the holocaust for the supposed "benefit" of society have never been forgotten. Therefore, "we are well advised to look back to the founder of scientific medicine, Hippocrates, whose oath is the basis of biomedical ethics" (Ganten 1997). Furthermore, in light of the past, psychiatrists and geneticists should be the first to fight against possible misuse of genetic information and to use gene technology to the benefit of patients and those who are disadvantaged.

5 References

Antonarakis SE, Blouin JL, Pulver AE et al (1995) Schizophrenia susceptibility and chromosome 6p24–22. Nat Genet 11: 235–236

Arinami T, Gao M, Hamaguchi H et al (1997) A functional polymorphism in the promoter region of the dopamine D_2 receptor gene is associated with schizophrenia. Hum Mol Genet 6: 577–582

Arolt V, Lencer R, Nolte A et al (1996) Eye tracking dysfunction is a putative phenotypic susceptibility marker of schizophrenia and maps to a locus on chromosome 6p in families with multiple occurrence of the disease. Am J Med Genet 67: 564–579

*Barondes SH, Alberts BM, Andreasen NC et al (1997) Workshop on schizophrenia. Proc Natl Acad Sci USA 94: 1612–1614

Barr CL, Kennedy JL, Pakstis AJ et al (1994) Progress in a genome scan for linkage in schizophrenia in a large Swedish kindred. Am J Med Genet 54: 51–58

Bassett AS, Husted J (1997) Anticipation or ascertainment bias in schizophrenia? Penrose's familial mental illness sample. Am J Hum Genet 60: 630–637

Bebbington P, Kuipers L (1994) The predictive utility of expressed emotion in schizophrenia: an aggregate analysis. Psychol Med 24: 707–718

Beckmann H, Franzek E (1992) Deficit of birthrates in winter and spring months in distinct subgroups of mainly genetically determined schizophrenia. Psychopathology 25: 57–64

Beckmann H, Franzek E, Stöber G (1996) Genetic heterogeneity in catatonic schizophrenia: a family study. Am J Med Genet 67: 289–300

Blackwood DH, Muir WJ, Stephenson A et al (1996) Reduced expression of HLA-B35 in schizophrenia. Psychiatr Genet 6: 51–59

Bleuler E (1911) Dementia praecox or the group of schizophrenias. International Universities Press, New York

Botstein D, White RL, Skolnick M et al (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am J Hum Genet 32: 314–331

Cai X, Fallin D, Stanton J et al (1997) ApoE is linked to Alzheimer's disease in a large pedigree. Am J Med Genet 74: 365–369

Cannon TD, Marco E (1994) Structural brain abnormalities as indicators of vulnerability to schizophrenia. Schizophr Bull 20: 89–102

Catalano M, Nobile M, Novelli E et al (1993) Distribution of a novel mutation in the first exon of the human dopamine D_4 receptor gene in psychotic patients. Biol Psychiatry 34: 459–464

Cao Q, Martinez M, Zhang J et al (1997) Suggestive evidence for a schizophrenia susceptibility locus on chromosome 6q and a confirmation in an independent series of pedigrees. Genomics 43: 1–8

Cardno AG, McGuffin P (1996) Aetiological theories of schizophrenia. Curr Opin Psychiatry 9: 45–49

Cohen D (1993) Les gènes de l'espoir. Laffont, Paris

Cohen SN, Chang AC, Boyer HW et al (1973) Construction of biologically functional bacterial plasmids in vitro. Proc Natl Acad Sci USA 70: 3240–3244

Collinge J, DeLisi LE, Boccio A et al (1991) Evidence for a pseudo-autosomal locus for schizophrenia using the method of affected sibling pairs. Br J Psychiatry 158: 624–629

Collins FS (1992) Positional cloning: let's not call it reverse anymore. Nature Genet 1: 3–6

*Collins FS (1997) Preparing health professionals for the genetic revolution. JAMA 278: 1285–1286

Collins FS, Guyer MS, Charkravarti A et al (1997) Variations on a theme: cataloging human DNA sequence variation. Science 278: 1580–1582

Cookson WOCM, Sharp PA, Faux J et al (1989) Linkage between immunoglobulin-E responses underlying asthma and rhinitis at chromosome 11q. Lancet 337: 1292–1295

Coon H, Jensen S, Holik J et al (1994) Genomic scan for genes predisposing to schizophrenia. Am J Med Genet 54: 59–71

Crow TJ (1986) The continuum of psychosis and its implications for the structure of the gene. Br J Psychiatry 149: 419–429

Crow TJ (1997) Current status of linkage for schizophrenia: polygenes of vanishingly small effect or multiple false positives? Am J Med Genet 74: 99–103

Crow TJ, Poulter M, Lofthouse R et al (1993) Male siblings with schizophrenia share alleles at the androgen receptor above chance expectation. Am J Med Genet 48: 159–160

Crow TJ, DeLisi LE, Lofthouse R et al (1994) An examination of linkage of schizophrenia and schizoaffective disorder to the pseudoautosomal region. Br J Psychiatry 164: 159–164

Crowe RR (1993) Candidate genes in psychiatry: an epidemiological perspective. Am J Med Genet 48: 74–77

Dann J, DeLisi LE, Devoto M et al (1997) A linkage study of schizophrenia to markers within Xp11 near the MAOB gene. Psychiatry Res 70: 131–143

*Dausset J (1986) Prospects and ethics of predictive medicine. Pathol Biol (Paris) 34: 812–813

DeLisi LE, Devoto M, Lofthouse et al (1994a) Search for linkage for schizophrenia on the X and Y chromosomes. Am J Med Genet 54: 113–121

DeLisi LE, Friedrich U, Wahlstrom J et al (1994b) Schizophrenia and sex chromosome anomalies. Schizophr Bull 20: 495–505

Dupont A, Moeller-Jensen O, Strömgren E, Jablensky A (1986) Incidence of cancer in patients diagnosed as schizophrenic in Denmark. In: ten Horn GHMM, Giel R, Gulbinat W, Henderson JH (eds) Psychiatric case registers in public health. Elsevier, Amsterdam, pp 229–239

Eaton WW, Tien AY, Poeschla BD (1995) Epidemiology of schizophrenia. In: Den Boer JA, Westenberg HGM, van Praag HM (eds) Advances in the neurobiology of schizophrenia. Wiley, Chichester, pp 27–57

Ebstein RP, Macciardi F, Heresco-Levi U et al (1997) Evidence for an association between the dopamine D_3 receptor gene DRD3 and schizophrenia. Hum Hered 47: 6–16

Edwards E, Konrich W, van Houtten P, Henn FA (1992) In vitro neurotransmitter release in an animal model of depression. Neurochem Int 21: 29–35

Elmslie FV, Rees M, Williamson MP et al (1997) Genetic mapping of a major susceptibility locus for juvenile myoclonic epilepsy on chromosome 15q. Hum Mol Genet 6: 1329–1334

Essen-Möller E (1946) The concept of schizoidea. Monatschr Psychiatr Neurol 112: 258–271

Farmer AE, McGuffin P, Gottesman II (1987) Twin concordance for DSM-III schizophrenia. Scrutinizing the validity of the definition. Arch Gen Psychiatry 44: 634–641

Fölster-Holst R, Moises HW, Yang L et al (1998) Linkage between atopy and the IgE high-affinity receptor gene at 11q13 in atopic dermatitis families. Hum Genet 102: 236–239

Frangou S, Sharma T, Alarcon G et al (1997) The Maudsley family study. II. Endogenous event-related potentials in familial schizophrenia. Schizophr Res 23: 45–53

Freedman R, Coon H, Myles-Worsley M et al (1997) Linkage of a neurophysiological deficit in schizophrenia to a chromosome 15 locus. Proc Natl Acad Sci USA 94: 587–592

Galton F (1875) The history of twins, as a criterion of the relative powers of nature and nurture. Fraser's Magazine 12: 566–576

Ganten D (1997) James D. Watson at the Congress of Molecular Medicine. J Mol Med 75: 615–617

Gejman PV (1997) Ernst Rüdin and Nazi euthanasia: another stain on his career. Am J Med Genet 74: 455–456

*Ghosh S, Collins FS (1996) The geneticist's approach to complex disease. Annu Rev Med 47: 333–353

Gill M, Vallada H, Collier D et al (1996) A combined analysis of D22S278 marker alleles in affected sib-pairs: support for a susceptibility locus for schizophrenia at chromosome 22q12. Am J Med Genet 67: 40–45

**Gottesman II (1991) Schizophrenia genesis: the origins of madness. Freedman, New York

Gottesman II, Bertelsen A (1996) Legacy of German psychiatric genetics: hindsight is always 20/20. Am J Med Genet 67: 317–322

Gottesman II, Shields J (1967) A polygenic theory of schizophrenia. Proc Natl. Acad Sci USA 58: 199–205

Gottesman II, Shields J (1972) Schizophrenia and genetics: a twin study vantage point. Academic, New York

Gottesman MM, Collins FS (1994) The role of the human genome project in disease prevention. Prev Med 23: 591–594

Greenberg DA (1993) Linkage analysis of "necessary" disease loci versus "susceptibility" loci. Am J Hum Genet 52: 125–143

Guyer MS, Collins FS (1993) The Human Genome Project and the future of medicine. Am J Dis Child 147: 1145–1152

Guyer MS, Collins FS (1995) How is the Human Genome Project doing, and what have we learned so far? Proc Natl Acad Sci USA 92: 10841–10848

Green MF, Breitmeyer B, Nuechterlein KH (1997) Backward masking performance in unaffected siblings of schizophrenia. Evidence for a vulnerability indicator. Arch Gen Psychiatry 54: 465–472

Häfner H (1987) Epidemiology of schizophrenia. In: Häfner H, Gattaz WF, Janzarik W (eds) Search for the causes of schizophrenia. Springer, Berlin Heidelberg New York, pp 47–74

Häfner H, an der Heiden W (1997) Epidemiology of schizophrenia. Can J Psychiatry 42: 139–151

Heath AC, Neale MC, Hewitt J et al (1989) Testing structural equation models for twin data using LISREL. Behav Genet 19: 9–35

Henn FA (1995) Neurobiology of schizophrenia. Schweiz Arch Neurol Psychiatr 146: 224–229

Heston LL (1966) Psychiatric disorders in foster home reared children of schizophrenic mothers. Br J Psychiatry 112: 819–825

Holzman PS (1992) Behavioral markers of schizophrenia useful for genetic studies. J Psychiatr Res 26: 427–445

Inayama Y, Yoneda H, Sakai T et al (1996) Positive association between a DNA sequence variant in the serotonin 2A receptor gene and schizophrenia. Am J Med Genet 67: 103–105

Jablensky A (1995) Schizophrenia: recent epidemiologic issues. Epidemiol Rev 17: 10–20

Johnson JE, Clearly J, Ahsan H et al (1997) Anticipation in schizophrenia: biology or bias? Am J Med Genet 74: 275–280

Kallmann FJ (1938) The genetics of schizophrenia. Augustin, New York

Kay DWK (1963) Late paraphrenia and its bearing on the aetiology of schizophrenia. Acta Psychiat Scand 39: 159–169

Kendler KS, Diehl SR (1993) The genetics of schizophrenia: a current, genetic-epidemiologic perspective. Schizophr Bull 19(2): 261–285

*Kendler KS, Gruenberg AM, Kinney DK (1994) Independent diagnoses of adoptees and relatives as defined by DSM-III in the provincial and national samples of the Danish Adoption Study of Schizophrenia. Arch Gen Psychiatry 51: 456–468

Kendler KS, MacLean CJ, O'Neill FA et al (1996) Evidence for a schizophrenia vulnerability locus on chromosome 8p in the Irish Study of High-Density Schizophrenia Families. Am J Psychiatry 153: 1534–1540

Kety SS, Wender PH, Jacobsen B et al (1994) Mental illness in the biological and adoptive relatives of schizophrenic adoptees. Replication of the Copenhagen Study in the rest of Denmark. Arch Gen Psychiatry 51: 442–455

*Kidd KK (1993) Associations of disease with genetic markers: Déjà vu all over again. Am J Med Genet 48: 71–73

Kretschmer E (1925) Physique and character. Harcourt Brace, New York

*Lander ES (1996) The new genomics: global views of biology. Science 274: 536–539

Lander E, Kruglyak L (1995) Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. Nature Genet 11: 241–247

Lasseter VK, Pulver AE, Wolyniec PS et al (1995) Follow-up report of potential linkage for schizophrenia on chromosome 22q, part 3. Am J Med Genet 60: 172–173

Leonard S, Adams C, Breese CR et al (1996) Nicotinic receptor function in schizophrenia. Schizophr Bull 22: 431–441

Leonhard K (1979) Classification of endogenous psychoses. Irvington, New York

Levinson DF, Mahtani MM, Nancarrow DJ et al (1998) A genome scan of schizophrenia. Am J Psychiatry 155: 741–750

Lin MW, Curtis D, Williams N et al (1995) Suggestive evidence for linkage of schizophrenia to markers on chromosome 13q14.1-q32. Psychiatr Genet 5: 117–126

Lin MW, Sham P, Hwu HG et al (1997) Suggestive evidence for linkage of schizophrenia to markers on chromosome 13 in Caucasian but not oriental populations. Hum Genet 99: 417–420

Machon RA, Mednick SA, Schulsinger F (1983) The interaction of seasonality, place of birth, genetic risk and subsequent schizophrenia in a high risk sample. Br J Psychiatry 143: 383–388

Maier W, Lichtermann D, Minges J et al (1993) Continuity and discontinuity of affective disorder and schizophrenia. Results of a controlled family study. Arch Gen Psychiatry 50: 871–883

Maier W, Frank P, Kopp B et al (1994) Reaction time paradigms in subjects at risk for schizophrenia. Schizophr Res 13: 35–43

Maziade M, Raymond V, Cliche D et al (1995) Linkage results on 11q21–22 in Eastern Quebec pedigrees densely affected by schizophrenia. Am J Med Genet 60: 522–528

Maziade M, Bissonnette L, Rouillard E et al (1997) 6p24–22 region and major psychoses in the Eastern Quebec population. Le Groupe IREP. Am J Med Genet 74: 311–318

McGue M, Gottesman II, Rao DC (1983) The transmission of schizophrenia under a multifactorial threshold model. Am J Hum Genet 35: 1161–1178

McGuffin P, Owen MJ (1996) Molecular genetic studies of schizophrenia. Cold Spring Harb Symp Quant Biol 61: 815–822

McGuffin P, Asherson P, Owen M, Farmer A (1994) The strength of the genetic effect. Is there room for an environmental influence in the aetiology of schizophrenia? Br J Psychiatry 164: 593–599

Mellemkjaer L, Linet MS, Gridley G et al (1996) Rheumatoid arthritis and cancer risk. Eur J Cancer 32A: 1753–1757

*Moises HW (1995) Genetic models of schizophrenia. In: Den Boer JA, Westenberg HGM, Van Praag HM (eds) Advances in the neurobiology of schizophrenia. Wiley, Chichester, pp 59–86

Moises HW, Gelernter J, Giuffra LA et al (1991) No linkage between D_2 dopamine receptor gene region and schizophrenia. Arch Gen Psychiatry 48: 643–647

Moises HW, Yang L, Kristbjarnarson H et al (1995a) An international two-stage genome-wide search for schizophrenia susceptibility genes. Nature Genet 11: 321–324

Moises HW, Yang L, Li T et al (1995b) Potential linkage disequilibrium between schizophrenia and locus D22S278 on the long arm of chromosome 22. Am J Med Genet 60: 465–467

Moldin SO (1997) The maddening hunt for madness genes. Nature Genet 17: 127129

*Moldin SO, Gottesman II (1997) Genes, experience, and chance in schizophrenia: positioning for the 21st century. Schizophr Bull 23: 547–561

Moldin SO, Rice JP, Gottesman II, Erlenmeyer-Kimling L (1990) Transmission of a psychometric indicator for liability

to schizophrenia in normal families. Genet Epidemiol 7: 163–176

Morris AG, Gaitonde E, McKenna PJ et al (1995) CAG repeat expansions and schizophrenia: association with disease in females and with early age-at onset. Hum Mol Genet 4: 1957–1961

Morton NE (1992) Major loci for atopy? Clin Exp Allergy 22: 1041–1043

Neubauer BA, Fiedler B, Himmelein B et al (1998) Centrotemporal spikes in families with rolandic epilepsy: linkage to chromosome 15q14. Neurology 51: 1608–1612

Nimgaonkar VL, Rudert WA, Zhang X et al (1997) Negative association of schizophrenia with HLA DQB1*0602: evidence from a second African-American cohort. Schizophr Res 23: 81–86

Norman RM, Malla AK (1993) Stressful life events and schizophrenia. I. A review of the research. Br J Psychiatry 162: 161–166

Nuechterlein KH, Dawson ME et al (1990) Testing vulnerability models. In: Häfner H, Gattaz WF, Janzarik W (eds) Search for the causes of schizophrenia, vol 2. Springer, Berlin Heidelberg New York, pp 177–191

O'Donovan MC, Guy C, Craddock N et al (1996) Confirmation of association between expanded CAG/CTG repeats and both schizophrenia and bipolar disorder. Psychol Med 26: 1145–1153

Ott J (1996) Complex traits on the map. Nature 379: 772–773

Pawlowtzki IH, Edwards JH, Thompson E (eds) (1997) Genetic mapping of disease genes. Academic, New York

Peltonen L (1995) All out for chromosome six. Nature 378: 665–666

Persico AM, Macciardi F (1997) Genotypic association between dopamine transporter gene polymorphisms and schizophrenia. Am J Med Genet 74: 53–57

Petronis A, Kennedy JL (1995) Unstable genes – unstable mind? Am J Psychiatry 152: 164–172

**Plomin R, Defries JC, McClearn GE, McGuffin P (2000) Behavioral genetics. Freeman, New York

Propping P (1983) Genetic disorders presenting as "schizophrenia". Karl Bonhoeffer's early view of the psychoses in the light of medical genetics. Hum Genet 65: 1–10

**Propping P (1989) Psychiatrische Genetik. Springer, Berlin Heidelberg New York

**Propping P (1992) Abuse of genetics in Nazi Germany. Am J Hum Genet 51: 909–910

Pulver AE, Moorman CC, Brown CH et al (1990) Age-incidence artefacts do not account for the season-of-birth effect in schizophrenia. Schizophr Bull 16: 13–15

Pulver AE, Karayiorgou M, Wolyniec PS et al (1994a) Sequential strategy to identify a susceptibility gene for schizophrenia: report of potential linkage on chromosome 22q12-q13.1, part 1. Am J Med Genet 54: 36–43

Pulver AE, Karayiorgou M, Lasseter VK et al (1994b) Follow-up of a report of a potential linkage for schizophrenia on chromosome 22q12-q13.1, part 2. Am J Med Genet 54: 44–50

Pulver AE, Lasseter VK, Kasch L et al (1995) Schizophrenia: a genome scan targets chromosomes 3p and 8p as potential sites of susceptibility genes. Am J Med Genet 60: 252–260

Rao DC, Morton NE, Gottesman II, Lew R (1981) Path analysis of qualitative data on pairs of relatives: application to schizophrenia. Hum Hered 31(6): 325–333

Risch N (1990) Linkage strategies for genetically complex traits. I. Multilocus models. Am J Hum Genet 46: 222–228

Risch N, Baron M (1984) Segregation analysis of schizophrenia and related disorders. Am J Hum Genet 36: 1039–1059

*Risch N, Merikangas K (1996) The future of genetic studies of complex human diseases. Science 273: 1516–1517

Rowen L, Mahairas G, Hood L (1997) Sequencing the human genome. Science 278: 605–607

Schuler GD, Boguski MS, Stewart EA et al (1996) A gene map of the human genome. Science 274: 540–546

Schulz B (1932) Zur Erbpathologie der Schizophrenie. Z Ges Neurol Psychiatr 143: 175–293

Schwab SG, Lerer B, Albus M et al (1995a) Potential linkage for schizophrenia on chromosome 22q12-q13: a replication study. Am J Med Genet 60: 436–443

Schwab SG, Albus M, Hallmayer J et al (1995b) Evaluation of a susceptibility gene for schizophrenia on chromosome 6p by multipoint affected sib-pair linkage analysis. Nature Genet 11: 325–327

Schwab SG, Eckstein GN, Hallmayer J et al (1997) Evidence suggestive of a locus on chromosome 5q31 contributing to susceptibility for schizophrenia in German and Israeli families by multipoint affected sib-pair linkage analysis. Mol Psychiatry 2: 156–160

Sherrington R, Brynjolfsson J, Petursson H et al (1988) Localization of a susceptibility locus for schizophrenia on chromosome 5. Nature 336: 164–167

Shirakawa T, Li A, Dubowitz M et al (1994) Association between atopy and variants of the β subunit of the high-affinity immunoglobulin E receptor. Nature Genet 7: 125–130

Silverman JM, Greenberg DA, Altstiel LD et al (1996) Evidence of a locus for schizophrenia and related disorders on the short arm of chromosome 5 in a large pedigree. Am J Med Genet 67: 162–171

Sing CF, Haviland MB, Reilly SL (1996) Genetic architecture of common multifactorial diseases. Ciba Found Symp 197: 211–232

SLCG (Schizophrenia Linkage Collaborative Group for Chromosomes 3, 6 and 8) (1996) Additional support for schizophrenia linkage on chromosomes 6 and 8: a multicenter study. Am J Med Genet 67: 580–594

St Clair D, Blackwood D, Muir W et al (1990) Association within a family of a balanced autosomal translocation with major mental illness. Lancet 336: 13–16

Straub RE, MacLean CJ, O'Neill FA et al (1995) A potential vulnerability locus for schizophrenia on chromosome 6p24-22: evidence for genetic heterogeneity. Nature Genet 11: 287–293

Straub RE, MacLean CJ, Kendler KS (1996) The putative schizophrenia locus on chromosome 6p: a brief overview of linkage studies. Mol Psychiatry 1: 84–92

Straub RE, MacLean CJ, O'Neill FA et al (1997) Support for a possible schizophrenia vulnerability locus in region 5q22–31 in Irish families. Mol Psychiatry 2(2): 148–155

*Suarez B, Hampe CL, van Eerdewegh P (1994) Problems of replicating linkage claims in psychiatry. In: Gershon ES, Cloninger CR, Barrett JE (eds) Genetic approaches to mental disorders. American Psychiatric Press, Washington DC, pp 23–46

Suslow T, Arolt V (1996) Disorders of early information processing and vigilance as vulnerability markers for schizophrenia. Fortschr Neurol Psychiatr 64: 90–104

Thibaut F, Martinez M, Petit M et al (1995) Further evidence for anticipation in schizophrenia. Psychiatry Res 59: 25–33

Tienari P (1991) Interaction between genetic vulnerability and family environment: the Finnish adoptive family study of schizophrenia. Acta Psychiatr Scand 84: 460–465

Tienari P, Wynne LC, Moring J et al (1994) The Finnish adoptive family study of schizophrenia. Implications for family research. Br J Psychiatry Suppl 23: 20–26

Thorlacius S, Tryggvadottir L, Olafsdottir GH et al (1995) Linkage to BRCA2 region in hereditary male breast cancer. Lancet 346: 544–545

Torrey EF (1992) Are we overestimating the genetic contribution to schizophrenia? Schizophr Bull 18: 159–170

Tsuang MT, Lyons MJ, Faraone SV (1990) Heterogeneity of schizophrenia. Conceptual models and analytic strategies. Br J Psychiatry 156: 17–26

Vallada H, Curtis D, Sham PC et al (1995) Chromosome 22 markers demonstrate transmission disequilibrium with schizophrenia. Psychiatr Genet 5: 127–130

Vinogradov S, Gottesman II, Moises HW, Nicol S (1991) Negative association between schizophrenia and rheumatoid arthritis. Schizophr Bull 17: 669–678

**Vogel F, Motulsky AG (1996) Human genetics: problems and approaches, 3rd edn. Springer, Berlin Heidelberg New York

Wahlberg KE, Wynne LC, Oja H et al (1997) Gene-environment interaction in vulnerability to schizophrenia: findings from the Finnish adoptive family study of schizophrenia. Am J Psychiatry 154: 355–362

Wang S, Detera-Wadleigh SD, Coon H et al (1996) Evidence of linkage disequilibrium between schizophrenia and the SCA1 CAG repeat on chromosome 6p23. Am J Hum Genet 59: 731–736

*Watson JD (1997) Genes and politics. J Mol Med 75: 624–636

Weber MM (1996) Ernst Rüdin, 1874–1952: a German psychiatrist and geneticist. Am J Med Genet 67: 323–331

Weeks DE, Lathrop GM (1995) Polygenic disease: methods for mapping complex disease traits. Trends Genet 11: 513–519

Wei J, Xu HM, Ramchand CN, Hemmings GP et al (1997) Is the polymorphic microsatellite repeat of the dopamine beta-hydroxylase gene associated with biochemical variability of the catecholamine pathway in schizophrenia? Biol Psychiatry 41: 762–767

Weissenbach J, Gyapay G, Dib C et al (1992) A second-generation linkage map of the human genome. Nature 359: 794–801

Weyand CM, Goronzy JJ (1995) Inherited and noninherited risk factors in rheumatoid arthritis. Curr Opin Rheumatol 7: 206–213

Wildenauer DB, Hallmayer J, Schwab SG et al (1996) Searching for susceptibility genes in schizophrenia by genetic linkage analysis. Cold Spring Harb Symp Quant Biol 61: 845–850

Williams J, Spurlock G, McGuffin P et al (1996) Association between schizophrenia and T102C polymorphism of the 5-hydroxytryptamine type 2a-receptor gene. European Multicentre Association Study of Schizophrenia (EMASS) Group. Lancet 347: 1294–1296

Williams NM, Cardno AG, Murphy KC et al (1997) Association between schizophrenia and a microsatellite polymorphism at the dopamine D_5 receptor gene. Psychiatr Genet 7: 83–85

Woolf CM (1997) Does the genotype for schizophrenia often remain unexpressed because of canalization and stochastic events during development? Psychol Med 27: 659–668

Wright P, Donaldson PT, Underhill JA et al (1996) Genetic association of the HLA DRB1 gene locus on chromosome 6p21.3 with schizophrenia. Am J Psychiatry 153: 1530–1533

Zamani MG, De Hert M, Spaepen M et al (1994) Study of the possible association of HLA class II, CD4, and CD3 polymorphisms with schizophrenia. Am J Med Genet 54: 372–377

CHAPTER

4

A. Heinz, D.R. Weinberger

Schizophrenia: The Neurodevelopmental Hypothesis

1 Historical Background

1.1 Schizophrenia as a Premature Dementia

Kraepelin conceptualized schizophrenia as a premature dementia, presenting in the second or third decade of life and deteriorating over the course of the illness (Kraepelin 1899). Both Bleuler (1911) and Kraepelin (1899, 1919) noticed that childhood characteristics such as seclusion, withdrawal, or irritability preceded the manifestation of psychotic symptoms in some schizophrenic patients. However, the prevailing biological perspective suggested that schizophrenia results from a pathological degenerative process that occurs in early adult life shortly before the onset of manifest symptoms. It was generally assumed that, in most cases, the brain was relatively normal until the illness struck and that the pathological changes would increase as the illness progressed (Weinberger 1995).

1.2 Schizophrenia as a Neurodevelopmental Disorder

In recent years, a fundamental shift in the perception of the underlying neurobiological process has occurred. Most researchers now see schizophrenia as a neurodevelopmental disorder and suggest that many or most cases of schizophrenia are caused by a defect in early brain development (Weinberger 1987, 1995; Murray et al. 1988; Crow et al. 1989). It is hypothesized that the neurodevelopmental abnormality will not manifest in a diagnostically recognizable form until after a considerable postnatal delay period and that normal developmental processes or additional pathogenic factors trigger the manifestation of the illness (Weinberger 1995).

1.3 Why the Shift in Neurobiological Disease Concepts Occurred

The reasons for this conceptual shift come from different areas of schizophrenia research. In neuropathological and brain imaging studies of schizophrenics, adult-onset pathological cerebral changes and indicators of pathological progress have remained elusive, while replicable evidence of early cortical maldevelopment was found (Bachneff 1991; Weinberger 1995). Studies of childhood behavior in subsequent schizophrenics have indicated a considerable degree of psychomotor anomalies, which has been attributed to abnormal brain development (Walker et al. 1994). Furthermore, neurobiological models exist which may explain the relationship between brain maldevelopment and the clinical features of schizophrenia (Deutch 1992; Weinberger and Lipska 1995). This evidence links schizophrenia to neuropathological changes in early brain development.

2 Indicators of Developmental Neuropathology

In 1982, Feinberg suggested that schizophrenia may be due to a fault in synaptic elimination during adolescence. While this hypothesis did not account for the neurodevelopmental changes that have later been associated with schizophrenia, it was an impetus to examine the question of neurodevelopmental problems in schizophrenia. Indicators of abnormal intrauterine development range from rather weak and circumstantial findings such as a slight overrepresentation of minor physical abnormalities (MPA) to compelling evidence of in vivo and ex vivo cerebral morphometric abnormalities. While the sum of these findings can be interpreted as a coherent story of developmental abnormalities, inconsistencies and methodological problems exist at the level of individual findings (Weinberger 1996).

2.1 Minor Physical Abnormalities

2.1.1 Abnormalities as an Indication of Developmental Pathology

Abnormal intrauterine events are expected to affect the development of extracerebral tissue and hence MPA such as deformed ears or high palate have been found in patients with neurodevelopmental disorders. In schizophrenic patients, increased frequency of MPA was associated with greater dermatoglyphic asymmetry (total finger ridge count and left–right ridge count), which is supposed to reflect second-trimester asymmetry (Torrey et al. 1994a). In monozygotic twins discordant for schizophrenia, a subgroup of patients displayed increased frequencies of MPA, dermatoglyphic asymmetry, and perinatal complications or low birth weight. These findings were interpreted as indicators of prenatal events, inducing neurodevelopmental abnormalities (Bracha et al. 1992; Davis and Bracha 1996). Several studies reported an excess of MPA in schizophrenics. The occurrence of MPA in schizophrenic patients has therefore been interpreted

as an indication of intrauterine developmental pathology which may have affected neuronal organization in the second or third trimester of gestation (O'Connell et al. 1997).

2.1.2 Inconsistencies in Studies

In spite of the plausibility of second-trimester insults in the etiopathogenesis of schizophrenia, fundamental problems with the data exist. No consistent pattern of association between certain MPA and schizophrenia has been found. Studies that report increases in MPA tend to lump them together as if they all signify a similar neurodevelopmental pathology, when, in fact, their relative frequencies are not correlated with each other (Torrey et al. 1994b). Moreover, association studies usually lacked well-matched control samples, and none of the studies seems to have been carried out as a blind comparison. Finally, an association between certain MPA and schizophrenia was often noticed only after the sample of schizophrenic patients had been broken down post hoc into putative subgroups. Currently, the existing data do not support the notion that MPA are good indicators of cerebral maldevelopment in schizophrenia (Weinberger 1995).

2.2 Ventricular Enlargement and Reduction of Cerebral Volume

2.2.1 Ventricular Enlargement in Neuroimaging Studies

Enlarged cerebral ventricles in postmortem specimens from hebephrenic patients were described by Hecker as early as 1871. In vivo imaging studies show that ventricular enlargement is the most frequently confirmed neurobiological finding in schizophrenic patients (Van Horn and McManus 1992). The interindividual variation in ventricular size is large and effect size of the finding in schizophrenics is relatively small; therefore, negative reports have been published, especially in studies with a small sample size. However, in studies of monozygotic twins discordant for schizophrenia, even affected twins with relatively small ventricles had larger ventricles than their healthy co-twins (Reveley et al. 1982; Suddath et al. 1990). These results suggest that the underlying pathological process is a subtle but commonplace feature of the disorder. They do not, however, address the question of whether these changes are due to neurodevelopmental pathology.

2.2.2 Lack of Brain Tissue in Neuroimaging Studies

Enlarged brain ventricles indicate a nonproduction or reduction of brain tissue. In fact, morphometric assessment of brain structures with high-resolution magnetic resonance imaging (MRI) techniques have provided evidence for an overall reduction in gray matter volume by 5%–10% in schizophrenic patients compared to healthy control subjects (Andreasen et al. 1994a; Zipursky et al. 1994). The loss of cortical volume was most frequently reported for the mesial temporal cortex in the area of the rostral hippocampus and for the superior temporal gyrus (Bogerts et al. 1990; Suddath et al. 1990; Shenton et al. 1992). Reductions in the volumes of other brain areas such as the prefrontal and parietal cortices, the basal ganglia, and the thalamus have also been reported, but the findings were less consistent (Andreasen et al. 1994b; Schlaepfer et al. 1994).

2.2.3 Indicators of Lateralized Brain Pathology

Some controversy exists around the question of whether the pathological changes are pronounced in the left hemisphere, especially in the left temporal lobe (Crow et al. 1989). However, most studies describe bilateral differences between patients and controls (Crow et al. 1989; Shenton et al. 1992). A related issue involves the question of whether schizophrenia is associated with a pathology in the development of normal human cerebral asymmetries, e.g. in the width of the occipital lobes or the lengths of the sylvian fissures. Normal human brain asymmetries are formed in the second trimester of gestation (Chi et al. 1977), so that abnormalities in these asymmetries might indicate second-trimester insults (Roberts 1991). These observations are interesting in light of findings indicating that patients with schizophrenia may be less completely lateralized than normal individuals when handedness, dichotic listening asymmetries, and lateralized cognitive tasks are assessed (Gruzelier et al. 1988). However, negative studies are common (Bartley et al. 1993; Kulynych et al. 1996). A reason for these contradictory results might be that asymmetries are less obvious in individuals who are incompletely right-handed and that mixed dominance appears to be more frequent in schizophrenic patients (Myslobodsky and Weinberger 1987). Incomplete anatomical asymmetries may therefore be a nonspecific sign of developmental variance which is found independent of schizophrenia and is not associated with schizophrenia per se (Weinberger 1995).

2.2.4 Attempts to Confirm the Reductions in Cerebral Volume by Postmortem Studies

Postmortem studies of brains of schizophrenic patients found increased ventricular size and reductions in various cortical regions, especially the hippocampal formation, confirming the results of the in vivo imaging studies (Bogerts et al. 1985; Crow et al. 1989). In accordance with these observations, neuronal counts were reduced in certain cortical and periventricular regions (Falkai and Bogerts 1995). Altogether, a fairly widespread neuropathological process was implicated. Some studies, however, did not confirm these findings (Heckers et al. 1990; Benes et al. 1991a); the reasons for these inconsistencies are not known (Weinberger 1995).

2.3 Clinical Correlates of Brain Abnormalities in Neuroimaging Studies

2.3.1 Lack of Correlation Between Ventricular Enlargement and Reductions in Brain Tissue and the Duration of Illness

The observation of ventricular enlargement and reduction of cerebral tissue in schizophrenic patients does not address the question of when these abnormalities were acquired. A comparison of brain imaging and clinical data shows that ventricular enlargement does not correlate with the duration of illness, which would have been expected if the progression of the illness were accompanied by an ongoing pathological process causing brain atrophy and ventricular enlargement (Weinberger et al. 1979; Raz and Raz 1990). Studies in first-break patients show that ventricular enlargement and reduced hippocampal volume were present at the onset of the clinical illness (Bogerts et al. 1990; Lieberman et al. 1993). These findings indicate that ventricular enlargement and reduction in brain tissue predate illness onset, perhaps by many years.

2.3.2 Clinical Correlates of Ventricular Enlargement

When the pathopsychological correlates of ventricular enlargement in adult schizophrenic patients were studied, an association with poor premorbid social and educational adjustment during early childhood was found (Weinberger et al. 1980). It was this rather surprising finding which first alerted researchers that the neuropathological abnormalities may have occurred years before the onset of psychotic symptoms (Weinberger 1995). Further studies showed correlations between ventricular enlargement and obstetric complications, supporting the hypothesis that the underlying pathological process may originate from early in life (Owen et al. 1988).

2.4 Clinical Indicators of Early Neurodevelopmental Pathology

2.4.1 Premorbid Neuropsychological Tests and Social Functioning

If a prenatal neuropathological process occurred in subsequent schizophrenics, subtle indications of abnormal neural function would be expected to be present during childhood. Several studies examined premorbid neuropsychological tests and found that individuals who later became schizophrenic did worse than their healthy siblings, although negative results have also appeared (Torrey et al. 1994a,b). In the prospective British National Child Development Study, reduced premorbid social functioning and increased social anxiety during childhood was found in patients later admitted for treatment of schizophrenia (Done et al. 1994). In a second study of childhood neurofunctional development that examined the birth cohort of 1946, gross motor and speech milestones were significantly delayed and educational test scores were significantly lower in patients later treated for schizophrenia (Jones et al. 1994).

2.4.2 Indicators of Abnormal Neurological Functioning in Childhood

Walker and coworkers examined home movies filmed during the childhood of adult-onset schizophrenic patients and their siblings. In a blind comparison, the affected family member was invariably identified by subtle neuromotor abnormalities such as slight posturing of the hand or transient choreoathetoid movements. The deficits were quite subtle, occurred primarily on the left side of the body, and tended to disappear after 2 years of life (Walker and Lewine 1990; Walker et al. 1994). Follow-up studies showed that these abnormalities were pronounced only in a subgroup of children who later developed schizophrenia (Neumann et al. 1995).

2.4.3 Association of Childhood Neuromotor Abnormalities and Ventricular Enlargement

In adult patients, childhood neuromotor abnormalities and negative affect were associated with ventricular

enlargement (Walker et al. 1996). This finding supports the hypothesis that ventricular enlargement in schizophrenics results from early neurodevelopmental malfunction and not from a disease process active during the clinical onset of manifest psychotic symptoms (Weinberger 1995).

2.5 Indicators of Neurodevelopmental Pathology in Postmortem Findings

Postmortem studies generally reported a lack of gliosis in neocortex, hippocampus, and parahippocampal cortices. This finding is inconsistent with adult-onset neurodegenerative conditions, but has been found in neuropathologial events that occur early in development, i.e., before the third trimester of gestation (Knable and Weinberger 1995). The presence of cortical changes without gliosis has been interpreted as circumstantial evidence that the neuropathological changes observed in brain tissue of schizophrenics occur prenatally. As a general disruption of the earliest events in cortical maturation is associated with severe abnormalities not found in schizophrenia, it is unlikely that the observed neuropathological changes result from insults before the sixth fetal week (Rakic 1988). The neuropathological findings therefore implicate a time of insult around the second trimester of gestation.

2.6 Cytoarchitectural Abnormalities

Laminar patterns of cortical neurons, their orientation, and internal relationships are formed during the second trimester of gestation and are assumed not to change during life. Abnormalities of cytoarchitecture would therefore strongly implicate prenatal pathological neurodevelopment. Different studies of cytoarchitecture in schizophrenic brain tissue found structural abnormalities, which implicate pathological development.

2.6.1 Orientation of Hippocampal Pyramidal Cells

Abnormal orientation of hippocampal pyramidal cells was found by Kovelman and Scheibel (1984) in the left hemispheres of chronic schizophrenic patients. Similar alterations were found in the right hemisphere by Conrad et al. (1991). This finding was interpreted as indicating a defect in neuronal migration. However, several studies failed to replicate these findings (Weinberger 1995).

2.6.2 Laminar Organization in the Entorhinal Cortex

What is potentially a landmark study on laminar organization in the entorhinal cortex was published by Jakob and Beckmann in 1986. They studied the entorhinal cortex in the region of the amygdala and pes hippocampus, where the most consistent reductions in brain tissue have been found in postmortem and in vivo studies (Weinberger 1995). Using Nissl-stained sections, Jakob and Beckmann (1986) found attenuation of cellularity in layers I and II, incomplete clustering of neurons into normal glomerular structures in layer II, and the presence of such clusters in deeper layers, where they are normally not found. This observation may indicate a disruption of neuronal migration to cortex in the second trimester of gestation, in which neurons proceed outward from the periventricular zone to reach their cortical destination. In this process, younger neurons have to pass over older ones on their way to more superficial layers (Rakic 1988). Jakob and Beckmann postulated that their findings may indicate an arrest of migration, in which a younger generation of neurons was held up in deeper layers.

2.6.3 Attempts to Replicate Evidence for a Developmental Pathology in the Entorhinal Cortex

Using the same technique, Arnold et al. (1991) found similar abnormalities as described by Jakob and Beckmann (1986). Moreover, they observed anomalous entorhinal sulci in their specimens. Like Jakob and Beckmann (1986), they interpreted their findings as indications of an abnormality of entorhinal cortex development, hypothetically due to a migration failure. Several studies, however, failed to replicate these findings. Hyde and Saunders (1991) observed that more caudal areas of the entorhinal cortex in normal controls look increasingly like the findings reported by Jakob and Beckmann (1986) in schizophrenic brains, indicating that slight differences in the examined entorhinal areas can produce substantially different cytoarchitectural results. A study by Heinsen and Gössman (1996) supports the hypothesis that the variability in the human entorhinal cortex may confound the detection of neurodevelopmental abnormalities in the brains of schizophrenic patients. The authors argued that in the rostromedial entorhinal area, regionally circumscribed features of so-called pre-beta clusters were probably misinterpreted as neurodevelopmental disturbances by Jakob and Beckmann

(1986). Moreover, Heinsen and Gössman (1996) did not find an association between "anomalous" entorhinal sulci and schizophrenia, as suggested by Arnold et al. (1991). Two attempts to directly replicate the findings of Jakob and Beckmann (1986) failed and supported the notion that disordered rows of neurons in the superficial layers of schizophrenic patients are characteristic of the olfactory region of normal brains (Akil and Lewis 1997; Krimer et al. 1997). These studies do not support the hypothesis that a migration failure of cortical neurons in the entorhinal cortex is associated with schizophrenia, although it does not rule out the existence of other abnormalities of neuronal circuitry in this area.

2.6.4 Cytotarchitectural Abnormalities in the Frontal Cortex

Akbarian et al. (1993a) studied primarily GABAergic neurons that express the enzyme nicotinamide adenine dinucleotide phosphate diaphorase (NADPH-d) in the superior frontal gyrus region of the dorsolateral prefrontal cortex. They found a reduction in the number of neurons in cortical layers I–III and increased numbers in deep layers, a finding which resembled the results of Jakob and Beckmann (1986) in the entorhinal cortex. Similar results were reported by Benes et al. (1991b), who studied small, presumably GABAergic cells in the prefrontal and cingulate cortex of schizophrenic patients. They found a reduction in small neurons in the superficial layers and larger numbers in deeper layers. These findings might indicate a failure of neuronal migration to superficial layers during the second trimester of gestation (Weinberger 1995).

2.6.5 Attempts to Replicate Other Evidence for a Cortical Developmental Pathology

An attempt to replicate the observations of Benes et al. (1991b) failed (Bunney et al. 1993). In a study of the lateral temporal neocortex and the mesial limbic cortex of schizophrenics, Akbarian et al. (1993b) observed similar cytoarchitectural abnormalities in NADPH-d neurons as in the prefrontal cortex (Akbarian et al. 1993b). However, while these abnormalities were present in the lateral temporal cortex and the hippocampus, the entorhinal cortex was normal (Akbarian et al. 1993b). It is possible that this failure to replicate the observations of Jakob and Beckmann (1986) is due to the fact that Akbarian et al. (1993b) studied primarily GABAergic neurons, while the layer II entorhinal neurons examined by Jakob and Beckmann (1986) are primarily glutamatergic. On the other hand, the presence of abnormalities in the cytoarchitectural distribution of NADPH-d neurons in other cortical areas (Akbarian et al. 1993a) and the failure of other groups to replicate cytoarchitectural abnormalities in the entorhinal cortex (Akil and Lewis 1997; Krimer et al. 1997) do not support the hypothesis of a developmental defect in the entorhinal cortices of schizophrenic patients.

2.6.6 Overall Outcome of Studies Assessing Neurodevelopmental Abnormalities in Schizophrenic Patients

Several studies of cortical cytoarchitecture in schizophrenics suggest subtle multifocal or diffuse anatomical deviations that are hard to explain on the basis of anything that might happen to a brain after birth (Weinberger 1996). Specifically, they point to a disruption of neuronal migration to superficial cortical layers in the second trimester of gestation. However, various inconsistencies exist and attempts to replicate important findings have repeatedly failed. Further studies are therefore needed before the uncertainties can be resolved (Weinberger 1995).

3 Etiologic Considerations

The etiology of the potential developmental defect is unknown. It is reasonable to assume that several genetic and environmental factors play a part. Recent studies implicated environmental influences such as obstetric complications, in utero virus infections, and malnutrition, at least in some patients (O'Connell et al. 1997).

3.1 Obstetric Abnormalities

Abnormal fetal development is often associated with abnormal delivery. The literature on this topic is difficult to interpret, as the same methods are rarely used and the same findings are rarely replicated. Sibship studies are important because uncertainty about the validity of maternal recall is less problematic (Weinberger 1995). Within these limitations, a majority of studies have found a trend toward more frequent complications during both pregnancy and delivery in births of individuals who subsequently suffered from schizophrenia (McNeil 1988). However, obstetric complications are a poor predictor of schizophrenia, increasing the risk of schizophrenia by at most 1% (Goodman 1988).

3.2
Prenatal Fetal Abnormalities as a Possible Cause of Obstetric Complications

Perinatal injuries are typically characterized by gliosis. It is therefore unlikely that perinatal obstetric complications are directly associated with the neuropathological findings of schizophrenic brain tissue described above (Knable and Weinberger 1995). Instead, it has been hypothesized that preexisting fetal abnormalities predispose to an increased frequency of obstetric complications (Weinberger 1995), an interpretation first proposed in 1897 by Freud (1968) in reference to cerebral palsy.

3.3
Prenatal Viral Exposure

Mednick et al. (1988) found an association between the manifestation of schizophrenia and the Helsinki influenza A2 epidemic of 1957. Patients who had been in their second trimester during the height of the epidemic had a significantly higher percentage of subsequent schizophrenic psychoses. This result stimulated a series of subsequent studies examining the risk of schizophrenia after potential intrauterine exposure to influenza in Europe and the United States. A slight majority of studies claimed to find such an association; however, it accounted for at most 4% of the variance (Weinberger 1995). A limitation of these studies is that they usually did not document actual maternal infection and reported associations only between the general risk of intrauteral exposure to influenza and manifestation of schizophrenia in adulthood. One study assessed children of mothers who had been diagnosed as having influenza during their second trimester of pregnancy. No increased risk of schizophrenia was found in the children of these mothers (Crow and Done 1992). However, the study was criticized for possibly having a low ascertainment rate of cases of schizophrenia (Weinberger 1995). In conclusion, the perinatal exposure literature is provocative but inconclusive. Even if validated by further studies, viral exposure will account for at most a small percentage of cases (Weinberger 1995).

3.4
Prenatal Malnutrition

Another environmental risk factor was addressed by the study carried out by Susser and Lin (1992), who examined a Dutch birth cohort exposed to starvation caused by the Nazi blockade of the Netherlands. They observed a twofold increase in the subsequent manifestation of schizophrenia in probands who had been exposed to severe famine during the first 2 months of gestation. As in the case of influenza exposition, severe malnutrition in utero is an unusual example of prenatal environmental stress and probably accounts for at most a very small number of cases. Nevertheless, the statistical association of these intrauterine insults with an increased risk for schizophrenia supports the hypothesis that certain disturbances of the intrauterine environment can help set the stage for the manifestation of schizophrenia in adulthood (Weinberger 1995).

3.5
Genetic Factors

It is widely accepted that genetic factors convey susceptibility to schizophrenia. Since approximately 30% of the genome is expressed in the brain, and many genes are turned on and off during discrete phases of brain development, it has been speculated that genetic factors may induce cerebral maldevelopment in schizophrenia. Different mechanisms such as a primary genetic defect or an increased vulnerability to intrauterine or perinatal environmental stresses have been suggested. As no current data link schizophrenia with a defect in any known gene related to brain development, these hypotheses require further research (Weinberger 1995; O'Connell et al. 1997).

4
Hypothetical Mechanisms of Delayed Manifestation of Psychotic Symptoms

4.1
Normal Brain Development Versus a Second Pathological Process

If schizophrenia is related to an abnormality of early brain development, potential mechanisms of delayed onset of the clinical illness have to be addressed. It has been hypothesized that the manifestation of psychotic symptoms is either due to an additional, independent pathological process or to an interaction between a static neurodevelopmental defect and normal development programs. Postulating an additional pathological process that is active around the age of illness onset requires the unlikely scenario of a second primary pathology (Weinberger 1995).

4.2 Neurological Analogies

It has been observed in certain neurological diseases that the likelihood of psychotic symptoms is related to the state of brain development, even if the neuropathological changes do not vary with age. Metachromatic leukodystrophy, for example, results in a "dysconnection" of cortical regions, and patients present with psychotic symptoms nearly exclusively between the ages of 13 and 30 years. When metachromatic leukodystrophy appears outside of this critical age, psychotic symptoms are almost never found. This example shows that a dysconnection of cortical regions can lead to the manifestation of psychotic symptoms if it interacts with a certain stage of brain development (Weinberger 1995).

4.3 Prefrontal Dysfunction

Brain imaging studies under certain conditions such as the Wisconsin Card Sorting Test showed a hypoactivity of the prefrontal cortex, which was reversed when amphetamine was applied, hypothetically reversing a prefrontal dopamine deficit (Daniel et al. 1991). Weinberger (1987) hypothesized that a dysfunction of the prefrontal cortex is associated with cognitive deficits and negative symptoms and induces a disinhibition of subcortical dopamine release, which may be associated with the manifestation of positive symptoms. Thus a prefrontal dopamine deficit may be due to neurodevelopmental pathology and result in a failure to regulate subcortical dopaminergic transmission.

4.4 Animal Models of Cortico-subcortical Dopaminergic Dysregulation

In certain animal models, a prefrontal dopaminergic lesion disinhibits subcortical dopamine release, especially under stress conditions. However, the effect is usually not persistent, and no long-term delay in the onset of dopaminergic dysregulation has been observed (Deutch 1992). A neonatal lesion of the mesial temporolimbic cortex in rats is, on the other hand, clinically silent until puberty. After puberty, the neonatally lesioned rats become hyperresponsive to dopaminergic drugs and to environmental stresses, behaviors that are ameliorated by antidopaminergic drugs (Weinberger and Lipska 1995). These findings indicate that a neonatally acquired lesion of the mesial temporal cortex can lead to a delayed onset of a stress-related dysfunction of dopaminergic transmission.

4.5 Stimulation of Dopaminergic Transmission by Stress

Studies in rhesus monkeys illustrate potential mechanisms by which environmental stress factors can trigger a dysfunction of dopaminergic transmission in subjects with a neonatal mesiotemporal lesion. Environmental stress can trigger dopamine release, especially in the prefrontal cortex (Abercrombie et al. 1989). Prefrontal monoaminergic stimulation reduces subcortical dopamine release in normal adult monkeys and in monkeys with an adult lesion of the mesial temporal cortex (Kolachana et al. 1995). Adult monkeys with a neonatal lesion of the mesial temporal cortex, on the other hand, show signs of a dysregulation of striatal dopaminergic transmission which seems to be due to a failure of prefrontal regulation of subcortical dopamine release (Saunders et al. 1998; Heinz et al. 1999). A failure of the prefrontal regulation of subcortical dopamine release might involve a dysfunction of inhibitory GABAergic interneurons in the prefrontal cortex (Bachneff 1991) and a subsequent disinhibition of glutamatergic projections to midbrain dopaminergic neurons (Taber et al. 1995). Olney and Farber (1996) hypothesized that a dysfunction of GABAergic interneurons may be due to *N*-methyl-D-aspartate (NMDA) receptor hypofunction present in the developing human brain at birth. However, GABAergic medication fails to improve schizophrenic symptoms, and animal models of neonatally induced NMDA receptor dysfunction displayed gliosis, which is absent in schizophrenia (Olney and Farber 1995; Knable and Weinberger 1995). Whatever the cause, an early developmental dysconnection between the temporolimbic and prefrontal cortex may lead to a stimulus-dependent disinhibition of subcortical dopaminergic transmission in adolescence, when cortico-subcortical connections are restructured (Anderson et al. 1995) and the stresses of independent adult life are present (Weinberger 1987).

5 Conclusions

The neurodevelopmental hypothesis of schizophrenia has developed over the past years as a heuristically useful approach to study schizophrenia. It was stimulated by the static nature of ventricular enlargement and cerebral brain reduction in in vivo studies, the correlations with early life adaptation, and the absence of gliosis in neuropathological studies. The neurodevelopmental hypothesis has stimulated new lines of research such as developmental neurobiology and clinical epidemiology and is responsible for animal

models of core neurobiological constellations of the syndrome which attest to the neurobiological plausibility of the hypothesis. Inconsistencies exist within the data implicating developmental neuropathology, which are either indirect or based on inconclusive studies. Further research will be necessary to support or falsify this cornerstone of the neurodevelopmental hypothesis of schizophrenia.

6 References

Abercrombie ED, Keefe KA et al (1989) Differential effects of stress on in vivo dopamine release in striatum, nucleus accumbens, and medial prefrontal cortex. J Neurochem 52: 1655–1658

*Akbarian S, Bunney WE, Potkin SG et al (1993a) Altered distribution of nicotine-adenine dinucleotide phosphate-diaphorase cells in frontal lobe of schizophrenics implies disturbance of cortical development. Arch Gen Psychiatry 50: 169–177

Akbarian S, Vinuela A, Kim JJ et al (1993b) Distorted distribution of nicotine-adenine dinucleotide phosphate-diaphorase cells in temporal lobe of schizophrenics implies anomalous cortical development. Arch Gen Psychiatry 50: 178–187

Akil M, Lewis DA (1997) Cytoarchitecture of the entorhinal cortex in schizophrenia. Am J Psychiatry 154: 1010–1012

Andreasen NC, Flashman L, Flaum M et al (1994a) Regional brain abnormalities in schizophrenia measured with magnetic resonance imaging. JAMA 272: 1763–1769

Andreasen NC, Arndt S, Swayze V 2nd et al (1994b) Thalamic abnormalities in schizophrenia visualized through magnetic resonance image averaging. Science 266: 294–298

Anderson SA, Classey JD, Conde F et al (1995) Synchronous development of pyramidal neuron dendritic spines and parvalbumin immuno-reactive chandelier neuron axon terminals in layer III of monkey prefrontal cortex. Neuroscience 67: 19–22

Arnold SE, Hyman BT, Van Hoesen GW, Damasio AR (1991) Some cytoarchitectural abnormalities of the entorhinal cortex in schizophrenia. Arch Gen Psychiatry 48: 625–632

Bachneff SA (1991) Positron emission tomography and magnetic resonance imaging: a review and a local circuit neurons hypo(dys)function hypothesis of schizophrenia. Biol Psychiatry 30: 857–886

Bartley AJ, Jones DW, Torrey EI et al (1993) Sylvian fissure asymmetries in monozygotic twins: a test of laterality in schizophrenia. Biol Psychiatry 34: 869–874

Benes FM, Sorensen I, Bird ED (1991a) Reduced neuronal size in posterior hippocampus of schizophrenic patients. Schizophr Bull 17: 597–608

Benes FM, McSparren J, Bird ED et al (1991b) Deficits in small interneurons in prefrontal and cingulate cortices of schizophrenic and schizoaffective patients. Arch Gen Psychiatry 48: 996–1001

Bleuler E (1911) Dementia praecox oder die Gruppe der Schizophrenien. Springer, Berlin

Bogerts B, Meertz E, Schonfeldt-Bausch R (1985) Basal ganglia and limbic system pathology in schizophrenia. A morphometric study of brain volume and shrinkage. Arch Gen Psychiatry 42: 784–791

*Bogerts B, Ashtari M, Degreef G et al (1990) Reduced temporal limbic structure volumes on magnetic resonance images in first-episode schizophrenia. Psychiatr Res Neuroimag 35: 1–13

Bracha HS, Torrey EF, Gottesman II et al (1992) Second-trimester markers of fetal size in schizophrenia: a study of monozygotic twins. Am J Psychiatry 149: 1355–1361

Bunney WE, Akbarian S et al (1993) Gene expression for glutamatergic acid decarboxylase is reduced in prefrontal cortex of schizophrenics. Neurosci Abs 19: 199

Chi JG, Dooling EC, Gilles FH (1977) Gyral development of the human brain. Ann Neurol 1: 86–93

Conrad AJ, Abebe T, Austin R et al (1991) Hippocampal pyramidal cell disarray in schizophrenia as a bilateral phenomenon. Arch Gen Psychiatry 48: 413–417

Crow TJ, Done DJ (1992) Prenatal exposure to influenza does not cause schizophrenia. Br J Psychiatry 161: 390–393

Crow TJ, Ball J, Bloom SR et al (1989) Schizophrenia as an anomaly of development of cerebral asymmetry. A postmortem study and a proposal concerning the genetic basis of the disease. Arch Gen Psychiatry 46: 1145–1150

Daniel DG, Weinberger DR, Jones DW et al (1991) The effect of amphetamine on regional cerebral blood flow during cognitive activation in schizophrenia. J Neurosci 11: 1907–1917

Davis JO, Bracha HS (1996) Famine and schizophrenia: first-trimester malnutrition or second-trimester beriberi. Biol Psychiatry 40: 1–3

Deutch AY (1992) The regulation of subcortical dopamine systems by the prefrontal cortex: interactions of central dopamine systems and the pathogenesis of schizophrenia. J Neural Transm [Suppl] 36: 61–69

Done DJ, Crow TJ, Johnstone EC, Sacker A (1994) Childhood antecendents of schizophrenia and affective illness: social adjustment at ages 7 and 11. Br Med J 309: 699–703

Falkai P, Bogerts B (1995) The neuropathology of schizophrenia. In: Hirsch SR, Weinberger DR (eds) Schizophrenia. Blackwell, Oxford, pp 477–493

Feinberg I (1982–1983) Schizophrenia: caused by a fault in programmed synaptic elimination during adolescence? J Psychiatr Res 17: 319–334

Freud S (1968) Infantile cerebral palsy. University of Miami Press, Miami

Goodman R (1988) Are complications of pregnancy and birth causes of schizophrenia? Dev Med Child Neurol 30: 391–395

Gruzelier J, Seymour K, Wilson L, Jolley A, Hirsch S (1988) Impairments on neuropsychologic tests of temporohippocampal and frontohippocampal functions and word fluency in remitting schizophrenia and affective disorders. Arch Gen Psychiatry 45(7): 623–629

Hecker E (1871) Die Hebephrenia. Arch Pathol Anat Physiol Klin Med 52: 394

Heckers S, Heinsen H, Heinsen YC, Beckmann H (1990) Limbic structures and lateral ventricle in schizophrenia. Arch Gen Psychiatry 47: 1016–1022

Heinsen H, Gössmann E (1996) Variability in the human entorhinal region may confound neuropsychiatric diagnoses. Acta Anat 157: 226–237

Heinz A, Saunders RC, Kolachana BS et al (1999) Striatal dopamine receptors and transporters in monkeys with neonatal temporal limbic damge. Synapse 32: 71–79

Hyde TM, Saunders RC (1991) The entorhinal cortex in humans: a cytoarchitectonic and comparative study with non-human primates. Neurosci Abs 17: 143

**Jakob H, Beckmann H (1986) Prenatal development disturbances in the limbic allocortex in schizophrenics. J Neural Transm 65: 303–326

Jones P, Rodgers B, Murray R, Marmot M (1994) Child development risk factors for adult schizophrenia in the British 1946 birth cohort. Lancet 344: 1398–1402

Knable MB, Weinberger DR (1995) Are mental diseases brain diseases? The contribution of neuropathology to understanding of schizophrenic psychoses. Eur Arch Psychiatry Clin Neurosci 245: 224–230

Kolachana BS, Saunders RC, Weinberger DR (1995) Augmentation of prefrontal cortical monoaminergic activity inhibits dopamine release in the caudate nucleus: an in vivo neurochemical assessment in the rhesus monkey. Neurosci 69: 859–868

Kovelman JA, Sheibel AB (1984) A neurohistological correlate of schizophrenia. Biol Psychiatry 19: 1601–1621

Kraepelin E (1899) Psychiatrie. Ein Lehrbuch für Studirende und Aerzte. Barth, Leipzig

Kraepelin E (1919) Dementia praecox and paraphrenia. Krieger, New York (facsimile edition, 1971)

Krimer LS, Herman MM, Saunders RC et al (1997) A qualitative and quantitative analysis of the entorhinal cortex in schizophrenia. Cereb Cortex 7: 732–739

Kulynych JJ, Vladar K, Fautic BD et al (1996) Normal asymmetry of the planum temporale in patients with schizophrenia: three-dimensional cortical morphometry with MRI. Br J Psychiatry 166: 742–749

Lieberman J, Jody D, Geisler S et al (1993) Time course and biological correlates of Treatment response in first-episode schizophrenia. Arch Gen Psychiatry 50: 369–376

Mednick SA, Machon RA, Huttunen MO, Bonnett D (1988) Adult schizophrenia following prenatal exposure to an influenza epidemic. Arch Gen Psychiatry 45: 189–192

McNeil (1988) Obstetric factors and perinatal injuries. In: Tsuang MT, Simpson JC (eds) Handbook of schizophrenia, vol. 3. Nosology, epidemiology and genetics. Elsevier, Amsterdam, pp 319–344

Murray RM, Lewis SW et al (1988) The neurodevelopmental origins of dementia praecox. In: Bebbington P, McGuffin P (eds) Schizophrenia: the major issues. Heinman, London, pp 90–107

Myslobodsky MS, Weinberger DR (1987) Brain CT asymmetry in schizophrenia and sighting dominance. In: Takahashi R, Flor-Henry P, Gruzelier J, Niwa S (eds) Cerebral dynamics, laterality and psychopathology. Elsevier, Amsterdam, pp 439–448

Neumann CS, Grimes K, Walker EF, Baum K (1995) Developmental pathways to schizophrenia: behavioral subtypes. J Abnorm Psychol 104: 558–566

*O'Connell P, Woodruff PWR, Wright I et al (1997) Developmental insanity or dementia praecox: was the wrong concept adopted? Schizophr Res 23: 97–106

Olney JW, Farber NB (1996) Glutamate receptor dysfunction and schizophrenia. Arch Gen Psychiatry 52: 998–1007

Owen MJ, Lewis SW, Murray RM (1988) Obstetric complications and schizophrenia: a computed tomographic study. Psychol Med 18: 331–339

Rakic P (1988) Specification of cerebral cortical areas. Science 241: 170–176

Raz S, Raz N (1990) Structural brain abnormalities in the major psychoses: a quantitative review of the evidence from computerized imaging. Psychol Bull 108: 93–108

Reveley AM, Reveley MA, Clifford CA, Murray RM (1982) Cerebral ventricular size in twins discordant for schizophrenia. Lancet ii: 540–541

Roberts GW (1991) Schizophrenia: a neuropathological perspective. Br J Psychiatry 158: 8–17

**Saunders RC, Kolachana BS et al (1998) Neonatal lesions of the mediotemporal lobe disrupt prefrontal cortical regulation of striatal dopamine. Nature 393: 169–171

Schlaepfer TE, Harris GJ, Tien AJ et al (1994) Decreased regional cortical gray matter volume in schizophrenia. Am J Psychiatry 151: 842–848

Shenton ME, Kikinis R, Jolesz FA et al (1992) Abnormalities of the left temporal lobe and thought disorder in schizophrenia. N Engl J Med 327: 604–612

**Suddath RL, Christison GW, Torrey EF et al (1990) Cerebral anatomical abnormalities in monozygotic twins discordant for schizophrenia. N Engl J Med 322: 789–794

Susser ES, Lin SP (1992) Schizophrenia after prenatal exposure to the Dutch Hunger Winter of 1944–1945. Arch Gen Psychiatry 49(12): 983–988

Taber MT, Das S, Fibiger HC (1995) Cortical regulation of dopamine release: mediation via the ventral tegmental area. J Neurochem 65: 1407–1410

Torrey EF, Taylor EH, Bracha HS et al (1994a) Prenatal origin of schizophrenia in a subgroup of discordant monozygotic twins. Schizophr Bull 20: 423–432

Torrey EF, Bowler AE et al (1994b) Schizophrenia and manic depression disorders: the biological roots of mental illness as revealed by a landmark study of identical twins. Basic, New York

Van Horn JD, McManus JC (1992) Ventricular enlargement in schizophrenia: a meta-analysis of studies of the ventricular brain ration (VBR). Br J Psychiatry 160: 687–697

**Walker E, Lewine R (1990) Prediction of adult-onset schizophrenia from childhood home movies of the patients. Am J Psychiatry 147: 1052–1056

Walker EF, Savoie T, Davis D (1994) Neuromotor precursors of schizophrenia. Schizophr Bull 20: 441–451

Walker EF, Lewine R, Neumann L (1996) Childhood behavioral characteristics and adult brain morphology in schizophrenia. Schizophr Res 22: 93–101

*Weinberger DR (1987) Implications of normal brain development for the pathogenesis of schizophrenia. Arch Gen Psychiatry 44: 660–669

Weinberger DR (1995) Schizophrenia as neurodevelopmental disorder. In: Hirsch SR, Weinberger DR (eds) Schizophrenia. Blackwell, Oxford, pp 293–323

Weinberger DR (1996) On the plausibility of 'The Neurodevelopmental Hypothesis' of schizophrenia. Neuropsychopharmacology 14: 1S–11S

Weinberger DR, Lipska BK (1995) Cortical maldevelopment, antipsychotic drugs, and schizophrenia: a search for common grounds. Schizophr Res 16: 87–110

Weinberger DR, Torrey EF, Neophytides AN, Wyatt RJ (1979) Lateral cerebral ventricular enlargement in chronic schizophrenia. Arch Gen Psychiatry 36: 735–738

Weinberger DR, Cannon-Spoor E, Potkin SG, Wyatt RJ et al (1980) Poor premorbid adjustment and CT scan abnormalities in chronic schizophrenia. Am J Psychiatry 137: 1410–1413

Zipursky RB, Marsh L, Kom KO et al (1994) Volumetric MRI assessment of temporal lobe structures in schizophrenia. Biol Psychiatry 35: 501–516

CHAPTER

5

H. Sauer, M. Weisbrod

Schizophrenia: Disturbances of Hemispheric Lateralization

Translator: E. Taub

1 Introduction

Hemispheric asymmetry is likely to have developed over the course of evolution and is demonstrable in humans, primates, rats, and birds. A further projection of the right hemisphere anteriorly and of the left hemisphere posteriorly has been found in both Peking man, who lived some 500,000 years ago, and Neanderthal man, who lived some 50,000 years ago. The increasing specialization of each hemisphere presumably led to the development of a greater repertoire of functions than would have been possible with symmetrical hemispheres (Kertesz and Naeser 1994). The earliest demonstration of a lateralized function of the brain was by Broca (1863), who recognized the relationship between left hemispheric lesions and aphasia.

Many studies on the lateralization of brain function have been carried out in recent decades. It has become clear that language and handedness are the most highly lateralized functions. Hemispheric specialization has also been studied with respect to attention, visual and auditory perception, affectivity, and autonomic regulation (Hellige 1993). Even though a direct comparison of left- and right-hemisphere functioning appears problematic because of the complexity of the subject, many groups of researchers have concluded that the left hemisphere more effectively performs cognitive tasks requiring a comparison of stimuli, while the right hemisphere is better at comparing and integrating different features. A brief overview of the functions ascribed to the left and right hemispheres in the past century is given in Table 1.

It was already postulated in the nineteenth century that schizophrenic manifestations were related to left-sided, rather than right-sided, cerebral disturbances (see, e.g. Crichton-Browne 1879). Such hypotheses played a relatively small role in psychiatric thought until the mid-1970s, when neuropsychological studies (Flor-Henry 1976; Gruzelier and Venables 1974) and, later, neuroimaging and neuropathological studies revealed the presence of left-hemisphere abnormalities and a reduction of the normal physiological asymmetry of the brain. These structural and functional findings were incorporated into the etiopathogenetic concepts of schizophrenia by a number of authors (see, e.g. Crow et al. 1989; Jaynes 1976). This led to the performance of many further informative studies and to the generation of considerable debate.

The structural findings are discussed in Chap. 10 (Vol. 1, Part 1). In this chapter, we will provide an overview of the functional findings, particularly with respect to handedness, language, and spatial and emotional perception, and the results of functional neuroimaging studies in the schizophrenias. Because of space restrictions, only the most important studies can be discussed.

Table 1. Hypotheses concerning the different functions of the two hemispheres over the last 130 years (after Cutting 1990)

Period	Left hemisphere	Right hemisphere
1860–1910	Language, motor function	Undetermined, complex perception
1950	Language	Spatial
1960	Verbal and analytic	Nonverbal and synthetic
1970–present	Analytic Serial processing Detail-oriented Logical Rational	Holistic Parallel processing Gestalt-oriented Creative Emotional, affective

2 Handedness

Handedness has been used as a parameter to assess possible alterations of hemispheric asymmetry. The earlier studies yielded inconsistent findings, possibly because they made a dichotomous distinction between right- and left-handedness. It is much more likely, however, that handedness is a continuously distributed variable, ranging from strong right-handedness, through a spectrum of mixed states, to strong left-handedness. Mixed handedness was, therefore, considered as a third category in later studies, and this led to a higher degree of agreement across studies.

It has since been established that mixed handedness is more common among schizophrenics than among normal control individuals. Nelson et al. (1993), for example, found mixed handedness in 43% of patients and only 14.3% of controls, while Cannon et al. (1995) arrived at figures of 34.4% versus 11.6%. Schizophrenics with mixed handedness more commonly switch hands during the performance of a task (Gorynia and Uebelhack 1992).

Satz and Green (1999), using a different methodological approach, determined the direction of the displacement of handedness in schizophrenic patients. They found that 14 studies had shown a leftward displacement, seven no definite displacement, and two a rightward displacement. These findings suggest that the prevalence of strong right-handedness is diminished among schizophrenics and that there is a displacement of handedness toward the left side. In complex motor tasks, patients with mixed handedness

tend to lateralize to the right more than to the left. When this is taken into account, mixed handedness is found to be reduced by as much as 50% (Nelson et al. 1993). Fleminger et al. (1977), in a very large study in which 800 patients and 800 control subjects were examined, found that male schizophrenics are less often right-handed than female schizophrenics.

There is a relationship between left-handedness and clinical findings and disease manifestations. Tyler et al. (1995) found that left-handed schizophrenics ($n = 94$), in comparison to right-handed ones ($n = 592$), were more likely to have had perinatal difficulties and severe cognitive and behavioral abnormalities in childhood and more frequently suffered from persistent auditory hallucinations. Cannon et al. (1995) found that mixed handedness was associated with the chronification of psychosis, and the genetic load tended to be less than that seen in right-handed patients. This led the authors to attribute mixed handedness to a disturbance of neural development. Further studies revealed an association between mixed handedness and cognitive abnormalities, as well as an increased ventricle-to-brain ratio (see Satz and Green 1999). Manoach (1994) found that left-handedness in schizophrenics is associated with formal thought disorders (in men) and with disturbances of language function. Their interpretation was that atypical handedness is a marker of a left-hemispherec disturbance.

3 Further Aspects of Motor Function

The study of motor function with neuroimaging techniques is particularly informative, because the greater localizing ability of these techniques can be used to address the question of lateralization more precisely. Schröder et al. (1995), using functional magnetic resonance imaging (fMRI), found that normal subjects performing a finger–thumb opposition task had an activation of the contralateral supplementary motor area and of the sensorimotor area bilaterally (left more than right). Schizophrenics, in contrast, had a lesser degree of activation and an inverted lateralization effect. In another fMRI study on schizophrenic patients, Mattay et al. (1997) found an increased degree of ipsilateral activation in the primary sensorimotor and lateral premotor regions and, as in the earlier study, a reduction of the lateralization quotient. The authors explained these findings as representing a failure of interhemispheric inhibition, which normally takes place by way of glutamatergic transcallosal projections impinging on the γ-aminobutyric acid (GABA)-ergic system and makes functional lateralization possible.

4 Language

Language, like handedness, has been used as a parameter of functional lateralization, because it is generally considered to be a function of the left hemisphere. The role of the right hemisphere in language has attracted attention only in recent years. Flor-Henry (1976), one of the first authors to postulate a left-hemisphere disturbance in schizophrenic patients, did so because of the finding of language-related disturbances in neuropsychological performance. This hypothesis has been tested in many subsequent studies. Many investigators made use of the fact that, in most normal individuals, linguistic stimuli are processed more rapidly and/or accurately by the left hemisphere, regardless of whether they are presented visually or acoustically. The fact that linguistic stimuli are processed more efficiently when presented to the right ear (and thus predominantly to the left hemisphere) is known as the right-ear advantage.

Most of the relevant studies have shown a reduction of the right-ear advantage in schizophrenics, but this reduction is related to the presence of florid symptoms, particularly auditory hallucinations (Bruder et al. 1995). This finding is in accordance with the demonstration, by functional neuroimaging techniques, that language centers in the left temporal lobe are activated during auditory hallucinations (Dierks et al. 1999). On the other hand, the fact that the right-ear advantage is also reduced in patients in remission (Wexler et al. 1991), "high-risk" children (Hallet and Green 1983), and parents implies that this finding is a genetically controlled "trait."

Further evidence for a left-hemisphere disorder comes from the fact that schizophrenic patients often produce unusual semantic associations (Bleuler 1911) and that the relationship of linguistic expression to context is loosened, particularly in patients with thought disorders. Weisbrod et al. (1998) used tachistoscopic testing to show that this excessive activation of semantic associations in schizophrenics is, in fact, attributable to a deficient ability to focus semantic activation in the left hemisphere.

More recent studies have shown, however, that language is not exclusively a left-hemisphere function. The right hemisphere recognizes and generates prosody and plays a major role in the interpretation of the emotional aspects of language and in the understanding of metaphors and jokes. Right-hemisphere linguistic

functioning in schizophrenics has been investigated, to our knowledge, in only one study to date (Spitzer 1993). In this study, normal subjects and schizophrenic patients were presented with proverbs and, immediately afterward, with words related to the abstract and metaphorical semantic content of the proverbs. The recognition of these words was facilitated in normal subjects but not in schizophrenics, which implies the presence of a right-hemisphere disorder.

Language requires the integrated functioning of the two hemispheres, each of which is specialized for different functions. In this context, the older finding that schizophrenic patients perform poorly in assessing the agreement of binaurally presented verbal stimuli (Beaumont and Diamond 1973) remains of interest. Patients were also better able to understand stories when they were presented in a single ear rather than both (Green and Kotenko 1980). These findings suggest that a disturbance of the integration of the linguistic functions of the two hemispheres may also contribute to the linguistic abnormalities of schizophrenic patients.

5 Spatial Perception and Facial Recognition

Right-hemisphere function has frequently been assessed by neuropsychological testing of spatial perception and facial recognition, because the right hemisphere is known to perform better than the left in spatial perception, pattern recognition (Warrington and Rabin 1970), and the judgment of spatial orientation and distance (Benton and Tranel 1993). Borod et al. (1993), for example, found that schizophrenic patients with negative symptoms were as severely impaired in the recognition of regular abstract patterns (Visual Matrices Test) as patients with right-hemisphere brain injury. Two further studies employing interhemispheric comparisons, however, failed to demonstrate a loss of the right-hemisphere advantage in spatial tasks (Connolly et al. 1979; Gur 1978).

Facial recognition is also considered to be predominantly a function of the right hemisphere (Rizolatti et al. 1971). Conrad (1958) had already described the "physiognomization" of the environment in early schizophrenic psychosis. In two recent studies, the normal advantage of the right hemisphere over the left in the recognition of faces was, indeed, not found (Borod et al. 1993; White et al. 1998); this can be interpreted as suggesting a right-hemisphere functional disturbance. Two other studies, however, came to the opposite conclusion (Ellis et al. 1993; George and Neufeld 1987).

The recognition of facial emotional expression seems to be, more clearly than the simple recognition of faces, a function of the right hemisphere (Borod et al. 1986). Numerous studies have shown a corresponding deficit in schizophrenic patients, independent of their ability to recognize faces or general spatial patterns (Borod et al. 1993). This disturbance probably represents a "trait," as suggested, for example, by the findings of Wölwer et al. (1996) and of Addington and Addington (1998), who studied not only symptomatic patients, but also patients in remission.

The results of neuropsychological studies in which both linguistic and visuospatial abilities were tested can be summarized as follows (see also Chap. 8, this volume): linguistic function is usually more severely impaired than visuospatial ability in schizophrenic patients. This implies that the neuropsychological deficit resides predominantly in the left hemisphere (Flor-Henry 1976; Ragland et al. 1999). In view of the presence of sex-specific differences in laterality shown by the study by Ragland et al. (1999), sex will have to be taken into account in future studies.

6 Early Auditory Information Processing

Abnormalities of the electroencephalogram and of event-correlated potentials in schizophrenics indicating the presence of a disturbance of hemispheric lateralization have been discussed in the chapter on neurophysiological techniques (Chap. 9, Vol. 1). In this section, we will discuss more fully three studies employing magnetoencephalography (MEG) that yielded further, highly indicative findings. In all three of them, interhemispheric functional differences were detected by measuring neuromagnetic fields after stimulation with a binaurally delivered tone, with special attention to the M100 peak, which is generated after about 100 ms in the area of the primary auditory cortex. The M100 can be characterized as a dipole, and the interhemispheric comparison of dipoles can be used as an index of hemispheric lateralization. These studies thus dealt with early auditory information processing in schizophrenics.

Reite et al. (1997) first determined the localization of the M100 dipole in normal subjects using a seven-channel MEG system and found that the neuroanatomical source of the M100 in the right hemisphere was anterior to that in the left hemisphere and that this asymmetry was more pronounced in men than in women. No asymmetry was found in male schizophrenics, while female schizophrenics had a more pronounced asymmetry than female control subjects. Reite and colleagues attributed these findings to a sex-specific disturbance of hemispheric lateralization,

which may occur during intrauterine development under estrogenic influence. In another study employing a 31-channel system (Hajek et al. 1997a,b), the findings were less clear with regard to dipole localization, but comparable with regard to dipole orientation, which can be measured more reliably: A deviation of the dipole in comparison to same-sex controls was demonstrated in the left hemisphere in male schizophrenics and in the right hemisphere in female schizophrenics. The possible explanations for these phenomena are discussed elsewhere in greater detail (see Sauer et al. 1998).

In the third study (Tiihonen et al. 1998), evidence was found for a reversal of cerebral asymmetry, but in only 32% of patients; further calculations revealed that this finding was predominantly due to left-hemisphere alterations. Nonetheless, 16% of patients had a stronger asymmetry than the control subjects; the authors concluded that the development of cerebral asymmetry in schizophrenics is subject to faulty regulatory control. Reversed asymmetry was found to be associated with a higher general psychopathology score.

With regard to the interpretation of these studies, it must be emphasized that only the cortical response to stimuli is detected by the MEG. Taken together, the three studies indicate that there are differences of laterality between men and women, and between schizophrenics and control subjects. The etiopathogenetic significance of the change in the M100 source has, however, not yet been explained. It may be no more than an anatomical index; if so, then the findings presumably reflect a disturbance of hemispheric lateralization occurring in utero under the influence of sex hormones between the 20th and 30th weeks of gestation, as hypothesized by Reite. On the other hand, it remains possible that the dipole abnormalities come about as a result of the disease itself or its treatment. This question could be answered by a study of "high-risk" children or of patients showing the earliest manifestations of the disease.

7 Attention

Disturbances of attention have been of central importance for the understanding of schizophrenia since the time of Kraepelin and Bleuler and are a particularly complex set of phenomena. In this section, we will discuss high-resolution fMRI studies in which changes of cerebral blood flow were measured during the performance of a task. A remaining problem in the interpretation of the results obtained with such paradigms is that the changes measured are usually due not only to attention, but also to other cognitive functions at the same time.

The complexity of attention processes was demonstrated in an fMRI study performed on normal subjects by Peterson et al. (1999). While the subjects were performing the Stroop Test, a word–color interference test that also contains a component relating to impulse control, activations were found primarily in the anterior cingulum and the mesiofrontal cortex. These were correlated with activity in several other brain areas that have been implicated in the processes of sensory perception, language comprehension, vigilance, working memory, selection of responses, and the planning and performance of motor functions. Evidence was also found in favor of the existence of several distinct attention systems, which are probably coordinated and integrated by the cingulate gyrus (see Chap. 7, this volume). There was no evidence for a lateralization of attention in normal subjects.

In a further fMRI study employing a single-slice technique, Häger et al. (1998) applied the Continuous Performance Test (CPT), which had been used in "high-risk" studies to determine which of a group of currently healthy individuals would later develop schizophrenia. The CPT is a test of attention span, but the version used in the study also tests working memory. The largest activations were found in the mesiofrontal cortex, cingulate gyrus, and dorsolateral prefrontal cortex on the right side and in the thalamus bilaterally. The same group of researchers, in a later publication, compared the activation patterns of normal subjects and schizophrenics (Volz et al. 1999). Schizophrenics had a lesser degree of activation than normal controls in the right mesiofrontal cortex, the right cingulate gyrus, and the left thalamus. Further analysis revealed that the smaller increase in cerebral blood flow was largely independent of the subject's performance on the test.

Even if only a few preliminary studies are currently available, it is clear that multiple brain areas are activated during complex processes such as attention and that distinct attention systems exist. It seems that one of these systems is more strongly localized to the right hemisphere than the other and is subject to a disturbance of activation, predominantly in the right hemisphere, in schizophrenic patients.

8 Conclusion

The studies discussed above did not yield consistent results on the question of hemispheric lateralization in the schizophrenias. The findings with regard to brain morphology (see Chap. 10, Vol. 1, Part 1),

the electroencephalogram (Chap. 9, Vol. 1, Part 1), and brain function, especially as revealed by studies of handedness and language (this chapter), all suggest the presence of a left-hemisphere disturbance. Further evidence for a left-hemisphere disturbance and an ensuing loss of the normal physiologic asymmetry of the hemispheres comes from quantitative MEG studies of early auditory information processing. On the other hand, a right-hemisphere disturbance is suggested by findings regarding attention, spatial perception, and the recognition of facial and linguistic emotional expression; these findings are less consistent than the left-hemisphere findings. There have also been several methodologically sound studies of right-hemisphere function in which schizophrenic patients were not found to have any abnormality. There is thus more evidence overall for a left-hemisphere than for a right-hemisphere functional disturbance.

When a disturbance is found in a particular brain function attributed to one hemisphere, or when functional imaging studies reveal an abnormality in one hemisphere, the inference that the seat of the dysfunction actually lies in that hemisphere remains provisional, because the possible effects of normal and disturbed interhemispheric connectivity are not yet adequately understood. Moreover, it has become evident in recent years that many variables can influence the results of studies of hemispheric lateralization. For example, lateralization abnormalities detected by MEG are dependent on sex: left-hemisphere abnormalities were found in male schizophrenics, and possible right-hemisphere abnormalities in female schizophrenics. Hemispheric lateralization is also influenced by psychopathological features, as seen, e.g. in the effect of auditory hallucinations on the reduction of the right-ear advantage. The acuity of the illness, too, modulates the disturbance of hemispheric lateralization: mixed handedness, for example, is more common in acute psychosis than during remission.

These abnormalities of hemispheric lateralization have been explained in various ways. Crow (1997), for example, postulates that a gene localized in the homologous region of the X chromosome, and responsible for the ontogenic development of hemispheric lateralization, bears an etiopathogenetic relationship to schizophrenia. This hypothesis can only be considered speculative at present. It seems more plausible that abnormal lateralization results from the same disturbance of neural development that is currently thought to cause the later cerebral morphologic and functional abnormalities of the schizophrenias in general (Chap. 4, Vol. 3, Part 1). Several lines of evidence point to such a developmental disturbance, most likely occurring in the second trimester of pregnancy, which is the period in which cerebral lateralization develops. The more severe involvement of the left hemisphere than of the right hemisphere may be attributable to the fact that it matures more slowly during fetal development (Saugstad 1998) and is thus more vulnerable to genetically based disturbances or exogenous injury (Bracha 1991).

Pathogenetic models of schizophrenia have now become more complex and are no longer based solely on circumscribed functional abnormalities, such as hypofrontality or temporofrontal disconnection. The "cognitive dysmetria" model (Andreasen et al. 1998), for example, postulates a disturbance of interconnected frontal, thalamic, cerebellar, and (in its most recent version; Andreasen 1999) temporal systems. Yet even these current, more complex models fail to integrate the findings on hemispheric lateralization. This is a difficult task, and it will remain so because, as the studies discussed here show, abnormal lateralization can no longer be understood as a single alteration equally affecting all regions of the brain and all cerebral functions. Future studies of hemispheric lateralization must be carried out more specifically, with respect to individual cerebral structures and functions, and future models of schizophrenia are thus likely to be more complex than current models. Such investigations may lead the way to a better understanding of those clinical findings and disease manifestations whose occurrence and degree of severity are related to an abnormality of cerebral lateralization.

9 References

Addington J, Addington D (1998) Facial affect recognition and information processing in schizophrenia and bipolar disorder. Schizophr Res 32: 171–181

Andreasen N (1999) Prefrontal circuitry in the normal brain and in schizophrenia. Presented at the World Congress of Psychiatry, Hamburg, 6–11 August 1999

*Andreasen NC, Paradiso S, O'Leary DS (1998) "Cognitive dysmetria" as an integrative theory of schizophrenia: a dysfunction in cortical-subcortical-cerebellar circuitry? Schizophr Bull 24: 203–218

Beaumont JG, Diamond SJ (1973) Brain disconnection and schizophrenia. Br J Psychiatry 123: 661–663

Benton A, Tranel D (1993) Visuoperceptual, visuospatial, and visuoconstructive disorders. In: Heilman KM, Valenstein EV (eds) Clinical neuropsychology. Oxford University Press, New York, pp 165–213

Bleuler E (1911) Dementia praecox. Deuticke, Leipzig

Borod J, Koff E, Lorch MP, Nicholas M (1986) The expression and perception of facial emotion in brain-damaged patients. Neuropsychologia 24: 169–180

Borod JC, Martin CC, Alpert M, Brozgold A, Welkowitz J (1993) Perception of facial emotion in schizophrenic and right brain-damaged patients. J Nerv Ment Dis 181: 494–502

Bracha HS (1991) Etiology of structural asymmetry in schizophrenia: an alternative hypothesis. Schizophr Bull 17: 551–552

Broca P (1863) Localisation des fonctions cérébrales. Siège du language articulé. Bull Soc Anthropol Paris 4: 300–304

Bruder G, Rabinowicz E, Towey J et al (1995) Smaller right ear (left hemisphere) advantage for dichotic fused words in patients with schizophrenia. Am J Psychiatry 152: 932–935

Cannon H, Byrne M, Cassidy B, Larkin C, Horgan R, Sheppard Noel P, O'Callaghan E (1995) Prevalence and correlates of mixed-handedness in schizophrenia. Psychiatry Res 59: 119–125

Connolly JF, Gruzelier JH, Kleinman KM, Hirsch SR (1979) Lateralized abnormalities in hemisphere-specific tachistoscopic tasks in psychiatric patients and controls. In: Gruzelier J, Flor-Henry P (eds) Hemisphere asymmetries of function in psychopathology. Elsevier, Amsterdam, pp 491–509

Conrad K (1958) Die beginnende Schizophrenie: Versuch einer Gestaltanalyse des Wahns. In: Conrad K, Scheid W, Weitbrecht HJ (eds) Sammlung psychiatrischer und neurologischer Einzeldarstellungen. Thieme, Stuttgart, pp 1–165

Crichton-Browne J (1879) On the weight of the brain and its component in the insane. Brain 2: 42–67

Crow TJ (1997) Is schizophrenia the prize that homo sapiens pays for language? Schizophr Res 28: 127–141

*Crow TJ, Ball J, Bloom SR, Brown R (1989) Schizophrenia as an anomaly of development of cerebral asymmetry. Arch Gen Psychiatry 46: 1145–1150

Cutting J (1990) The right cerebral hemisphere and psychiatric disorders. Oxford University Press, Oxford

Dierks T, Linden DEJ, Jandl M, Formisano E, Goebel R, Lanfermann H, Singer W (1999) Activation of Heschl's gyrus during auditory hallucinations. Neuron 22: 615–621

Ellis HD, de Pauw KW, Christodoulou GN, Papageorgiou L, Milne AB, Joseph AB (1993) Responses to facial and non-facial stimuli presented tachistoscopically in either or both visual fields by patients with the Capgras delusion and paranoid schizophrenics. J Neurol Neurosurg Psychiatry 56: 215–219

Fleminger JJ, Dalton R, Standage KF (1977) Handedness in psychiatric patients. Br J Psychiatry 131: 44–452

*Flor-Henry P (1976) Lateralized temporal-limbic dysfunctions and psychopathology. Ann NY Acad Sci 280: 777–797

George L, Neufeld RWJ (1987) Attentional resources and hemispheric functional asymmetry in schizophrenia. Br J Clin Psychology 26: 35–45

Gorynia I, Uebelhack R (1992) Functional motor asymmetries correlated with clinical findings in unmedicated schizophrenic patients. Eur Arch Psychiatry Clin Neurosci 242: 39–45

Green P, Kotenko V (1980) Superior speech comprehension in schizophrenics under monaural versus binaural conditions. J Abnorm Psychology 89(3): 399–408

Grosh E, Docherty NM, Wexler BE (1995) Abnormal laterality in schizophrenics and their parents. Schizophr Res 14: 155–160

Gruzelier JH, Venables PH (1974) Bimodality and lateral asymmetry of skin conductance orienting activity in schizophrenics: replication and evidence of lateral asymmetry in patients with depression and disorders of personality. Biol Psychiatry 8: 55–73

Gur RE (1978) Left hemisphere dysfunction and left hemisphere overactivation in schizophrenia. J Abnorm Psychology 2: 226–238

Häger F, Volz HP, Gaser C, Mentzel HJ, Kaiser WA, Sauer H (1998) Challenging the anterior attentional system with a continuous performance task: a functional magnetic resonance imaging approach. Eur Arch Psychiatry Clin Neurosci 248: 161–170

Hajek M, Huonker R, Boehle C, Volz HP, Nowak H, Sauer H (1997a) Abnormalities of auditory evoked magnetic fields and structural changes in the left hemisphere of male schizophrenics – a MEG-MRI study. Biol Psychiatry 42: 609–616

Hajek M, Boehe C, Huonker R, Volz HP, Nowak H, Schrott P, Sauer H (1997b) Abnormalities of auditory evoked magnetic fields in the right hemisphere of schizophrenic females. Schizophr Res 24: 329–332

Hallet S, Green P (1983) Possible defects of interhemispheric integration in children of schizophrenics. J Nerv Ment Dis 171: 421–425

Hellige JB (1993) Hemispheric asymmetry. Harvard University Press, Cambridge/MA

Jaynes J (1976) The origin of consciousness in the breakdown of the bicameral mind. Houghton Mifflin, Boston

Kertesz A, Naeser MA (1994) Anatomical asymmetries and cerebral lateralization. In: Kertesz A (ed) Localization and neuroimaging in neuropsychology. Academic, San Diego, pp 213–244

Manoach DS (1994) Handedness is related to formal thought disorder and language dysfunction in schizophrenia. J Clin Exp Neuropsychol 16: 2–14

Mattay VS, Callicott JH, Bertolino A et al (1997) Abnormal functional lateralization of the sensorimotor cortex in patients with schizophrenia. Neuroreport 8: 2977–2984

Nelson LD, Satz P, Green M, Cicchetti D (1993) Re-examining handedness in schizophrenia: now you see it – now you don't! J Clin Exp Neuropsychol 15: 149–158

Peterson BS, Skudlarski P, Gatenby JC, Zhang H, Anderson AW, Core JC (1999) An fMRI study of stroop word-color interference: evidence for cingulate subregions subserving multiple distributed attentional systems. Biol Psychiatry 45: 1237–1258

*Ragland D, Gur RE, Klimas BC, McGrady N, Gur RC (1999) Neuropsychological laterality indices of schizophrenia: interactions with gender. Schizophr Bull 25: 79–89

Reite M, Sheeder J, Teale P et al (1997) Magnetic source imaging evidence of sex differences in cerebral lateralization in schizophrenia. Arch Gen Psychiatry 54: 433–440

Rizolatti B, Umlita C, Berlucchi G (1971) Opposite suoeriorities of the right and left cerebral hemispheres in discriminative reaction time to physiognomical and alphabetic material. Brain 94: 431–442

Satz P, Green MF (1999) Atypical handedness in schizophrenia: some methodological and theoretical issues. Schizophr Bull 25: 63–78

Sauer H, Rosburg T, Kreitschmann-Andermahr I, Volz HP, Huonker R, Nowak H, Hajek M (1998) Geschlechtsspezifische Unterschiede der Hemisphärenlateralisation bei Schizophrenien? Nervenarzt 69: 249–256

Saugstad LF (1998) Cerebral lateralisation and rate of maturation. Int J Psychophysiol 28: 37–62

Schröder J, Wenz F, Baudendistel K, Schad LR, Knopp MV (1995) Sensorimotor cortex supplement motor area changes in schizophrenia. Br J Psychiatry 167: 197–201

Spitzer M (1993) The psychopathology, neuropsychology, and neurobiology of associative and working memory in schizophrenia. Eur Arch Psychiatry Clin Neurosci 243: 57–70

Spitzer M, Braun U, Maier S, Hermle L, Maher BA (1993) Indirect semantic priming in schizophrenic patients. Schizophr Res 11: 71–80

Tiihonen J, Katila H, Pekkonen E et al (1998) Reversal of cerebral asymmetry in schizophrenia measured with magnetoencephalography. Schizophr Res 30: 209–219

Tyler M, Diamond J, Lewis S (1995) Correlates of left-handedness in a large sample of schizophrenic patients. Schizophr Res 18: 37–41

Volz HP, Gaser C, Häger F et al (1999) Decreased frontal activation in schizophrenics during stimulation with the continuous performance test – a functional magnetic resonance imaging study. Eur Psychiatry 14: 17–24

Warrington EK, Rabin P (1970) Perceptual matching in patients with cerebral lesions. Neuropsychologia 8: 475–487

Weisbrod M, Maier S, Harig S, Himmelsbach U, Spitzer M (1998) Lateralized semantic and indirect semantic priming effects in people with schizophrenia. Br J Psychiatry 172: 142–146

Wexler BE, Giller EI, Southwick S (1991) Cerebral laterality, symptoms and diagnosis in psychotic patients. Biol Psychiatry 29: 103–116

White MS, Maher BA, Manschreck TC (1998) Hemispheric specialization in schizophrenics with perceptual aberration. Schizophr Res 32: 161–170

Wölwer W, Streit M, Polzer U, Gaebel W (1996) Facial affect recognition in the course of schizophrenia. Eur Arch Psychiatry Clin Neurosci 246: 165–170

CHAPTER

6

H. Beckmann

Neuropathology of the Endogenous Psychoses

Translator: E. Taub
Dedicated to Hermann Jakob for his 80th birthday.
I thank Mrs. Sabine Voss for her competent and thorough assistance with the preparation of the manuscript.

1 Introduction

Although research into the etiology of the schizophrenic diseases is still largely dominated by genetic, biochemical, (socio-)psychological, and analytical methods, an increasing number of neuropathological studies have also been published recently. The search for "specific" anatomic and histologic alterations in the psychoses had already begun by the late nineteenth century. The existence of focal areas of neuronal loss (*Lückenfelder*, i.e. "lacunar fields") noted by the earlier authors (Alzheimer 1897; Josephy 1923; Fünfgeld 1952) was later confirmed in adequately controlled studies by Dunlap (1928), Spielmeyer (1930), and Peters (1937), but was considered thereafter to be merely a reflection of age-related changes in the nervous system, intercurrent illnesses, and terminal conditions. Questions concerning the histopathology of schizophrenia were not raised again until the First International Congress of Neuropathology was held in Rome in 1952.

At this congress, Cécile and Oskar Vogt and their students presented some of the earlier findings again and introduced the so-called dwindling disease (*Schwunderkrankung*) and fatty degeneration as new criteria for the pathology of schizophrenia (Vogt and Vogt 1952). Peters (1956, 1967) was the most prominent critic of these reported findings, which related to the central nuclei, including the dorsomedial thalamic nuclei, the pallidum, and the anterior portion of the cingulate gyrus. He regarded these alterations as artifacts of preparation or as agonal cellular changes. It was primarily because of his influence that research in the specialized area of the pathological anatomy of the psychoses came to an almost complete halt in the institutions in which it had been carried out up to that time. Pathological anatomy was no longer mentioned, even as a possible point of departure for basic research in schizophrenia. Schizophrenia remained a subject of neuropathological research in a small number of centers, but none of the findings obtained were consistent and specific enough to be considered a basic process with clinical relevance (Jellinger 1985).

The critical evaluation of pathoanatomic findings requires a thorough understanding of the histologic techniques applied. All pressure on the brain must be avoided from the moment it is removed, including during positioning. Faulty positioning of the brain and the addition of water can lead to homogeneous neuronal alterations, including cloudy swelling (Cammermeyer 1967; Lindenberg and Haymaker 1982), that are often mistaken for pathological processes by inexperienced investigators (Hyden and Hartelius 1948). The practice of fixing the brain in formalin for several months, which has recently become standard, has the advantage of leaving brain volume nearly unchanged, but the long fixation time leads to an obscuring of structure and to artifacts, at least on Nissl staining (Benes 1988; Casanova and Kleinman 1990), and this material is therefore suitable only for volumetry. Artifactual changes in the parenchyma must be excluded by means of histologic study. The postmortem interval, correction factors for shrinkage, and variations in the cutting angle may also lead to considerable differences in quantitative studies.

The mutual inconsistency of many findings may have several causes in addition to those already mentioned. Differences in the origin of the histologic preparations are critically important. The two most important large collections of tissue that are used for such studies are the Vogt Collection in Düsseldorf and the Yakovlev Collection in Washington, DC. The methods of tissue preparation in these two collections differ considerably: in the former collection, the brains are embedded in paraffin, in the latter in celloidin. The celloidin embedding process is much less damaging to the tissue than the paraffin embedding process. Because the process of dehydration is very slow, shrinkage and other artifacts are nearly completely absent, and lipids are preserved relatively well. In contrast, in the paraffin process, the tissue shrinks more extensively, and lipids are almost completely dissolved during embedding (Blackwood 1976).

These complex methodological considerations may serve to explain why anatomic and histologic research in the area of the so-called endogenous psychoses is carried out by only a small number of scientists. The earliest findings of morphological abnormalities in the schizophrenic brain that are now held to be securely confirmed were made in radiological studies. Studies employing pneumoencephalography, echoencephalography, and, later, computed tomography yielded the finding of cerebral atrophy or degeneration of a previously unknown type, which was taken to be an important pathogenetic factor. This consisted of a usually asymmetrical, regionally more pronounced dilatation of the lateral ventricles and the third ventricle, and a similarly locally variable dilatation of the cerebral sulci.

2 Macroscopic Alterations

Macroscopic alterations of the brain in schizophrenic patients are rather mild and are not found in all cases. There is usually an overall reduction in volume of the entire brain in patients with affective disorders, as

compared to normal control subjects. This volume reduction is associated with a reduction in the overall length of the brain and with dilatation of the lateral ventricles (Crow et al. 1989). The weight of the brain may be either reduced (Pakkenberg 1987) or normal (Bogerts et al. 1990), perhaps because different macroscopic forms exist. Stereological methods also usually reveal volume reduction in schizophrenia (Pakkenberg 1987, 1989). When leukotomized and non-leukotomized brains were compared, there was a reduction of volume only in the white matter and the central nuclei, but the number of neurons in the neocortex was not reduced (Pakkenberg 1993). The cortical volume was reduced, however, and it was proposed that this was due to changes in the neuropil. On the other hand, the same method applied to the brains of schizophrenics revealed no significant changes of this nature in either the cortex or the white matter (Heckers et al. 1991).

A quantitative postmortem study of the brains of 35 schizophrenic patients revealed a significant reduction in the length of the left sylvian fissure, which in the normal brain is markedly longer than the right sylvian fissure. This finding and the observed dilatation of the inferior horn of the left lateral ventricle may be interpreted as expressions of a developmental anomaly (Crow et al. 1992; Falkai et al. 1992).

In 108 cases of chronic schizophrenia, including four cases of manic-depressive illness, macroscopic abnormalities were found mainly in the middle and inferior temporal gyri, far more frequently on the left side than on the right (Jakob and Beckmann 1989; Beckmann and Jakob 1991). The brains were found to be divisible into two groups on the basis of their pathological features:

1. Brains of *type 1* had normal mass and no macroscopic abnormality other than an abnormal gyral pattern of the left temporal lobe, in which the inferior and middle temporal gyri ran not sagittally, but diagonally, from caudal and inferior to rostral and superior, toward the superior temporal gyrus. These two gyri were usually interrupted by several vertically oriented grooves.
2. Brains of *type 2* were of mildly reduced mass (approx. 1100–1150 g) and were characterized by an abnormal configuration. They were usually smaller, with blunting of the frontal poles and a tendency toward micro- or brachyencephaly. The temporal gyri were generally full and the surface smooth, without the otherwise typical fine relief pattern. These rather nonspecific anomalies are also found in other disorders, such as oligophrenia or multiple congenital anomalies (Le Mire et al. 1975; Jakob and Beckmann 1989).

These findings were later confirmed in vivo by magnetic resonance imaging (MRI) with three-dimensional surface rendering of the brain in 15 schizophrenic patients and 15 control subjects. The qualitative analysis was primarily based on measurements of the sulci and revealed a more vertical orientation of the left temporal sulci, with an interrupted course, resulting from a diagonal orientation of the gyri. The control subjects, in contrast, had more horizontally oriented gyri and sulci, without interruptions. It was quantitatively confirmed that there were significantly more sulcal lines with interrupted courses, predominantly on the left side, and particularly in the left temporal lobe (Kikinis et al. 1994).

It is of interest in this context that the pattern of gyri and sulci in the lateral inferior portion of the human temporal lobe appears relatively late in fetal development, from the seventh to the eighth month of gestation (Le Mire et al. 1975). The gyri develop at different times. The superior temporal gyrus, along with the parahippocampal gyrus, appears relatively early, at about 23 weeks. The development of a finer pattern in the superior temporal gyrus occurs only in the last month of gestation (Dooling et al. 1983). After the macroscopic formation of these structures, further time is needed for the formation of afferent and associative pathways and for synaptogenesis.

These macroscopic abnormalities of cortical development are probably to be grouped within the overall complex of developmental disorders. Nonetheless, the macroscopic features of the two types of brains described above are highly variable and were found only in most, but not all, of the cases with chronic schizophrenia. The question of a possible correlation between these different macroscopic types and the clinical subgroups of the schizophrenias has not yet been adequately addressed.

A significant reduction in the volume of the amygdala, the hippocampus, the parahippocampal gyrus, and the internal pallidal segment on both sides was found by Bogerts and collaborators in serial sections from the Vogt Collection and was interpreted by them as focal atrophy of unknown etiology (Bogerts 1986; Bogerts et al. 1985). Doubt was cast on this hypothesis by later radiologic studies demonstrating that the hypothetical atrophic process is present even before the earliest clinical manifestations of disease.

Gliosis in the periventricular area around the third ventricle, in the hypothalamus, in several frontobasal areas, in periaqueductal areas of the midbrain, and in the rostral pons was demonstrated quite early in paraffin and frozen sections with the Holzer method. Granulations were also found on the ventricular ependyma. These abnormalities were most pro-

nounced in pathways and nuclei of the limbic system: the centromedial nuclei of the amygdala, the dorsomedial thalamic nuclei, the caudate nucleus, and the nucleus accumbens (Stevens 1982). The presence of gliosis could not be confirmed, however, by quantitative methods with the glial fibrillary acid protein (GFAP) immunoperoxidase technique (Roberts et al. 1986, 1987). Stevens and Casanova (1988) pointed out that gliosis is not a reliable finding and is not associated with specific areas, but, rather, differs from brain to brain. According to these authors, the GFAP-immunohistochemical methods are more suitable for tumors and experimental pathology and are applicable only under certain conditions (Hirano 1985) that were not fulfilled in Roberts's study.

Gliosis, such as has been described in several areas of the brain, need not be associated with major parenchymal injury. The severity of tissue injury and the density of gliosis are not always proportional to each other. According to Scholz (1957a,b), meningeal or ependymal inflammation (among other causes) may give rise to gliotic reactions without damaging nervous tissue to a comparable extent. Mild edema, especially protein-rich exudates, may result in the production of fibrous precipitates that resemble fibrous gliosis, without destruction of the parenchyma. This type of gliosis formation may occur either on the outer or on the inner surface of the brain, i.e. the ventricular wall, where subependymal glial fiber deposits may form. Surface phenomena and tension effects play an important role in this process.

Concepts of the pathogenesis of schizophrenia slowly changed as a result of the discovery of a number of local abnormalities in central areas of the limbic system. The findings obtained to date in schizophrenia and its subgroups will now be presented according to their anatomic locations.

3 Hippocampus

The hippocampus is the central structure of the limbic system. It consists of supramodal association cortex and projects to many other complex supramodal association areas, to visceromotor control systems, and to the nucleus accumbens (Swanson 1983). The perforant pathway links the hippocampus with the entorhinal area, with which it forms a functionally important regulatory circuit ("entorhino-hippocampal loop"; Swanson et al. 1978). This fiber bundle passes from the entorhinal fields, primarily layer II pre-α, directly to the neurons of CA1–CA3 and also, as an axodendritic connection, to the granule cells of the dentate gyrus (Segal and Landis 1974; Fifkova 1974). The Papez regulatory circuit connects the hippocampus by way of the fornix, the cingulum, and the mamillary body with the anterior thalamic nuclei and, by way of the septal nuclei, with the hypothalamus (Hassler 1964; Stephan 1975). The hippocampus thus occupies a central position in which it integrates information from all sensory modalities (O'Keefe and Nadel 1978; Turner et al. 1980) and influences somatomotor, visceral, voluntary, and cognitive mechanisms at the cortical level (Swanson et al. 1978; Braitenberg and Schüz 1983).

Scheibel and Kovelman (1981; Kovelman and Scheibel 1984), using Nissl and Golgi methods, found pronounced alterations of the orientation of the pyramidal cells and their dendrites. The degree of axial disorientation of the pyramidal cells was as high as 180° in places. Disorientation was most severe in the intermediate areas between the subiculum and CA1 and between CA1 and CA2, but was present only in some of the cases studied. It was stated that the histologic appearance of this abnormality implies that it must arise at a relatively early stage of fetal development (Scheibel and Kovelman 1981). These findings were confirmed in material from the Yakovlev Collection (Altshuler et al. 1987) and also quantitatively, as a possibly bilateral phenomenon, by Conrad et al. (1991), who made similar findings in the right hippocampus.

These findings remained controversial, however, because they could not be confirmed in later studies of tissue from the Yakovlev Collection (Christison et al. 1989). Arnold et al. (1995) stated that they had found no significant abnormalities of this kind in a thorough investigation of 14 cases of chronic schizophrenia. They found only reduced neuronal size in the pyramidal cell layer, reaching significance in the subiculum and CA1 as well as in layer II of the posterior portion of the entorhinal cortex.

It is possible that these abnormalities are not generally present, and it must also be borne in mind that the Golgi method was not used in this area. These findings in distinct sectors of the pyramidal cell layer lead us to consider a possible relation to the adjacent entorhinal area, to which it is anatomically and functionally linked and which occupies a central position in the limbic system (Stephan 1975). The types of injury found there (see Sect. 5) may be caused, among other factors, by the same disturbances of fetal development in the second gestational trimester that give rise to neuronal disorientation. It is hypothesized that these pathological alterations may also affect the synaptic pattern and may thereby result in functional disturbances of information processing (Scheibel and Kovelman 1981; Kovelman and Scheibel 1984).

It is very difficult to make a judgment of reduced cell number in individual sectors of the hippocampus,

Table 1. Postmortem abnormalities of hippocampal neurons in patients with chronic schizophrenia

Authors	Origin of material, methods	Abnormalities
Bogerts et al. (1985, 1986)	Vogt Collection (paraffin, Nissl)	Reduced cell number in CA1 through CA4, also in the pallidum in cases with catatonia
Falkai and Bogerts (1986)	Vogt Collection (paraffin, Nissl; structural and quantitative studies)	Reduced cell number in CA1 through CA4 in internal and external pallidal segments
Scheibel and Kovelman (1981), Kovelman and Scheibel (1984)	The authors' own cases (Nissl and Golgi technique; Davenport)	Disorientation ("twisting") of pyramidal cells from the subiculum to CA1, as well as CA3 and CA4, in the left hippocampus, quantitatively demonstrated
Altshuler et al. (1987)	Yakovlev Collection (celloidin, Nissl)	"Pyramidal cell disarray" in the left hippocampus, pronounced abnormality of migration
Conrad et al. (1991)	Los Angeles Veterans Administration Medical Center (frozen section method, Nissl)	Abnormalities of hippocampal pyramidal cell orientation, quantitatively demonstrated on both sides
Christison et al. (1989)	Yakovlev Collection (celloidin, Nissl)	No "pyramidal cell disarray" in CA1 of the left hippocampus
Heckers et al. (1991)	The authors' own cases (own frozen section method; stereological methods in serial hippocampal sections)	No abnormality of cell number in CA1 through CA4
Benes et al. (1991)	McLean-Hospital, Belmont, Massachusetts (vibratome sections after brief fixation, cresyl violet; Nissl)	Reduced number of pyramidal cells in CA1; reduced size of pyramidal cells in sectors of the pyramidal cell layer; no "pyramidal cell disarray," no reduction of volume

because the number of cells depends on the age of the patient, on intercurrent illnesses, and on the histologic method used. The reported qualitative differences in neuronal density among different hippocampal sectors are, therefore, controversial (Table 1). The hippocampus – particularly sector CA1, the so-called Sommer sector – is highly susceptible to hypoxic damage. Hypoxia in the setting of intercurrent illnesses often leads to neuronal injury, ranging from a reduction in cell number to total neuronal loss (hippocampal sclerosis); vascular factors may also play a role in this process (Scholz 1957a).

The data with respect to reductions in hippocampal volume in schizophrenia are not entirely consistent (Table 2). Jeste and Lohr (1989), working with material from the Yakovlev Collection, Altshuler et al. (1990), using their own case material, and similarly Arnold et al. (1995) were unable to find a general reduction in volume at any site, with the exception of CA4. Heckers et al. (1990a,b) also found no significant reduction in volume in either the parahippocampal gyrus or the hippocampus, but there was a suggestion of volume reduction below the level of statistical significance on the left side (6%), which was limited to the anterior portion of the hippocampus. In view of the preservation of parenchyma in the different sectors of the pyramidal cell layer of Ammon's horn, this mild reduction in white matter volume might be attributable to the perforant pathway (see Sect. 5). This abnormality is also consistent with the occurrence of developmental disturbances nearly simultaneously in the entorhinal area and the hippocampus.

Benes (1989) studied myelin-stained sections of the frontal cortex, the parahippocampal gyrus, the perforant pathway, the cingulate gyrus, and the hippocampus to detect changes in myelination during development. Increased formation of myelin in late adolescence as compared to earlier times was detected in hippocampal areas and in the subiculum and presubiculum, where fibers of the perforant pathway and distal portions of the cingulate fiber bundle are distributed. The author relates these findings to the findings in schizophrenics discussed above and advances the hypothesis that certain early manifestations of schizophrenia in adolescence may bear a causal relationship to this relatively late development of myelination. These disease manifestations are seated in the anatomic connections of the limbic system and may be the result of inadequate integration of corticolimbic functional circuits.

Table 2. Postmortem findings regarding alterations of hippocampal volume in chronic schizophrenia

Authors	Origin of material, statistical and volumetric methods	Alterations of volume
Bogerts (1984, 1985, 1986)	Vogt Collection (paraffin, Nissl)	Reduced volume of hippocampus, parahippocampal gyrus, amygdala, and internal pallidal segment
Falkai and Bogerts (1986)	Vogt Collection (paraffin, Nissl)	Reduced hippocampal volume (CA1–CA4)
Bogerts (1989)	Vogt Collection (paraffin, Nissl)	Reduced volume of the hippocampal formation
Bogerts et al. (1990)	The authors' own cases (long fixation in formalin; paraplast, Nissl, and Heidenhain-Woelcke technique)	Reduced hippocampal volume; approximately 10% smaller volume of the internal pallidal segment in women compared to men, here also reduced volume of the nucleus accumbens
Jeste and Lohr (1989)	Yakovlev Collection (celloidin, Nissl)	No volume reduction of the hippocampus overall; volume reduction only in CA4, with reduction of pyramidal cell density
Altshuler et al. (1990)	Neuropathological Section of the National Institute of Mental Health, Bethesda, Maryland	No alteration of hippocampal volumes; reduced volume of the parahippocampal gyrus
Heckers et al. (1990a)	The authors' own cases (thick sections, serial Nissl sections)	No reduction of hippocampal or amygdalar volume
Heckers et al. (1990b, 1991)	The authors' own cases (thick sections, stereological methods in serial Nissl sections)	Reduced volume of the left hippocampus only (not significant); white matter changes
Arnold et al. (1995)	The authors' own cases (paraffin, Nissl, 15 μm)	No volume reduction of the hippocampus or the parahippocampal gyrus; reduced neuronal size in subiculum, CA1, and layer II of the entorhinal cortex (caudal portion)

4 Parahippocampal Gyrus

Macroscopic findings to date in the parahippocampal gyrus, which is located at the base of the temporal lobe and is cytoarchitecturally nearly identical to the entorhinal cortex, are highly variable; they are listed in Table 2. The topography of cytoarchitectural changes in the rostral portion of the entorhinal area within the parahippocampal gyrus, in the ventral portion of the claustrocortex (insular cortex), and in the rostral portion of the cingulate gyrus is depicted in Fig. 1 (hatched areas).

5 Entorhinal Area

The entorhinal area is an important, differentiated "association center" within the allocortex (Braak 1980). It is intimately connected to the hippocampus by way of the perforant pathway and thus forms, together with it, a multineuronal regulatory circuit at the center of the limbic system (see Sect. 3). Signals arriving in the entorhinal cortex proceed to the hippocampus, pass through several synapses, and return, in part, to the entorhinal cortex. This regulatory circuit seems to be of major importance for the storage of orientation and for memory (Braitenberg and Schüz 1983).

Studies in primates have shown that primary cortical fields and all secondary cortical fields with visual, auditory, and somatosensory functions have reciprocal connections with the entorhinal cortex, either directly or by way of the perirhinal area (Jones and Powell 1970; van Hoesen 1982; Pandya and Yeterian 1985). The multisensory areas in caudal portions of the orbitofrontal region, and the rostral and ventral fields of the claustrocortex, project mainly onto the rostral fields of the entorhinal area. The olfactory cortex is the only unimodal field that projects onto rostral entorhinal areas (Insausti et al. 1987). Furthermore, as extensive studies in the cat have shown, there are well-developed systems of both

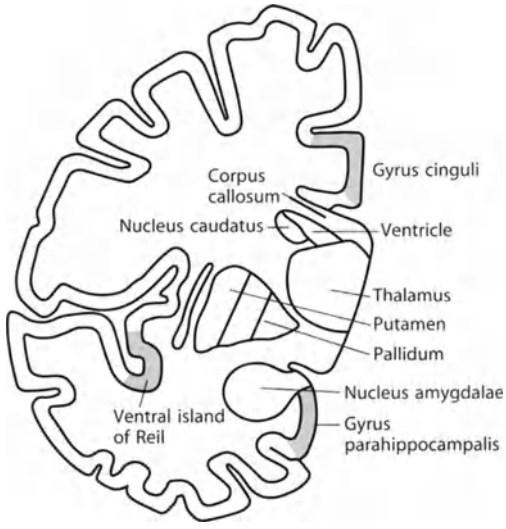


Fig. 1. Coronal section of the left hemisphere passing through frontal and temporal lobes at the level of the rostral portion of the amygdala; *hatching* indicates areas of cytoarchitectural alteration. (Modified from Jakob and Beckmann 1986)

longitudinal and transverse connections that enable the activity of systems within the entorhinal cortex to be integrated with the complex of afferent information. Sensory afferent information is delivered to the hippocampus by way of the upper layers of the perirhinal area and the entorhinal cortex. Efferent projections arise from the lower layers of the perirhinal and entorhinal areas. The entorhinal cortex thus integrates information from all sensory modalities from both the interior and the exterior of the organism (O'Keefe and Nadel 1978; Turner et al. 1980; Witter et al. 1986).

The allocortex is roughly divided, on a topographical basis, into medial, lateral, and perirhinal portions. The last-named area, which is also called the transentorhinal subregion (Braak 1980), lies between the lateral part of the entorhinal cortex and the isocortex of the temporal lobe (Stephan 1975). It displays the full array of cortical layers in a coronal section passing through the central portion of the amygdala, in the entorhinal central medial and lateral area (and the interpolar medial area); it is most highly differentiated in the central lateral area. The upper portion of the cortex, the principal external lamina, is divided according to Rose (1927) into the zonal lamina (I) and the Pre-α (II), Pre-β, and Pre-γ (III) layers, and the adjacent acellular layer, the lamina dissecans (layer IV). The lower portion of the cortex, the principal internal lamina, is subdivided into the Pri-α, Pri-$\alpha\beta$, and Pri-$\alpha\gamma$ sublayers (Pri layers), which consist mainly of pyramidal cells. In the central fields, layer II Pre-α consists mainly of islands of similar-appearing, medium-sized to large multipolar "modified" pyramidal cells with long axons extending into the white matter. This type of neuron is characteristic of this area and is not found elsewhere (Braak 1980).

Detailed observations concerning the development of this area were made in autoradiographic studies of rhesus monkey embryos at different developmental stages (Nowakowski and Rakic 1981; Rakic and Nowakowski 1981). These findings may be regarded as a model of the developmental process in humans (Nowakowski 1987; Sidman and Rakic 1973). The matrix for archi- and peri-archicortical areas corresponding to the hippocampus and the entorhinal cortex lies in the medial wall of the hemispheric vesicle, the initial embryonic precursor of the cerebral hemisphere. The neuroblasts destined to form the cerebral cortex are already determined at this stage (Rakic 1988a,b). While the neuroblasts of the ventricular zone form the lower layers of what will later become the entorhinal cortex, the subventricular zone gives rise to its upper layers. This is also the site where, after the last cell division, active movement of the neuroblasts (i.e. cell migration) begins. Neurons at this stage have a leading process, a fusiform, bipolar shape with an ovoid nucleus, and a long trailing process and are called "young neurons." The leading process is essential for movement of the migrating neuron (Rakic 1978; Caviness et al. 1981).

"Cohorts" of closely spaced young neurons migrate along the course of previously laid down glial fibers outward to the cortical plate, passing by groups of neurons that had reached their destinations earlier, and proceed to the outermost surface of the cortical plate, thus forming "vertical or ontogenetic columns" (Rakic 1988a). The young neurons do not assume their mature pyramidal or polygonal shapes until shortly before they reach their cortical destinations. Later, as further columns migrate to the surface, they become resubmerged in the deeper layers. Thus the outer cortical layers are formed last. This so-called inside-to-outside spatiotemporal gradient is operative for all neocortical and most allocortical areas of the human brain (Sidman and Rakic 1973; Nowakowski 1987). Even at early developmental stages, distant regions begin to make afferent connections, which radiate into the lower zone of the cortical plate (Rakic 1988a).

The development of the entorhinal area in humans is similar to the development in the rhesus monkey just described in many important respects, although it is not yet known in equivalent detail. Compared to other cortical areas, the entorhinal area develops in a relatively brief period of migration. The earliest evidence of a germinal epithelium, or matrix, in the developing fetus is found in the third month at the base of a caudal area of the lateral ventricle. The first signs of migration are demonstrable in embryos aged

approximately 10 weeks (Kostóvic et al. 1990). At the end of the third gestational month, entorhinal and presubicular areas can already be distinguished in the cortical plate (Macchi 1951; Kahle 1969). Positions in the cortex are occupied in the sixth month, and the matrix is entirely depleted in the seventh (Stephan 1975). Further details of the early development of the entorhinal cortex in humans remain controversial.

Among 108 cases of schizophrenia of various types studied histologically, including four cases of manic-depressive illness, 78 cases showed major cytoarchitectural abnormalities in the rostral portion of the entorhinal cortex. These abnormalities extended, in the anteroposterior direction, to the frontobasal area rostrally, but caudally only to the level of a section through the inferior horn of the lateral ventricle and the anterior portion of the hippocampus, where the cytoarchitecture became increasingly normal. The most pronounced abnormalities were found in the anterior sections. An increased number of glia was not observed anywhere (Jakob and Beckmann 1985, 1986, 1994).

As for technique, standard neuropathological methods were used (see Sect. 1), and the inferior portion of the left hemisphere, sectioned at the level of the amygdalar nucleus, was embedded in celloidin; 20-μm-thick celloidin sections were stained with Nissl and Heidenhain-Woelcke stains for histologic investigation, and 16 cases with other clinical diagnoses were selected for use as controls. Graded series of cases and controls were used to obtain an overall view of the extent of the histologic abnormalities.

The most pronounced findings were cytoarchitectural abnormalities in layers Pre-α and Pre-β; abnormalities of Pre-α in the central regions often consisted of only a few characteristic island-like formations. These layers were irregularly constructed. Because the structural abnormalities were variable, a uniform pathological picture could not be obtained. While only the Pre-α and Pre-β layers were affected in "mild" cases, the entire cortex was affected in "severe" cases. In the latter cases, layers III and IV (the Pri-layers) were depleted of approximately 20% and 40% of their neurons, respectively, in comparison with controls (Jakob and Beckmann 1986).

The most commonly encountered abnormalities of layers II Pre-α and III Pre-β appeared to be less of a quantitative than of a structural nature (Jakob and Beckmann 1994). The authors described two basic types of abnormality:

1. *Type 1:* absence of layer Pre-α, with only a few atypical neurons.
2. *Type 2:* here the insular formations of Pre-α were also absent.

Together with the upper portion of layer Pre-β, layer Pre-α had the appearance of a "double row." This consisted of a narrow upper layer, composed of a row of small neurons lying adjacent to one another, and a lower row of tightly spaced collections of groups of atypical neurons that normally do not appear at this site, clearly distinct from the pyramidal cells of layer Pre-β. The authors regarded these as malformed heterotopies. The entire layer thus takes on a markedly "spotted" appearance.

Two distinct, well-defined types of neuronal formation were found among the malformed heterotopies of layer Pre-β just discussed. One of these consisted of atypical pyramidal neurons, of considerably reduced volume, usually lying so close together that no separation between them was visible under the light microscope (Fig. 2). The other type of abnormal neuronal formation consisted of groups of loosely scattered fusiform or bipolar neurons in layer Pre-β, which were markedly smaller and were often arranged in columns (Fig. 3).

The histologically evident reduction of neuronal volume in layer III in schizophrenic patients can be documented with the aid of a computed analytical method. The two types of neuronal groups were encountered in alternating fashion in sections at different levels. The authors assumed that these atypical neurons, which are reminiscent of "young neurons" (Rakic 1975, 1988a) in their shape and arrangement, had become stuck, as it were, in the last phase of migration and stayed in place as "ectopic neurons" unable to reach their pre-assigned destinations in layer Pre-α. It thus appears that there may be a local disturbance of neuronal development and/or migration restricted to the rostral portion of the entorhinal area in a late phase of development (Jakob and Beckmann 1986, 1994; Beckmann and Jakob 1994).

In four cases of manic-depressive illness, there were cytoarchitectural abnormalities here and in the rostral ventral portion of the insular cortex that were similar to those described in the schizophrenic psychoses, agreeing in all histologic details. The common features of these two types of psychosis have been pointed out many times (Beckmann and Jakob 1991).

We still possess inadequate knowledge, however, of the possible differences between the overall constellations of neuropathological abnormalities in these two types of disease. At present, only scattered observations indicate that such differences may exist. Correlations of the pathological features of these two types of psychosis with their clinical manifestations will presumably be possible in the future through further, detailed histologic studies of potential differences in the severity and extent of alterations in the entorhinal cortex of the two hemispheres.

The existence of structural architectural abnormalities in the rostral entorhinal cortex of patients with either type of major psychosis was confirmed by two

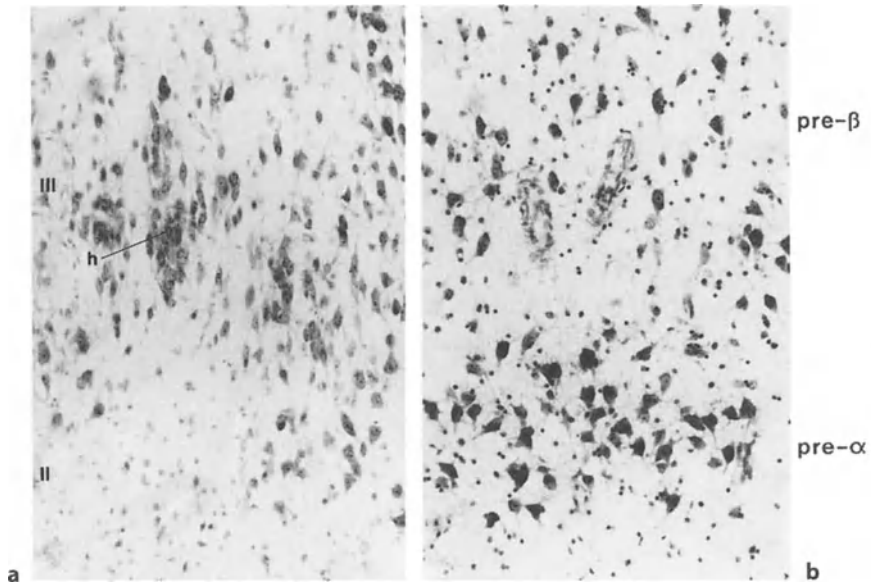


Fig. 2. a Focal malformation in the entorhinal area in a patient with chronic schizophrenia. Rostral cortical fields in a serial histologic study, fourth stage of the series, layers II/III Pre-α and Pre-β; layer II Pre-α shows irregularly scattered neurons without characteristic formation of cell islands; heterotopic groups (*h*) of presumably immature neurons in columnar configuration, in layer III Pre-β; and, below these areas, a neuron-free zone. **b** Control sections at the same level: layers Pre-α and Pre-β with normal cytoarchitectural structures. Nissl (20 μ), × 125. (After Jakob and Beckmann 1994)

further groups of investigators who studied cases taken from the Yakovlev Collection (Arnold et al. 1991). Quantitative studies yielded striking findings: in eight cases of schizophrenia, five cases of cyclothymia, and eight control subjects, the number of neurons in layer II Pre-α varied from zero, in severely affected cases, to normal. In such cases, the cytoarchitecture of the other layers was also markedly abnormal. Neurons of layer Pre-α were displaced into layer Pre-β. No significant differences were found between the two types of major psychosis (Casanova et al. 1991).

Krimer et al. (1997) studied the entorhinal area of schizophrenic patients and controls using inadequate methods. The fixation time was excessively long (up to 1 year), and the postmortem interval (average, 36 h) was unsuitable for sophisticated cytoarchitectural studies. Nonetheless, even in these authors' unclear illustrations, cytoarchitectural differences between control subjects and schizophrenics are evident. Senitz and Kalus (in press) were recently able to confirm the findings of Jakob and Beckmann (1986, 1994) in a series of 20 cases.

Cortical malformations of this type may have either of two possible causes:

1. The neurons are unable to begin migrating.
2. The migrating neurons remain in an ectopic position along the way to the cortex (Rakic 1988a,b).

The atypical neurons that do not belong in layer Pre-β seem to have encountered the second type of difficulty. Many of them are of obviously reduced volume when compared to other neurons in the same layer and to those of control subjects. These small neurons, which are often marked by a bipolar shape (see Fig. 3) or lie in layer Pre-β more as heterotopic clusters or as columns containing densely arrayed, undifferentiated neurons (see Fig. 2), seem to have become stuck along their way to the upper layer, Pre-α (Jakob and Beckmann 1994; Beckmann and Jakob 1994). A specific histologic demonstration of these neurons is not yet possible at present; they can be characterized only with the aid of an optimal staining technique.

These findings seem to imply that the malformation arises at a relatively late time in development. It is possible that there is a defect in the ontogenetic columns (second category; Rakic 1988b). The lower Pri-layers are heavily depopulated of cells in only a few cases. In view of the spectrum of alterations seen, and the time at which migration begins in the corresponding region of the human brain, a fetal injury at some time between the late third month and the fifth month of gestation can be presumed. A time of injury between the late fourth month and the fifth month is likely in most cases, because the upper cortical layers are often exclusively affected (Beckmann and Jakob 1994).

6
Temporal and Frontal Cortex

Evidence for a more generalized developmental disorder was obtained from the study of the temporal and frontal lobes of schizophrenics and control subjects. A particular type of neuron was studied that contains the

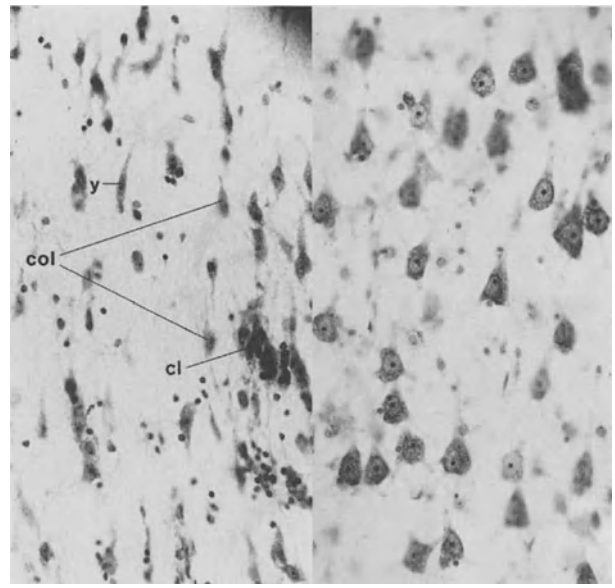


Fig. 3. a Sixth stage of the same serial histologic study as in Fig. 2 (left entorhinal area in a patient with chronic schizophrenia). Layer III Pre-β of the rostral portion of the left entorhinal area. Groups of small, undifferentiated neurons, either as clusters (*cl*) or as columnar arrays of fusiform cells (*col*); a few neurons clearly possess the characteristic features of "young" neurons (*y*). **b** Control: layer III of the left entorhinal region at the same level; considerable variation in the size and density of Pre-β pyramidal cells. Nissl (20 μ), × 200. (After Jakob and Beckmann 1994)

enzyme nicotinamide adenine dinucleotide phosphate diaphorase (NADPH-d) and is normally found in the subcortical white matter. The number of nonpyramidal neurons of this type was significantly reduced in schizophrenic patients in the lower cortical layers and in the directly underlying white matter, both in the temporal isocortex and in the dorsolateral prefrontal cortex, but they were present at increased density in the deep white matter. This local difference was regarded by the authors as due to a disturbance in the development of the "subplate zone." It was postulated that these neurons fail to reach their destination while migrating toward the cortical plate and, instead, become stuck in the deep white matter as a consequence of a disturbance or inhibition of some type.

The "subplate zone" plays a vital role in the establishment of afferent pathways from other areas during the final phase of neuronal migration. A developmental disturbance in this area may thus affect widely dispersed frontal association fields. Defective functioning of the frontal lobe, particularly of association systems, is known to occur as a negative manifestation of schizophrenia (Akbarian et al. 1993a,b). Immunocytochemical studies employing neuron-specific antibodies also appear to indicate a disorder of migration in the frontal cortex of schizophrenics. The Cajal-Retzius cells (CRC) normally found in the molecular layer were distributed differently in schizophrenics and normal control subjects. The number of CRC in layer I did not differ in the two groups, but schizophrenics had significantly more of them in the lowest third of the layer than control subjects, in whom more CRC were found in the middle and uppermost thirds. The authors interpreted this finding as an expression of a disturbance of neuronal migration (Kalus et al. 1997a).

7 Orbitofrontal Area

Senitz and colleagues (Senitz et al. 1979; Senitz and Winkelmann 1981, 1991) used the Golgi technique as a routine method for the study of Brodmann areas 19 and 11 and were the first to find neuropathological abnormalities in the orbitofrontal region of schizophrenic patients. Because their method demonstrated the overall structure of the neurons, they were able to describe particularly striking and unusual neuronal forms:

- So-called "triangle cells" in layer VI were found to be more numerous and irregularly organized than in control cases.
- Many pyramidal cells were demonstrated in layer V that had forked major dendrites that could be followed all the way to layer II (Fig. 4). Dendritic duplication of this type can occur only during cortical development.
- Pyramidal cells were found in layer III that had relatively thick dendrites and atypically long, unusually shaped and thick spines. The number of spines was quantitatively measured and found to be significantly elevated on a large proportion of pyramidal cells. They often lay in tufts on the surface of the major dendrite or had several forked spine heads.

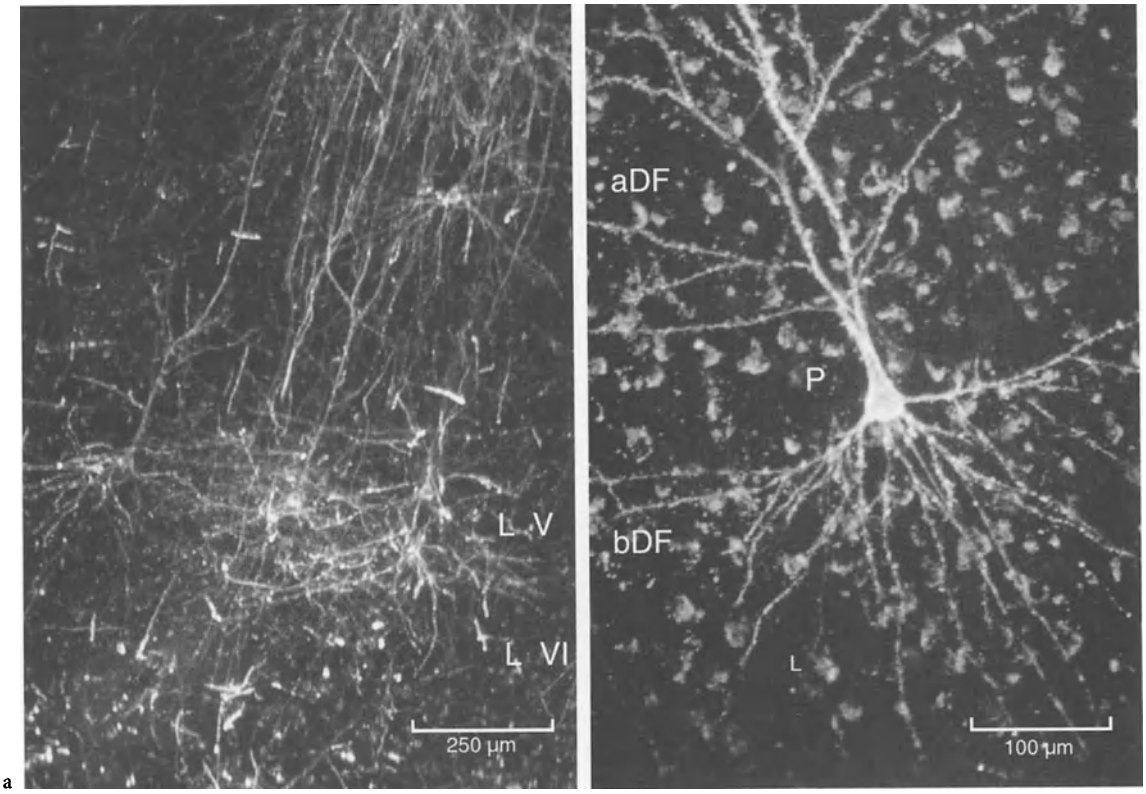


Fig. 4a,b. Orbitofrontal cortex (Brodmann area 11) in a schizophrenic patient. **a** Low-magnification view: histologic demonstration of a group of layer V pyramidal cells with multiply forked major dendrites. *LV*, *LVI*, cortical layers V and VI. **b** Detail: pyramidal cell with a singly forked major dendrite. *aDF*, apical dendritic field; *bDF*, basal dendritic field; *P*, perikaryon (cell body); *L*, lipofuscin in neuronal perikaryon. Confocal microscopy. Postmortem fluorescence technique with DiI (1,1′-dioctadecyl-3,3,3′,3′-tetramethylindocarbocyanine perchlorate). (Modified from Senitz and Winkelmann 1981; see also Supprian et al. 1993)

The findings in these neurons presumably reflect ordinary, nonpathological histologic features that are to be regarded as plastic alterations in the area of the dendritic trunk (Geinisman et al. 1989). They may be interpreted as an expression of altered functioning. Because this area is tightly linked to the rostral cortex (Insausti et al. 1987), it is possible that the abnormalities noted are due to a developmental disturbance occurring in the migratory phase as a result of the malformation in the entorhinal area. It is clear that the orbitofrontal cortex is subject both to independent, mild disturbances of development running in parallel to the entorhinal disturbance and to secondary compensation effects leading to plastic alterations in the pyramidal neurons of layer III. The topography of the areas projecting to the entorhinal area is depicted in Fig. 5. Orbitofrontal cortex and the rostral portion of the claustrocortex (insula) project onto rostral portions of the entorhinal cortex, while the cingulate gyrus and the fields of the superior temporal gyrus project onto caudal portions. The different fields of the perirhinal and parahippocampal cortical areas send their projections uniformly throughout the entire entorhinal area.

8 Claustrocortex

The cytoarchitectural abnormalities of the rostral ventral portion of the insular cortex are rather mild. They are restricted to a particular region and occur in tandem with the cytoarchitectural abnormalities of the rostral portion of entorhinal area (Jakob and Beckmann 1986; see Fig. 1). In this cortical region, the number of neurons in layers II and III is fairly uniformly reduced by approximately 30%. Their volume is also reduced. The neurons are not radially oriented, but rather diagonally oriented or twisted. These findings and the complete absence of gliosis are evidence against atrophy or hypoxic parenchymal injury (Jakob and Beckmann 1989). The fields of this region of the insula are characterized by a particular

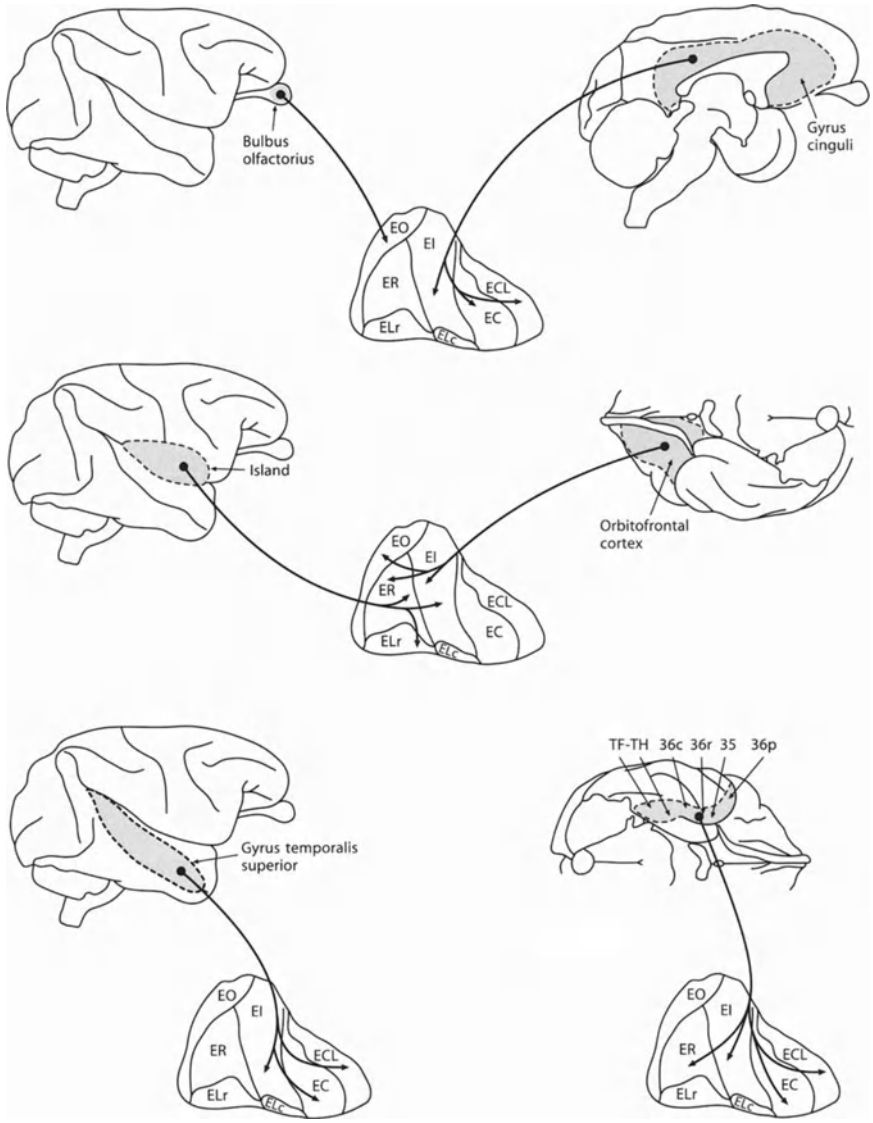


Fig. 5. Summary diagram of the (mostly cortical) afferent projections of the primate entorhinal area. *EO*, olfactory field; *ER*, rostral field; *Elr*, rostral-lateral field; *EI*, intermediate field; *EC*, caudal field; *Elc*, caudal-lateral field; *ECL*, caudal limiting field. (After Insausti et al. 1987)

cellular architecture (agranular region; Stephan 1975) and project onto rostral portions of the entorhinal cortex (see Fig. 5). It is not clear whether the alterations in this area are secondary ones occurring in the presence of a malformation.

9 Cingulate Gyrus

Neuronal clusters, which occur primarily in layer II of cingulate cortex but are also found in the prefrontal region, are thought to constitute the most important specific functional component of the cingulate gyrus. These groups of tightly associated cells are smaller and farther apart from each other in patients with endogenous psychoses than in normal control subjects. Furthermore, studies of the cingulate gyrus of patients with endogenous psychoses using immunohistochemical methods with neuron-specific antibodies have revealed that it contains a significantly increased number (25% increase) of long association fibers with a vertical course, particularly in the vicinity of blood vessels. This finding probably reflects an increase in associative neurons from frontal cortical areas (Benes

et al. 1992; Benes 1993). There were also cytoarchitectural changes of a more general nature in both cingulate and frontal cortex. There was a marked reduction in interneurons in layer II of both regions, more pronounced in the cingulate gyrus. Furthermore, the number of neurons in layer V of the cingulate gyrus, in layer VI of the frontal cortex, and in layer III of the motor cortex was significantly reduced in the absence of evidence of degenerative changes (Benes et al. 1986, 1991).

On the other hand, there was an increased number of pyramidal cells in layer V of the frontal cortex, which was thought to be related to an increase in associative afferent input. It is assumed that the increased number of vertical axons in the cingulate gyrus may arise from layer V and from the pyramidal cells of the prefrontal area (Benes et al. 1992; Benes 1993). The various types and subtypes of γ-aminobutyric acid (GABA)-ergic cortical interneurons in the rostral portion of the cingulate gyrus were analyzed by an immunohistochemical technique with antibodies against parvalbumin (PV). Most of the PV-positive interneurons were found in layers III and Vb of Brodmann area 24c (Kalus and Senitz 1996). Postmortem studies on the brains of schizophrenic patients at first revealed the same distribution of interneurons in the cerebral cortex as in normal controls. Nonetheless, even though the number of Nissl-stained neurons was unchanged, there was a significant increase in the number of neurons in layers Va and Vb. This finding, which must represent an increased number of interneurons at the expense of other neurons, was interpreted as a reflection of a developmental disturbance. The increased number of neurons might result in an increased inhibition of projection neurons, which could, in turn, influence the neuronal output of the rostral portion of the cingulate gyrus and, through it, the limbic functional circuit (Kalus et al. 1997b).

The fact that these cytoarchitectural changes occur precisely in the rostral portion of the cingulate gyrus, as has since been confirmed in multiple replicating studies, leads us to consider the connections of this area to the entorhinal cortex (see Fig. 5). Although the different groups of investigators arrived at partly conflicting conclusions, there is a consensus that neuronal disturbances occur in the lower layers of the cortex. This would be an unusual finding for a primary developmental disorder. It thus seems reasonable to assume that there is a secondary neuronal disorder resulting from the malformation in the entorhinal cortex.

The cytoarchitectural abnormalities of the entorhinal area and the anterior portion of the cingulate gyrus are among the best-validated findings in schizophrenia research (Talamini et al. 1995). Both of these structures, like the hippocampus, are essential components of the limbic system. There are connections between the cingulate gyrus and the frontal isocortex that enable an exchange of information at all levels (Lopez da Silva et al. 1990). Thus the anterior portion of the cingulate gyrus is a second connection between the frontal cortex and the hippocampal formation. The most important exit pathway of this system is via the nucleus accumbens, which is located in the ventral striatum (Sorensen and Witter 1983). The nucleus accumbens projects, in turn, by way of the internal pallidal segment and the mediodorsal thalamic nucleus back to the prefrontal cortex (Groenewegen and Russchen 1984; Groenewegen et al. 1990; Haber et al. 1985). Thus frontal cortical association areas are incorporated into the limbic regulatory circuit.

10 Other Cortical Areas

In a study of the brains of patients with schizophrenia and schizoaffective disorder, the cell density in Brodmann areas 9 and 46 of the frontal cortex and in area 17 of the occipital cortex (i.e. the primary visual cortex) was analyzed with respect to cytoarchitecture. In schizophrenic patients, the neuron density was significantly increased in the individual layers of area 9 (17%) and area 46 (layers II and III) of frontal cortex and in the visual cortex (10%). In area 9, the neuron density was increased in layers III–VI because of increased numbers of both pyramidal and nonpyramidal neurons. The neuron density was not parallel to either of the two parameters postmortem interval (PMI) or time in formaldehyde (TF). Cortical thickness, however, was reduced in schizophrenics to a mild but not statistically significant extent, combined with a relatively marked reduction of layer V in frontal area 9. In the frontal lobe, volume reduction was found only in the neurons, and not in the glia (Selemon et al. 1993; Rajkowska et al. 1994; Selemon et al. 1995).

The authors first postulated neuronal atrophy as a possible anatomic substrate for faulty information processing in schizophrenia. This was made less likely, however, by the finding that the lessened cortical thickness was not attributable to a uniform reduction of all of the layers. Considerable differences between individual layers were found; layer V was the most strongly affected in the brains of schizophrenic patients. This finding may be of functional significance precisely at this location in the frontal lobe, because these layers contain the neurons of origin for the corticocortical and corticostriatal projections; moreover, layer V contains a high concentration of D_1- and D_2-type dopamine receptors (Goldman-Rakic et al. 1990).

The greatest amount of variation in neuronal density among layers was found in the brains of patients with schizoaffective disorders. Densities were found to range from extreme values to an entirely normal distribution. Nonetheless, only one subgroup of patients with schizoaffective disorders also had an elevation of neuron density reminiscent of that seen in schizophrenia. These alterations in the layers may be attributable to a reduction in neuropil (Selemon et al. 1995; Watson and Meador-Woodruff 1995). In this case, these alterations, too, could be regarded as only secondary in the overall scheme of disturbances of cortical development that has been constructed to date; the developmental disturbances of the entorhinal cortex and of the cingulate gyrus would retain their primary importance (Talamini et al. 1995). It is reasonable to conclude that the malformation of the rostral portion of the entorhinal area, which has extensive connections to frontal areas and to all sensory areas, including the visual cortex (van Hoesen 1982; Insausti et al. 1987; Pandya and Yeterian 1985), is the cause of the reduction in neuropil.

With respect to its function, the prefrontal association cortex is a highly differentiated structure. This cortical area carries out the task of planning motor activity, after first assessing the consequences of possible future actions to the individual. Studies employing positron emission tomography (PET) and single photon emission computed tomography (SPECT) have shown that the prefrontal cortex functions abnormally in schizophrenic patients, particularly those with predominantly "negative" disease manifestations. Furthermore, lesions of frontal cortical areas, including limbic association areas, usually lead to deficits of attention and initiative, both of which are "negative" manifestations of schizophrenia (Talamini et al. 1995).

11 Subcortical Nuclei

In addition to the hippocampus, the basal ganglia and the thalamus have also been subjects of interest in schizophrenia research. The findings obtained to date remain controversial. Volumetric studies of the basal ganglia using material from the Vogt Collection revealed a reduction in volume of the internal pallidum and the amygdalar nucleus (Bogerts 1984, 1985; Bogerts et al. 1985); in cases of catatonic schizophrenia, the number of neurons was reduced in both the internal and the external pallidal segments (Bogerts et al. 1985). The nucleus accumbens was also reduced in volume in male schizophrenics (Bogerts et al. 1990). Unfortunately, these volumetric findings could not be confirmed in further studies employing stereological methods on serial Nissl sections of entire hemispheres. The volume of the amygdalar nucleus was not significantly reduced, while the volume of the striatum on the left side and that of the pallidum on the right side were significantly increased (Heckers et al. 1991).

Extensive postmortem studies of entire cerebral hemispheres were recently carried out on nine brains of schizophrenic patients and nine of normal control subjects, all of whom were less than 60 years old. Individual and regionally specific shrinkage factors and the ratio of the size of the basal ganglia to that of the entire hemisphere were taken into account in each case by the application of sophisticated statistical volumetric methods. A relative increase in volume – also designated volume density – was found for the striatum of both sides and for the right nucleus accumbens. In the same study, the neurons and glial cells in the striatum of both sides were counted by optical dissection under a stereomicroscope. A significantly increased number of neurons was found in the caudate nucleus–nucleus accumbens complex of the right side, but not in the right putamen. There was no alteration in the number of glial cells. An influence of the developmental disorder of limbic cortical areas on the course of the normally occurring process of apoptosis is under discussion as a possible pathogenetic mechanism (Lauer and Beckmann 1997; Beckmann and Lauer 1997). "Cell loss" in the intermediate and posterior thalamic nuclei was reported in early studies (Hempel and Treff 1959; Dom et al. 1981).

An improvement in stereological technique was brought about by the introduction of the so-called Cavalieri dissector combination, which employs two microscopes to differentiate neurons and glial cells from one another precisely, in three dimensions. This method allows both the numerical density and the total number of neurons to be determined, while avoiding double counting. No significant alterations were found in the numbers of neurons and glial cells in the inner pallidal segment or in the basolateral nucleus of the amygdala (Gundersen and Jensen 1987). The number of neurons was reduced by 40% in the mediodorsal thalamic nucleus and by 50% in the nucleus accumbens (Pakkenberg 1990). The mean volumes of the left thalamic nucleus and the nucleus accumbens were studied with similar methods and found to be reduced (Pakkenberg and Gundersen 1988; Pakkenberg 1990).

In functional anatomic terms, the subcortical nuclei are connected in many different ways with the limbic areas and functional circuits. Frontal, temporal, and parietal cortical association areas and the cingulate gyrus project principally onto the caudate nucleus (Selemon and Goldman-Rakic 1985). Limbic cortical

areas and the hippocampus provide afferent input to the nucleus accumbens (Pennartz et al. 1994). Particular attention has been paid to the position of the mediodorsal thalamic nuclei because of their intimate connections to the prefrontal cortex within the limbic system (Divac et al. 1978; Markowitsch 1982; Goldman-Rakic et al. 1984; Goldman-Rakic and Porrino 1985). They derive afferent input from frontal NADPH neurons (Giguere and Goldman-Rakic 1988).

The anterior thalamic nuclei derive afferent input from the hippocampal formation and project back to it by way of the cingulum (Papez regulatory circuit). They belong to a large group of cells that, among other functions, provide afferent input to the limbic system. Information is relayed from the brainstem to the limbic Papez regulatory circuit by way of the mamillary bodies, which have the function not only of regulating the forward flow of information, but also of relaying it back to the brainstem by way of the mamillotegmental tract (Swanson 1983). There are also connections to the nucleus accumbens within the Papez regulatory circuit. Descending fibers of the precommissural fornix arising in the subiculum divide into individual efferent connections to the lateral septum, medial portions of the nucleus accumbens, the frontal cortex, and the gyrus rectus (Rosene and van Hoesen 1977).

Thus both the anterior and the dorsomedial thalamic nuclei, the striatum, and the nucleus accumbens are either wholly incorporated into the limbic system or have multiple connections, not only with "ascending" fibers from the hypothalamus and parts of the brainstem, but also with "descending" fibers from the limbic area.

12 Brainstem

The anatomic structures under discussion here include areas of the midbrain, the pons, and the reticular formation, which extends downward into the medulla oblongata, and also the periventricular nuclei in the walls of the third ventricle.

A study of serial myelin-stained sections of the brains of schizophrenics and control subjects obtained from the Vogt Collection revealed a significant reduction in volume of the periventricular gray matter (Lesch and Bogerts 1984). The authors assumed that this reflected a loss of diencephalic gray matter and related these anatomic changes to the diencephalic dysfunction that can be observed in chronic schizophrenics. It remains questionable, however, whether such far-reaching conclusions can be drawn from myelin-stained sections alone.

More recently, the brainstem reticular formation was studied by immunocytochemical methods in a small group of schizophrenics and control subjects. The pedunculopontine nucleus (PPN), the lateral dorsal tegmental nucleus (LDT), and the locus ceruleus were studied as completely as possible at several sectional levels with the aid of three-dimensional reconstruction. Using NADPH diaphorase, the authors found that the PPN of schizophrenics contained an increased number of cholinergic neurons and was reduced in volume by approximately 30%. These changes in the gray matter were attributed to the effect of a central neuronal developmental disturbance (Karson et al. 1991).

As for the function of these nuclei, it is known that disturbances of non-rapid eye movements (NREM) and a reduction of the REM latency can occur in the sleep of schizophrenics (Feinberg et al. 1969; Itil et al. 1972; Zarcone et al. 1975; Tandon and Greden 1989). There are also specific anomalies of movement and posture (King 1974). These and other functions are modulated by the important cholinergic cell groups, the PPN and the LDT (Mesulam et al. 1984; Satoh and Fibiger 1985; Isaacson and Tanaka 1986; Woolf and Butcher 1986; Jones and Beaudet 1987). There are projections from the dorsolateral prefrontal cortex to the reticular formation and to the pontine nuclei in the primate brain, and it is therefore supposed that a functional disturbance of the reticular formation may provide the basis for the "hypofrontality" seen in schizophrenia (Arnsten and Goldman-Rakic 1984).

13 Conclusion

The neuropathological findings obtained since 1979 relating to cortical and subcortical developmental disturbances in the so-called endogenous psychoses, and particularly schizophrenia, have partly been confirmed in recent years, but have also led to the raising of new questions. As far as can be said today, the focal cortical alterations that have been confirmed to exist in a small number of specific areas share certain common features. Numerous variations in the involved areas, however, make each individual case unique. This, in turn, is reflected in the variety of clinical pictures encountered among the so-called endogenous psychoses. The findings available to date thus give rise to a mosaic.

The anterior portion of the entorhinal area, intimately connected with the hippocampus in the "entorhino-hippocampal loop," and the cingulate gyrus, which is connected to the hippocampus by the

Papez regulatory circuit, are the structures mainly affected by developmental disturbances of cytoarchitecture and are thus important areas of limbic functional disturbance (Talamini et al. 1995). Histologic abnormalities are seen more frequently in anterior than in posterior regions of the temporal lobe and cingulate gyrus, and it is therefore of interest that these rostral areas are much more closely connected to the prefrontal cortex than the caudal areas of these structures. Furthermore, the cytoarchitectural abnormalities are mainly or exclusively localized in the upper cortical layers, which, in general, are connected to other areas by way of afferent or efferent fibers. If histologic data of relevance to functional disorders and their consequences are to be obtained, this should be done by the study of the entire neuron, including its processes, spines, and synapses.

The question of why the course of endogenous psychoses may be phasic, stuttering, or chronic cannot yet be answered satisfactorily. It seems, however, that disturbances in the structures mentioned serve as "foci," which, in a process similar to "kindling," may progressively sensitize or "trigger" certain intracerebral regulatory circuits until these finally discharge and thereby release specific psychotic or affective manifestations (Racine et al. 1989). This hypothesis is supported by a number of findings discussed by Post et al. (1975, 1988), among others, and has a certain degree of plausibility.

Nothing conclusive can be said about the origin of the prenatal developmental disturbances that occur in the second trimester of gestation. Faulty genetic programming and toxic influences on the developing brain are under intense discussion as possible etiologies and are the object of further research that is currently being planned (Beckmann and Franzek 1992).

14 References

*Akbarian S, Bunney WE, Potkin SG, Wigal SB, Hagman JO, Sandman CA, Jones EG (1993a) Altered distribution of nicotinamide-adenine dinucleotide phosphate-diaphorase cells in frontal lobe of schizophrenics implies disturbances of cortical development. Arch Gen Psychiatry 50: 169–177

Akbarian S, Vinuela A, Kim JJ, Potkin SG, Bunney WE, Jones EG (1993b) Distorted distribution of nicotinamide-adenine dinucleotide phosphate-diaphorase neurons in temporal lobe of schizophrenics implies anomalous cortical development. Arch Gen Psychiatry 50: 178–187

Altshuler LL, Conrad A, Kovelman JA, Scheibel A (1987) Hippocampal pyramidal cell orientation in schizophrenia: a controlled neurohistological study of the Yakovlev collection. Arch Gen Psychiatry 44: 1094–1098

Altshuler LL, Casanova MF, Goldberg TE, Kleinman JE (1990) The hippocampus and parahippocampus in schizophrenic, suicide, and control brains. Arch Gen Psychiatry 44: 1094–1098

Alzheimer A (1897) Beiträge zur pathologischen Anatomie der Hirnrinde und zur anatomischen Grundlage einiger Psychosen. Monatsschr Psychiatr Neurol 2: 1763–1769

Arnold SE, Hyman BT, van Hoesen GW, Damasio AR (1991) Cytoarchitectural abnormalities of the entorhinal cortex in schizophrenia. Arch Gen Psychiatry 48: 625–632

Arnold SE, Franz BR, Gur RC, Gur RE, Shapiro RM, Moberg PJ, Trojanowski JQ (1995) Smaller neuron size in schizophrenia in hippocampal subfields that mediate cortical-hippocampal interactions. Am J Psychiatry 152: 738–748

Arnsten AFT, Goldman-Rakic PS (1984) Selective prefrontal cortical projections to the region of the locus coeruleus and raphe nuclei in the rhesus monkey. Brain Res 306: 9–18

*Beckmann H, Franzek E (1992) Deficit in birthrates in winter and spring months in distinct subgroups of mainly genetically determined schizophrenia. Psychopathology 25: 57–64

**Beckmann H, Jakob H (1991) Prenatal disturbances of nerve cell migration in the entorhinal region: a common vulnerability factor in functional psychoses? J Neural Transm Gen Sect 84: 155–164

**Beckmann H, Jakob H (1994) Pränatale Entwicklungsstörungen von Hirnstrukturen bei schizophrenen Psychosen. Nervenarzt 65: 454–463

*Beckmann H, Lauer M (1997) The human striatum in schizophrenia. II. Increased number of striatal neurons in schizophrenics. Psychiatry Res 68: 99–109

Benes FM (1988) Post-mortem structural analysis of schizophrenic brain: study design and the interpretation of data. Psychiatr Dev 3: 213–226

*Benes FM (1989) Myelination of cortical-hippocampal relays during late adolescence. Schizophr Bull 15: 585–593

Benes FM (1993) Neurobiological investigations in cingulate cortex of schizophrenic brain. Schizophr Bull 19: 537–549

Benes FM, Davidson J, Bird ED (1986) Quantitative cytoarchitectural studies of the cerebral cortex of schizophrenics. Arch Gen Psychiatry 43: 31–35

Benes FM, McSparren J, Bird ED, SanGiovanni JP, Vincent SL (1991) Deficits in small interneurons in prefrontal and cingulate cortices of schizophrenic and schizoaffective patients. Arch Gen Psychiatry 48: 996–1001

Benes FM, Sorensen I, Vincent SL, Bird ED, Sathi M (1992) Increased density of glutamate-immunoreactive vertical processes in superficial laminae in cingulate cortex of schizophrenic brain. Cereb Cortex 2: 503–512

*Blackwood W (1976) Neuronal structure and cellular pathology of the nerve call and neuroglia. In: Blackwood W, Corsellis JAN (eds) Greenfield's neuropathology, 3rd edn. Arnold, London, pp 1–42

Bogerts B (1984) Zur Neuropathologie der Schizophrenien. Fortschr Neurol Psychiatr 52: 428–437

Bogerts B (1985) Schizophrenien als Erkrankungen des limbischen Systems. In: Huber G (ed) Basisstadien endogener Psychosen und das Borderline-Problem. Schattauer, Stuttgart, pp 163–179

Bogerts B (1986) Hirnatrophische Prozesse bei Schizophrenen. Ein quantitativer Vergleich mit Parkinson- und Huntington-

Erkrankung. In: Keup W (ed) Biologische Psychiatrie. Springer, Berlin Heidelberg New York, pp 270–275

*Bogerts B (1989) Limbic and paralimbic pathology in schizophrenia: interaction with age- and stress-related factors. In: Schulz SC, Tamminga CA (eds) Schizophrenia: scientific progress. Oxford University Press, Oxford, pp 216–227

Bogerts B, Meertz E, Schönfeldt-Bausch R (1985) Basal ganglia and limbic system pathology in schizophrenia: a morphometric study of brain volume and shrinkage. Arch Gen Psychiatry 42: 784–791

Bogerts B, Falkai P, Tutsch J (1986) Cell numbers in the pallidum and hippocampus of schizophrenics. In: Shagass C, Josiassen RC, Bridger WH, Weiss KJ, Stoff D, Simpson GM (eds) Biological psychiatry. Elsevier, New York, pp 1178–1180

Bogerts B, Falkai P, Haupts M, Greve B, Ernst S, Tapernon-Franz U, Heinzmann U (1990) Post-mortem volume measurements of limbic system and basal ganglia structures in chronic schizophrenics. Initial results from a new brain collection. Schizophr Res 3: 295–301

**Braak H (1980) The allocortex. The entorhinal region. In: Braitenberg V, Barlow HB, Bizzi E, Florey E, Grüsser OJ, van der Loos H (eds) Studies of the brain function, vol 4. Architectonics of the human telencephalic cortex. Springer, Berlin Heidelberg New York, pp 37–48

Braitenberg V, Schüz A (1983) Some anatomical comments on the hippocampus. In: Seifert W (ed) Neurobiology of the hippocampus. Academic, London, pp 21–37

Cammermeyer J (1967) Artificial displacement of neuronal nucleoli in paraffin sections. J Hirnforsch 9: 209–224

Casanova MF, Kleinman JE (1990) The neuropathology of schizophrenia: a critical assessment of research methodologies. Biol Psychiatry 27: 353–362

Casanova MF, Saunders R, Altshuler L, Goldberg T, Armstrong E, Weinberger DR, Kleinman JE (1991) Entorhinal cortex pathology in schizophrenia and affective disorders. In: Racagni G, Brunello N, Fukuda T (eds) Biological psychiatry. Elsevier, Amsterdam, pp 504–506

Caviness VS Jr, Pinto-Lord MC, Evrard P (1981) The development of laminated pattern in the mammalian neocortex. In: Connelly TG, Brinkley LL, Carlson BM (eds) Morphogenesis and pattern formation. Raven, New York, pp 103–126

Christison GW, Casanova MF, Weinberger DR, Rawlings R, Kleinmann JE (1989) A quantitative investigation of hippocampal pyramidal cell size, shape and variability of orientation in schizophrenia. Arch Gen Psychiatry 46: 1027–1032

*Conrad AJ, Abebe T, Austin R, Forsythe S, Scheibel AB (1991) Hippocampal pyramidal cell disarray in schizophrenia as a bilateral phenomenon. Arch Gen Psychiatry 48: 413–417

Crow TJ, Ball J, Bloom SR et al (1989) Schizophrenia as an anomaly of development of cerebral asymmetry: a postmortem study and a proposal concerning the genetic basis of the disease. Arch Gen Psychiatry 46: 1145–1150

*Crow TJ, Brown R, Bruton CJ, Frith CD, Gray V (1992) Loss of Sylvian fissure asymmetry in schizophrenia: findings in the Runwell 2 series of brains. Schizophr Res 6: 152–153

Divac I, Kosnal A, Björklund A, Lindvall O (1978) Subcortical projections to the prefrontal cortex in the rat as revealed by its horseradish peroxidase. Neuroscience 3: 785–796

Dom R, de Saedeler J, Bogerts B, Hopf A (1981) Quantitative cytometric analysis of basal ganglia in catatonic schizophrenics. In: Perris C, Struwe G, Jansson B (eds) Biological psychiatry. Elsevier, Amsterdam, pp 723–726

Dooling EC, Chi JG, Gilles FH (1983) Telencephalic development: changing gyral patterns. In: Gilles FH, Leviton A, Dooling EC (eds) The developing human brain: growth and epidemiologic neuropathology. Wright, Boston, pp 94–104

Dunlap CB (1928) The pathology of the brain in schizophrenia. Proc Assoc Res Nervous Ment Dis 5: 371–381

Falkai P, Bogerts B (1986) Cell loss in the hippocampus of schizophrenics. Eur Arch Psychiatry Neurol Sci 236: 154–161

Falkai P, Bogerts B, Greve B et al (1992) Loss of Sylvian fissure asymmetry in schizophrenia. A quantitative post-mortem study. Schizophr Res 7: 23–32

Feinberg I, Braun M, Koresko RL, Gottlieb F (1969) Stage 4 sleep in schizophrenia. Arch Gen Psychiatry 21: 262–266

Fifkova E (1974) Two types of terminal degeneration in the molecular layer of the dentate fascia following lesions of the entorhinal cortex. Brain Res 96: 169–175

Fünfgeld EW (1952) Pathologisch-anatomische Untersuchungen am Nucleus anterior thalami bei Schizophrenie. In: Rosenberg H, Sellier V (eds) First International Congress of Neuropathology, vol 3. Turin, pp 648–659

Geinisman Y, Morrell F, de Toledo-Morrell L (1989) Perforated synapses on double-headed dendritic spines: a possible structural substrate of synaptic plasticity. Brain Res 480: 326–329

Giguere M, Goldman-Rakic PS (1988) Mediodorsal nucleus: areal, laminar and tangential distribution of afferents and efferents in the frontal lobe of rhesus monkeys. J Comp Neurol 277: 195–213

Goldman-Rakic PS, Porrino LJ (1985) The primate mediodorsal (MD) nucleus and its projection to the frontal lobe. J Comp Neurol 252: 535–560

Goldman-Rakic PS, Selemon LD, Schwarz ML (1984) Dual pathways connecting the dorsolateral prefrontal cortex with the hippocampal formation and parahippocampal cortex in the rhesus monkey. Neuroscience 12: 719–743

Goldman-Rakic, Lidow M, Gallager DW (1990) Overlap of dopaminergic, adrenergic, and serotinergic receptors and complementarity of their subtypes in primate prefrontal cortex. J Neurosci 10: 2125–2138

Groenewegen HJ, Russchen FT (1984) Organization of the efferent projections of the nucleus accumbens to pallidal, hypothalamic and mesencephalic structures: a tracing and immunohistochemical study in the cat. J Comp Neurol 223: 347–367

Groenewegen HJ, Berendse, HW, Wolters JG, Lohman AHM (1990) The anatomical relationship of the prefrontal cortex with the striatopallidal system, the thalamus and the amygdala: evidence for a parallel organization. Progr Brain Res 85: 95–118

Gundersen HJG, Jensen EB (1987) The efficiency of systematic sampling in stereology and its prediction. J Microsc 147: 229–263

Haber SN, Groenewegen, HJ, Grove, EA, Nauta WJH (1985) Efferent connections of the ventral pallidum: evidence of a dual striato-pallidofugal pathway. J Comp Neurol 235: 322–335

*Hassler R (1964) Limbische und diencephale Systeme der Affektivität und Psychomotorik. In: Hoff H, Tschabitscher H, Kryspin-Exner K (eds) Muskel und Psyche. Karger, Vienna, pp 1–31

Heckers S, Heinsen H, Heinsen Y, Beckmann H (1990a) Limbic structures and lateral ventricle in schizophrenia. A quantitative post-mortem study. Arch Gen Psychiatry 47: 1016–1022

Heckers S, Heinsen H, Heinsen Y, Beckmann H (1990b) Morphometry of the parahippocampal gyrus in schizophrenics and controls. Some anatomic considerations. J Neural Transm Gen Sect 80: 151–155

*Heckers S, Heinsen H, Geiger B, Beckmann H (1991) Hippocampal neuron number in schizophrenia. Arch Gen Psychiatry 48: 1001–1008

Hempel KJ, Treff WM (1959) Über "normale Lücken" und "pathologische Lückenbildungen" in einem subcorticalen Griseum (mediodorsaler Thalamuskern). Beitr Pathol 121: 287–300

Hirano A (1985) Neurons, astrocytes and ependyma. In: Davis RL, Robertson DM (eds) Textbook of neuropathology. Williams and Wilkins, Baltimore, pp 1–91

Hyden H, Hartelius H (1948) Stimulation of nucleoprotein production in nerve cells by malonitrile and its effect on psychic functions in mental disorders. Acta Psychiatr Neurol [Suppl] 48: 1

Insausti R, Amaral DG, Cowman WM (1987) The entorhinal cortex of the monkey. II. Cortical afferents. J Comp Neurol 264: 365–395

Isaacson LG, Tanaka D (1986) Cholinergic and non-cholinergic projections from the canine pontomesencephalic tegmentum (Ch5 area) to the caudal intralaminar thalamic nuclei. Exp Brain Res 62: 179–188

Itil TM, Hsu W, Klingenberg H, Saletu B, Gannon P (1972) Digital-computer analyzed all-night sleep EEG patterns (sleep prints) in schizophrenics. Biol Psychiatry 4: 3–16

Jakob H, Beckmann H (1985) Clinical-neuropathological studies of developmental disorders in the limbic system in chronic schizophrenia. In: Cazzullo CL (ed) Schizophrenia: an integrative view. Libbey, London, p 81

**Jakob H, Beckmann H (1986) Prenatal developmental disturbances in the limbic allocortex in schizophrenics. J Neural Transm 65: 303–326

Jakob H, Beckmann H (1989) Gross and histological criteria for developmental disorders in brains of schizophrenics. J Roy Soc Med 82: 466–469

**Jakob H, Beckmann H (1994) Circumscribed malformation and nerve cell alterations in the entorhinal cortex of schizophrenics. J Neural Transm Gen Sect 98: 83–106

Jellinger K (1985) Neuromorphological background of pathochemical studies in major psychoses. In: Beckmann H, Riederer P (eds) Pathochemical markers in major psychoses. Springer, Berlin Heidelberg New York, pp 1–23

Jeste DV, Lohr JB (1989) Hippocampal pathologic findings in schizophrenia: a morphometric study. Arch Gen Psychiatry 46: 1019–1024

Jones BE, Beaudet A (1987) Distribution of acetylcholine and catecholamine neurons in the cat brainstem: a choline acetyltransferase and tyrosine hydroxylase immunohistochemical study. J Comp Neurol 261: 15–32

Jones EG, Powell TPS (1970) An anatomical study of converging sensory pathways within the cerebral cortex of the monkey. Brain 93: 793–820

Josephy EG (1923) Beiträge zur Histopathologie der dementia praecox. Z Ges Neurol Psychiatr 86: 391–485

Kahle W (1969) Die Entwicklung der menschlichen Großhirnhemisphäre. In: Bauer HJ, Gänshirt H, Spatz H, Vogel P (eds) Neurology series, vol 1. Springer, Berlin Heidelberg New York, pp 6–107

Kalus P, Senitz D (1996) Parvalbumin in the human anterior cingulate cortex: morphological heterogeneity of inhibitory interneurons. Brain Res 729: 45–54

Kalus P, Senitz D, Beckmann H (1997a) Cortical layer I changes in schizophrenia: a marker for impaired brain development? J Neural Transm 104: 549–559

Kalus P, Senitz D, Beckmann H (1997b) Altered distribution of parvalbumin immunoreactive local circuit neurons in the anterior cingulate cortex of schizophrenic patients. Psychiatry Res Neuroimaging 75: 49–59

Karson CN, Garcia-Rill E, Biedermann J, Mrak RE, Husain MM, Skinner RD (1991) The brain stem reticular formation in schizophrenia. Psychiatr Res 40: 31–48

Kikinis R, Shenton ME, Gerig G et al (1994) Temporal lobe sulcogyral pattern anomalies in schizophrenia: an in vivo MR three-dimensional surface rendering study. Neurosci Lett 182: 7–12

King LJ (1974) A sensory-integrative approach to schizophrenia. Am J Occup Ther 28: 529–536

Kostóvic I, Petanjek Z, Judas M (1990) The earliest areal differentiation of the human cerebral cortex: entorhinal area. Soc Neurosci Abstr 16: 351–356

Kovelman JA, Scheibel AB (1984) A neurohistological correlate of schizophrenia. Biol Psychiatry 19: 1601–1619

Krimer LS, Herman MM, Saunders RC et al (1997) A qualitative and quantitative analysis of the entorhinal cortex in schizophrenia. Cereb Cortex 7: 732–739

**Lauer M, Beckmann H (1997) The human striatum in schizophrenia. I. Increase in overall relative striatal volume in schizophrenics. Psychiatr Res Neuroimaging Sect 68: 87–98

Le Mire RJ, Laeser RD, Leech RW, Alvord ED (1975) Normal and abnormal development of the human nervous system. In: Hagerstown ML (ed) The forebrain cortex. Harper and Row, Hagerstown, pp 231–259

Lesch A, Bogerts B (1984) The diencephalon in schizophrenia: evidence for reduced thickness of the periventricular grey matter. Eur Arch Psychiatry Neurol Sci 234: 212–219

Lindenberg R, Haymaker W (1982) Tissue reactions in the grey matter of the central nervous system. In: Haymaker W, Adams RD (eds) Histology and histopathology of the nervous system. Thomas, Springfield

Lopez da Silva FH, Witter MP, Boeijinga PH, Lohman AHM (1990) Anatomic organisation and physiology of the limbic cortex. Physiol Rev 70: 453–511

Macchi G (1951) The ontogenetic development of the olfactory telencephalon in man. J Comp Neurol 95: 245–305

Markowitsch HJ (1982) Thalamic mediodorsal nucleus and memory: a critical evaluation of studies in animals and man. Neurosci Biobehav Rev 6: 351–380

Mesulam MM, Mufson EF, Levey AI, Wainer BH (1984) Atlas of cholinergic neurons in the forebrain and upper brainstem of the macaque based on monoclonal choline acetyltransferase immunohistochemistry and acetylcholinesterase histochemistry. Neuroscience 12: 669–686

Nowakowski RS (1987) Basic concepts of CNS development. Child Dev 58: 568–595

Nowakowski RS, Rakic P (1981) The site of origin and route and rate of migration of neurons to the hippocampal region of the rhesus monkey. J Comp Neurol 196: 129–154

O'Keefe I, Nadel L (1978) The hippocampus as a cognitive map. Clarendon, Oxford

Pakkenberg B (1987) Post-mortem study of chronic schizophrenic brains. Br J Psychiatry 151: 744–752

Pakkenberg B (1989) What happens in the leucotomised brain? A post-mortem morphological study of brains from schizophrenic patients. J Comp Neurosurg Psychiatry 52: 156–161

Pakkenberg B (1990) Pronounced reduction of total neuron number in mediodorsal thalamic nucleus and nucleus accumbens in schizophrenia. Arch Gen Psychiatry 47: 1023–1028

Pakkenberg B (1992) The volume of the mediodorsal thalamic nucleus in treated and untreated schizophrenics. Schizophr Res 7: 95–100

Pakkenberg B (1993) Leucotomized schizophrenics lose neurons in the mediodorsal thalamic nucleus. Neuropathol Appl Neurobiol 19: 373–380

Pakkenberg B, Gundersen HJG (1988) Total number of neurons and glia cells in human brain nuclei estimated by the disector and the fractionator. J Microsc 150: 1–20

Pandya DN, Yeterian EH (1985) Architecture and connections of cortical association areas. In: Peters A, Jones GE (eds) Cerebral cortex, vol 4. Association and auditory cortices. Plenum, New York, pp 3–61

Pennartz CM, Groenewegen HJ, Lopez da Silva FH (1994) The nucleus accumbens as a complex of functionally distinct neuronal ensembles: an integration of behavioural, electrophysiological and anatomical data. Prog Neurobiol 42: 719–761

Peters G (1937) Zur Frage der pathologischen Anatomie der Schizophrenie. Z Ges Neurol Psychiatr 160: 361–380

Peters G (1956) Dementia praecox und manisch-depressives Irresein. In: Scholz W (ed) Handbuch der speziellen pathologischen Anatomie und Histologie, vol XIII/4. Nervensystem. Springer, Berlin Göttingen Heidelberg, pp 1–57

Peters G (1967) Neuropathologie und Psychiatrie. In: Gruhle HW, Jung R, Mayer-Gross W (eds) Psychiatrie der Gegenwart, vol 1. Springer, Berlin Heidelberg New York, pp 286–324

Post RM (1975) Cocaine psychoses: a continuum model. Am J Psychiatry 132: 225–231

Post RM (1988) Time course of clinical effects of carbamazepine: implications for mechanisms of action. J Clin Psychiatry 49: 35–46

Racine RJ, Ivy GO, Milgram NW (1989) Kindling: clinical relevance and anatomical substrate. In: Bolwig TG, Trimble MR (eds) The clinical relevance of kindling. Wiley, Chichester, pp 15–34

Rajkowska G, Selemon LD, Goldman-Rakic PS (1994) Reduction in neuronal sizes in prefrontal cortex of schizophrenics and Huntington patients. Soc Neurosci Abstr 20: 620

Rakic P (1975) Cell migration and neuronal ectopias in the brain. In: Bergsma D (ed) Morphogenesis and malformations of the face and brain. Liss, New York, pp 95–129

Rakic P (1978) Neuronal migration and contact guidance in primate telencephalon. Postgrad Med J 54: 25–40

Rakic P (1988a) Defects of neuronal migration and the pathogenesis of cortical malformations. Prog Brain Res 73: 15–37

Rakic P (1988b) Specification of cerebral cortical areas. Science 241: 170–176

Rakic P, Nowakowski RS (1981) The time of origin of neurons in the hippocampal region of the rhesus monkey. J Comp Neurol 196: 99–128

Roberts GW, Colter N, Lofthouse R, Bogerts B, Zech M, Crow TJ (1986) Gliosis in schizophrenia: a survey. Biol Psychiatry 21: 1043–1050

Roberts GW, Colter N, Lofthouse R, Johnstone ED, Crow TJ (1987) Is there gliosis in schizophrenia? Investigation of the temporal lobe. Biol Psychiatry 22: 1459–1468

Rose M (1927) Der Allocortex bei Tier und Mensch, part 1. J Psychol Neurol 34: 1–111

Rosene D, van Hoesen GW (1977) Hippocampal efferents reach widespread areas of cerebral cortex and amygdala in the rhesus monkey. Science 198: 315–317

Satoh K, Fibiger HC (1985) Distribution of central cholinergic neurons in the baboon (Papio papio). J Comp Neurol 236: 197–214

Scheibel AB, Kovelman JA (1981) Disorientation of the hippocampal pyramidal cells and its processes in the schizophrenic patients. Biol Psychiatry 16: 101–102

Scholz W (1957a) Für die allgemeine Histopathologie degenerativer Prozesse bedeutsame morphologische, histochemische und strukturphysiologische Daten. In: Scholz W (ed) Nervensystem. Handbuch der speziellen pathologischen Anatomie und Histologie, vol XIII/1A. Springer, Berlin Göttingen New York, pp 212–225

Scholz W (1957b) An nervöse Systeme gebundene (topistische) Kreislaufschäden. Die Ammonshornsklerose. In: Scholz W (ed) Nervensystem. Handbuch der speziellen pathologischen Anatomie und Histologie, vol XIII/1B. Springer, Berlin Göttingen New York, pp 1364–1373

Segal M, Landis S (1974) Afferents to the hippocampus of the rat studied with the method of retrograde transport of horseradish peroxidase. Brain Res 78: 1–15

Selemon LD, Goldman-Rakic PS (1985) Longitudinal topography and interdigitation of corticostriatal projections in the rhesus monkey. J Neuroscience 5: 776–794

Selemon LD, Rajkowska G, Goldman-Rakic PS (1993) A morphometric analysis of prefrontal areas 9 and 46 in the schizophrenic and normal human brain. Schizophr Res 9: 151

Selemon LD, Rajkowska G, Goldman-Rakic PS (1995) Abnormally high neuronal density in the schizophrenic cortex: a morphometric analysis of prefrontal area 9 and occipital area 17. Arch Gen Psychiatry 52: 805–818

Senitz D, Kalus P. Differentiated criteria of cytoarchitectural disturbances of the rostral entorhinal cortex in schizophrenis. J Neural Transm (in press)

**Senitz D, Winkelmann E (1981) Morphologische Befunde in der orbitofrontalen Rinde bei Menschen mit schizophrenen Psychosen. Eine Golgi- und elektronenoptische Studie. Psychiatr Neurol Med Psychol (Leipz) 33: 1–9

Senitz D, Winkelmann E (1991) Neuronale Struktur-Anomalität im orbito-frontalen Cortex bei Schizophrenien. J Hirnforsch 32: 149–158

Senitz D, Winkelmann E, Brauer K (1979) Anwendung von Silbermethoden in der pathologischen Anatomie am Beispiel der Schizophrenie. Zbl Allge Pathol Path Anat 123: 128

Sidman RL, Rakic P (1973) Neuronal migration with special reference to developing human brain: a review. Brain Res 62: 1–35

Sorensen KE, Witter MP (1983) Entorhinal efferents reach the caudato-putamen. Neurosci. Lett 35: 259–264

Spielmeyer W (1930) Die anatomische Krankheitsforschung in der Psychiatrie. In: Bumke O (ed) Handbuch der Psychiatrie, vol XI/7. Springer, Berlin, pp 1–41

Stephan H (1975) Allocortex. Regio entorhinalis. In: Bargmann W (ed) Handbuch der mikroskopischen Anatomie des Menschen, vol IV/9. Springer, Berlin Heidelberg New York, pp 642–715

Stevens JR (1982) Neuropathology of schizophrenia. Arch Gen Psychiatry 39: 1131–1139

Stevens JR, Casanova MF (1988) Is there a neuropathology of schizophrenia? Biol Psychiatry 121: 259–264

Supprian T, Senitz D, Beckmann H (1993) Presentation of human neocortical neurons stained with the carbocyanine dye DiI compared to the Golgi silver impregnation technique. J Hirnforsch 34: 403–406

Swanson LW (1983) The hippocampus and the concept of the limbic system. In: Seifert W (ed) Neurobiology of the hippocampus. Academic, London, pp 1–19

Swanson LW, Wyss JM, Cowan WM (1978) An autoradiographic study of the organization of intrahippocampal association pathways in the rat. J Comp Neurol 181: 681–716

Talamini LM, Louwerens JW, Slooff CJ, Korf J (1995) PET versus postmortem studies in schizophrenia research: significance for the pathogenesis and pharmacotherapy. In: den Boer JA, Westenberg HGM, van Praag HM (eds) Advances in the neurobiology of schizophrenia. Wiley, Chichester, pp 158–187

Tandon R, Greden JF (1989) Cholinergic hyperactivity and negative schizophrenic symptoms. Arch Gen Psychiatry 46: 745–753

Turner BH, Mishkin M, Knapp M (1980) Organization of the amygdalopetal projections from modality-specific cortical association areas in the monkey. J Comp Neurol 191: 515–544

van Hoesen GW (1982) The parahippocampal gyrus: its cortical connections in the monkey. Trends Neurosci 5: 345–350

Vogt C, Vogt O (1952) Alterations anatomiques de la schizophrénie et d'autres psychoses dites fonctionelles. In: Rosenberg H, Sellier V (eds) First International Congress of Neuropathology, vol 1. Turin, pp 515–532

Watson SJ, Meador-Woodruff JH (1995) Neocortical abnormalities in schizophrenia. Arch Gen Psychiatry 52: 819–820

Witter MP, Room P, Groenewegen HJ, Lohmann AHM (1986) Connection of the parahippocampal cortex in the cat. V. Intrinsic connections; comments on input/output connections with the hippocampus. J Comp Neurol 252: 78–94

Woolf NJ, Butcher LL (1986) Cholinergic systems in the rat brain. III. Projections from the pontomesencephalic tegmentum to the thalamus, tectum, basal ganglia, and basal forebrain. Brain Res Bull 16: 603–637

Zarcone V, Azumi K, Dement W, Gulevich G, Kraemer H, Pivik T (1975) REM phase deprivation and schizophrenia. Arch Gen Psychiatry 32: 1431–1436

CHAPTER

7

C.A. Tamminga, A.C. Lahti, D.R. Medoff, H.H. Holcomb

The Functional Involvement of the Anterior Cingulate Cortex in Schizophrenic Psychosis

1 Introduction: Schizophrenia

Schizophrenia is a serious, lifelong psychotic illness afflicting up to 1% of the world's population. Because the illness begins in late adolescence and continues unabated throughout life, affected years are many and medical need exceeds the mere disease incidence. It is an illness whose pathophysiology and etiology remain unknown. While its study is a challenge, it is an exercise with few critical signposts. What makes the serious biologic study of schizophrenia possible at all is the accelerated rate of discovery within basic neuroscience, providing details of mammalian brain anatomy, neurochemistry, and function to inform schizophrenia hypotheses. Studies of the schizophrenic brain have not yet uncovered the disease mechanisms. Clues from pathology have been veiled. The biology we know to look for and can measure is normal. Thus schizophrenia research is even more dependent on new neuroscience information to advance its disease concepts than many other brain diseases. In addition, the illness is a human one, without an animal model or useful surrogate preparation. Thus clinical studies in schizophrenia are called upon to contribute the critical observations and strategies for effective disease study, while basic neuroscience must supply the concepts and explanatory mechanisms underlying the hypotheses. With a promising formulation in hand, given the long history of previous attempts to study pathophysiology in this area, clinical research systems are primed for hypothesis testing with a view to identifying a vital mechanism.

Already, characteristic signs and symptoms of the illness have been identified by many astute clinicians (Sartorius et al. 1974). The data from a 1974 WHO study of schizophrenia asked the question of whether schizophrenia looks the same around the world and answered it in the affirmative (Sartorius et al. 1974). The descriptive list of common symptoms generated during this WHO study illustrates the groups of symptoms common to schizophrenia in order of their prevalence (Table 1). The most frequent symptom, lack of insight, is in many ways the most crippling aspect of the illness, prohibiting affected individuals from distinguishing their disordered mental symptoms from reality-based thinking.

Since the early descriptive studies, several large clinical data bases have been queried to discover whether schizophrenic subgroups exist, sorted by clinical characteristics (for a review, see Carpenter and Buchanan 1994; Andreasen 1995; Barnes and Liddle 1990; Liddle 1987). These studies have uniformly reported that certain symptoms of schizophrenia appear to run together. Positive symptoms (hallucinations, delusions, thought disorder) cluster together, as do negative symptoms (poverty of thought, asociality, alogia); both cognitive symptoms and depression form a symptom background for the variable positive and negative symptom expressions of the illness. Whether these analyses indicate that schizophrenia is a disease with multiple mechanisms and multiple etiologies (like, for example, anemia) or is a single illness with multiple manifestations (like, for example, diabetes) remains an unanswered question around which there has been much discussion and controversy (Carpenter et al. 1993; Carpenter and Buchanan 1989).

Table 1. Frequency of psychotic symptoms in schizophrenia (international pilot study)

Symptom	Frequency (%)
Lack of insight	97
Auditory hallucinations	74
Verbal hallucinations	70
Ideas of reference	70
Suspiciousness	65
Flatness of affect	65
Voices speaking	65
Paranoid state	64
Thought alienation	52
Thoughts spoken aloud	50

Adapted from Sartorius et al. (1974).

2 Localization of Schizophrenia

Brain scientists all presume that the mechanism of schizophrenia resides within the brain. The fundamental subsequent question is where. Various answers have been advanced, including the medial temporal cortex (Benes 1993; Heckers et al. 1998; Tamminga 1997; Weinberger et al. 1992), the middle frontal cortex (Andreasen et al. 1996; Gur et al. 1987; Carpenter and Buchanan 1989; Liddle and Morris 1991), or the anterior cingulate cortex (reviewed here). Methodologic challenges exist in applying techniques such as human brain imaging to this clinical question in schizophrenia (including the questions of subgrouping and of chronic medication effects); therefore, we began with an initial regional survey of schizophrenia using functional brain imaging in our own laboratory. Thereafter, we followed up these initial findings in subsequent clinical imaging and postmortem tissue experiments.

An initial "screening" study addressed a localization question in schizophrenia; it compared 12 drug-free schizophrenic volunteers with 12 healthy ones using

positron emission tomography (PET) and fluorodeoxyglucose (FDG). This was a crude, but initial screening study where a single question was addressed: where does regional neuronal activity differ between matched patient and healthy volunteers? The 12 schizophrenic volunteers were relatively young and not institutionalized; all were drug withdrawn and demonstrated positive psychotic symptoms. The schizophrenic volunteers were matched to 12 healthy volunteers with no immediate family history of schizophrenia. Each was scanned with PET/FDG in an at-rest condition (Tamminga et al. 1992). From these data, the answer to the study question emerged: schizophrenic volunteers differ from healthy volunteers in their limbic cortex metabolism. Specifically, the schizophrenic volunteers differed in having reduced neuronal activity in the anterior cingulate cortex and in the parahippocampal gyrus/hippocampus (Table 2). No other central nervous system (CNS) areas demonstrated significant differences between these two groups. Because these two brain regions are functionally connected in the mammalian CNS, the results were even more plausible. All patient volunteers in this study had clear positive symptoms. In a later functional imaging study with a new patient group, a significant correlation emerged among the patient volunteers between positive psychotic symptoms – Brief Psychiatric Rating Scale (BPRS) psychosis scores – and neuronal activity in the limbic cortex anterior cingulate and hippocampus (rCMRglu) (Fig. 1; Tamminga 1997). These data taken together suggest that the limbic system is involved in the generation or the mediation of positive psychotic symptoms in schizophrenia.

The crude localization study described above included patient volunteers who did ($n = 4$) and did not ($n = 7$) have persistent negative symptoms. These two patient groups were compared to see where neuronal activity differed in negative-symptom schizophrenia. The volunteers with negative symptoms in the first study demonstrated a profound decrease in cortical neuronal activity in the middle frontal and the inferior parietal cortices and in the thalamus. The schizophrenics with only positive symptoms had normal levels of neuronal

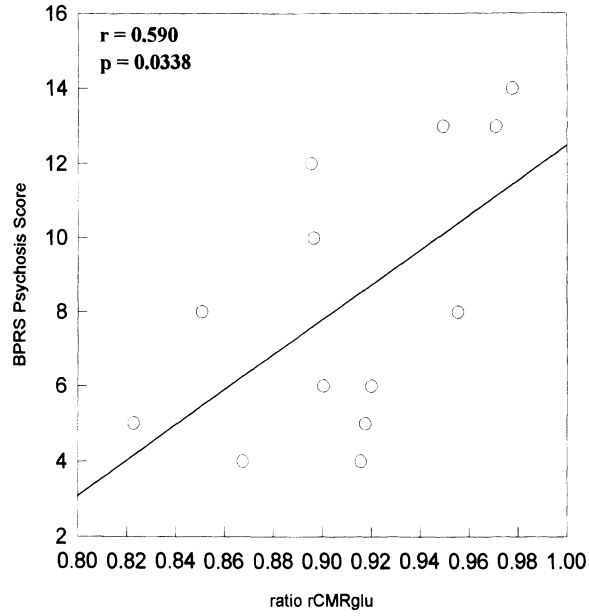


Fig. 1. Correlation between regional metabolism in the limbic cortex (hippocampal and cingulate) and the level of psychosis in 13 schizophrenic volunteers free from antipsychotic drugs. The correlation when drug free is $r = 0.59$, whereas no correlation was obtained in any brain region when the volunteers were treated with drugs and rescanned. *rCMRglu*, neuronal activity in the limbic cortex, anterior cingulate and hippocampus

Table 2. Regional levels of glucose metabolism in schizophrenic and healthy volunteers measured using positron emission tomography (PET) with fluorodeoxyglucose (mg glucose/100 g tissue per min)

Brain region	Schizophrenic		Normal control		*p* value
	Mean	SD	Mean	SD	
Frontal cortex	10.4	2.1	10.9	2.0[a]	.57
Parietal cortex	9.3	2.5	10.6	2.6	.23
Temporal cortex	8.4	1.6	9.6	2.0	.89
Caudate	10.4	2.2	11.7	3.6	.30
Putamen	10.4	2.4	12.3	4.0	.17
Thalamus	10.1	1.9	11.2	3.3	.33
Anterior cingulate	9.4	2.0[b]	12.5	3.3	.0035
Hippocampus	6.9	1.5[b]	9.4	2.0	.01

[a]Metabolic rate of glucose: mg/100 g tissue per min.
[b]Significant at the $p < 0.05$ level.

activity in these "negative-symptom" brain regions. Both groups (with and without negative symptoms) showed functional limbic cortical dysfunction, however. A subsequent study from this laboratory using a prospective evaluation of regional cerebral blood flow (rCBF) (Lahti et al. 1998) in primary negative-symptom schizophrenic volunteers has confirmed this localization of persistent negative symptoms to the middle frontal and inferior parietal cortices bilaterally. Whether schizophrenia is a single pathology expressed in multiple CNS regions, associated with characteristic clinical syndromes, or multiple primary diseases in these regions is a question which is still unanswered by these data.

To pursue the line of research directed by these localization data, additional experiments were indicated. The simplistic working hypothesis at this time associated cingulate and hippocampal dysfunction with the positive symptoms of schizophrenia and frontal parietal dysfunction with negative symptoms. A study of the activity of the limbic cortex during psychosis exacerbation, during psychosis inhibition, and during cognitive task performance could further test this hypothesis. If psychosis is mediated by or generated within the limbic cortex, the anterior cingulate cortex dysfunction could be associated not only with increased psychosis, but also with behavioral dysfunction and alterations in cognition. Experimentally increasing psychosis (with ketamine), decreasing psychosis (with haloperidol), or provocation with a cognitive task would contribute to localizing schizophrenia pathophysiology across different functional states.

3 Psychosis

Ketamine is a commonly used pediatric anesthetic. Pharmacologically, the anesthetic is a mild, noncompetitive *N*-methyl-D-aspartate (NMDA) receptor antagonist which acts as an agonist at the phencyclidine (PCP) receptor within the NMDA ionophore. It is a chemical congener of PCP with lower affinity and potency (White et al. 1982). Ketamine retains the pharmacologic and behavioral properties of PCP, albeit with mild potency. Hence, based on known PCP pharmacology, mild psychotomimetic activity was predicted in patients and healthy volunteers with ketamine administration (Luby et al. 1959; Krystal et al. 1994). To evaluate this, the behavioral activity of a subanesthetic dose range of ketamine (0.1–0.5 mg/kg IV) was tested in these volunteer groups. A mild (<50% of baseline) and short-lived (25–45 min) symptomatic psychosis exacerbation occurred with the ketamine in the patient volunteers (Lahti et al. 1995). This response was distinctive in several respects. First, ketamine's action was not blocked by the traditional antipsychotic haloperidol, and not even blunted. Second, and perhaps more importantly, ketamine affected the patient's own symptoms, be they hallucinations, delusions, or thought disorder; it did not merely produce its own characteristic psychotomimetic effects, but mildly exacerbated the existent symptoms. This observation suggests that glutamatergic inhibition at the NMDA-sensitive glutamate receptor can provoke schizophrenic symptoms themselves, not just a general psychosis. These observations implicate glutamatergic transmission at some neuronal level in the mechanisms of schizophrenia, possibly close to the pathology of the illness. These clinical observations with ketamine fit into an extensive body of research which has already implicated glutamate in schizophrenia (Tamminga 1998).

The behavioral response to ketamine in schizophrenic volunteers suggested the importance of localizing this drug action and its behavioral correlates in the CNS, to identify brain regions involved with this drug response and with psychosis. Again, rCBF studies were carried out with both healthy and schizophrenic volunteers at regular time intervals over 60 min after ketamine administration. Ketamine, in a dose-sensitive fashion, produced a significant increase in rCBF in the anterior cingulate and contiguous medial frontal cortex as well as in a smaller right inferior frontal region. The largest rCBF decrease appeared in the cerebellum and essentially involved the entire cerebellar cortex (Fig. 2, Table 3). These rCBF changes induced by ketamine were not significantly different between the two subject groups tested (schizophrenia vs. healthy), even though the schizophrenic group had numerically higher activation levels. There were three brain regions where the behavioral response (total BPRS score) correlated with the drug-induced rCBF changes: (1) the anterior cingulate cortex, (2) the right inferior frontal cortex, and (3) an area in the brain stem, near the substantia nigra (Fig. 3). These correlational data suggest that these three brain regions may be involved in mediating this ketamine-induced behavioral response.

It is noteworthy that the brain region whose rCBF is prominently stimulated by ketamine, the anterior cingulate cortex, is the very region identified in our first PET/FDG psychosis "screening" study as being dysfunctional in schizophrenia. Glucose utilization (as a marker of neuronal activity) is lower in schizophrenic individuals at rest relative to normal subjects in the initial screening study, whereas with ketamine, psychosis is positively associated with anterior cingulate metabolism and rCBF activation. This apparent contradiction may be related to cortical behavior during the resting, unstimulated cognitive state of the initial scan.

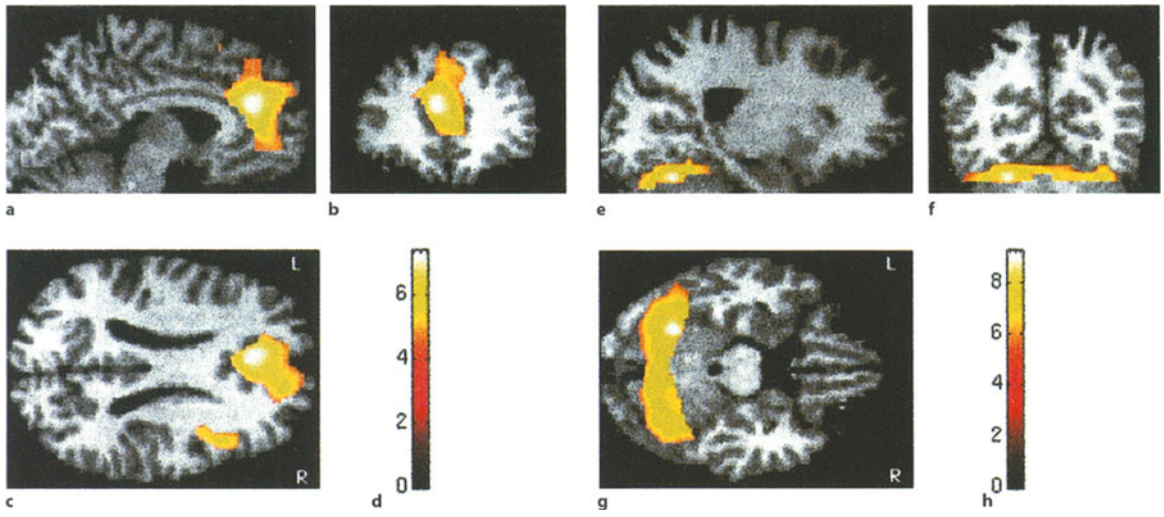


Fig. 2a–h. Statistical parametric mapping (SPM) analyses of ketamine-induced changes in regional cerebral blood flow (*rCBF*) in human brain ($n = 17$) at a dose of 0.3 mg/kg IV. **a–d** Areas of significant activation; these areas include the anterior cingulate cortex and the right inferior frontal cortex. **e–h** Area of significant inhibition; this is the cerebellum throughout its entire extent. **a,e** Saggital view. **b,f** Coronal view. **c,g** Transverse view. **d,h** *Z* values. See also Table 3

Table 3. Results of the statistical parametric mapping analysis of the images in Fig. 2

Size {*k*}	$P(N^{max} > k)$	*Z*	$P(Z^{max} > u)$	{*x,y,z* mm}		
1899[a]	0.000	7.37	0.000	−8	32	20
163[a]	0.000	6.20	0.000	38	14	16
1918[b]	0.000	9.25	0.000	−22	−64	−20

[a]Stimulation. Threshold, 4.50; volume (S), 45645 voxels; df, 143, FWHM, (22.8 22.4 19.0) mm (i.e. 75 RESELS). See also Fig. 2a–d.
[b]Suppression. Threshold, 5.50; volume (S), 45645 voxels; df, 143, FWHM, (22.8 22.4 19.0) mm (i.e. 75 RESELS). See also Fig. 2e–h.
k is the blob size corresponding to the identified *x,y,z* coordinates; *p* values are given both for blob size and for blob *Z* score.

4 Psychosis Reduction

The next study evaluated the localization of action of the traditional antipsychotic drug haloperidol to discover which brain regions are involved in antipsychotic action. Haloperidol was once the most commonly used antipsychotic drug (Klein and Davis 1969) and has its major action at the D_2 family of dopamine receptors in striatum, especially at the clinically effective dose levels used in schizophrenia. Moreover, haloperidol is a potent drug with no active metabolite, making pharmacokinetic and pharmacodynamic studies more informative. The question addressed by this experiment is which brain areas show altered neuronal activity when a psychotic person is treated with a standard and effective antipsychotic drug, suggesting which brain areas are involved in mediating that effect.

The action of haloperidol to displace dopaminergic ligands with an antagonist action at the striatal D_2 family dopamine receptor is well known (Wong et al. 1986). Animal model experiments suggest that potent pharmacologic actions at dopamine receptors in striatum have distant functional and neurochemical effects which are mediated through the well-known frontal cortical–subcortical segregated neuronal circuits (Alexander et al. 1986); these antipsychotic-induced functional and neurochemical changes are apparent throughout all regions of basal ganglia, thalamus, and cortex (Shirakawa and Tamminga 1994; Abercrombie and DeBoer 1997). The prediction in this study was that haloperidol would have both direct actions in striatum and indirect (projected) functional actions throughout basal ganglia, thalamus, and limbic and neocortical regions. This prediction is consistent with the distribution of the parallel, segregated cortical–subcortical prefrontal circuits already well described in primate brain (Alexander et al. 1986; Nauta 1989).

Volunteers with schizophrenia agreed to a clinical study where they received 0.3 mg haloperidol/kg per day for 4 weeks and then underwent a PET scan with

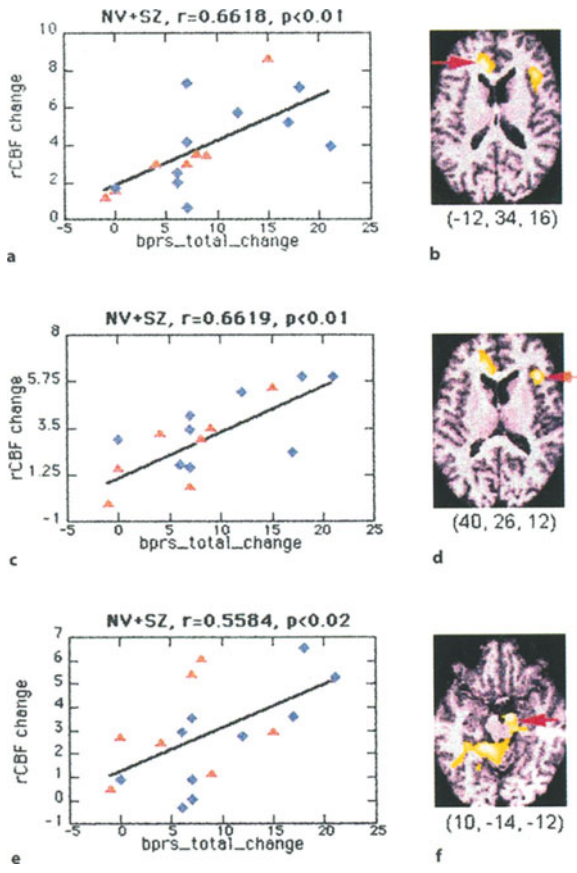


Fig. 3a–f. The three brain regions where the extent of the ketamine-induced regional cerebral blood flow (*rCBF*) change correlated with the extent of the psychotic symptoms induced. **a,b** Anterior cingulate cortex. **c,d** Right inferior frontal cortex. **e,f** A small region in the midbrain which could be the substantia nigra. The schizophrenic (*triangles, SZ*) and normal volunteers (*diamonds, NV*) are intermixed in these plots, suggesting that the process may be similar in both groups. The numbers in parentheses beneath the right-hand figures (**b,d,f**) indicate the Talairach atlas coordinates (*x, y, z*)

FDG. Next, the volunteers received a matched placebo regimen for the same time period and had a second PET/FDG scan. In an intra-subject analysis, the regional differences in activation between the two scans by the same individual were analyzed. Haloperidol, even with chronic administration, increases glucose metabolism (rCMRglu) in the caudate and putamen (Holcomb et al. 1996). This finding had been previously reported (Gur et al. 1987; Buchsbaum et al. 1982) and is readily explained on the basis of haloperidol's action in striatum to block dopamine, an inhibitory transmitter, hence producing activation by disinhibition. What was more surprising, and also a new finding with this study, was the haloperidol-induced activation in thalamus, greater in the anterior than posterior portion. One surprise in this response derives from knowing that the density of dopamine receptors is very low in human thalamus, suggesting an indirect action in that region. In cortical areas, only two delimited cortical regions demonstrated significant change with haloperidol: the frontal cortex and the anterior cingulate region. Moreover, this change was a decrease in rCMRglu, not an increase as seen in the basal ganglia and thalamus, again suggesting the possibility of an indirect action (Fig. 4). This distinctive pattern of overall brain activation changes could not easily be explained by a single common drug action of haloperidol in multiple brain regions. The most parsimonious explanation appeared to involve a single effect of haloperidol exerted in striatum with this effect transmitted to additional brain regions through the segregated cortical–subcortical neuronal pathways. The metabolic changes predicted by blocking striatal dopamine receptors using this hypothesis closely matched the data actually collected in this experiment. Thus these haloperidol-induced actions in the CNS can most cleanly be explained by a single, direct, and potent drug action in the striatum, with the functional effects in other regions being transmitted in a secondary and tertiary fashion through the basal ganglia to thalamus and on to the neocortex and limbic cortex. Thalamic activation and cortical deactivation with dopamine blockade in striatum has been previously described in animals (Abercrombie and DeBoer 1997).

Again, it is striking in this experiment that the anterior cingulate cortex is one of the few cortical regions functionally affected by chronic haloperidol treatment. These results strengthened the ongoing working hypothesis that this rCBRglu change in the anterior cingulate is associated with the drug-induced reduction in psychotic and cognitive symptoms. Additional haloperidol pharmacodynamic imaging studies in our laboratory have more fully demonstrated the action of haloperidol in stimulating rCBF in the caudate and thalamus and reducing rCBF in frontal and cingulate cortices data consistent with the initial rCMRglu findings (Fig. 5; Lahti et al. 1998). An association between the haloperidol-induced psychosis change and the rCBF alterations, also drug induced, in the anterior cingulate cortex is hypothesized and being further tested.

5 Cognitive Task Activation

Task-associated activation patterns provide information about cerebral strategies employed to carry out a defined mental task. Comparing task-related regional activation patterns between two subject groups is legitimate if performance characteristics of the task are

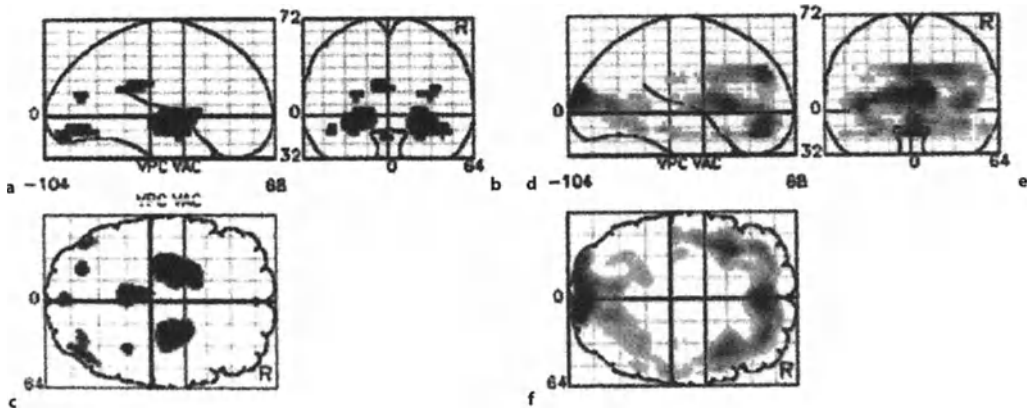


Fig. 4a–f. SPM analyses of haloperidol action in the human brain ($n = 12$). **a–c** Haloperiodol minus placebo. Activation of regional metabolism in the caudate/putamen (bilaterally) and in the thalamus (midline). **d–f** Placebo minus haloperidol. Regional metabolism significantly reduced in the middle frontal cortex and in the anterior cingulate region. The prominent reduction in the occipital region is not significant. **a,d** Sagittal. **b,e** Coronal. **c,f** Transverse. *VPC*, vertical posterior commissure; *VAC*, ventrical anterior commissure

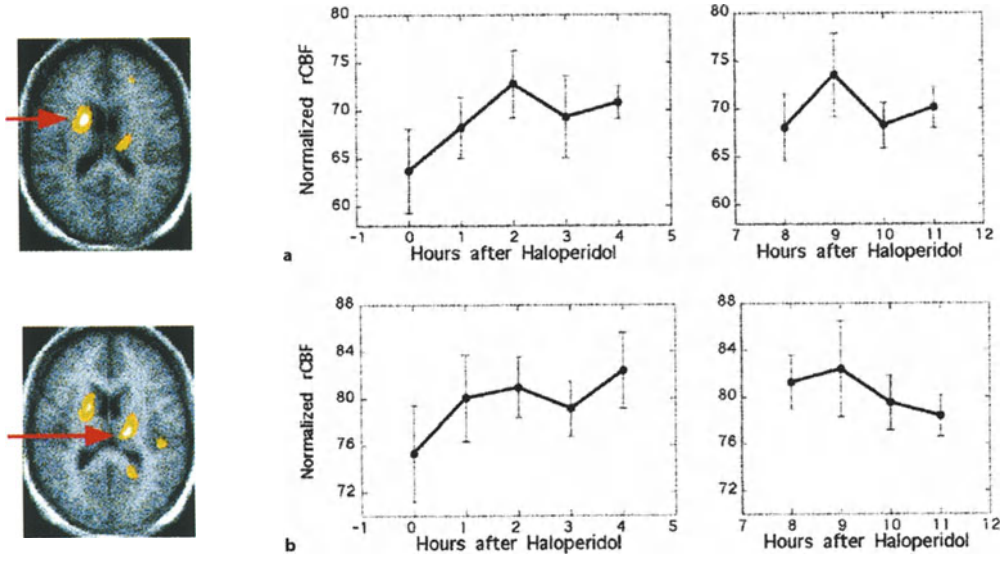


Fig. 5a,b. Pharmacodynamic characterization of haloperidol-induced regional cerebral blood flow (*rCBF*) change over time in **a** the caudate (−18, 4, 16) and **b** the thalamus (16, −18, 12) (numbers in parentheses indicate the Talairach atlas coordinates). The same patient group was analyzed in the initial set of scans (0–4 h) and in the late scan set (8–11 h), with the scan sets separated by 1 week for feasibility reasons

similar across individuals. This is difficult when studying schizophrenia because general task performance of volunteers with the illness is often diminished. In this experiment, however, schizophrenic and healthy volunteers were selected and trained to perform an auditory recognition task with performance fixed at 80% task accuracy by means of varying slightly the auditory stimuli. rCBF with PET and ^{15}O-water was carried out using this task condition under rest, sensorimotor control, and active task conditions. Usual analysis with SPM 96 (statistical parametric mapping) and hierarchical subtraction was performed to arrive at distinct activations in the control and in the task state by group (Holcomb et al. 1998). During the sensorimotor control task, individuals demonstrated activations in their primary auditory cortex and in the left postcentral gyrus (motor cortex). In the auditory task condition, with the control activation areas subtracted away, activations in the healthy volunteers were apparent in the right middle and inferior frontal cortex and in the anterior cingulate cortex, areas consistent with the

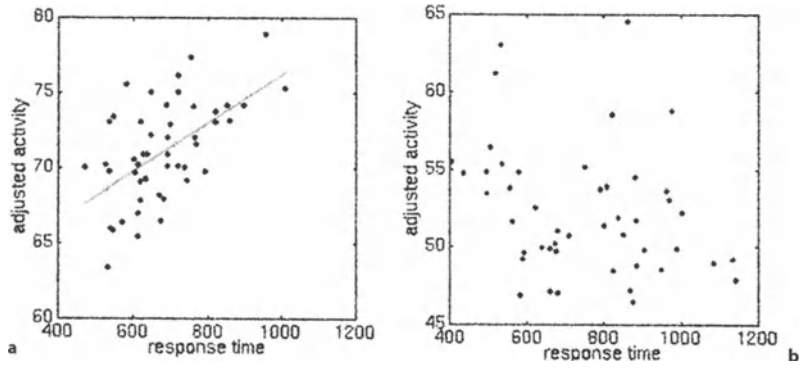


Fig. 6a,b. Correlation between a measure of task difficulty (reaction time) on the auditory discrimination task and regional cerebral blood flow (rCBF) in the anterior cingulate cortex. **a** Healthy volunteers. A positive correlation is apparent, suggesting that the harder the task becomes, the greater the rCBF activation in the anterior cingulate is. **b** The schizophrenic group. No correlation was obtained among these individuals, despite the fact that they were performing the same task at the same level

demands of the recognition task. When activation patterns of the schizophrenic volunteers were compared to patterns described in the healthy group, only one area showed an overall altered activation with task performance in schizophrenia: the anterior cingulate cortex (activation was reduced in schizophrenia) (Holcomb et al. 1999). Moreover, whereas there was a significant positive relationship between rCBF and reaction time (a measure of difficulty) in the anterior cingulate cortex in the healthy volunteer group, no such relationship was present or even suggested in the patient group (Fig. 6). This was the case even though the schizophrenic volunteers were successfully performing the task.

An additional correlational abnormality in schizophrenia was noted in the right frontal cortex, an area where, in an additional analysis, there were no activation differences between groups. Here, the healthy subject group showed a significant positive relationship between rCBF and reaction time (a measure of task difficulty), consistent with the rCBF behavior of the anterior cingulate cortex. However, the schizophrenic subject group showed a significant negative correlation in the frontal cortex between rCBF and task difficulty, a correlation opposite to that found in the normal subjects. This suggests that the two volunteer groups, who were both performing the task equivalently, each employed their own distinctive mental strategy for carrying out the work. The schizophrenic group, although able to perform the task well, was not able to engage their anterior cingulate or right middle frontal cortex normally or to vary the behavior performance of these areas with the difficulty of the task.

Of note again is the abnormal performance of the anterior cingulate cortex in this cognition task condition, associated with schizophrenia. In early studies, we noted that the abnormal rCBF response to task difficulty in the anterior cingulate cortex tends to normalize with antipsychotic treatment (Holcomb et al. 1999).

6 Conclusion

These functional imaging studies lead to the conclusion that the anterior cingulate cortex may be involved in schizophrenia, especially in the mediation of positive symptoms. The region is functionally involved in psychosis modulation and works abnormally during cognitively demanding tasks in schizophrenia. Not only the anterior cingulate, but also related cortical areas, such as the hippocampus, the middle frontal cortex, and the insular cortex also function abnormally in volunteers with schizophrenia.

Our current working hypothesis, based on these imaging data as well as results from animal studies with phencyclidine and MK801 (Gao and Tamminga 1995; Gao et al. 1998), and on human postmortem tissue analysis (Gao et al. 1999), suggests the possibility that the hippocampus may be a site of primary dysfunction in schizophrenia (or in any psychosis), with reduced hippocampal efferent signals aversely affecting each of its projection targets, possibly most prominently the anterior cingulate cortex. The anterior cingulate dysfunction can be associated with both positive and cognitive symptoms of the illness. This hypothesis involves areas of schizophrenic brain where prominent tissue pathology has already been reported (for a review, see Tamminga 1997) and abnormal functional imaging signals have been noted (Heckers et al. 1998). This working hypothesis provides a parsimonious and testable hypothesis for the positive symptoms of psychosis, i.e. a failure of hippocampal efferent activity,

whether induced by schizophrenia, a hallucinogen, or another psychotic illness such as bipolar disease. Such an explanation, no matter what the specific tissue pathology, would have the ability to explain several psychotic symptoms, particularly those mediated by a hippocampal target area, e.g. attention, as mediated by the anterior cingulate cortex, or short-term memory, as mediated by the middle frontal cortex.

The idea that overall limbic cortical function is impaired in schizophrenia, with the initiating defect located within hippocampus, is supported by our current data in schizophrenic postmortem tissue studies. Here we have found a normal binding of ^{3}H-glutamate to the NMDA-sensitive glutamate receptors; however, the expression of the NR_1 subunit of that receptor is reduced, and the NR_{2B} subunit expression is increased. This abnormality could imply reduced functional activity of the hippocampal NMDA receptor and possibly a reduced efferent glutamatergic signal, assuming that the reduced NR_1 expression results in more NMDA receptors with no critical NR_1 subunit. The testing of this straightforward hypothesis requires paired (i.e. from the same individual brain) hippocampal, anterior cingulate, and anterior thalamus tissue samples comparing healthy and schizophrenic tissue. These tissue samples should be examined both histologically and neurochemically for evidence consistent with or contradictory to this formulation. Because these regional defects can show variability by location, a complete examination (anterior to posterior) of the structures would be important.

The successful marriage of animal preparation experiments, clinical probe studies, in vivo human brain imaging, and postmortem tissue techniques in pursuing hypothesis-testing studies in schizophrenia may be important in arriving at clear answers to questions of etiology and mechanism. Such approaches hold promise for broadly informative research results in this area.

7 References

Abercrombie ED, DeBoer P (1997) Substantia nigra D_1 receptors and stimulation of striatal cholinergic interneurons by dopamine: a proposed circuit mechanism. J Neurosci 17: 8498–8505

**Alexander GE, DeLong MR, Strick PL (1986) Parallel organization of functionally segregated circuits linking basal ganglia and cortex. Ann Rev Neurosci 9: 357–381

Andreasen NC (1995) Symptoms, signs, and diagnosis of schizophrenia. Lancet 346(8973): 477–481

*Andreasen NC, O'Leary DS, Cizadlo T, Arndt S, Rezai K, Boles Ponto LL, Watkins GL, Hichwa RD (1996) Schizophrenia and cognitive dysmetria: a positron-emission tomography study of dysfunctional prefrontal-thalamic-cerebellar circuitry. Proc Natl Acad Sci USA 93: 9985–9990

Barnes TR, Liddle PF (1990) Evidence for the validity of negative symptoms. Mo Prob Pharmacopsychiatry 24: 43–72

Benes FM (1993). Neurobiological investigations in cingulate cortex of schizophrenic brain. Schizophr Bull 19(3): 537–549

Buchsbaum MS, Ingvar DH, Kessler R, Waters RN, Cappelletti J, van Kammen DP, King AC, Johnson JL, Manning RG, Flynn RW, Mann LS, Bunney WE Jr, Sokoloff L (1982) Cerebral glucography with positron tomography. Arch Gen Psychiatry 39: 251–259

Carpenter WT Jr, Buchanan RW (1989) Domains of psychopathology relevant to the study of etiology and treatment in schizophrenia. In: Schulz SC, Tamminga CA (eds) Schizophrenia: scientific progress. Oxford University Press, New York, pp 13–22

*Carpenter WT Jr, Buchanan RW (1994) Schizophrenia. N Engl J Med 330: 681–690

Carpenter WT Jr, Buchanan RW, Kirkpatrick B, Tamminga CA, Wood F (1993) Strong inference, theory testing, and the neuroanatomy of schizophrenia. Arch Gen Psychiatry 50(10): 825–831

Gao XM, Tamminga CA (1995) MK801 induces late regional increases in NMDA and kainate receptor binding in rat brain. J Neural Transm Gen Sect 101(1–3): 105–113

Gao XM, Hashimoto T, Tamminga CA (1998) Phencyclidine (PCP) and dizocilpine (MK801) exert time-dependent effects on the expression of immediate early genes in rat brain. Synapse 29: 14–28

*Gao XM, Sakai K, Roberts RC, Conley RR, Dean B, Tamminga CA (2000) Ionotropic glutamate receptors and NMDA subunit expression in subregions of human hippocampus: effects of schizophrenia. Am J Psychiatry (in press)

Gur RE, Resnick SM, Alavi A, Gur RC, Caroff S, Dann R, Silver FL, Saykin AJ, Chawluk JB, Kushner M, Reivich M (1987) Regional brain function in schizophrenia. Arch Gen Psychiatry 44: 119–125

*Heckers S, Rauch SL, Goff D, Savage CR, Schacter DLFAJ, Alpert NM (1998) Impaired recruitment of the hippocampus during conscious recollection in schizophrenia. Nature 1(4): 318–323

*Holcomb HH, Cascella NG, Thaker GK, Medoff DR, Dannals RF, Tamminga CA (1996) Functional sites of neuroleptic drug action in the human brain: PET/FDG studies with and without haloperidol. Am J Psychiatry 153(1): 41–49

Holcomb HH, Caudill PJ, Medoff DR, Zhao Z, Lahti AC, Ravert T, Dannals RF, Tamminga CA (1998) Cerebral blood flow relationships associated with a difficult tone recognition task in trained normal volunteers. Cerebral Cortex 8: 534–542

Holcomb HH, Lahti AC, Weiler M, Medoff DR, Tamminga CA (1999) Neuroleptic treatment of schizophrenic patients: how do haloperidol and clozapine normalize brain blood flow patterns associated with a difficult tone recognition task? In: Gattaz WF, Häfner H (eds) Search for the causes of schizophrenia, vol IV. Balance of the century. Steinkopff, Darmstadt, pp 355–365

Klein DF, Davis JM (1969) Diagnosis and drug treatment of psychiatric disorders. Williams and Wilkins, Baltimore

Krystal JH, Karper LP, Seibyl JP, Freeman GK, Delaney R, Bremner D, Heninger GR, Bowers MB, Charney DS (1994) Subanesthetic effects of the noncompetitive NMDA antagonist, ketamine, in humans: psychotomimetic, perceptual, cognitive, and neuroendocrine responses. Arch Gen Psychiatry 51: 199–214

Lahti AC, Holcomb HH, Weiler MA, Kile I, Tamminga CA (1998) Time course of rCBF changes after acute haloperidol in patients with schizophrenia. Schizophr Res 29: 173

Lahti AC, Koffel B, LaPorte D, Tamminga CA (1995) Subanesthetic doses of ketamine stimulate psychosis in schizophrenia. Neuropsychopharmacology 13(1): 9–19

Liddle PF (1987) The symptoms of chronic schizophrenia: a re-examination of the positive-negative dichotomy. Br J Psychiatry 151: 145–151

*Liddle PF, Morris DL (1991) Schizophrenic syndromes and frontal lobe performance. Br J Psychiatry 158: 340–345

Luby ED, Cohen BD, Rosenbaum G, Gottlieb JS, Kelley R (1959) Study of a new schizophrenomimetic drug: serenyl. Arch Neurol Psychiatry 71: 363–369

*Nauta WJH (1989) Reciprocal links of the corpus striatum with the cerebral cortex and limbic system: a common substrate for movement and thought? In: Mueller J (ed) Neurology and psychiatry: a meeting of minds. Karger, Basel, pp 43–63

Sartorius N, Shapiro R, Jablensky A (1974) The international pilot study of schizophrenia. Schizophr Bull 1: 21–34

Shirakawa O, Tamminga CA (1994) Basal ganglia GABAA and dopamine D1 binding site correlates of haloperidol-induced oral dyskinesias in rat. Exp Neurol 127(1): 62–69

Tamminga CA (1997) Neuropsychiatric aspects of schizophrenia. In: Yudofsky SC, Hales RE (eds) American Psychiatric Press textbook of neuropsychiatry, 3rd edn. American Psychiatric Press, Washington, DC, pp 855–882

*Tamminga CA (1998) Schizophrenia and glutamatergic transmission. Crit Rev Neurobiol 12(1–2): 21–36

*Tamminga CA, Thaker GK, Buchanan R, Kirkpatrick B, Alphs LD, Chase TN, Carpenter WT (1992) Limbic system abnormalities identified in schizophrenia using positron emission tomography with fluorodeoxyglucose and neocortical alterations with deficit syndrome. Arch Gen Psychiatry 49(7): 522–530

Weinberger DR, Berman KF, Suddath R, Torrey EF (1992) Evidence of dysfunction of a prefrontal-limbic network in schizophrenia: a magnetic resonance imaging and regional cerebral blood flow study of discordant monozygotic twins. Am J Psychiatry 149(7): 890–897

White PF, Way WL, Trevor AJ (1982) Ketamine – its pharmacology and therapeutic uses. Anesthesiology 56 119–136

Wong DF, Wagner HN Jr, Tune LE, Dannals RF, Pearlson GD, Links JM, Tamminga CA, Broussolle EP, Ravert HT, Wilson AA (1986) Positron emission tomography reveals elevated D2 dopamine receptors in drug-naive schizophrenics. Science 234(4783): 1558–1563

CHAPTER

8

T.W. Weickert, T.E. Goldberg

Neuropsychology of Schizophrenia

1 Introduction

Schizophrenia is a debilitating neuropsychiatric disorder which can result in variable behavioral symptomatology, including hallucinations, delusions, and flat affect. Additionally, schizophrenia has been characterized by impaired performance on a variety of neuropsychological measures which assess the cognitive domains of attention, higher-level executive function, memory, and motor control. While symptoms such as hallucinations and delusions may be independent of the classic cognitive domains such as memory and working memory, this is not to say that these symptoms do not have cognitive underpinnings, but rather that the cognitive underpinnings have not been well elucidated (see, e.g. our work below on thought disorder in which we attempt to put thought disorder in context; see also Frith 1996). Deficits in these cognitive domains broadly implicate the regions of the prefrontal cortex and the medial temporal lobes in the neuropsychopathology of schizophrenia. Moreover, the clinical psychopathology and cognitive domains may be relatively independent.

2 Symptoms, Motivation, and Cognition

It has been argued that the cognitive deficits observed in schizophrenia may occur as a result of the distracting influence of psychotic symptoms, such as hallucinations, distortions, avolition, and apathy. This argument would suggest that the cognitive deficits observed in schizophrenia are secondary to these symptoms and that remediation of the primary symptoms would, therefore, alleviate the cognitive abnormalities. In fact, antipsychotic medication has been demonstrated to ameliorate the cognitive deficits associated with schizophrenia to only a limited degree despite producing marked improvement in psychiatric symptoms (for reviews, see Medalia et al. 1988; Spohn and Strauss 1989). In fact, the two domains may be dissociable. For instance, Goldberg et al. (1993b) showed that, while psychiatric symptoms in schizophrenic patients improved dramatically with administration of conventional and atypical neuroleptic medication, attention, memory, and higher-level problem-solving abilities did not improve. Numerous studies have also addressed the issue of impairment in cognitive performance due to an overall lack of cooperation and motivation (Stuss et al. 1983; Schneider and Asarnow 1987; Levin et al. 1989). In general, these studies have shown that, while either instructions and/or reinforcers may improve schizophrenic patients' performance on tests of problem solving and set shifting described in further detail below, their performance still failed to normalize (Bellack et al. 1990; Goldberg et al. 1987; Summerfelt et al. 1991; Tompkins et al. 1991; Green et al. 1992).

Additionally, there has been a paucity of findings of strong correlations between neuropsychological deficits and psychotic symptoms. Faustman et al. (1988) found no relationship between cognitive performance and psychiatric symptoms in unmedicated patients. Likewise, Goldberg et al. (1993a) were unable to uncover a relationship between cognition and psychiatric symptoms in schizophrenic patients. In this study, symptom factors accounted for 15%–30% of the variance in cognitive performance of unipolar and bipolar depressed patients, whereas symptoms accounted for less than 5% of the variance in cognitive abilities in schizophrenic patients. Although Kibel et al. (1993) showed a moderate relationship between negative symptoms and cognitive deficits, negative symptoms accounted for less than half of the variance observed in the cognitive performance of schizophrenic patients. Neuropsychological testing may be a more reliable and precise method to characterize the disorder than symptoms, which are variable and subject to change with treatment.

In summary, the majority of cognitive deficits observed in schizophrenia tend to remain stable in spite of treatment with neuroleptic medication. Additionally, improvements in symptoms which follow neuroleptic treatment often fail to return patients to normal levels of cognitive performance. Thus the cognitive deficits of schizophrenia are independent of psychotic symptoms and appear to be central and enduring features of the disease.

2.1 Validity

A pertinent question arises as to whether such neurocognitive deficits are the product of the disease process or merely side effects of the neuroleptic medications administered to the schizophrenic patients. On the basis of "first break" studies, "medication-free" studies, and evidence from the preneuroleptic era, the pattern of results on neuropsychological batteries of tests administered to schizophrenic patients suggests that the neuropsychological deficits observed in schizophrenia are a product of the disease rather than an effect of treatment (for a detailed discussion, see Goldberg and Gold 1995). For example, Saykin et al. (1994) demonstrated no differences in performance on a neuropsychological battery

of tests between a group of neuroleptic naïve, first-episode schizophrenic patients and a group of previously treated, unmedicated schizophrenic patients. In this study, verbal learning and memory measures, attention, speeded visual–motor processing, and visual search tasks were impaired in both neuroleptic naïve and previously treated groups, and these groups also differed from a group of normal controls. In "medication-free" or dosage-reduction studies, in which patients have been withdrawn from neuroleptics and placed on placebo, usually in a double-blind design, investigators have demonstrated no effect of neuroleptic medication relative to the placebo condition on tests of higher cognitive functioning (Seidman et al. 1993; Cleghorn et al. 1990). However, in the Continuous Performance Test (CPT), which requires both vigilance and response readiness, improvements in performance following neuroleptic treatment in comparison to placebo have been frequently observed (see, e.g. Goldberg and Weinberger 1996; Oltmanns et al. 1978). Finally, Rapaport et al. (1945/1946) reported deficits of judgement, concentration, planning, anticipation, memory, and concept formation all prior to the era in which neuroleptic medications were introduced.

2.2 Frequency

In general, and perhaps unexpectedly, the distribution of specific neuropsychological test scores in the schizophrenic population appears to represent a negative "shift" in performance for the schizophrenic population as a whole relative to normal. While studies in which binary cutoffs were used to examine the cognitive deficits displayed in schizophrenia generally demonstrate that less than 40% of patients can be considered to be "abnormal" (Braff et al. 1991; Goldberg et al. 1988) and that patients generally fall within one to two standard deviations below the mean of the normal population on the majority of neuropsychological variables investigated (Sullivan et al. 1994), studies of monozygotic twins discordant for schizophrenia yield a different picture. In these studies, the affected twin routinely performed at a level that was lower in comparison to their unaffected co-twin irrespective of absolute level (Goldberg et al. 1993c, 1994). The twin findings therefore strongly suggest that nearly every patient would display cognitive impairment relative to the performance of their ideal control (in this case, a co-twin that would be matched for age, sex, genetic compliment, socioeconomic status, and education level, yet discordant for the presence of disease).

2.3 Predictive Validity

Delineation of the neuropsychological deficits associated with schizophrenia may also provide a better understanding of the impairment exhibited by schizophrenic patients with respect to their ability to perform routine activities of daily living. Deficits in the cognitive domains of attention, memory, and executive function would suggest severe functional impairment. Impairment in various cognitive domains displayed by psychiatric patients and patients with circumscribed brain lesions result in impairment of self-care, independent living, academic achievement, and vocational function (Heaton et al. 1978; Heaton and Pendleton 1981; Newnan et al. 1978). In a review of the literature pertaining to predictive and correlative relationships between various neuropsychological measures and functional outcome in schizophrenic patients, Green (1996) found relationships between verbal memory measures and all forms of functional outcome, vigilance and social problem solving and skill acquisition, the Wisconsin Card Sorting Test (WCST) and community functioning, and between negative symptoms and social problem solving. Similarly, Addington and Addington (1993) obtained significant correlations between neuropsychological measures (verbal reasoning and concept formation) and outcome. In a study that investigated neuropsychological differences between long-term hospitalized schizophrenic patients and schizophrenic patients successfully residing in the community for more than 3 years, Perlick et al. (1992) demonstrated that neuropsychological measures of motor coordination, perseveration, memory, and attention successfully discriminated between the two groups. Goldberg et al. (1993c) obtained significant correlations between Global Assessment Scale (GAS) scores (a scale used to measure the level of social and vocational functioning) and paired associates learning, memory for stories, verbal fluency, and the three-disk version of the Tower of Hanoi (a problem-solving task). In a study of monozygotic twins that were both discordant and concordant for schizophrenia, Goldberg and colleagues (1995) determined that cognitive measures were significant predictors of intrapair differences in GAS scores. Memory quotient (MQ) from the Wechsler Memory Scale–Revised (WMS-R), the number of categories obtained on the WCST, the easy (A) version of the trail-making task (a test of speed and scanning), and intelligence quotient (IQ) as measured by the Wechsler Adult Intelligence Scale–Revised (WAIS-R) accounted for over 90% of the variance. Taken together, these findings would suggest that neuropsychological measures are reliable predictors of functional abilities.

Furthermore, associations between positive symptoms and outcome measures have been weak. Jonsson and Nyman (1991) failed to obtain a relationship between psychotic symptoms and outcome variables. While Addington et al. (1991) and Breier et al. (1991) report associations between cognitive deficits and negative symptoms in schizophrenia, they also observed low correlations between neuropsychological measures and positive symptoms. These findings suggest that the factors underlying the neuropsychological impairment observed in schizophrenia accounts, in part, for the inability of these patients to execute routine functions of daily living and that these deficits are characteristic of the disease in the sense that they are better predictors of functional outcome than psychotic symptoms.

2.4 Course

No chapter on the neuropsychology of schizophrenia would be complete without a discussion addressing the nature and extent of the cognitive deficits throughout the course of the illness. Toward this end, it has been shown that the neuropsychological deficits observed in schizophrenia are maintained throughout the course of the illness without marked deterioration or improvement. Hyde et al. (1994) administered a battery of neuropsychological tests, routinely used to assess degenerative neurologic disorders, to schizophrenic patients grouped into five cohorts based on age (i.e. patients in their twenties, thirties, forties, fifties, and sixties). The results suggested that no progressive increase in cognitive impairment in schizophrenic patients was present across the five cohorts. Similarly, Heaton et al. (1994) found no difference between three schizophrenic groups (a young group with early onset, an elderly group with early onset, and a late-onset group) on the basis of age-adjusted neuropsychological tests that have been demonstrated to be sensitive to changes that occur with aging, schizophrenia, and dementia. In this sense, the course of schizophrenia has been portrayed as a static encephalopathy rather than as a progressive dementing illness. This view is not without controversy, and other investigators have found declines in cognitive function (Davidson et al. 1995; Waddington et al. 1997). However, declines were very small and not consistent with a progressive dementia. The remainder of this chapter will provide detailed neuropsychological interpretations of the impaired cognitive domains which characterize schizophrenia.

3 Deficit Profile

3.1 Attention

There is a large body of literature pertaining to attentional impairments in schizophrenia. In this chapter, we will only consider some of the more relevant issues. Attentional processes are thought to be composed of numerous subcomponents and may be under the direction of a distributed network within the brain. According to Posner (1995), there are three major attentional processes: orienting to sensory stimuli, detecting target events, and maintaining an alert state. Although Posner et al. (1988) demonstrated a deficit in schizophrenia patients for the processing of targets in the right visual field when attention was diverted, subsequent studies failed to obtain similar results (Strauss et al. 1991; Gold et al. 1992b).

Maruff et al. (1996) showed that, similar in extent to controls, medicated schizophrenic patients display the beginning of orienting responses. However, unlike controls, these patients were unable to make use of information (provided by the experimenter) which would inhibit directing attention to inappropriate peripheral cues. Thus, when informed of the correct location of the target prior to initiation of the task, normal controls were capable of inhibiting the urge to orient to an inappropriately located peripheral cue and use the occurrence of the inappropriate cue to orient to the correct contralateral target location. Schizophrenic patients, however, were incapable of ignoring the inappropriate peripheral cue and correctly orienting to the contralateral target location. These studies suggest that schizophrenic patients may have difficulty applying a strategy which conflicts with a previously learned response.

Schizophrenic patients typically display an increase in the number of omission errors or failures to respond to the presence of a target during the CPT. In one version of the CPT, a signal is given prior to the presence of the target, which is intended to prepare the subject for the possible occurrence of a target. Schizophrenic patients are not capable of effectively employing this "ready" signal, which would suggest that schizophrenic patients experience difficulty in maintaining a readiness to respond. Mirsky et al. (1992) have demonstrated that schizophrenic patients exhibit deficits in sustained attention on the CPT. Additionally, using the CPT, Servan-Schreiber et al. (1996) have shown that increasing the delay between cue and target stimuli results in a decrease in the number of correct responses and an increase in the number of commission errors (the tendency to report the presence of the

cue-target pair when none was present) in unmedicated schizophrenic patients relative to medicated schizophrenic patients and controls. Additionally, Servan-Schreiber et al. (1996) demonstrated that schizophrenic patients were unable to make use of a cue which would alert them to suppress an overlearned response. These findings contribute further evidence to the belief that schizophrenic patients are unable to maintain a readiness to respond. Additionally, since the performance of schizophrenic patients declined with an increase in the interval between cue and target, the finding by Servan-Schreiber et al. (1996) suggests that the CPT may not be a test of "attention" per se, but rather a test of working memory. The additional time between cue and target would appear to provide time during which the relationship between cue and target may decay.

3.2 Memory

Learning and memory, in the sense of the acquisition and retention of new information, is a domain that has been reliably demonstrated to be adversely affected in schizophrenia. Memory has been operationally divided into numerous subtypes (for reviews, see Squire 1992; Schacter and Tulving 1994; see also Vol. 1, Chap. 13). One mnemonic distinction that has particular relevance for schizophrenia is the differential performance observed between immediate and delayed recall.

Schizophrenic patients routinely display deficits with respect to immediate recall. The California Verbal Learning Test (CVLT) is a test of memory in which a list of 16 words is repeatedly presented over five trials with immediate recall occurring after each trial as well as delayed recall after a brief interval and again after a long interval of 20 min. Schizophrenic patients routinely score one to two standard deviations below the mean of the normal population on virtually all measures of the CVLT. The schizophrenic population as a whole also yields a similar performance on many of the mnemonic variables measured in the WMS-R, a battery of tests that assesses various types of immediate and delayed memory, e.g. paragraph recall, visual and verbal paired associate learning, and simple two-dimensional figure recall. Although schizophrenic patients display a relative deficit on immediate recall measures, they routinely exhibit normal savings over long-term delays (Goldberg et al. 1993c; Paulsen et al. 1995).

Memories, in general, are believed to be a product of three processes: encoding, storage, and retrieval. Encoding pertains to placing information to be remembered into a preexisting structure, storage refers to the process that maintains the information in the structure over time, and retrieval refers to the process of selecting information and putting it to use. The memory changes associated with schizophrenia have been demonstrated to be due to inefficient encoding and retrieval strategies. In a list-learning study in which schizophrenic patients and normal controls were required to memorize three lists of 20 words that were either unrelated, categorically related but not grouped by category, or categorically related and grouped by category, schizophrenic patients were impaired under all conditions relative to normal controls (Gold et al. 1992a). Though displaying improved performance during recall of words that were grouped by category relative to the other conditions (suggesting that they experience impaired encoding processes), the schizophrenic patients still demonstrated markedly impaired performance relative to controls. Moreover, in this study, recognition and recall were correlated, indicating that storage of the lists was problematic and the results were not solely due to inefficient encoding or retrieval. Additionally, in a large study in which the CVLT was administered to 175 schizophrenic patients and 229 normal controls, Paulsen et al. (1995) demonstrated that the schizophrenic patients displayed a retrieval deficit, as measured by disproportionate recall to recognition impairment (although recognition was still abnormal).

The neuropsychological findings of generalized impairment in mnemonic function in schizophrenia would suggest that schizophrenic patients experience at least some disruption of the connectivity between the hippocampal formation and other cortical regions involved in memory. Abnormalities in the schizophrenic performance on the CVLT and the logical memory segment of the WMS-R suggest a deficit in the rapid acquisition of new information, which has been demonstrated to be a function of the hippocampal formation (for a review, see Squire 1992). The demonstration of a correlation between the impaired performance of schizophrenic patients on the logical memory portion of the WMS-R and volumetric reductions of the left hippocampal region, as measured by magnetic resonance imaging (Goldberg et al. 1994), provides further evidence in support of the proposal that alterations of the hippocampal formation may be at least partly responsible for the memory deficit observed in schizophrenia.

3.2.1 Working Memory

Working memory is another form of memory that may be impaired in schizophrenic patients. Working memory has been defined as the ability to retain and manipulate information over short periods of time in the absence of external cues (for a review, see Baddeley and Hitch 1994). Working memory is believed to be

composed of three components: a central executive system (CES), an articulatory loop, and a visuospatial scratch pad. Executive function refers to complex thought processing and is defined as the ability to solve problems, plan, abstract, form concepts, and make adjustments or "shift sets" when older response sets are no longer adaptive. The CES is also thought to allocate resources among concurrent processes. The articulatory loop actively maintains verbal information, while the visuospatial scratch pad actively maintains visually presented information. Schizophrenic patients exhibit deficits in each of the systems that comprise working memory.

The CES is thought to provide a controlling mechanism during tasks that require multiple operations upon primary or immediate memory. Deficits during dual-task procedures are thought to represent CES impairment. Using a modified Brown-Peterson dual task in which participants are required to remember information, such as a list of words, while performing a distracter task, such as counting backwards, Fleming et al. (1995) obtained a decline in the performance of schizophrenic patients when the distracter task called for counting forward. This result would support the notion of a CES working memory impairment in schizophrenia, since the patients were unable to perform at a normal level even while performing a relatively simple distracter task. Additionally, when performing the dual task, schizophrenic patients displayed impaired performances on those distracter tasks which were more related to the memory task and thereby increased the load on working memory as opposed to tasks which were less related to the memory task (Fleming et al. 1995). However, while performing the simplest form of distracter task, finger tapping, which made little or no demands on verbal memory, schizophrenic patients performance on the recall task was no different than controls. Increasing the load on the working memory system of schizophrenic patients therefore results in impaired performance relative to normal controls.

Gold et al. (1997) demonstrated a strong positive correlation between impaired performance on a putative measure of working memory, the Letter–Number Span test, and impaired performance on the WCST in the schizophrenic population. The Letter–Number Span test is a test of immediate recall in which series of alternating letters and numbers are presented for memorization with an increasing series of digits presented after each successful recall. Impaired performance on tasks involving executive function has long been considered a hallmark of schizophrenia. The WCST requires the patient to match a sample card containing simple geometric shapes to an array of four cards on the basis of one of three attributes – color, form, or number. The patient is not told on which dimension to match the card in advance, but rather only whether or not their choice was correct or incorrect. In this sense, patients are required to formulate the rules of the test, which alternate in a predetermined manner. The WCST is thus believed to be a test of problem solving and set shifting. Fey (1951) used the WCST to differentiate between schizophrenic patients and normal controls. The WCST has been used extensively and is thought to be a classic test of executive function (Milner 1963). Relatively few numbers of categories obtained and relatively high percentages of perseverative errors on the WCST is the characteristic performance observed in schizophrenic patients (Goldberg et al. 1988).

The integrity of the prefrontal cortex has long been associated with performance on the WCST (Milner 1963). Both lesion studies and studies of activation in normal controls undergoing functional neuroimaging have suggested that the prefrontal cortex is necessary for performance on the WCST (Milner 1963; Goldberg et al. 1994). Seidman et al. (1994) demonstrated strong, significant, inverse correlations between impaired performance on the WCST and volumetric reductions of the dorsolateral prefrontal cortex as measured by magnetic resonance imaging in schizophrenic patients. Using difference scores between monozygotic twins that were discordant for schizophrenia, Goldberg et al. (1994) demonstrated a strong correlation between prefrontal regional cerebral blood flow (rCBF) and perseveration on the WCST. With respect to the number of categories obtained and the number of perseverative errors committed on the WCST, Sullivan et al. (1994) demonstrated that the performance of schizophrenic patients fell between one and two standard deviation units below normal controls. These findings in schizophrenic patients suggest that working memory may be a critical component of WCST performance and that the function of the prefrontal cortex is compromised in schizophrenia.

In addition to displaying deficits of the CES of the working memory system, procedures which measure articulatory loop and visuospatial scratch pad performance also elicit deficits in the schizophrenic population. Using an occulomotor delayed-response task thought to be a test of visuospatial working memory in which the participant must make a delayed motoric response to the correct spatial location of a target stimulus, Park and Holzman (1992) demonstrated significant impairment in schizophrenic patients relative to controls. Additionally, Fleming et al. (1997) demonstrated impaired performance of schizophrenic patients in visuospatial tasks that required working memory: WMS-R visual memory span and a spatial delayed-response task. In the same study, the schizophrenic patients did not display deficits on a visuo-

spatial task which does not involve a delay, the judgement of line orientation, which selectively measures perceptual abilities and does not draw upon the resources of working memory. This pattern of results would implicate prefrontal cortical regions as opposed to occipital and temporal regions as being responsible, at least in part, for the observed working memory deficit in schizophrenia.

Perhaps the construct which best captures schizophrenic patients' inability to maintain and manipulate information in working memory is that of capacity limitations on the amount of information that working memory is capable of processing. For example, in the letter-number span study of Gold et al. (1997), schizophrenic patients performance declined relative to controls with increased span length. Additionally, Cohen and Servan-Schreiber (1992) observed a decrease in comprehension by schizophrenic patients when the distance between phrases increased. This increase of distance may represent an increase in the "load" or amount of information being handled by working memory and would suggest an impairment in the ability to manipulate such increases in schizophrenia.

3.2.2 Thought Disorder and Semantic Memory

Schizophrenic patients frequently exhibit formal thought disorder, as manifested in disorganized speech characterized by poverty of content of speech, derailment or tangentiality, illogicality, use of word approximations or neologisms, and loss of goal (Andreasen 1986). In schizophrenia, thought disorder appears to arise from an inability to produce coherent language rather than an inability to understand and comprehend language.

One hypothesis that addresses the neuropsychological basis for thought disorder in schizophrenia suggests that the speech production errors observed in schizophrenia are the result of a malfunction in the semantic organization of schizophrenic patients. Evidence in support of this hypothesis comes from work done on priming and fluency in schizophrenia. With respect to fluency, Gourovitch et al. (1996) demonstrated that schizophrenic patients displayed a superior performance for the generation of words (in 1-min epochs) from phonologic categories, (i.e. the letters *f*, *a*, and *s*), relative to semantic categories, (i.e. "animals," "fruits," and vegetables"), and the magnitude of the difference between these two measures of fluency was strongly correlated with the presence of thought disorder (Goldberg and Weinberger 2000). Additionally, Aloia et al. (1996) demonstrated that the "semantic space" in schizophrenia patients was disorganized relative to normal controls on the basis of a multidimensional scaling analysis of a verbal fluency task, suggesting an abnormal weighting of connection strengths between concepts or attributes in semantic memory.

Further evidence supporting the existence of semantic abnormalities in schizophrenia can be found in priming studies in which a word may "prime" or facilitate a future response on the basis of relatedness. For example, when the word "cat" follows the word "dog," it is responded to more rapidly than if it follows the word "stone," presumably due to increased connection strengths between the "nodes" representing the more frequently associated words. Work on semantic priming in schizophrenia is complex and has yielded mixed results, perhaps due to differences in paradigm or cohort (Barch et al. 1996; Chapin et al. 1989; Goldberg and Weinberger 2000; Ober et al. 1995; Spitzer 1997).

Although Chapin et al. (1989) and Ober et al. (1995) displayed similar semantic priming effects between schizophrenic patients and controls during a lexical decision task which required patients to make a decision as to whether or not a word was a member of a particular category, these studies also obtained slower reaction times for schizophrenic patients. Additionally, Chen et al. (1994) obtained slowed reaction times in schizophrenic patients versus controls during a task requiring determination of the degree to which words were exemplars of specific categories. These results would suggest that category boundaries are vague and overinclusive in schizophrenia, perhaps due to overactivation of marginally related concepts.

In an attempt to determine the cognitive mechanism of the slowed and inaccurate processing by schizophrenic patients during semantic priming, in particular, with the hope of being able to determine a cognitive mechanism for thought disorder in schizophrenia in general, we (Aloia et al. 1998) developed a priming task in which the prime and the target were either strongly, moderately, or weakly related words from a single semantic category. In this study, thought-disordered patients displayed "negative" priming, i.e. their responses were slower to strongly and moderately related words, which directly contrasts with the performance of normal controls producing faster responses to the strongly and moderately related words. This differential performance between thought-disordered schizophrenic patients and normal controls would suggest that, when a patient attempts to produce a word automatically on the basis of spreading activation, closely related words are not activated. Speculatively, in speech, an inappropriate but somewhat related word might be produced instead. The result would be interpreted by the clinician as a "loose," tangential, or irrelevant response on the part of the patient.

3.3 Motor Control

Schizophrenic patients also display deficits of motor control. Slowed reaction times on both simple and complex tests have been demonstrated in schizophrenia (Vrtunski et al. 1986). Additionally, impaired performance in schizophrenic patients has been demonstrated on tests of fine motor control such as the grooved peg board peg placement test and finger tapping (Saykin et al. 1994).

One problematic finding relates to the demonstration of the effect of neuroleptic medications on motor control and responses. The extrapyramidal side effects of neuroleptic medications can produce a retardation of reaction time. Thus the question arises of whether the slowing of motoric responses by the administration of neuroleptic medication in schizophrenia is accompanied by a generalized slowing of neuronal processing that results in the cognitive changes described above. As mentioned previously, on the basis of "first break" or neuroleptic naive studies (Saykin et al. 1994), drug-free studies, and comparisons with patients with movement disorder (Goldberg et al. 1990), the cognitive impairments demonstrated in schizophrenic patients tend to deteriorate in the absence of neuroleptic treatment and would therefore appear to occur independently of the presence of neuroleptic medications (for a review, see Goldberg and Weinberger 1996).

Indeed, instead of a neuroleptic effect, most of the cognitive changes described above appear to be enduring effects of the disease and dissociable from psychomotor slowing in general. Thus, when Hanes et al. (1996) compared schizophrenic patients to patients with known psychomotor slowing due to basal ganglia pathology (i.e. Huntington's disease), marked differences on several higher cognitive tasks were observed. Moreover, studies of procedural learning in schizophrenia have demonstrated that schizophrenic patients are capable of improving their motor skill learning at a rate which is analogous to that observed in normal control participants (Goldberg et al. 1993c). These results would suggest that the cognitive impairments observed in schizophrenia are not a product of a generalized motoric slowing.

4 Conclusions

Schizophrenia can be characterized by impaired performances in the areas of attention, episodic memory, working memory, executive function, semantic organization, and motor control. Impairment of these cognitive domains has serious consequences for the functional capacities of the schizophrenic patient, including social and vocational skills. Additionally, the characteristic neuropsychological profile of schizophrenic patients suggests cortical dysfunction in the regions of the frontal lobe, particularly in the motor, premotor, and prefrontal cortex, as well as the medial temporal lobe.

5 References

Addington J, Addington D (1993) Premorbid functioning, cognitive functioning, symptoms and outcome in schizophrenia. J Psychiatry Neurosci 18: 18–23

Addington J, Addington D, Maticka-Tyndale E (1991) Cognitive functioning and positive and negative symptoms in schizophrenia. Schizo Res 5: 123–134

Aloia MS, Gourovitch ML, Weinberger DR, Goldberg TE (1996) An investigation of semantic space in patients with schizophrenia. J Int Neuropsychol Soc 2: 267–273

*Aloia MS, Gourovitch ML, Missar D, Pickar D, Weinberger DR, Goldberg TE (1998) Cognitive substrates of thought disorder. II. Specifying a candidate cognitive mechanism. Am J Psychiatry 155(12): 1677–1684

Andreasen NC (1986) Scale for the assessment of thought, language, and communication (TLC). Schizophr Bull 12(3): 473–482

Baddeley AD, Hitch GJ (1994) Developments in the concept of working memory. Neuropsychology 8: 485–493

Barch DM, Cohen JD, Servan-Schreiber D, Steingard S, Cohen JD, Steinhauer SS, van Kammen DP (1996) Semantic priming in schizophrenia: an examination of spreading activation using word pronunciation and multiple SOA's. J Abnorm Psychol 105: 592–601

Bellack AS, Mueser KT, Morrison RL, Tierney A, Podell K (1990) Remediation of cognitive deficits in schizophrenia. Am J Psychiatry 147: 1650–1655

Braff DL, Heaton R, Kuck J, Munro C, Moranville J, Grant I, Zisook S (1991) The generalized pattern of neuropsychological deficits in outpatients with chronic schizophrenia with heterogenous Wisconsin Card Sorting Test results. Arch Gen Psychiatry 48: 891–898

Breier A, Schreiber JL, Dyer J, Pickar D (1991) National Institute of mental health longitudinal study of chronic schizophrenia: prognosis and predictors of outcome. Arch Gen Psychiatry 48: 239–246

Chapin K, Vann LE, Lycaki H, Josef N, Meyendorff E (1989) Investigation of the associative network in schizophrenia using the semantic priming paradigm. Schizophr Res 2: 35–360

Chen EYH, Wilkins AJ, McKenna PJ (1994) Semantic memory is both impaired and anomalous in schizophrenia. Psychol Med 24: 193–202

Cleghorn JM, Kaplan RD, Szechtman B, Szechtman H, Brown GM (1990) Neuroleptic drug effects on cognitive function in schizophrenia. Schizophr Res 3: 211–219

Cohen JD, Servan-Schreiber D (1992) Context, cortex and dopamine: a connectionist approach to behavior and biology in schizophrenia. Psychol Rev 99: 45–77

Davidson M, Harvey PD, Powchik P, Parrella M, White L, Knobler HY, Losonczy MF, Keefe RSE, Katz S, Frecska E (1995) Severity of symptoms in chronically institutionalized geriatric schizophrenic patients. Am J Psychiatry 152: 197–207

Faustman WO, Moses JA, Csernansky JB (1988) Luria-Nebraska performance and symptomatology in unmedicated schizophrenic patients. Psychiatry Res 26: 29–34

Fey ET (1951) The performance of young schizophrenics and young normals on the Wisconsin Card Sorting Test. J Consult Psychol 15: 311–319

Fleming K, Goldberg TE, Gold JM, Weinberger DR (1995) Verbal working memory dysfunction in schizophrenia: use of a Brown-Peterson paradigm. Psychiatry Res 56: 155–161

Fleming K, Goldberg TE, Binks S, Randolph C, Gold JM, Weinberger DR (1997) Visuospatial working memory in patients with schizophrenia. Biol Psychiatry 41: 43–49

Frith C (1996) Neuropsychology of schizophrenia: what are the implications of intellectual and experiential abnormalities for the neurobiology of schizophrenia? Br Med Bull 52: 618–626

Gold JM, Randolph C, Carpenter CJ, Goldberg TE, Weinberger DR (1992a) Forms of memory failure in schizophrenia. J Abnorm Psychol 101: 487–494

Gold JM, Randolph C, Coppola RC, Carpenter C, Goldberg TE, Weinberger DR (1992b) Visual orienting in schizophrenia. Schizophr Res 7: 203–209

*Gold JM, Carpenter C, Randolph C, Goldberg TE, Weinberger DR (1997) Auditory working memory and Wisconsin Card Sorting Test performance in schizophrenia. Arch Gen Psychiatry 54: 159–165

*Goldberg TE, Gold JM (1995) Neurocognitive deficits in schizophrenia. In: Hirsch SR, Weinberger DR (eds) Schizophrenia. Blackwell, London

Goldberg TE, Weinberger DR (1996) Effects of neuroleptic medications on the cognition of patients with schizophrenia: a review of recent studies. J Clin Psychiatry 57[Suppl 9]: 62–65

Goldberg TE, Weinberger DR (2000) Thought disorder: a critical reappraisal of older studies and recent formulations. Cogn Neuropsychiatry 5: 1–19

Goldberg TE, Weinberger DR, Berman KF, Pliskin NH, Podd MH (1987) Further evidence for dementia of the prefrontal type in schizophrenia? A controlled study of teaching the Wisconsin Card Sorting Test. Arch Gen Psychiatry 44: 1008–1014

Goldberg TE, Kelsoe JR, Weinberger DR, Pliskin NH, Kirwin PD, Berman KF (1988) Performance of schizophrenic patients on putative neuropsychological tests of frontal lobe function. Int J Neurosci 42: 51–58

Goldberg TE, Berman KF, Mohr E, Weinberger DR (1990) Regional cerebral blood flow and cognitive function in Huntington's disease and schizophrenia: a comparison of patients matched for performance on a prefrontal-type task. Arch Neurol 47: 418–422

Goldberg TE, Gold JM, Greenberg R, Griffin S, Schulz SC, Pickar D, Kleinman JE, Weinberger DR (1993a) Contrasts between patients with affective disorders and patients with schizophrenia on a neuropsychological test battery. Am J Psychiatry 150: 1355–1362

Goldberg TE, Greenberg RD, Griffin SJ, Gold JM, Kleinman JE, Pickar D, Schulz SC, Weinberger DR (1993b) The effect of clozapine on cognition and psychiatric symptoms in patients with schizophrenia. Br J Psychiatry 162: 43–48

Goldberg TE, Torrey EF, Gold JM, Ragland JD, Bigelow LB, Weinberger DR (1993c) Learning and memory in monozygotic twins discordant for schizophrenia. Psychol Med 23: 71–85

Goldberg TE, Torrey EF, Berman KF, Weinberger DR (1994) Relations between neuropsychological performance and brain morphological and physiological measures in monozygotic twins discordant for schizophrenia. Psychiatry Res 55: 51–61

Goldberg TE, Torrey EF, Gold JM, Bigelow LB, Ragland RD, Taylor E, Weinberger DR (1995) Genetic risk of neuropsychological impairment in schizophrenia: a study of monozygotic twins discordant and concordant for the disorder. Schizophr Res 17: 77–84

Gourovitch ML, Goldberg TE, Weinberger DR (1996) Verbal fluency deficits in patients with schizophrenia: semantic fluency is differentially impaired as compared with phonologic fluency. Neuropsychology 10: 573–577

**Green MF (1996) What are the functional consequences of neruocognitive deficits in schizophrenia? Am J Psychiatry 153: 321–330

Green MF, Staz P, Ganzell S, Vaclav JF (1992) Wisconsin Card Sorting Test in schizophrenia: remediation of a stubborn deficit. Am J Psychiatry 149: 62–67

Hanes KR, Andrewes DG, Pantelis C, Chiu E (1996) Subcortical dysfunction in schizophrenia: a comparison with Parkinson's disease and Huntington's disease. Schizophr Res 19: 121–128

Heaton RK, Pendleton MG (1981) Use of neuropsychological tests to predict adult patients' everyday functioning. J Con Clin Psych 49: 807–821

Heaton RK, Chelune GJ, Lehman RAW (1978) Using neuropsychological and personality tests to assess the likelihood of patient employment. J Nerv Ment Dis 166: 408–416

Heaton R, Paulsen JS, McAdams LA, Kuck J, Zisook S, Braff D, Harris MJ, Jeste DV (1994) Neuropsychological deficits in schizophrenics: relationship to age, chronicity, and dementia. Arch Gen Psychiatry 51: 469–476

Hyde TM, Nawroz S, Goldberg TE, Bigelow LB, Strong D, Ostrem JL, Weinberger DR, Kleinman JE (1994) Is there cognitive decline in schizophrenia? A cross-sectional study. Br J Psychiatry 164: 494–500

Jonsson H, Nyman AK (1991) Predicting long-term outcome in schizophrenia. Acta Psychiatr Scand 83: 342–346

Kibel DA, Laffont I, Liddle PF (1993) The composition of the negative syndrome of chronic schizophrenia. Br J Psychiatry 162: 744–750

Levin S, Yurgelun-Todd D, Craft S (1989) Contributions of clinical neuropsychology to the study of schizophrenia. J Abnorm Psychology 98: 341–356

Maruff P, Pantellis C, Danckert J, Smith D, Currie J (1996) Deficits in the endogenous redirection of covert visual attention in chronic schizophrenia. Neuropsychologia 11: 1079–1084

Medalia A, Gold JM, Merriam A (1988) The effects of neuroleptics on neuropsychological test results of schizophrenics. Arch Clin Neuropsychol 3: 249–271

Milner B (1963) Effects of different brain lesions on card sorting: the role of the frontal lobes. Arch Neurol 9: 100–110

Mirsky AF, Lochhead SJ, Jones BP, Kugelmass S, Walsh D, Kemdler KS (1992) On familial factors in the attentional deficit in schizophrenia: a review and report of two new subject samples. J Psychiatr Res 26: 383–403

Newnan OS, Heaton RK, Lehman RAW (1978) Neuropsychological and MMPI correlates of patients' future employment characteristics. Percept Mot Skills 46: 635–642

Ober BA, Vinogradov S, Shenaut GK (1995) Semantic priming of category relations in schizophrenia. Neuropsychology 9: 220–228

Oltmanns TF, Ohayon J, Nezle JM (1978) The effect of antipsychotic medication and diagnostic criteria on distractibility in schizophrenia. J Psychiatr Res 14: 81–91

*Park S, Holzman PS (1992) Schizophrenics show spatial working memory deficits. Arch Gen Psychiatry 49: 975–982

Paulsen JS, Heaton RK, Sadek JR, Perry W, Delis DC, Braff D, Kuck J, Zisook S, Jeste DV (1995) The nature of learning and memory impairments in schizophrenia. J Int Neuropsychol Soc 1: 88–99

Perlick D, Mattis S, Stastny P, Teresi J (1992) Neuropsychological discriminators of long-term inpatients or outpatients in chronic schizophrenia. J Neuropsychiatr Clin Neurosci 4: 428–434

Posner MI (1995) Attention in cognitive neuroscience: an overview. In: Gazzaniga MS (ed) The cognitive neurosciences. MIT, Cambridge, MA

Posner MI, Early TS, Reiman E, Pardo PJ, Dhawan M (1988) Asymmetries in hemispheric control of attention in schizophrenia. Arch Gen Psychiatry 45: 814–821

Rapaport D, Gill M, Schafer R (1945/1946) Diagnostic psychological Testing. Year Book, Chicago

**Saykin AJ, Shtasel DL, Gur RE, Kester DB, Mozley LH, Stafiniak P, Gur RC (1994) Neuropsychological deficits in neuroleptic naïve patients with first-episode schizophrenia. Arch Gen Psychiatry 51: 124–131

Schacter DL, Tulving E (1994) What are the memory systems of 1994? In: Schacter DL, Tulving E (eds) Memory systems 1994. MIT, Cambridge, MA

Schneider SG, Asarnow RF (1987) A comparison of cognitive/neuropsychological impairments of nonretarded autistic and schizophrenic children. J Abnorm Child Psychol 15: 29–46

Seidman LJ, Pepple JR, Faraone SV, Kremen WS, Green AI, Brown WA, Tsuang MT (1993) Neuropsychological performance in chronic schizophrenia in response to neuroleptic dose reduction. Biol Psychiatry 33: 575–584

Seidman LJ, Yurgelun-Todd D, Kremen WS, Woods BT, Goldstein JM, Faraone SV, Tsuang MT (1994) Relationship of prefrontal and temporal lobe MRI measures to neuropsychological performance in chronic schizophrenia. Biol Psychiatry 35: 235–246

*Servan-Schreiber D, Cohen JD, Steingard S (1996) Schizophrenic deficits in the processing of context: a test of a theoretical model. Arch Gen Psychiatry 53: 1105–1112

Spitzer M (1997) A cognitive neuroscience view of schizophrenic thought disorder. Schizophr Bull 23: 29–50

Spohn HE, Strauss ME (1989) Relation of neuroleptic and anticholinergic medication to cognitive function in schizophrenia. J Abnorm Psychol 98: 367–380

Squire LR (1992) Memory and the hippocampus: a synthesis from findings with rats, monkeys, and humans. Psychol Rev 99: 195–231

Strauss ME, Novakovic T, Tien AY, Bylsma F Pearlson GD (1991) Disengagement of attention in schizophrenia. Psychiatry Res 37: 139–146

Stuss DT, Benson DF, Kaplan EF, Weir WS, Naeser MA, Lieberman I, Ferrill D (1983) The involvement of orbitofrontal cerebrum in cognitive tasks. Neuropsychologia 21: 235–248

*Sullivan EV, Shear PK, Zipursky RB, Sagar HJ, Pfefferbaum A (1994) A deficit profile of executive, memory, and motor functions in schizophrenia. Biol Psychiatry 36: 641–653

Summerfelt AT, Alphs LD, Wagman AMI, Funderberk FR, Hierholzer RM, Strauss ME (1991) Reduction of perseverative errors in patients with schizophrenia. Using monetary feedback. J Abnorm Psychol 100: 613–616

Tompkins L, Goldman RS, Axelrod BN (1991) Modifiability of neuropsychological dysfunction in schizophrenia. J Exp Clin Neuropsychol 14: 57

Vrtunski PB, Simpson DM, Weiss KM, Davis GC (1986) Abnormalities of fine motor control in schizophrenia. Psychiatry Res 18: 275–284

Waddington JL, Scully PJ, Youssef HA (1997) Developmental trajectory and disease progression in schizophrenia: the conundrum, and insights from a 12-year prospective study in the Monaghan 101. Schizophr Res 23: 107–118

CHAPTER 9

W. Rössler

Schizophrenia: Psychosocial Factors

Translator: E. Taub

1 Historical Overview

The effect of psychosocial factors on the occurrence of schizophrenia and its course has been a subject of scientific investigation for just over 100 years. According to Kraepelin's view (1896), schizophrenia was caused by a (biological) disease process inevitably ending in "dementia" (*Verblödung*). Nonetheless, by the mid-twentieth century, it had become clear that the observed disease manifestations and behavioral abnormalities often did not, in fact, evolve in a uniformly negative way, even though this had been the conclusion of studies of long-term hospitalized patients around the turn of the century. Rather, it was found that they were subject to considerable influence from the custodial living conditions under which the patients were cared for (Wing and Freudenberg 1961; Wing and Brown 1970). The recognition that the patient's social environment can have a major effect on the course of severe mental illnesses, particularly schizophrenia, was an important motivating factor for the worldwide reforms in the care of the mentally ill that took place in the second half of the twentieth century.

The influence of psycho- and sociodynamic factors on schizophrenic manifestations and behavioral abnormalities was already a central element of Eugen Bleuler's concept of schizophrenia, which incorporated both the psychopathology of dementia praecox, as described by Kraepelin, and the psychodynamic dimension of Freudian theory (E. Bleuler 1911). The relationship of schizophrenic patients to their living environment was described in further detail by Manfred Bleuler (1972), according to whom negative influences in childhood were of particular importance for the development of schizophrenia. Manfred Bleuler later recognized, however, that childhood influences were relatively nonspecific; in the aftermath of his own observations, they were found to have comparable etiological importance for psychiatric illnesses other than schizophrenia (Benedetti 1995).

As early as the nineteenth century, psychiatrists had already suspected that harmful influences from the social environment might promote the appearance of mental illnesses or have a deleterious effect on their course. In that era, the isolation of the mentally ill in the "healthful" milieu of institutions far removed from their original living environment was held to be an appropriate means of shielding them from the disease-producing influences of their surroundings (Rössler 1992). In contrast, the twentieth-century reform movement sought to (re-)integrate patients in their normal living environment as rapidly as possible, because it was thought that their natural social network would make a major supporting contribution to the healing process.

The long-term deinstitutionalization of the mentally ill, predominantly schizophrenic patients who in some cases had been hospitalized for decades, succeeded only because the reform movement occurred in parallel with the introduction of neuroleptic medications for the treatment of schizophrenia. It soon became clear that long-term pharmacotherapy was also of major importance for the prevention of relapses, even though approximately one third of patients relapse despite medication, and approximately one fifth have a spontaneous remission without medication (Leff 1987). The fundamental contribution of the neuroleptics to the treatment of schizophrenia is uncontested.

Research on schizophrenia has therefore turned more toward defining its biological underpinnings since the 1970s, initially with the major objective of understanding the mechanisms of action of antipsychotic medications. This shift toward biological psychiatry was further accelerated by the development of new imaging techniques, new laboratory methods, and the field of genetic research.

The influence of genetic factors on the development of schizophrenia is universally accepted; it is immediately evident in the (approximately) 48% concordance rate for schizophrenia among monozygotic twins, for example. Similarly, the risk of developing schizophrenia is approximately 46% for the children of two schizophrenic parents (Gottesman 1991). It is precisely these unambiguous findings, however – in which a large portion, but by no means all, of the variance is accounted for by genetic factors – that indicate that other factors must also play a role. Furthermore, despite major research efforts, the specific genes involved in the development of schizophrenia have not yet been identified (DeLisi 1999). It therefore seems worthwhile to devote some attention to possible influential factors of other types, including psychosocial factors.

2 Conception of Illness in Relation to the Environment

The history of research on schizophrenia reveals that biology and social environment are not a dichotomous pair, but rather interact with each other to exert an influence on the disease. This is particularly true with respect to the course of the illness, less so, however, with respect to its onset.

Most theories of the etiology and course of schizophrenia are based on the so-called vulnerability–stress

model (Zubin and Spring 1977), according to which a biological vulnerability of undetermined origin and variable degree, accompanied by short- or long-term social stressors or other environmental factors, leads to the initial onset of the illness or to a relapse (Nuechterlein 1987).

The interaction-development model, which evolved from the vulnerability-stress model, postulates an active mutual influence between affected individuals and their environment. While the original vulnerability-stress model saw affected individuals as passive recipients of various psychosocial influences, the interaction-development model views them as active participants in the formation of their own environment (Strauss et al. 1985).

A systematic account of psychosocial factors should include a distinction between the immediate and the broader social environment. The immediate social environment comprises the social network of the affected individual, including family, friends, workplace, etc. Critical life events and family climate should also be included among the immediate social environment. Furthermore, not only the present situation, but also the developmental perspective should be considered; for example, one area of research interest deals with the relationship between the early childhood milieu and the later development of mental disturbances. The broader social environment includes sociocultural influences (also in an intercultural perspective) and, most importantly, socioeconomic factors.

3 Sociocultural Factors

3.1 Cultural Influences

The influence of culture on schizophrenia has long attracted considerable interest. A number of studies have shown (Pfeiffer 1994) that the disease takes on characteristic culture-specific forms. Comparative studies across cultures also suggest that the incidence and prevalence of schizophrenia are lower in developing countries than in industrial countries (Murphy and Raman 1971; Waxler 1979; Torrey 1980).

The lack of standardized comparative studies led the World Health Organization (WHO) to conduct multicenter comparative studies with uniform study instruments and follow-up examinations, starting in the 1960s (Sartorius et al. 1972; WHO 1974, 1975). The results of these studies imply that the incidence of the core manifestations of schizophrenia is approximately the same all over the world, but that the course of schizophrenic illnesses is highly variable and, in particular, differs between the developing and the industrial countries. Schizophrenics in developing countries have similar initial disease manifestations to those of patients in industrial countries, but they have a less chronic course of the illness, fewer relapses, and better social adaptation (WHO 1979; Sartorius et al. 1987; Jablensky et al. 1992). Further significant psychosocial factors were identified in addition to the developing versus industrial country factor, namely, family situation and social network (Sartorius et al. 1996).

It has been suggested that these differences in disease course between the developing and the industrial countries are attributable to the more easily mastered patterns of social interaction in less complex societies, in contrast to the complex, conflict-laden, and less easily grasped demands placed on the individual in modern industrial societies. Alternatively, it may be useful to discuss whether society, in the developing countries, places a less intense demand for autonomy and competitive behavior on vulnerable individuals and enables them to live in smaller, more stable, and longer-lasting social networks.

Migration studies also suggest an effect of the environment on the risk of developing schizophrenic illnesses. British studies, for example, have amply documented that immigrants from Trinidad and Jamaica, particularly in the second generation, have an elevated incidence of schizophrenic illnesses compared to the general population (e.g. Davies et al. 1995). Nonetheless, even if it seems plausible that unfavorable environmental conditions elevate the risk of developing disease, a number of confounding factors must also be taken into account.

Selection effects in migration are relevant in this connection. For example, Ødegaard (1932) showed, in a classic study, that Norwegian emigrants had an elevated incidence of schizophrenia, but Häfner (1980) showed that Turks in Germany were less likely to develop schizophrenia than the native population. This is presumably explained by the fact that strict health criteria were applied in the selection of guest workers coming to Germany.

Meanwhile, a number of more recent epidemiologic studies have shown no elevated risk of becoming ill among the populations of origin in Trinidad and Jamaica (Hickling and Rodgers-Johnson 1995; Bhugra et al. 1997). Now that other possible factors have also been studied (Hutchinson et al. 1997), it seems likely that the environment is a major risk factor for the development of schizophrenia, at least in this population group.

This is no longer an isolated finding. A Dutch study revealed that immigrants to the Netherlands from the

former colonies of Surinam and the Netherlands Antilles have a fourfold elevated risk of developing schizophrenia, as compared to the general Dutch population (Selten et al. 1997). Selection processes among the populations of origin are unlikely to have played a major role in this phenomenon, because large portions of the populations of origin took part in the migration to the Netherlands.

3.2 Socioeconomic Influences

The pioneering epidemiologic studies by Faris and Dunham (1939) of the ecological distribution of schizophrenia in Chicago in 1935 revived the discussion of the influence of social factors on the occurrence and course of schizophrenia. They found that the Chicago slums had the highest rates of individuals undergoing initial hospitalization for schizophrenia. The same pattern was found thereafter in many different cities and countries: in Bristol by Hare (1956), in Liverpool by Castle and Gittus (1957), in Nottingham by Giggs (1986), and in Mannheim (Germany) by Häfner and Reimann (1970). The Mannheim findings were later found to be essentially stable after a follow-up interval of 15 years (Weyerer and Häfner 1989).

The question of a relationship between the urban living environment and the risk of mental illness exercises a certain fascination even today. In a recent study, Torrey et al. (1997) analyzed U.S. census data from the year 1880 and found that the risk of developing psychosis was elevated by a factor of 1.6 in urban areas. Marcelis et al. (1998) used the Dutch national registry of psychiatric cases to study the relationship between place of birth and the risk of becoming ill and reported a moderately strong, but statistically significant association between urban birthplace and elevated incidence of disease. Mortensen et al. (1999), too, in their analysis of the Danish national registry, found a 2.4-fold elevation in the incidence of schizophrenia among individuals born in the national capital. At the same time, they found a ninefold elevation in incidence among subjects with a positive family history.

Evidence for a heterogeneous ecological distribution of schizophrenia is found not only within cities, but also in relation to urban–rural differences (Dohrenwend and Dohrenwend 1974). The largest North American epidemiologic study of the last few decades, however, revealed no urban–rural differences in the prevalence of schizophrenia when factors such as age, sex, and race were taken into account (Shapiro et al. 1984).

The heterogeneous ecological distribution of schizophrenia is in accordance with the demonstrated differences in the frequency of schizophrenia among different social classes. Numerous studies have documented an overrepresentation of schizophrenics in lower social classes (Clark 1948; Stein 1957; Hollingshead and Redlich 1958; Myers and Bean 1968; Eaton et al. 1988). In the review articles by Dohrenwend and Dohrenwend (1969) and Eaton (1974), it was reported that five of seven and 15 of 17 studies, respectively, had yielded this result. The explanation was sought mainly in the greater social isolation of the affected individuals and in the lack of social support in lower social classes.

Although, at the time of the reform movement in psychiatry in the 1960s and 1970s, it was thought that research findings of the type just discussed had conclusively demonstrated the social origin of schizophrenia, this interpretation is no longer accepted as valid today. These findings are now generally explained by the hypotheses of social decline and social selection. Social decline refers to the social consequences that follow the onset of illness, and social selection refers to the lack of social rise even before the onset of illness (Häfner 1992).

The social decline of schizophrenics has been amply documented. Marneros et al. (1991), for example, in a long-term study, found that 71% of schizophrenics studied had suffered a downward change of occupation, usually associated with a shift into a lower social class, and often also with a move into a neighborhood with social problems of many different kinds. The prevailing anonymity of such neighborhoods may sometimes be a positive feature for schizophrenics with disturbances of communication. Not least, a greater offering of social services is usually available in inner-city neighborhoods, which may be more attractive to affected individuals for this reason as well.

It is more difficult to find empirical support for the hypothesis of social selection, which is presumed to be primarily the effect of premorbid personality changes. Malmberg et al. (1998) studied a cohort of Swedish army recruits from the years 1969 and 1970 who developed schizophrenia over the following 15 years and found significant premorbid deficiencies of social adaptation.

The so-called break in the performance curve before the overt appearance of the disease is also well known. Ødegaard (1971), in an analysis of the Norwegian case registry, found that occupations requiring lower qualifications were significantly overrepresented among initially hospitalized schizophrenics. Goldberg and Morrison (1963), in a controlled study, showed that initially hospitalized schizophrenics were employed in occupations requiring lower qualifications than their own fathers' occupations.

Nonetheless, these findings, too, have alternative explanations. The largest German study of patients in

the earliest stages of schizophrenia revealed that initial hospitalization occurs on average approximately 4.5 years after the appearance of the first manifestations of any kind and approximately 2 years after the appearance of the first psychotic manifestations (Häfner et al. 1998). This suggests that the social abnormalities found in the period before the disease becomes overt ("social selection") are actually early signs of social decline.

4
Influences of the Immediate Social Environment

4.1
Early Childhood Environment

The presumed influence of the early childhood environment on the risk of developing schizophrenia played a major role in the psychosocial theories of the 1960s and 1970s.

According to Bateson and colleagues (Bateson 1972; Bateson et al. 1956), the cognitive and affective disturbances in schizophrenia are caused by an abnormal parent-child relationship. One aspect of this idea, the so-called double-bind hypothesis, became widely known: contradictory messages from parents to their children were said to lead inevitably to schizophrenic reactions. On the other hand, in the views of Wynne and Singer (1966) or Lidz (1975), particular types of conflicts between the parents were held to cause schizophrenia in their children.

The major weakness of these explanatory approaches – aside from the inappropriateness of constructing theories without adequate empirical support – lies in their failure to consider whether the observed phenomena represent causes or effects of illness. This deficiency might be remedied by long-term prospective studies. Such studies have been performed on individuals with an elevated risk of developing schizophrenia, so that an adequate number of schizophrenics would eventually be included in the study population; the risk-conferring feature usually chosen for this purpose has been genetic load.

Among five such long-term studies of subjects at elevated risk of developing schizophrenia, two showed that unfavorable family circumstances were an additional risk factor (Cornblatt and Obuchowski 1997). In the Copenhagen risk study, perinatal complications and unstable family circumstances in early childhood were found to be important risk factors (Cannon and Mednick 1993; Cannon et al. 1994). In the Finnish adoption study, schizophrenia appeared almost exclusively in individuals with a genetic risk, but, within this group, mainly in those who had grown up in difficult family circumstances (Tienari et al. 1989, 1994).

4.2
Family Atmosphere

Almost in parallel to the development of psychosocial theories discussed previously regarding unfavorable family influences in early childhood as risk factors for the development of schizophrenia, a branch of empirical research came into being that was concerned primarily with the influence of family circumstances on the course of the disease. The point of departure for this area of research, which came to be known under the heading of "expressed emotion," was the observation that schizophrenics discharged from inpatient treatment into the care of their families had a high rate of relapse (Brown 1959). The ensuing study conducted by Brown et al. (1962) showed that 76% of patients living in a family atmosphere characterized by criticism and hostility experienced a relapse, as compared to only 26% of those who experienced relatively little criticism and hostility.

In the years thereafter, many empirical studies were performed on the question of a relapse-promoting family atmosphere, most of which confirmed the findings of the original study (Leff and Vaughn 1985; Lam 1991; Bebbington and Kuipers 1994). Bebbington (1995), in a meta-analysis, found that a favorable family atmosphere had a more significant relapse-preventing effect than pharmacotherapy.

Even though these research findings regarding the influence of family atmosphere on the course of schizophrenia are relatively robust, it should not be forgotten that family atmosphere is also an expression of the course of the illness. Criticism and hostility tend to be expressed in families confronted with an unfavorable course. Like many of the findings discussed in the preceding section, these findings, too, are difficult to interpret because they provide no way of differentiating unambiguously between cause and effect.

4.3
Critical Life Events

A number of studies have investigated the extent to which difficult life situations may be involved in the development of schizophrenia. Steinberg and Durell (1968), for example, found a significant elevation of the incidence of schizophrenia in the months following induction into the army. The most important study in this category was that carried out by Brown and Birley (1968), which revealed that a significantly higher number of critical life events occurred in the

3 weeks preceding the onset of the illness. In the following years, a number of studies on this question were carried out in widely varying cultural environments and failed to provide an unambiguous answer (Bebbington 1995).

The fact that no clear statement can be made on this issue at present is largely due to the complexity of the object under study. Among the numerous methodological difficulties involved, we may mention just a few: first of all, there is no generally accepted definition of a critical life event. Critical life events are significant only in the individual context. Furthermore, as is known from studies in social psychology, possible psychosocial stress factors tend to be overemphasized in retrospective explanatory approaches, e.g. the attempt to explain what might have triggered the onset of schizophrenia. Moreover, the perception of such events and factors may already be distorted by the illness itself (Rössler and Lackus 1986). Finally, it must be considered that a possible clustering of critical life events shortly before the onset of disease might itself be produced by the disease in the offing.

5 Discussion

The scientific discourse of the 1950s and 1960s was mainly characterized by attempts to provide psycho- and sociodynamic explanations for the onset and course of schizophrenia. It was recognized that schizophrenia was most likely due to abnormal biological functions of some kind, but, on the whole, the possible variables of biological factors at work were, instead, devalued by the common designation of all such factors as "constitutional." The term "psychosocial" today is at risk of suffering the same fate. In current scientific discourse, psychosocial factors are often relegated to the status of a single variable of merely peripheral interest, without further specification. Neuroscience has, indeed, made enormous progress toward an understanding of the biological causes of schizophrenia, but it should not be forgotten that research in the social sciences has also yielded a considerable body of knowledge (even if, in part, contradictory) with respect to the psychosocial factors that influence it.

As mentioned above, the complexity of the social environment has often led researchers to resort to the use of socio-structural indicators. Such indicators as the percentage of unmarried or unemployed individuals, those raising children alone, or those receiving disability compensation, or the classification of the population by age, sex, and so forth, are routinely measured and used as bases for extrapolation (Rössler and Salize 1996). Figures of these types are quantitatively reliable, yet they qualify for modified interpretation only. Future research in the social sciences on schizophrenia will face the challenge of again finding more direct means of evaluating the social environment. In this context, Eaton and Harrison (1998) have suggested that a project might be undertaken to classify the human environment systematically, much as the Human Genome Project aims to catalogue all human genes.

Lastly, it should be mentioned that the psychosocial environment plays a paramount role in the practical care of schizophrenic patients. Most of the rehabilitative treatment approaches in use today are intended to influence the psychosocial environment in some way, be it by the elevation of the vulnerability threshold in a demanding environment or by providing for an appropriate living and working environment in which the individual is subject to less stress (Rössler, in press).

In a value system according the highest priority to the personal freedom and autonomy of affected individuals, these individuals themselves have an important voice in determining how their living environment is to be organized. A constructive dialogue with the affected individuals requires a still further deepening of our knowledge of the psychosocial factors influencing schizophrenia.

6 References

Bateson G (1972) Steps to an ecology of the mind. Paladin, London

Bateson G, Jackson D, Haley J, Weakland J (1956) Towards a theory of schizophrenia. Behav Sci 1: 251–264

Bebbington L (1995) The content and context of compliance. Int Clin Psychopharmacol 9[Suppl 5]: 41–50

Bebbington P, Kuipers L (1994) The predictive utility of expressed emotion in schizophrenia. Psychol Med 24: 707–718

Benedetti G (1995) Die Bleulersche Tradition der Schizophrenielehre und das Burghölzli als Stätte der Psychotherapie bei Schizophrenen. Schw Arch Neurol Psychiatr 146: 195–199

Bhugra D, Leff J, Mallett R, Der G, Corridan B, Rudge S (1997) Incidence and outcome of schizophrenia in Whites, African-Caribbeans and Asians in London. Psych Med 27: 791–798

Bleuler E (1911) Dementia Praecox oder Gruppe der Schizophrenien. Deuticke, Leipzig

Bleuler M (1972) Die schizophrenen Geistesstörungen im Lichte langjähriger Kranken- und Familiengeschichten. Thieme, Stuttgart

Brown G (1959) Experiences of discharged chronic schizophrenic mental hospital patients in various types of living group. Milbank Memorial Fund Q 37: 105–131

Brown G, Birley J (1968) Crisis and life changes and the onset of schizophrenia. J Health Soc Behav 9: 203–214

Brown G, Monck E, Carstairs G, Wing J (1962) Influence of family life on the course of schizophrenic illness. Br J Prev Soc Med 16: 55–68

Cannon T, Mednick S (1993) The schizophrenia high-risk project in Copenhagen: three decades of progress. Acta Psychiatr Scand 370[Suppl]: 33–47

Cannon T, Mednick S, Parnas J, Chulsinger F, Praesthol J, Vestergaard A (1994) Development of brain abnormalities in the offspring of schizophrenic mothers. II. Structural brain characteristics of schizophrenia and schizotypal personality disorder. Arch Gen Psychiatry 51: 955–962

Castle I, Gittus E (1957) The distribution of social defects in Liverpool. Sociol Rev 5: 43–64

Clark R (1948) The relationship of schizophrenia to occupational income and occupational prestige. Am Sociol Rev 13: 325–330

Cornblatt B, Obuchowski M (1997) Update of high-risk research: 1987–1997. Int Rev Psychiatry 9: 347–447

Davies S, Thornicroft G, Leese M, Higgingbotham A, Phelan M (1995) Ethnic differences in risk of compulsory psychiatric admission among representative cases of psychosis in London. Br Med J 312: 533–537

DeLisi L (1999) A critical overview of recent investigations into the genetics of schizophrenia. Curr Opin Psychiatry 12: 29–39

Dohrenwend B, Dohrenwend B (1969) Social status and psychological disorder: a causal inquiry. Wiley, New York

Dohrenwend BP, Dohrenwend BS (1974) Psychiatric disorders in urban settings. In: Arieti S, Caplan G (eds) American handbook of psychiatry, 2nd edn. Basic Books, New York

Eaton W (1974) Residence, social class, and schizophrenia. J Health Soc Behav 15: 289–299

Eaton W, Harrison G (1998) Epidemiology and social aspects of the human environment. Curr Opin Psychiatry 11(2): 165–168

Eaton W, Day R, Kramer M (1988) The use of epidemiology for risk factor research in schizophrenia: an overview and methodologic critique. In: Tsuang M, Simpson J (eds) Handbook of schizophrenia, vol 3. Elsevier, Amsterdam, pp 169–204

Faris R, Dunham H (1939) Mental disorders in urban areas. University of Chicago Press, Chicago

Giggs J (1986) The distribution of schizophrenics in Nottingham. Trans Int Br Geogr 59: 55–76

Goldberg E, Morrison S (1963) Schizophrenia and social class. Br J Psychiatry 109: 785–802

Gottesman I (1991) Schizophrenia genesis. The origin of madness. Freeman, New York

Häfner H (1980) Psychiatrische Morbidität von Gastarbeitern in Mannheim – Epidemiologische Analyse einer Inanspruchnahmepopulation. Nervenarzt 51: 672–683

Häfner H (1992) The epidemiology of schizophrenia. Triangle 31(4): 133–154

Häfner H, Reimann H (1970) Spatial distribution of mental disorders in Mannheim. In: Hare E, Wing J (eds) Psychiatric epidemiology. Oxford University Press, London, pp 341–354

Häfner H, Maurer K, Löffler W et al (1998) The ABC schizophrenia study: a preliminary overview of the results. Soc Psychiatry Psychiatr Epidemiol 330: 380–386

Hare E (1956) Family settings and the distribution of schizophrenia. J Ment Sci 102: 753–760

Hickling F, Rodgers-Johnson P (1995) The incidence of first contact schizophrenia in Jamaica. Br J Psychiatry 167: 193–196

Hollingshead A, Redlich F (1958) Social class and mental illness. Wiley, New York

Hutchinson G, Takei D, Bhugra T, Fahy A, Gilvary C, Mallett R (1997) Increased rate of psychosis among African-Caribbeans in Britain is not due to an excess of pregnancy and birth complications. Br J Psychiatry 171: 145–147

Jablensky A, Sartorius N, Ernberg G et al (1992) Schizophrenia: manifestations, incidence and course in different cultures – a World Health Organization ten-country study. Cambridge University Press, Cambridge

Kraepelin E (1896) Psychiatrie. Barth, Leipzig

Lam D (1991) Psychosocial family intervention in schizophrenia: a review of empirical studies. Psychol Med 21: 423–441

Leff J (1987) A model of schizophrenic vulnerability to environmental factors. In: Häfner H, Gattaz W, Janzarik W (eds) Search for the causes of schizophrenia. Springer, Berlin Heidelberg New York, pp 317–330

Leff J, Vaughn C (1985) Expressed emotion in families. Guilford, New York

Lidz T (1975) The origin and treatment of schizophrenic disorders. Hutchinson, London

Malmberg A, Lewis G, David A, Allebeck P (1998) Premorbid adjustment and personality in people with schizophrenia. Br J Psychiatry 172: 308–313

Marcelis M, Navarro-Mateu F, Murray R, Selten J, van Os J (1998) Urbanization and psychosis: a study of 1942–1978 birth cohorts in the Netherlands. Psychol Med 28: 871–879

Marneros A, Deister A, Rhode A (1991) Affektive, schizoaffektive und schizophrene Psychosen. Eine vergleichende Langzeitstudie. Springer, Berlin Heidelberg New York

Mortensen PB, Pedersen CB, Westergaard T et al (1999) Effects of family history and place and season of birth on the risk of schizophrenia. N Engl J Med 340: 603–608

Murphy H, Raman A (1971) The chronicity of schizophrenia in indigenous tropical peoples. Br J Psychiatry 118: 489–497

Myers J, Bean L (1968) A decade later: A follow-up study of social class and mental illness. Wiley, New York

Nuechterlein K (1987) Vulnerability models of schizophrenia: state of the art. In: Häfner H, Gattaz W, Janzarik W (eds) Search for the causes of schizophrenia. Springer, Berlin Heidelberg New York, pp 297–316

Ødegaard Ø (1932) Emigration and insanity: a study of mental disease among the Norwegianborn population of Minnesota. Levin and Munksgaards, Copenhagen

Ødegaard Ø (1971) Hospitalized psychoses in Norway; time trends 1926–1965. Soc Psychiatry 6: 53–58

Pfeiffer W (1994) Transkulturelle Psychiatrie, 2nd edn. Thieme, Stuttgart

Rössler W (1992) Wilhelm Griesinger und die gemeindenahe Versorgung. Nervenarzt 63: 257–261

Rössler W. Rehabilitation techniques. In: Gelder M, Lopez-Ibor J, Andreasen N (eds) New Oxford textbook of psychiatry. Oxford University Press, Oxford (in press)

Rössler W, Lackus B (1986) Cognitive disorders in schizophrenics viewed from the attribution theory. Eur Arch Psychiatry Neurol Sci 235(6): 382–387

Rössler W, Salize H (1996) Die psychiatrische Versorgung chronisch psychisch Kranker. Daten, Fakten, Analysen. Nomos, Baden-Baden

Sartorius N, Shapiro R, Kimura M, Barett K (1972) WHO international pilot study of schizophrenia. preliminary communication. Psychol Med 2: 422–425

Sartorius N, Jablensky A, Ernberg G, Leff JAK, Gulbinat WH (1987) Course of schizophrenia in different countries. Some results of a WHO comparative 5-year follow-up study. In:

Häfner H, Gattaz W, Janzarik W (eds) Search for the causes of schizophrenia. Springer, Berlin Heidelberg New York, pp 107–113

Sartorius N, Gulbinat W, Harrison G, Laska E, Siegel C (1996) Long-term follow-up of schizophrenia in 16 countries. A description of the international study of schizophrenia conducted by the World Health Organization. Soc Psychiatry Psychiatr Epidemiol 31: 249–258

Selten J, Slaets J, Kahn R (1997) Schizophrenia in Surinamese and Dutch Antillean immigrants to The Netherlands: evidence of an increased incidence. Psychol Med 27: 807–811

Shapiro S, Skinner E, Kessler L et al (1984) Utilization of health and mental health services. Three epidemiologic catchment area sites. Arch Gen Psychiatry 41: 971–978

Stein L (1957) Social class gradient in schizophrenia. Br J Prev Soc Med 11: 181–195

Steinberg H, Durell J (1968) A stressful social situation as a precipitant of schizophrenic symptoms: an epidemiological study. Br J Psychiatry 114: 1097–1105

Strauss JS, Hafez H, Lieberman P, Harding CM (1985) The course of psychiatric disorder. III. Longitudinal principles. Am J Psychiatry 142: 289–296

Tienari P, Lahti I, Sorri A et al (1989) The Finnish adoptive family study of schizophrenia: possible joint effects of genetic vulnerability and family environment. Br J Med Suppl 5: 29–32

Tienari P, Wynne LC, Moring J et al (1994) The Finnish adoptive study of schizophrenia. Implications for family research. Br J Psychiatry 164[Suppl 23]: 20–26

Torrey J (1980) Schizophrenia and civilisation. Aronson, New York

Torrey E, Bowler A, Clark K (1997) Urban birth and residence as risk factors for psychoses: an analysis of 1880 data. Schizophr Res 25: 169–176

Waxler N (1979) Is outcome for schizophrenia better in non-industrial societies? The case of Sri Lanka. J Nerv Ment Dis 167: 144–158

Weyerer S, Häfner H (1989) The stability of the ecological distribution of the incidence of treated mental disorder in the city of Mannheim. Soc Psychiatry Psychiatr Epidemiol 24: 57–62

WHO (1974) International pilot study of schizophrenia. World Health Organization, Geneva

WHO (1975) Schizophrenia: a multi-national study. Summary of the initial evaluation phase of the international pilot study of schizophrenia. World Health Organization, Geneva

WHO (1979) Schizophrenia: an international follow-up study. Wiley, Chichester

Wing J, Brown G (1970) Institutionalism and schizophrenia. Cambridge University Press, London

Wing J, Freudenberg R (1961) The response of severely ill chronic schizophrenic patients to social stimulation. Am J Psychiatry 118: 311–322

Wynne L, Singer M (1966) Communication styles in parents of normals, neurotics and schizophrenics. Psychiatr Res Rep 20: 25–38

Zubin J, Spring B (1977) Vulnerability – a new view of schizophrenia. J Abnorm Psychol 86: 103–126

CHAPTER

10

W. Gaebel

General Principles of the Treatment of Schizophrenic Disorders

Translator: E. Taub

1 Introduction

The treatment of patients with schizophrenic disorders is based on clear, empirically derived principles. Guidelines for treatment may be found in a number of published works (see below; Gaebel, in press). In countries with a structured health care sector, schizophrenic patients are treated in a system comprising multiple components. The psychiatrist is not necessarily the caregiver of first resort for such patients, but is, as a practically universal rule, consulted as a specialist – at least upon the initial presentation of the illness – and directs further treatment measures, in whatever component of the mental health care system they may take place. In the context of multidisciplinary treatment, the psychiatrist bears the ultimate responsibility for the development and implementation of an overall treatment plan.

2 Therapeutically Relevant Aspects of Disease Course

Schizophrenic disorders are generally characterized by a lifelong course, with the exception of rare cases of monoepisodic disease. E. Bleuler (1911) drew attention to the heterogeneity of the manifestations and course of schizophrenia by referring to the disease as "the group of schizophrenias" – in contrast to Kraepelin (1896), for whom "dementia praecox" was characterized by a uniform, unfavorable course.

Types of schizophrenia are diagnostically differentiated and encoded in the ICD-10 system according to the acuity of onset (acute, insidious, primary chronic), the pattern of acute episodes (phasic-remitting, episodic with stable or increasing residua), and the long-term outcome (e.g. remission, persisting positive and/or negative manifestations, psychosocial disability). Both the pattern of acute episodes and the long-term trend of disease course, which becomes evident after multiple episodes have occurred, display considerable intercultural as well as inter- and intraindividual variability. This fact is regarded as evidence for the psychosocial plasticity of the disease course and against its being purely biologically determined. A number of longitudinal studies have shown that the long-term course of schizophrenia is rather favorable (e.g. M. Bleuler et al. 1976; Tsuang et al. 1979; Harding et al. 1987).

2.1 Phases

With the exception of primarily continuous forms of schizophrenia, the course of the disease can generally be subdivided into the following phases, each of which requires specific therapy (APA 1997):

- The *acute decompensation phase* (lasting weeks or months) begins with nonspecific prodromal manifestations and/or specifically psychotic manifestations, sometimes with endangerment of the patient or others, which resolve more or less completely under treatment.
- The *postacute stabilization phase* (approx. 3–6 months) is characterized by further regression of positive manifestations and psychopathological stabilization, but negative manifestations and cognitive and sociocommunicative deficits often persist, and there is an increased likelihood of relapse.
- The stable (partial) *remission phase* (months or years) is characterized by the following (depending on the type of the disease course): disappearance or stable residua of positive manifestations, negative manifestations, and stable or increasingly impaired social competence, corresponding to the premorbid level.

2.2 Stages

The sequence of individual phases constitutes the overall disease course, which is itself divided into therapeutically relevant stages:

In the first 5–10 years of illness, there is an increased frequency of relapses, suicide attempts, and suicides (M. Bleuler et al. 1976; McGlashan 1991). This stage is characterized by frequent inpatient hospitalizations, in which the course is set for rehabilitative efforts, in accordance with the patient's overall prognosis, life situation, and coping ability.

After these turbulent years, a plateau phase is often reached in which relapses are less common and a certain degree of social stabilization takes place (M. Bleuler et al. 1976; McGlashan 1991).

2.3 Treatment and Course

The phenomenology of schizophrenic disorders is divided into positive, negative, and social manifesta-

tions (Strauss et al. 1974). Rehabilitative efforts are directed against so-called impairments, disabilities, and handicaps. This differentiated view of the disease necessitates a multidimensional concept for the assessment of treatment outcome. Thus, alongside the clinical dimension, rehabilitative and humanitarian dimensions and a dimension of public safety may be distinguished (McGlashan 1994). Individual aspects of disease course, such as disease manifestations, role function, relapse rate, and quality of life, are only moderately related. The treatment of schizophrenic disorders must be phase- and stage-specific and must take account of the varying objectives of treatment over the course of disease in the individual patient.

Several studies indicate that acute episodes respond less well to treatment the later it is instituted (Wyatt 1991; Loebel et al. 1992). It follows that treatment should be begun as early as possible, even though this does not seem to have a major influence on overall prognosis.

In cases of schizophrenia with a relapsing course, a decline in the therapeutic responsiveness of acute episodes to neuroleptics has been observed (Lieberman 1993). It has been speculated that repeated episodes of disease may exert a "toxic" or "kindling"-like effect and thus pave the way for further episodes that are less responsive to therapy.

2.4
Predictors of Course and Therapeutic Response

Many different aspects of schizophrenic patients have been investigated as possible predictors of the course of disease and of therapeutic responsiveness (Gaebel and Awad 1994). A well-developed premorbid personality, the presence of provoking factors, acute onset of illness, accompanying affective manifestations, absence of impoverishment of affect, and a psychologically favorable life situation have long been considered generally favorable prognostic signs (Langfeldt 1937). More recently, further characteristics of patients, of the disease, and of the environment (sometimes quantified in prognostic scales) and, increasingly, biological factors have also been extensively studied (Awad 1994; Möller 1994; Lieberman 1994). To summarize the results of these studies, it may be stated that no single feature or combination of features yet enables a reliable prediction upon which differential indications for therapy can be based in individual cases.

3
Diagnosis and Treatment

3.1
Diagnostic Concepts

Operational systems such as ICD-10 and DSM-IV can be used for the diagnosis of schizophrenic disorders. These systems require the presence or absence of certain manifestations, a deterioration of social functioning, and a specified minimum duration of illness as prerequisites for the diagnosis of schizophrenia. An organic etiology must be ruled out for the diagnosis to be made (in spite of the accumulating findings of biological abnormalities in schizophrenia). Additional diagnostic procedures must also be used. Drug screening and clinical laboratory testing, including complete blood count and urinalysis, are recommended in all cases. Pregnancy tests, electrocardiography (ECG), electroencephalography (EEG), computed tomography or magnetic resonance imaging, and neuropsychological or general psychological examination are recommended only if special questions are to be answered (Frances et al. 1996). The descriptive diagnostic criteria do not yield differential indications for therapy.

A biopsychosocial systems approach to diagnosis and treatment is appropriate for the development of a multidimensional model (Engel 1980). Such an approach affords a hypothetical framework in which potential biological and psychosocial causative factors and corresponding therapeutic measures may be critically considered.

The vulnerability-stress model (Nuechterlein 1987) goes a step further and posits hypothetical relationships between an elevated vulnerability to disease and factors that promote or inhibit the appearance of disease manifestations. It represents a heuristic framework in which vulnerability, stressors, and protective factors may be targets for therapeutic intervention.

3.2
Therapeutic Concepts

3.2.1 General Objectives of Treatment

The objective of treatment is to free patients from the manifestations of disease and the accompanying disabilities to the extent that they can lead their lives as they themselves see fit, with minimal dependence on clinical institutions (Helmchen 1978). Early, acute treatment of exacerbations and consistently maintained prophylaxis against relapses are the essential prerequisites for stable remission and may help

prevent the disease from becoming chronic. The major types of intervention are intended to reduce disease manifestations, vulnerability factors, and stressors and to strengthen coping strategies. Interventions intended to stabilize primarily or secondarily affected psychosocial functions round out the treatment program.

3.2.2 (Re-)Assessment and Monitoring

The time-consuming diagnostic process that is undertaken upon the initial manifestation of disease generally need not be repeated later, as long as no uncertainty exists about the diagnosis. Nonetheless, the various dimensions of disease course must be assessed at regular intervals so that the treatment can be correspondingly adjusted. The use of standardized instruments of assessment is sometimes beneficial. If patients do not respond to antischizophrenic medication, their compliance should be tested, e.g. by determining plasma levels, and their motivation for possible noncompliance should be clarified, above all by consideration of side effects.

The fine-grained (daily) monitoring of patient status performed in the acute phase can be reduced to weekly or monthly visits to the physician in the postacute phase; later, in the stable phase, some patients may see the physician as rarely as once every 3–6 months. The most important factor is the availability of a caregiver to the patient and the family, in case an acute need should arise.

3.2.3 Specific Therapeutic Interventions

The treatment of schizophrenic disorders has a multidimensional orientation. This means that biological/somatic, psychological/psychotherapeutic, and sociotherapeutic/rehabilitative aspects must all be considered whenever treatment and supportive care are offered. Different aspects may receive different emphases in the individual case, depending on the particular phase and stage of the disease.

The prevention or elimination of positive manifestations is the major objective of neuroleptic treatment, but negative manifestations, too, respond partially to such treatment (Carpenter et al. 1985; Goldberg 1985). No primary effects of neuroleptic treatment are expected on the social axis. Although a possible negative effect of classical neuroleptic treatment on the social dimension cannot be ruled out (May and Goldberg 1978), the patients' quality of life may be improved secondarily through the amelioration of disease manifestations and the prevention of relapses (Barnes et al. 1983; Awad 1992). The introduction of newer, atypical antipsychotic agents with more favorable side effect profiles has allowed still further improvement of these patients' quality of life and increased acceptance of pharmacotherapy.

The effectiveness of neuroleptic agents in the acute and long-term treatment of schizophrenic psychoses has been documented beyond doubt (Davis et al. 1980). Remission of manifestations and prevention of relapse can be achieved in approximately 70% of patients undergoing treatment in accordance with accepted recommendations. The monthly spontaneous relapse rate is 10% under placebo treatment and falls to approximately 3% under neuroleptic treatment (Davis 1985). Even though the spontaneous relapse rate tends to decrease over time, there is a lasting, significant difference between neuroleptic treatment and placebo (Hogarty and Ulrich 1977). Recommendations and guidelines exist for both acute and long-term neuroleptic treatment (Kissling 1991; Kane and Marder 1993; Gaebel and Marder 1996; Frances et al. 1996; APA 1997; DGPPN 1998; Lehman et al. 1998a). In long-term treatment, maintenance pharmacotherapy at the lowest possible neuroleptic dose, i.e. sufficient to suppress schizophrenic manifestations with minimal side effects in the individual case, is the preferred strategy, rather than intermittent neuroleptic treatment with early intervention (Schooler 1991; Pietzcker et al. 1993).

The options for treatment and the prognosis under long-term treatment are limited by a number of factors. As many as 50% of outpatients undergoing long-term pharmacotherapy fail to achieve drug levels in the therapeutic range because of inadequate compliance (Johnson 1984). Tardive dyskinesia in 10%–15% of patients (Gaebel 1993), inadequate response to treatment in 20%–30%, and an equally high rate of response to placebo (Hogarty et al. 1974) must all be taken into account in the risk–benefit analysis of long-term treatment. Currently available criteria for the indications for long-term treatment in the individual case are not reliable.

Antidepressant or phase-prophylactic agents such as lithium, carbamazepine, and valproate may be used in combination with neuroleptic agents, when indicated.

Pharmacotherapy is used in the framework of psychoeducational treatment, in which the causes and consequences of the disease, potential stressors (in the sense of the vulnerability-stress model), and possibilities for treatment are explained to patients and their families in a manner that will promote their cooperation. Indications for psychodynamically oriented individual psychotherapy are rare in the present-day treatment of schizophrenia (Fenton and Cole 1995).

The published results of this form of treatment are unconvincing, and it further harbors the risk of provoking disease manifestations. It should therefore be considered only in individual cases, in patients in

stable remission, with a solid therapeutic alliance, good drug compliance, and the capacity and motivation for introspective work. The goals of psychodynamic treatment are to improve control of self-destructive behavior and to promote emotional maturity (APA 1997).

Psychotherapy of schizophrenic patients must take account of the biological etiology of the disease and thereby focus on helping patients cope with the disease and its consequences (acceptance of chronic illness, self-management, problem-solving) (Coursey 1989). Essential elements of psychotherapy include psychoeducation, crisis intervention, support, practical counseling, and approaches to the normalization of inappropriate reactions related to illness; all of these must be carried out with due consideration of the patient's (neuro)psychological and sociocommunicative deficits. At present, these elements are usually incorporated into group therapy techniques, which may also include interactive elements, depending on individual condition (Fenton and Cole 1995).

It has been recognized, at least since the studies by Hogarty et al. (1974), that the combination of long-term neuroleptic treatment with psychosociotherapeutic techniques can further reduce the rate of relapse and favorably influence several aspects of long-term outcome (e.g. cognition, social adaptation).

Since then, a number of psychosocial treatment and rehabilitation techniques have been evaluated in individual, group, and family therapy interventions (Bellack and Mueser 1993; Fenton and Cole 1995), including the following:

- Family intervention
- Training of social skills
- Cognitive rehabilitation
- Coping skills training

General principles for psychosocial intervention include the following (McGlashan 1994):

- Evaluation
- Continuous reevaluation
- Timing
- Intensity
- Integration with psychopharmacotherapy

It is particularly important for the psychiatrist to consider the phase of disease and the individual patient's cognitive receptiveness and tolerance for stress when determining the timing and intensity of the psychosocial treatment and rehabilitation measures that are to be undertaken.

3.2.4 Principles of Psychiatric and Psychotherapeutic Treatment

The psychiatric and psychotherapeutic treatment ("therapeutic management") of the schizophrenic patient requires a comprehensive understanding of the patient and his or her needs, goals, intrapsychic conflicts, defense mechanisms, coping styles, and strengths (APA 1997). The psychiatrist must understand the biological, interpersonal, social, and cultural factors that influence both the course of disease and the patient's adaptive abilities. Specific components of psychiatric and psychotherapeutic treatment include the following (APA 1997):

- Establishment and maintenance of the therapeutic relationship
- Monitoring of the patient's psychiatric condition
- Education with regard to the disease and its treatment
- Establishment of indications for pharmacological and other specific treatment in the framework of an overall treatment plan
- Support of the patient in following the treatment plan
- Awakening of understanding of the psychosocial consequences of the disease and how they may be dealt with
- Early detection and treatment of new episodes of disease, including consideration of provocative or maintaining factors
- Measures to reduce family stress and to improve family interaction
- Facilitation of access to the care system, including coordination of resources in the mental and general health care sectors

3.2.5 Treatment Settings

The above-mentioned therapeutic techniques may be applied in a number of therapeutic settings. The German mental health care system possesses multiple components, each of which may be the appropriate setting for the treatment of a specific subset of schizophrenic patients, depending on disease phase, stage of progression, and prognosis. The available treatment settings include the following:

- Inpatient care (psychiatric wards, psychiatric hospitals, training hospitals)
- Partial hospitalization (day wards, night wards)
- Outpatient care (local general practitioners, psychiatrists, and neurologists; outpatient departments)
- Complementary settings (permanent residential homes, halfway houses, protected living arrangements, rehabilitation facilities)

These different treatment settings must be offered in a functionally integrated fashion, so that the patient can always receive treatment in the optimal setting, in accordance with changing individual needs (see below).

4
Principles of Phase-Specific Treatment

In this section, general principles of treatment will be discussed in relation to the individual phases of the course of schizophrenia.

4.1
Acute Phase

The primary objectives of treatment in the acute phase are as follows:

- Elimination of positive manifestations and prophylaxis against their recurrence
- Prevention of harm to the patient and others
- Preparation for the postacute stabilization phase

Treatment in the acute phase is intended not only to achieve remission of the acute psychopathological manifestations, but also to restore or improve the patient's ability to function in a social role on the highest possible level.

A complete diagnostic assessment is performed on initial presentation. If relapse occurs, the diagnosis should be reevaluated. The acute course under treatment should be monitored frequently. When the acute phase has passed, possible persistent cognitive and social deficits and ensuing needs for rehabilitation should be evaluated, and the appropriate measures taken.

Pharmacotherapy with neuroleptic agents is the first line of treatment in the acute phase, sometimes in combination with other drugs (e.g. benzodiazepines, antidepressants, antimanic agents). Electroconvulsive therapy may be indicated in patients with catatonia, accompanying severe depressive manifestations, or resistance to treatment (including with clozapine). The treatment continues until the end of the acute phase and the beginning of the postacute stabilization phase. Psychosocial techniques involving the patient's family, particularly psychoeducational techniques, are used as the situation becomes less acute.

The objectives of acute-phase treatment listed above can be achieved only after a therapeutic alliance has been established with the patient and his or her family. Short-term and medium-term treatment plans must be developed, and consideration must be taken of the modes of therapy and institutional settings that will be most suitable for future treatment.

Inpatient treatment plays an especially important role at the time of the initial presentation. Danger to the patient or others is an absolute indication for hospitalization. If the patient is not competent to give consent to hospitalization, the physician must act in accordance with the local law of involuntary commitment. In the case of later relapse, an attempt can be made to treat the patient in an outpatient setting. The choice between inpatient and outpatient treatment must be made in consideration of the patient's ability to cooperate with outpatient treatment and of the relatives' ability to accept the burden that it will place on them.

4.2
Postacute Stabilization Phase

The principal objectives of therapy in this phase are as follows:

- Achievement of stable remission
- Elimination of delusions and promotion of insight into the illness
- Elimination of disease-related deficits
- Education about concepts of the disease and its treatment
- Assurance of compliance with treatment
- Early recognition of threatened relapses
- Development of individual coping strategies
- Harmonization of family conflicts
- Preparation of rehabilitative measures (in the narrower sense)

In this phase, closely spaced discussions (approximately once per week, on an outpatient basis) with the treating specialist are necessary, particularly after a change of setting (see below), so that a new therapeutic alliance can be established and any changes in the patient's condition can be addressed rapidly with corresponding changes in therapy.

In the phase of postacute stabilization, pharmacotherapy is continued, the dosage is cautiously adjusted to the optimal level for the new outpatient setting, and psychoeducational interventions are carried out, as well as other nonpharmacological interventions, if needed (see above). Depot neuroleptic agents should be substituted if compliance is uncertain.

The therapeutic alliance must be maintained so that the patient can be supported throughout the process of social reintegration. Excessive "rehabilitation pressure" should be avoided, and any potential persistent risk of suicide must be recognized and openly discussed.

In this phase, inpatient stays (insofar as they were initially necessary) become progressively shorter, and treatment is often delivered in extramural settings, such as day clinics. During the transition from inpatient care to a partly or wholly outpatient setting, it must be realized that the patient remains more

irritable than normal and less able to cope with stress, and these problems may be exacerbated by changes of caregivers. Changes of setting must therefore be carefully planned and accompanied by appropriate treatment.

4.3 Remission Phase

The principal objectives of therapy in the remission phase are as follows:

- Suppression of manifestations
- Prophylaxis against relapse
- Social (re)integration

Contacts with caregivers can usually take place at less frequent intervals in this phase. Compliance with, and possible side effects of, pharmacotherapy should be assessed regularly, particularly tardive dyskinesia.

In addition to suppression of psychotic manifestations and pharmacotherapeutic prevention of relapse, supportive and rehabilitative measures are provided in this phase. The neuroleptic dose should be cautiously reduced, in steps, to the lowest possible maintenance dose. Small decreases should be made at intervals of 3–6 months in view of the latency between dosage changes and clinical changes in the remission phase. Attention must be paid to the possible appearance of prodromal manifestations to ensure that the dose does not drop below the critical threshold. If serious side effects should arise (e.g. weight gain, sexual dysfunction, tardive dyskinesia), a switch to another medication should be considered (such as an atypical neuroleptic agent or, if compliance is inadequate, a depot neuroleptic agent).

A special role is played by the repeated emergence of individual prodromal manifestations (even if these have not been shown to be definite predictors of relapse; see Gaebel et al. 1993) and of relapse behavior. As the patient stabilizes, there is greater motivation to initiate psychosocial and rehabilitative processes in case of need, and the patient's contact with his or her family is maintained.

The treatment of schizophrenic disorders in the remission phase is now usually carried out in the extramural setting, except in the case of acute reexacerbation. It has been shown that even chronic schizophrenics who have been hospitalized in psychiatric clinics for decades may be successfully "deinstitutionalized" into protected extramural living arrangements. Long-term care is thus now generally delivered on an outpatient basis, so that patients can be given the maximum degree of independence.

5 Principles of Phase-Independent Treatment

5.1 System of Care

The different forms of treatment discussed above can be applied sensibly only in a coordinated system of care. The cooperation of the patient's relatives is also indispensable.

Depending on the treatment setting, representatives of many different disciplines may participate in the care of the patient, including psychiatrists, psychologists, nurses, social workers, physical/occupational therapists, and others. The various methods of treatment and rehabilitation are applied in the context of a multidisciplinary team. The psychiatrist not only belongs to the team as one of the caregivers, but also bears the primary responsibility for establishing the treatment plan and seeing it through to its application. To do this, the psychiatrist must work alongside, advise, and supervise the other members of the patient care team.

Therapeutic continuity is an important general principle of treatment. It requires therapeutic information to be comprehensively exchanged, treatment measures to be coordinated among the members of the patient care team, and the patient to have a relationship of trust with "his" (or "her") therapist. The principle is imperfectly applied in practice because of discontinuities inherent in the system of care. On the other hand, the familiarity of the patient with a single caregiver or group of caregivers over the long term may, on occasion, be an obstacle to therapeutic innovation. If so, a change of setting or of caregiver may be indicated.

Deinstitutionalization and placement in the mainstream community are the guiding concepts of health policy with regard to the care of the mentally ill. There is controversy over the applicability of these principles to all patients (Häfner and an der Heiden 1991). Outpatient is generally preferable to inpatient treatment, whenever possible. On the other hand, for a subgroup of decompensated, homeless patients, the inpatient clinic may provide an "asylum" in the best sense and thereby constitute the more humane option, even today (McGlashan 1994).

A consequence of the trend toward shorter inpatient stays and prolonged outpatient treatment is that even acute treatment is now increasingly given on an outpatient basis, partly by mobile teams. Regardless of the setting in which acute treatment is delivered, the indicated diagnostic and therapeutic measures must still be carried out. Account must also be taken of the

family's need for an occasional period of respite, which may be provided by caring for the patient temporarily in a suitable setting.

The goal of treatment is a patient with insight into the diagnosis who has a say in determining what treatment will be provided and can consent to it of his or her own free will. In contrast to the paternalistic treatment model, this liberal model requires that the physician–patient relationship should be based on cooperation, information, and education. The "consumer perspective" deserves greater attention (Van Putten and May 1978); in other words, the patient's views and expectations (e.g. concerning the monitoring of side effects of medications) need to be taken into account. The liberal model of treatment is of limited applicability, however, when the patient's insight into the illness is transiently or permanently impaired to such an extent that the treating physician, having exhausted the available legal options, must act, for a time, on the patient's behalf.

The concept of case management is based on the observation that chronically mentally ill patients are liable to make inadequate use of the available therapeutic resources and that some work is required to coordinate these resources effectively. The case manager, usually a paraprofessional, has the task of guiding the patient through the mental health care system so that his or her needs are optimally met (Bachrach 1992). For schizophrenic disorders in particular, however, case management has not been found effective in preventing rehospitalization (Rössler et al. 1992).

5.2 Adaptive Versus Selective Indications

Two important goals are to make the validated methods of treatment more widely available and to tailor the treatment offered more closely to the needs of the patient. The latter goal has been called "subject–treatment matching" (Bellack and Mueser 1993).

5.3 Standardization of Interventions

It would be desirable for treatment to be given more often in accordance with empirically justified therapeutic guidelines. This can only occur if standardized guidelines are established and taught during residency training and continuing medical education. The various terms used in this connection should be explained: "directives" are binding rules for medical practice, "guidelines" are oriented toward providing a reference for diagnostic and therapeutic standards, and "recommendations" and "opinions" are merely of informational content, suggesting possible courses of action to the practitioner. Practice guidelines enable the practitioner to perform diagnosis and treatment in accordance with the prevailing rules of the discipline, while retaining the essential therapeutic freedom to make modifications in individual cases, whenever this is necessary or desirable.

Guidelines for the treatment of schizophrenia, based on empirical studies and a consensus of expert opinion are available (Kissling 1991; Frances et al. 1996; APA 1997; DGPPN 1998; Lehman et al. 1998a). In view of the finding that less than 50% of the current treatment of schizophrenia accords with published guidelines (Lehman et al. 1998b), the implementation of practice guidelines appears to be urgently necessary if treatment is to be optimized.

5.4 Combination Therapy

Most of the interventions discussed have been evaluated in combination with pharmacotherapy. As a rule, the success of one form of treatment is dependent on that of the other: psychosocial interventions are effective only against the background of effective pharmacotherapy. On the other hand, psychosocial interventions may improve the patient's motivation to participate in, and comply with, pharmacotherapy.

5.5 Timing and Duration of Interventions

Individual therapeutic interventions should be undertaken according to the phase of the illness. Pharmacotherapy should be instituted as early as possible in the course of the illness, before the initial manifestation or a relapse. Psychosocial interventions are mainly of use in the postacute phase.

A common feature of all interventions is that they are only effective while they are being applied. A lifelong need for treatment must therefore be assumed. Treatment must be given either continuously, as with pharmacotherapy, or in periodic installments, as with nonpharmacologic interventions.

5.6 Prevention

Potential methods for the primary prevention of schizophrenic disorders are attracting increasing attention (e.g. *British Journal of Psychiatry* 1998, vol. 172 Suppl. 33), but have not yet been adequately studied.

The major emphasis of therapy thus remains on secondary and tertiary prevention.

5.7 Quality Assurance

Awareness of the guidelines and preconditions for treatment discussed above and their promulgation in residency training and continuing medical education are necessary if schizophrenic patients are to be given the best possible treatment according to current scientific knowledge. At the institutional level, corresponding instruments must be introduced for the assessment of treatment outcomes, so that internal quality assurance and quality improvement may be carried out, on the basis of comparative analysis across institutions (Janssen et al. 1998). There is no doubt that the treatment of schizophrenic disorders would be optimized, and outcomes correspondingly improved, if these steps were taken today.

6 References

**APA (1997) Practice guideline for the treatment of patients with schizophrenia. American Psychiatric Association, Washington

Awad AG (1992) Quality of life of schizophrenic patients on medications and implications for new drug trials. Hosp Community Psychiatry 43: 262–265

Awad AG (1994) Prediction research of neuroleptic treatment outcome in schizophrenia – state of the art: 1978–1993. In: Gaebel W, Awad AG (eds) Prediction of neuroleptic treatment outcome in schizophrenia – concepts and methods. Springer, Berlin Heidelberg New York, pp 1–14

Bachrach LL (1992) Case management revisited. Hosp Community Psychiatry 43: 209–210

Barnes TRE, Milavic G, Curson DA, Platt SD (1983) Use of the social behaviour assessment schedule (SBAS) in a trial of maintenance antipsychotic therapy in schizophrenic outpatients: pimozide versus fluphenazine. Soc Psychiatry 18: 193–199

*Bellack AS, Mueser KT (1993) Psychosocial treatment of schizophrenia. Schizophr Bull 19: 317–336

*Bleuler E (1911) Dementia praecox oder Gruppe der Schizophrenien. In: Aschaffenburg G (ed) Handbuch der Psychiatrie. Deuticke, Leipzig, pp 1–420

Bleuler M, Huber G, Gross G, Schüttler R (1976) Der langfristige Verlauf schizophrener Psychosen. Nervenarzt 47: 477–481

Carpenter WT, Heinrichs DW, Alphs LD (1985) Treatment of negative symptoms. Schizophr Bull 11: 440–452

Coursey RD (1989) Psychotherapy with persons suffering from schizophrenia: the need for a new agenda. Schizophr Bull 15: 349–358

Davis JM (1985) Maintenance therapy and the natural course of schizophrenia. J Clin Psychiatry 11: 18–21

Davis JM, Schaffer CB, Killian GA, Kinard C, Chan C (1980) Important issues in the drug treatment of schizophrenia. Schizophr Bull 6: 70–87

*DGPPN (Deutsche Gesellschaft für Psychiatrie, Psychotherapie und Nervenheilkunde) (1998) Praxisleitlinien in Psychiatrie und Psychotherapie, vol 1. Behandlungsleitlinie Schizophrenie. Steinkopff, Darmstadt

Engel GL (1980) The clinical application of the biopsycho-social model. Am J Psychiatry 137: 535–544

Fenton WS, Cole SA (1995) Psychosocial therapies of schizophrenia: individual, group, and family. In: Gabbard GO (ed) Treatments of psychiatric disorders, vol I, 2nd edn. American Psychiatric Press, Washington DC, pp 987–1018

Frances A, Docherty JP, Kahn DA (1996) The expert consensus guideline series. Treatment of schizophrenia. J Clin Psychiatry 57[Suppl 12B]: 1–58

Gaebel W (1993) Tardive Dyskinesien unter Neuroleptika-Behandlung. Dtsch Ärzteblatt 90: 1041–1046

Gaebel W (in press) Internationale Leitlinien der Schizophreniebehandlung. In: Möller HJ, Müller N (eds) Behandlung mit atypischen Neuroleptika. Springer, Berlin Heidelberg New York

*Gaebel W, Awad AG (eds) (1994) Prediction of neuroleptic treatment outcome in schizophrenia – concepts and methods. Springer, Berlin Heidelberg New York

Gaebel W, Marder S (1996) Conclusions and treatment recommendations for the acute episode in schizophrenia. Int Clin Psychopharmacol 11[Suppl 2]: 93–100

Gaebel W, Frick U, Köpcke W et al (1993) Early neuroleptic intervention in schizophrenia: are prodromal symptoms valid predictors of relapse? Br J Psychiatry 163[Suppl 21]: 8–12

Goldberg SC (1985) Negative and deficit symptoms in schizophrenia do respond to neuroleptics. Schizophr Bull 11: 453–456

Häfner H, an der Heiden W (1991) Evaluating effectiveness and cost of community care for schizophrenic patients. Schizophr Bull 17: 441–451

Harding CM, Brooks GW, Ashikaga T et al (1987) The Vermont longitudinal study of persons with severe mental illness. II: long-term outcome of subjects who retrospectively met DSM-III criteria for schizophrenia. Am J Psychiatry 144: 727–735

Helmchen H (1978) Forschungsaufgaben bei psychiatrischer Langzeitmedikation. Nervenarzt 49: 534–538

Hogarty GE, Ulrich RF (1977) Temporal effects of drug and placebo in delaying relapse in schizophrenia. Arch Gen Psychiatry 36: 585–590

Hogarty GE, Goldberg SC, Schooler NR, Ulrich RF (1974) Drug and sociotherapy in the aftercare of schizophrenic patients: two year relapse rates. Arch Gen Psychiatry 31: 603–608

Janssen B, Burgmann C, Held T et al (1998) Qualitätsindikatoren der stationären Behandlung schizophrener Patienten. Ergebnisse einer Pilotstudie zur externen Qualitätssicherung mit Hilfe einer Tracer-Diagnose. Psychiatr Prax 25: 303–309

Johnson DAW (1984) Observations on the use of long-acting depot neuroleptic injections in the maintenance therapy of schizophrenia. J Clin Psychiatry 45: 13–21

Kane JM, Marder SR (1993) Psychopharmacologic treatment of schizophrenia. Schizophr Bull 19: 87–302

Kissling W (ed) (1991) Guidelines for relapse prevention in schizophrenia. Springer, Berlin Heidelberg New York

*Kraepelin E (1896) Lehrbuch der Psychiatrie. Barth, Leipzig

Langfeldt G (1937) The prognosis in schizophrenia and the factors influencing the course of the disease. Munksgaard, Copenhagen

Lehman AF, Steinwachs DM, the Survey Co-investigators of the PORT Project (1998a) Translating research into practice: the Schizophrenia Patient Outcomes Research Team (PORT) treatment recommendations. Schizophr Bull 24: 1–10

Lehman AF, Steinwachs DM, the Survey Co-investigators of the PORT Project (1998b) Patterns of usual care for schizophrenia: initial results from the Schizophrenia Patient Outcomes Research Team (PORT) client survey. Schizophr Bull 24: 11–20

Lieberman JA (1993) Prediction of outcome in first-episode schizophrenia. J Clin Psychiatry 54[Suppl 3]: 13–17

Lieberman JA (1994) Predictors of outcome in schizophrenia: the concept of time. In: Gaebel W, Awad AG (eds) Prediction of neuroleptic treatment outcome in schizophrenia – concepts and methods. Springer, Berlin Heidelberg New York, pp 43–49

Loebel AD, Lieberman JA, Alvir JMJ, Mayerhoff DJ, Geisler SH, Szymanski SR (1992) Duration of psychosis and outcome in first-episode-schizophrenia. Am J Psychiatry 149: 1183–1188

May PRA, Goldberg SC (1978) Prediction of schizophrenic patients' response to pharmacotherapy. In: Lipton MA, Dimascio A, Killam KF (eds) Psychopharmacology: a generation of progress. Raven, New York, pp 1139–1153

McGlashan TH (1991) Selective review of recent North American long-term follow-up studies of schizophrenia. In: Mirin SM, Gossett JT, Grob MC (eds) Psychiatric treatment: advances in outcome research. American Psychiatric Press, Washington DC, pp 61–105

McGlashan TH (1994) Psychosocial treatment of schizophrenia. The potential of relationships. In: Andreasen NC (ed) Schizophrenia. From mind to molecule. American Psychiatric Press, Washington DC, pp 189–215

Möller HJ (1994) General aspects of predictor research in schizophrenia and depression. In: Gaebel W, Awad AG (eds) Prediction of neuroleptic treatment outcome in schizophrenia – concepts and methods. Springer, Berlin Heidelberg New York, pp 27–36

*Nuechterlein KH (1987) Vulnerability models for schizophrenia: state of the art. In: Häfner H, Janzarik KW, Gattaz W (eds) Search for the causes of schizophrenia. Springer, Berlin Heidelberg New York, pp 297–316

Pietzcker A, Gaebel W, Köpke W, Linden M, Müller P, Müller-Spahn F, Tegeler J (1993) Continuous vs intermittent neuroleptic longterm treatment in schizophrenia – results of a German multicenter study. J Psychiatr Res 27: 321–339

Rössler W, Löffler W, Fätkenheuer B, Riecher-Rössler A (1992) Does case management reduce the rehospitalization rate? Acta Psychiatr Scand 86: 445–449

*Schooler NR (1991) Maintenance medication for schizophrenia: strategies for dose reduction. Schizophr Bull 17: 311–324

Strauss JS, Carpenter WT (1974) The prediction of outcome in schizophrenia. II. Relationships between predictor and outcome variables. Arch Gen Psychiatry 31: 37–42

Tsuang MT, Woolson RF, Fleming JA (1979) Long-term outcome of major psychosis. I. Schizophrenia and affective disorders compared with psychiatrically symptom-free surgical conditions. Arch Gen Psychiatry 39: 1295–1301

Van Putten T, May PRA (1978) Subjective response as a predictor of outcome in pharmacotherapy. Arch Gen Psychiatry 35: 477–480

Wyatt RJ (1991) Neuroleptics and the natural course of schizophrenia. Schizophr Bull 17: 325–351

CHAPTER

11

W.W. Fleischhacker

Drug Treatment of Patients with Schizophrenia

This is a modified and updated version of the chapter published in Maj and Sartorius (1999).

1 Introduction

Ever since antipsychotic drug treatment was introduced into clinical psychiatry almost half a century ago, establishing a breakthrough in the management of schizophrenic disorders, this new therapeutic area has been characterized by a continuing attempt to optimize the results of treatment efforts for these patients. A plethora of medications, called neuroleptics in classical terminology, a term increasingly replaced by the indication-driven word antipsychotics, has been synthesized and tested to this end. Molecules of different chemical structures, ranging from tricyclic phenothiazines to thioxanthenes, butyrophenones, dibenzazepines, substituted benzamides and benzisoxazole derivatives are now used in the treatment of schizophrenia. The development of clozapine was clearly a quantum leap in these efforts. This drug, available in some European countries since the early 1970s and introduced in the United States in the late 1980s, was the first antipsychotic to effectively treat the symptoms of schizophrenia with only a minimal risk of inducing extrapyramidal motor side-effects (EPS) (Fitton and Heel 1990; Kurz et al. 1995a). In addition to changing the mindset concerning clinical efficacy and adverse drug effects, the success of clozapine has also considerably influenced preclinical development strategies for new antipsychotics. By and large, this was fuelled by the recognition that clozapine has excellent antipsychotic efficacy without blocking nigrostriatal dopamine (D_2) receptors to a similar extent as classical neuroleptics (Farde et al. 1992). Up to this point, powerful D_2-blockade was seen as a prerequisite for an antipsychotic effect. Eventually, as a consequence of clozapine's differing pharmacological profile, a number of new points began to be considered in the preclinical screening of putative antipsychotic drugs. These include dopamine blockade in extrastriatal, mesofrontal and mesolimbic dopaminergic pathways, which have been studied using a host of different procedures, ranging from single neuron action potential recordings (Skarfeldt 1995) to the expression of neuropeptides (Bissette and Nemeroff 1995) and immediate early genes (Deutch and Duman 1996). Phases I and II in clinical psychopharmacology have also experienced a methodological boost, including modern neuroimaging techniques such as positron emission tomography (PET) and single photon emission computed tomography (SPECT), which have proved helpful to characterize the action of new drugs both qualitatively and quantitatively (Nordström et al. 1995; Nyberg et al. 1997; Pilowsky et al. 1997; Farde and Nyberg 1998; Travis et al. 1998). These strategies are complemented by experiments evaluating non-dopaminergic systems, such the serotonergic (Roth and Meltzer 1995) and glutamatergic (Bunney et al. 1995) pathways. Thus clinical psychopharmacology, just as after the introduction of chlorpromazine, has catalysed new research in the neurobiology of schizophrenia.

Clozapine has also helped to change the demands on the clinical management of schizophrenia. While there used to be a strong emphasis on positive symptom reduction in the early days of pharmacological treatment, new therapeutic targets have now been identified. In addition to treating delusions and hallucinations, negative, affective and cognitive symptoms and the dimensions of suicide prevention, quality of life and psychosocial reintegration are now very much in the foreground of our considerations. Safety research has also become a major new focus.

The following review is restricted to pharmacological interventions. Clearly, modern concepts of schizophrenia management include psychosocial and rehabilitative measures. Even though pharmacotherapy is still very much the backbone of our treatment efforts, it should always be embedded in integrative treatment procedures which include all levels of intervention.

In the following, we will make an artificial distinction between short-term and long-term management, well aware that these cannot always be clearly differentiated. When discussing short-term treatment, we will focus on symptom control and crisis intervention, while making maintenance of therapeutic effects and relapse prevention main topics in the section on long-term treatment, together with aspects of quality of life and psychosocial reintegration.

2 Short-Term Treatment (Acute Treatment)

2.1 Choice of Drug

Before the launch of clozapine, it was generally taken for granted, and supported by many comparative trials, that the available neuroleptics differed in terms of tolerability but not in terms of their efficacy. As a rule of thumb, high-potency drugs were associated with a greater risk for extrapyramidal motor side-effects, while low-potency medication was said to have more autonomic adverse events. The latter have also been said to be more sedative, even though this was never formally studied in clinical trials (Wirshing et al. 1995). These differences in safety profile meant that the choice of drug was mainly guided by side-effects.

Other criteria of choice include the potential for plasma level measurements, which will be discussed in a separate section, as well as the availability of depot preparations. In patients who have been pre-exposed to antipsychotics, it is advisable to fall back on medication that has shown a favourable risk–benefit profile during the treatment of previous episodes. It should also be mentioned in this context that the choice of drug is of particular importance when treating patients for the first time, as a dysphoric response to an antipsychotic drug is known to influence the attitude towards further pharmacological treatments (Van Putten 1974). Particularly stressful side-effects may lead to a long-lasting negative attitude towards medication, which is unfortunate given the fact that many patients will eventually require long-term treatment.

Clozapine was the first drug for which qualitative and quantitative differences to traditional antipsychotics could be convincingly demonstrated (Gerlach et al. 1974; Fischer-Cornelssen and Ferner 1976; Kane et al. 1988a). Not only is clozapine more efficacious in patients suffering from treatment-resistant schizophrenia, but it has also been shown to have advantages in terms of the treatment of negative symptoms (Claghorn et al. 1987; Kane et al. 1988a) and in various psychosocial domains, including quality of life (Meltzer et al. 1990).

The fact that clozapine has good antipsychotic efficacy and a minimal risk of inducing EPS has led to the term "atypical" neuroleptic being used. This classification was originally designed to differentiate clozapine from the older, traditional neuroleptics, which are sometimes referred to as "typical". Historically, this was based on the finding that clozapine, in contrast to all neuroleptics available at the time, did not show cataleptogenic or apomorphine antagonistic effects, which was originally termed an "anomaly" (Schmutz et al. 1967; Stille and Hippius 1971). Unfortunately, neither preclinical nor clinical psychopharmacology offers a succinct definition for "atypicality". As a categorical distinction between "typical" and "atypical" antipsychotic drugs is not possible, some psychopharmacologists have fallen back on a dimensional point of view in which various antipsychotics with more or less "atypical" properties can be found. These properties include a low incidence of EPS, a reduced tendency to increase prolactin levels and greater improvements in the negative symptoms of schizophrenia than traditional drugs. This is where clozapine and other new drugs, which will be described in more detail at a later stage, show advantages over traditional neuroleptics. As the term "atypical antipsychotic" does not describe a homogeneous group of medications, it is really of very limited usefulness. New antipsychotics that were mainly developed to mirror the efficacy of clozapine should more correctly be classified as "novel" antipsychotics or as antipsychotics of the second generation.

Since clozapine cannot be used as a first-choice drug, given world-wide legal restrictions that arise from the drug's high risk of inducing agranulocytosis (Alvir et al. 1993), there was a strong urge to develop similar antipsychotics without the propensity to induce blood dyscrasias. This led to the development and consequent licensing of olanzapine, quetiapine, risperidone, sertindole, ziprasidone and zotepine. Much can be said in favour of the use of these agents as first-line drugs. Merely the fact that they have a significantly lower EPS risk than traditional neuroleptics (Fleischhacker and Hummer 1997) would be argument enough. One of the disadvantages of these novel antipsychotics, however, is the lack of parenteral formulations, including depot preparations. Moreover, many of the new drugs appear to be less sedative than traditional neuroleptics, which poses some problems in the treatment of acutely ill, highly agitated and aggressive patients. As most of the published data on the novel drugs provide results from pre-registration phase II and III clinical trials performed in a highly selected group of schizophrenic patients, we would be well advised to await additional phase IV data and experience gathered under routine treatment conditions before giving definitive recommendations for the use of these drugs. Lastly, pharmaco-economic aspects need to be acknowledged, as the new drugs are substantially more expensive than their predecessors. However, these high drug costs have been shown to be counterbalanced by a reduction in overall treatment costs, through lower relapse rates and fewer subsequent rehospitalizations, as has been shown in various long-term studies (Aitchison and Kerwin 1997; Glazer and Johnstone 1997).

In conclusion, it appears to be justified to recommend novel antipsychotics as first-line treatment, especially in patients with initial manifestations of schizophrenic disorders. It is still unclear whether these new drugs will eventually replace the traditional antipsychotics, which currently still play an important role, especially in the treatment of multi-episode patients, who have shown a good response to these drugs and have tolerated them well in the past.

2.2 Dose, Plasma Level and Route of Administration

Ever since neuroleptics were introduced into clinical practice, dose recommendations have varied up to 100-fold. On the basis of a meta-analysis of studies using traditional drugs, Baldessarini et al. (1988) stated that 100–700 mg chlorpromazine equivalents per day rep-

resent an adequate dose range for most psychotic patients. This has been confirmed in various more recent dose-finding studies. These showed that doses in excess of 20 mg haloperidol (Rifkin et al. 1991; Van Putten et al. 1992) or fluphenazine (Levinson et al. 1990) or 6 mg risperidone (Marder and Meibach 1994) daily do not provide substantial additional therapeutic benefit. The lower end of the dose spectrum is less well defined. While McEvoy et al. (1991), for instance, reported that most of their patients responded well to a mean of 3.4 mg haloperidol daily, Van Putten et al. (1992) found improvement in only 6% of their patients treated with less than 5 mg haloperidol. This discrepancy can at least in part be explained by different patient selection criteria in the respective studies: in the first trial, almost half of the patients were either first-episode schizophrenics or patients suffering from schizoaffective disorders, while the second team studied mainly chronically ill schizophrenic patients, who most likely differ from the first group of patients in terms of treatment response.

Various authors have also reported that the incidence of side-effects increases with dose (Baldessarini et al. 1988; Levinson et al. 1990; Van Putten et al. 1992; Fleischhacker et al. 1994a). This is especially well documented for EPS and implies that a potential efficacy advantage of a higher dose may by nullified by reduced tolerability. Clearly, the dose should be adjusted to the needs of individual patients within a certain frame of reference. Doses will therefore also be dependent on target syndromes, meaning that higher doses will be used in more agitated patients, while lower doses are given to patients with predominantly negative symptoms, although these recommendations come from a clinical empirical background rather than from controlled trials.

For clozapine there is a distinct discrepancy between the United States and Europe. The doses used in clinical trials and in everyday practice in many centres in the United States are about double the mean dose administered in many European countries (200–300 mg daily) (Fleischhacker et al. 1994a). The reasons for this are not entirely clear. A possible explanation is that clozapine is used more restrictively in the United States and is therefore mainly given to patients who are more severely ill. These differences in dose are of great clinical relevance, as at least some of the side-effects of clozapine, such as seizures or confusion, are dependent on dose and plasma level (Fleischhacker et al. 1994a; Haring et al. 1994). It seems strange that more than 20 years after the registration of clozapine, a sound dose-finding study is still lacking.

Some antipsychotics can also be administered parenterally. In principle, this route of administration should be confined to emergency situations, where lack of insight or acute psychopathology lead to a high risk of patients harming themselves or others. Furthermore, parenteral application may be necessary in certain patients in whom well-documented pharmacokinetic problems make it difficult to reach adequate drug plasma levels on oral medication.

There are no well-documented plasma level–efficacy correlation studies for most of the traditional antipsychotics, with the exception of fluphenazine and haloperidol. A haloperidol plasma level of around 15 ng/ml has been described as optimal by various independent groups (Volavka et al. 1992; Janicak et al. 1997; Coryell et al. 1998). Research also suggests that no additional therapeutic benefit is to be expected at fluphenazine plasma levels beyond 1.5–2 ng/ml (Levinson et al. 1995).

Judging effective clozapine plasma levels is much more difficult and has to be seen in the light of the dosing discrepancies discussed above. Consequently, U.S. authors recommend considerably higher plasma levels (Miller et al. 1994; Potkin et al. 1994) than those that have been documented in successfully treated European patients (Haring et al. 1990; Kurz et al. 1995b). It must be noted, however, that American data stem from prospectively designed plasma level response studies, while European authors have reported clozapine plasma concentrations derived from naturalistic study designs. In the most recent American report, VanderZwaag et al. (1996) suggest that optimal plasma levels range between 200 and 250 ng/ml and report that quite a few patients have also responded well to levels below this range. These numbers come closer to those found by European groups.

VanderZwaag et al. (1996) also stress that the time of venopuncture and the distribution of clozapine doses over the day have a considerable impact on clozapine plasma levels. Giving the full daily dose in the evening before blood for plasma levels is drawn, for instance, results in higher levels than if patients are treated with divided daily doses.

Even though plasma level monitoring of antipsychotics cannot yet be recommended as a routine procedure, it may be of help in certain instances. These include non-response to an adequate antipsychotic dose, the suspicion of compliance problems and the use of pharmacological combination treatment which may lead to pharmacokinetic interactions, such as combinations between certain antipsychotics and specific serotonin re-uptake inhibitors. Clinically significant elevations of clozapine plasma levels have been found when this drug was combined with fluvoxamine (Hiemke et al. 1994). The possibility of monitoring plasma levels should also be entertained in very young or old patients and in patients that suffer from relevant concomitant medical diseases. Unusual side-effects, especially if they occur at low doses, may also justify checking antipsychotic plasma levels.

2.3 Duration

As there is generally a great deal of inter-individual variation in the response to the acute treatment of schizophrenia, the question of when to expect the first signs of this response or, alternatively, at what point in time the response may be judged to be insufficient and a treatment change should be initiated, is commonly asked. Recommendations range between 1 and 2 weeks up to 6 months. Some answers may be found in reports such as that by Levinson et al. (1992), which documents that patients who ultimately respond to antipsychotic treatment have shown an amelioration of various non-specific symptoms, such as sleep disturbance or anxiety and agitation, but also of positive symptoms within the first 2 treatment weeks. It may be cautiously concluded from such studies that ongoing treatment needs to be re-evaluated if patients show no response whatsoever within the first few treatment weeks.

There are some suggestions from clozapine trials that more patience is required in treatment-resistant patients (Melzer 1989). Treatment trials of up to 3 months have been recommended.

2.4 Insufficient Treatment Response

The concept of treatment-resistant schizophrenia is not a homogeneous one. Only rarely do patients not respond at all to psychopharmacological interventions. Most patients show at least a partial response in one or the other symptom of their disorder. It is not uncommon, for instance, for delusions or hallucinations to remit, while negative symptoms remain unchanged. Sometimes, the definition of treatment-resistant schizophrenia also includes patients who do respond to antipsychotic treatment but do not tolerate it due to significant side-effects (Bondolfi et al. 1998).

The definition of treatment-resistant or treatment-refractory schizophrenia usually follows two different lines of thought: one is based on the necessity of reproducible research, while the other is built upon the needs of everyday clinical practice.

Research calls for well-defined criteria that allow study of the issue in different research centres and ultimately make a comparison of different studies possible. In addition to obtaining an exact drug history, patients in such studies usually undergo at least one prospective antipsychotic treatment trial (Kane et al. 1988a). Only if the history of the patient and the well-documented results of this treatment attempt yield no adequate response is such a patient classified as treatment resistant.

In routine clinical care, the treating physician usually cannot fall back on all of this sophisticated information. Judgement must often rely on an incomplete history provided either by the patients themselves or their significant others and on far from perfect case notes. This information is then amalgamated with the clinical impression, which usually concerns overall social functioning and quality of life in addition to psychopathology.

It is not uncommon for the researcher's and the clinician's judgements to differ quantitatively. Many patients in whom the clinician is not satisfied with the treatment response will not fulfil strict research criteria, while on the other hand patients entering research studies may represent a select population and results provided by the studies of such patients are sometimes difficult to generalize for a more clinical population.

It used to be common practice to switch patients who had not adequately responded to a neuroleptic to another drug of a different chemical class. Similarly, this was recommended for patients who did not tolerate certain medications. Research on this issue has produced controversial findings (Kinon et al. 1993; Shalev et al. 1993). So far clozapine is the only drug for which this procedure is unequivocally supported (Fleischhacker 1999).

Taking all the published evidence into account, we recommend the following in the case of inadequate treatment response: if patients have not shown at least partial response to an adequate dose of a traditional antipsychotic after 2–3 weeks, compliance and plasma levels should be checked. This may lead to additional supportive psychosocial measures or to an adaptation of dose. If these modifications do not yield relevant results within the next 2–3 weeks, switching to clozapine (or to another one of the novel antipsychotics) is indicated. This new treatment trial should last at least 2–3 months. If response remains unsatisfactory, various options should be considered. These include concomitant administration of lithium (Growe et al. 1979; Collins et al. 1991; Wilson 1993) or carbamazepine (Schulz et al. 1990) or, alternatively, the use of benzodiazepines (Wolkowitz et al. 1990), specific serotonin re-uptake inhibitors (Goff et al. 1990; Decina et al. 1994; Silver et al. 1996) or serotonin antagonists (Duinkerke et al. 1993). Electroconvulsive therapy also still has its place in these last-resort treatment trials (Krueger and Sackeim 1995). All of these options result from clinical experience rather than from controlled double-blind prospective treatment trials.

Novel antipsychotics are increasingly being used as drugs of first choice. What should be done if patients do not respond to these agents is still an open question. Increasing the dose, sometimes beyond the

levels recommended by the manufacturers, switching to other novel antipsychotics or to traditional neuroleptics, but also combining novel and classical drugs (Shiloh et al. 1997), appear to be the most commonly chosen alternatives.

It should be made very clear that the treatment of patients suffering from treatment-resistant schizophrenia represents a highly complex clinical problem that needs to be taken care of by experienced and well-trained specialists.

2.5 Adverse Effects During Acute Treatment with Antipsychotics

Table 1 gives an overview of antipsychotic-induced adverse events. This table is a general list that deliberately omits prevalence or incidence rates as, given the huge methodological differences between the available studies, unjustified comparisons of relative risks may

Table 1. Adverse effects of antipsychotic drugs

Systems concerned		
Central nervous system	Extrapyramidal	Acute: dystonia, dyskinesia, akathisia, parkinsonism Tardive: dyskinesia, dystonia (akathisia)
	EEG alterations, seizures Sedation Confusion Neuroleptic malignant syndrome	
Cardiovascular	Orthostatic hypotension ECG alterations	 Tachycardia Tachyarrhythmia Depressed ST segments Flattened U waves Prolonged QT intervals
Haematological	Leucocytosis Eosinophilia Thrombocytopenia Leucopenia, agranulocytosis	
Gastrointestinal and hepatic	Gastrointestinal motility disturbances Liver enzyme increase Cholestatic jaundice	
Autonomic nervous system	Dry mouth	
	Constipation Loss of accommodation (blurred vision) Urinary retention Sialorrhoea	
Endocrine and sexual	Prolactin increase	Gynaecomastia Galactorrhoea Menstrual irregularities
	Weight gain Glucose and lipid metabolism disturbances	
	Sexual disturbances	Diminished libido Orgasmic dysfunctions Erectile dysfunctions, priapism Ejaculatory disturbances: reduced volume, delayed ejaculation
Cutaneous	Allergic reactions Photosensitivity Seborrhoeic dermatitis	
Ocular	Lenticular changes Pigmentary retinopathy	

lead to misinterpretation. There is also no reference to the clinical relevance of the respective side-effects, as this will need to be determined on an individual basis in most instances. For the classical drugs, the main focus has traditionally been on EPS (Casey 1996). As the novel antipsychotics induce EPS to a significantly lesser extent (Fleischhacker and Hummer 1997), non-EPS-associated side-effects have gained interest following their introduction into clinical practice.

Many side-effects can be successfully managed by dose reduction or switching from one drug to another (Csernansky and Newcomer 1995). So far, specific pharmacological interventions are only successful against EPS. Even though the efficacy of anticholinergics against acute dystonia or parkinsonism (Remington and Bezchlibnyk-Butler 1996) and that of beta blockers in treating akathisia (Fleischhacker et al. 1990) are impressive, the former drugs in particular should be used cautiously. Anticholinergics are potent psychotropic drugs (Fleischhacker et al. 1987) that may lead to memory deficits (Fayen et al. 1988), substance abuse (Smith 1980) and to a worsening of psychotic symptoms (Tandon et al. 1990). The general prophylactic use of anticholinergic drugs is therefore discouraged, a position also reinforced by a World Health Organization (WHO) recommendation (World Health Organization 1990). The exception to this rule are patients with a high risk for EPS, such as young male patients or patients with a history of significant EPS.

Concerning many of the non-EPS-associated adverse events, including, e.g. sedation, orthostatic hypotension, or laboratory abnormalities, to name but a few of the more common problems, individual tolerance levels have to be discussed with the patients. This follows the general rule that there is increasing emphasis on the subjective attitude of patients and their relatives to pharmacotherapy, as this has been found to be a crucial factor for compliance (Marder 1998). Consequently, drug-induced side-effects must be an ongoing topic even during the acute treatment phase.

The use of the novel antipsychotics as first-line drugs is an important prophylactic measure to reduce the EPS risk. In addition, a slow increase in dose and the use of lower doses prevent some of the acute side-effects. This approach is obviously limited in very acutely, severely ill patients.

3
Long-Term Treatment

In addition to maintaining the effects of acute treatment and relapse prevention, an improvement in quality of life and psychosocial integration are the goals of long-term treatment of patients suffering from schizophrenic disorders.

There is increasing evidence that the long-term outcome of schizophrenia is correlated to early pharmacological interventions and successful relapse prevention (Crow et al. 1986; Wyatt 1992; Loebel et al. 1995). It is one of the best-documented and most reproduced results in psychiatric outcome research that long-term treatment with antipsychotics is the major factor in preventing relapses and recurrences of the disorder (Kane and Lieberman 1987; Csernansky and Newcomer 1995; Gilbert et al. 1995). The relapse risk is reduced by about two-thirds if long-term antipsychotic medication is sustained (Kissling 1991). In order to optimize treatment response, pharmacological measures have to be complemented by psychosocial interventions. As discussed in the introductory section, the latter are not the topic of this review and will be summarized elsewhere.

Even given optimal treatment conditions, about 20% of all schizophrenic patients will experience a relapse despite antipsychotic prophylaxis (Steingard et al. 1994). On the other hand, another 20% will suffer a single episode of schizophrenia only, regardless of whether they are treated or not (Möller and van Zerssen 1995). Unfortunately, we do not have any way of predicting the outcome of schizophrenia in individual patients. All patients suffering from schizophrenia, including those with first episodes, are therefore advised to continue medication on a long-term basis (Kissling et al. 1991; American Psychiatric Association 1997; Gaebel and Falkai 1998; Lehmann et al. 1998). The rare patient with brief psychotic episodes without negative psychosocial consequences may be an exception to this rule, as would be a patient with an intolerance to all existing antipsychotics.

3.1
Choice of Drug

The choice of the antipsychotic will follow similar considerations to those outlined in the acute treatment section. It has to be kept in mind that switching from one antipsychotic to another may pose problems, as the efficacy and tolerability of the different antipsychotics show significant intra-individual variation. It is therefore preferable to already plan long-term treatment when initiating acute treatment, in order not to have to change drugs between treatment phases.

Unfortunately, pharmaco-economic considerations play an increasing role in the choice of antipsychotics in many countries. As the novel drugs are considerably more expensive than traditional neuroleptics, clinicians and patients are increasingly confronted with restrictive reimbursement policies. This is especially

disturbing when we consider that pharmaco-economic studies have convincingly shown that higher drug costs are outweighed by savings in other areas of treatment, such as reduced rehospitalization and relapse rates (Aitchison and Kerwin 1997; Rosenheck et al. 1997).

In addition to efficacy, tolerability and subjective acceptance of a drug by the patient are important factors in long-term treatment. The risk–benefit profile of an antipsychotic has to be a regular topic in treatment sessions with patients and their significant others.

3.2 Duration

Two types of studies give direct and indirect information on the length of treatment issue: firstly, prospective placebo-controlled long-term trials and, secondly, discontinuation studies. In the latter, antipsychotics are discontinued under controlled conditions in patients who have been prophylactically treated for varying amounts of time. While prospective studies usually extend over time periods from 1 to 2 years, discontinuation studies often provide information about considerably longer courses of treatment. Both study options unanimously document a high relapse risk without antipsychotic prophylaxis (Kane and Lieberman 1987; Csernansky and Newcomer 1995; Gilbert et al. 1995).

One- to 2-year maintenance treatment is usually recommended for patients suffering their first episode of schizophrenia. Multi-episode patients should be in remission for at least 5 years before the discontinuation of antipsychotic treatment is discussed (Kissling et al. 1991; American Psychiatric Association 1997; Lehmann et al. 1998; Gaebel and Falkai 1998).

The time frames recommended above must be seen in the light of the fact that there are no prospective relapse prevention studies which cover a time period of more than 2 years and that all discontinuation trials have shown high relapse rates, even if patients had been in remission for many years before stopping antipsychotics. These recommendations must therefore be considered a minimal standard. Especially in first-episode patients, they are also influenced by practical considerations, insofar as it is unrealistic to suggest life-long pharmacological relapse prevention, even though, when judging the available evidence, this would not appear unreasonable. It is also evident that such recommendations, although commonplace in various chronic somatic diseases, are still met by much irrational criticism when applied to the psychiatrically ill.

In the last decade, five independent research groups have evaluated the effects of so-called intermittent pharmacological treatment (Jolley et al. 1989; Carpenter et al. 1990; Herz et al. 1991; Pietzcker et al. 1993; Schooler et al. 1997). These studies were based on the assumption that it should be possible to educate patients and their significant others about early warning signs of an impending relapse. In such patients, antipsychotics could be stopped after successful acute treatment and reinitiated in the case of impending relapse, detected by the advent of early warning signs. This was hypothesized to lead to a reduced use of antipsychoties, thereby minimizing the risk of long-term side-effects such as tardive dyskinesia.

Patients randomized to this type of management showed significantly higher relapse rates than patients on continuous antipsychotic treatment. Even though patients on intermittent treatment received significantly fewer cumulative antipsychotic doses, there were no differences in the incidence of tardive dyskinesia between groups. In summarizing the available evidence, it can be concluded that intermittent treatment has not been shown to be a generally practical alternative to the current recommendation of continuous antipsychotic administration, especially when considering a separate study in which an increased tardive dyskinesia risk was found in patients with multiple interruptions of neuroleptic treatment (Van Harten et al. 1998).

The strategic goal of the long-term treatment of patients suffering from schizophrenic disorders remains to minimize the risk of a psychotic relapse in order to avoid all its negative biological and psychosocial consequences.

3.3 Dose, Plasma Level and Route of Administration

In general, the same antipsychotic doses that have been efficacious during the acute and the stabilization phases are also recommended at the beginning of relapse prevention. For most patients, these doses range between 5 and 15 mg daily oral haloperidol or a respective equivalent dose of another antipsychotic (Kissling et al. 1991). Dose–response relationships of the novel antipsychotics have not yet been sufficiently studied; the few published double-blind, long-term continuation trials (Daniel et al. 1998; Tran et al. 1998) indicate that the principle of maintaining patients on the dose that was used in acute treatment is also helpful with these drugs.

If a dose reduction is indicated, it should not be performed in steps larger than 20% of the previous dose. The intervals between these steps should be between 3 and 6 months, as it is well known that relapses following insufficient neuroleptic doses may appear with a significant time lag (Johnson 1979; Kane and Lieberman 1987).

It should obviously be attempted to treat patients with a minimal effective dose. However, in clinical

practice this is not always easy and, while this dose is sought, may lead to a risk of underdosing and subsequent relapse.

The administration of depot antipsychotics is an important treatment option during long-term management. These injectable drugs produce relatively constant plasma levels of the neuroleptics over a period of several weeks (J.M. Davis et al. 1994). The disadvantages of this type of treatment include the fact that some patients refuse intramuscular injections or develop irritations at the injection site. Another problem is that the dose of a depot antipsychotic cannot be reduced once administered. Advantages include the facilitation of management due to certainty concerning compliance and the fact that patients do not need to take medication every day. Various expert groups have therefore promoted the use of depot antipsychotics (Kissling et al. 1991; J.M. Davis et al. 1994; Kane et al. 1998).

If treatment with these drugs is anticipated, patients should first be treated with the oral form of the antipsychotic, in order to gather information about dose requirements and the risk–benefit profile of the drug. Ideally, patients should be switched from the oral to the depot route after successful stabilization has been achieved. This should be done in an overlapping fashion, as depot medications reach steady-state plasma levels only after a certain period of time (Altamura et al. 1989).

Dose-finding studies are available for haloperidol and fluphenazine. They are the basis of the following recommendations: 50–200 mg haloperidol decanoate every 4 weeks (J.M. Davis et al. 1993) or 12.5–50 mg fluphenazine decanoate in biweekly intervals (J.M. Davis et al. 1994) represent an optimal dose range for many patients.

3.4 Adverse Effects of Long-Term Treatment with Antipsychotics

Most acute antipsychotic-induced side-effects can become chronic. Clearly, clinicians will do their best to avoid this, although in some instances compromises may be necessary, especially when clear benefits of medication outweigh the relevance of certain side-effects. As discussed previously, determining this risk–benefit ratio must be an integral component of all treatment strategies.

Some side-effects may already be apparent during acute treatment, but become relevant only during long-term management. These include sedation, weight gain and sexual dysfunctions (Whitworth and Fleischhacker 1995). As these adverse events are thought to have a negative impact on compliance, they warrant special consideration.

Tardive dyskinesia is a specific long-term side-effect that is relatively common with traditional antipsychotics (American Psychiatric Association 1992). The annual cumulative incidence rate is around 5% (Kane et al. 1988b). Given the increased attention to tardive dyskinesia and the implementation of prophylactic measures, severe and irreversible manifestations have become considerably less frequent. It is important to acknowledge that tardive dyskinesia is not necessarily irreversible.

Observations over long time periods have shown that this motor side-effect remits spontaneously in about half of the afflicted patients despite continuous antipsychotic treatment (Gardos et al. 1994).

While no well-documented case of clozapine-induced tardive dyskinesia has been published so far, it is important to note that this does not hold true for the other novel compounds. Even if these drugs are associated with a certain risk for tardive dyskinesia, this is still significantly lower than that found with traditional antipsychotics (Tollefson et al. 1997a).

The management of chronic side-effects of antipsychotics, with the exception of tardive dyskinesia, follows the same principles as have been described in the acute treatment section. Treating manifest tardive dyskinesia is still unsatisfactory; prophylactic measures are therefore of the utmost importance. Patients have to be examined regularly with regard to incipient tardive dyskinesia. As soon as such symptoms are found, if continuous antipsychotic treatment is considered necessary, an attempt to reduce the dose should be the first step. If tardive dyskinesia progresses despite this, treatment must be changed to clozapine or another compound with a low tardive dyskinesia risk. Some patients with a shorter duration of tardive dyskinesia may also profit from an addition of tocopherol (vitamin E) (Adler et al. 1998), although a large recent study has failed to find a positive effect of vitamin E (Adler et al. 1988).

4 Novel Antipsychotics

The problem of artificially categorizing all novel antipsychotics as "atypical" has been alluded to earlier in this review. A critical review of these drugs clarifies that they are pharmacologically heterogeneous substances (Fleischhacker and Hummer 1997). This most likely results in diverging clinical profiles, although these are not yet fully understood.

New developments are briefly summarized below. As this is currently a very prosperous field of research, the

Table 2. Novel antipsychotics

Drug	Receptor profile[a]	Recommended dose (mg per day)[b]
Amisulpride	D	300–800 (50–1200)
Clozapine	5HT, D, α, M, H	200–450 (50–900)
Olanzapine	5HT, D, M, α, H	10 (5–20)
Quetiapine	H, 5HT, α, D	150–750
Risperidone	5HT, D, α, H	2–6 (1–16)
Sertindole	5HT, D, α	12–20 (4–24)
Ziprasidone	5HT, D	40–160
Zotepine	5HT, D, α, H, M	100–300 (50–450)

D, dopamine; α, α-adrenergic; M, muscarinic; H, histamine; 5HT, serotonin.
[a]Listed in order of descending affinity.
[b]Doses in general correspond to the recommendations of the manufacturer; doses listed in parentheses represent extremes sometimes justified in individual patients.

interested reader is referred to the most recent publications in scientific journals. Table 2 summarizes receptor affinities and dose recommendations for the novel antipsychotics. As some of the special features of amisulpride and clozapine are mentioned elsewhere in the chapter, they will not be discussed in the following.

4.1 Olanzapine

Olanzapine is similar to clozapine both in its chemical structure and in its pharmacological properties (Bymaster et al. 1996). It has a plasma half-life of about 30 h (Eli Lilly & Co., Indianapolis, USA); the manufacturer recommends doses of between 5 and 20 mg daily.

Before registration, olanzapine was compared to placebo and haloperidol in clinical trials (Beasley et al. 1996; Tollefson et al. 1997b). It was shown to be superior to placebo (Beasley et al. 1996) and at least equal to haloperidol (Beasley et al. 1996; Tollefson et al. 1997b) when treating patients with schizophrenic or schizoaffective disorder. It was more effective than haloperidol in treating depressive (Tollefson et al. 1998) and negative symptoms (Tollefson et al. 1997c) in schizophrenic patients. The problems in interpreting these findings will be discussed at the end of this chapter.

All therapeutic effects of olanzapine are maintained over longer periods of time, as 1-year extension studies comparing the drug with placebo and haloperidol have shown (Hamilton et al. 1998; Tran et al. 1998).

In all clinical trials and at all dose levels, olanzapine induced extrapyramidal symptoms that, with the exception of akathisia, were not higher than those found in patients randomized to placebo and amounted to about 20% (Tran et al. 1997). Olanzapine was shown to be associated with a significantly lower tardive dyskinesia risk than haloperidol (Tollefson et al. 1997a). As with most other antipsychotics with strong anti-serotonergic effects, the treatment with olanzapine leads to a significant weight gain in a considerable number of patients (Allison et al. 1999; Beasley et al. 1996; Tollefson et al. 1997b; Weiss et al. 1998). Despite its similarities to clozapine, there is currently no indication that blood count monitoring beyond the recommendations given for antipsychotic treatment in general is warranted.

4.2 Quetiapine

While quetiapine also structurally resembles clozapine, its preclinical pharmacology is different, especially with respect to a lack of anticholinergic effects (Saller and Salama 1993). It has a short half-life (about 3 h; Fulton and Goa 1995). Dose recommendations range between 150 and 750 mg daily.

In phase II and III studies, quetiapine was found to be superior to placebo and comparable to haloperidol (Borison et al. 1996; Arvanitis et al. 1997; Small et al. 1997) and chlorpromazine (Hirsch et al. 1996).

The risk of inducing acute EPS was not higher than that of placebo, while data concerning incidence rates of tardive dyskinesia have not yet been published. Transient elevations of liver enzymes, dizziness and orthostatic hypotension, especially in the first few days of treatment, have led to the recommendation of a slow dose increase during the initiation of treatment.

4.3 Risperidone

Risperidone is a new molecule. Its elimination half-life has been reported to be between 3.2 and 24 h (Byerly and DeVane 1996). Recent dose recommendations are 2–6 mg daily. Risperidone has also been compared with placebo (Marder and Meibach 1994) and haloperidol (Peuskens 1995). As with the other drugs, its

therapeutic effect was found to be better than placebo and similar to haloperidol. For negative symptoms, advantages similar to those of olanzapine have been found when comparing the drug to haloperidol (Möller 1995). Preliminary results have also shown positive effects when treating children and adolescents (Mandoki 1995; Sternlicht and Wells 1995).

Risperidone is also associated with a lower risk of inducing acute EPS than traditional neuroleptics (Lemmens et al. 1999). In contrast to the other drugs, the EPS risk of risperidone is dose dependent. Higher doses start to resemble traditional antipsychotics; above 12 mg daily, the EPS rate is similar to that of haloperidol (Marder and Meibach 1994; Peuskens 1995).

As for quetiapine, a slow dose increase is recommended to prevent orthostatic hypotension.

4.4 Sertindole

Sertindole has the longest half-life of all the novel drugs (close to 3 days) (Dunn and Fitton 1996). Doses between 4 and 24 mg daily are used to treat patients with schizophrenia. Pre-registration clinical trials comparing sertindole to placebo (Zimbroff et al. 1997) and haloperidol (Van Kammen et al. 1996; Zimbroff et al. 1997) have yielded similar results as for the previously mentioned novel antipsychotics. A long-term study found significantly lower rehospitalization rates with sertindole compared with haloperidol (Daniel et al. 1988). The same holds true for its EPS profile. An unusual side-effect reported for sertindole is a reduction of the ejaculatory volume. In addition, weight gain has been observed.

Sertindole leads to a prolongation of the QTC interval on electrocardiography (ECG) (Hale et al. 1996; Van Kammen et al. 1996; Zimbroff et al. 1997). This effect is not only more common, but also more pronounced than with any other of the novel antipsychotics and is still being discussed with respect to its clinical relevance. At the time of writing, the European regulatory agency was re-evaluating the safety profile of this drug.

4.5 Ziprasidone

Ziprasidone differs pharmacologically from the previously described drugs by its potent 5-HT_{1A} agonistic effect and by the inhibition of serotonin and noradrenaline re-uptake (Davis and Markham 1997; Tandon et al. 1997). Its half-life is reported to be between 3.2 and 10 h (Davis and Markham 1997); the dose range is 80–160 mg daily.

These doses were also evaluated in clinical trials in which ziprasidone was compared with placebo (Keck et al. 1998) and/or haloperidol (Goff et al. 1998; Keck et al. 1998). Again, positive and negative symptoms (at higher doses) were improved, as found with other novel antipsychotics. In a placebo-controlled 1-year trial in which stable patients were switched to either ziprasidone or placebo, the former had a significantly higher efficacy than placebo in reducing the risk of impending relapse (Arato et al. 1997).

The EPS risk is comparable to the other new medications. The most commonly found side-effect is sedation. Interestingly, and in contrast to all the drugs discussed above and below, ziprasidone has not been found to induce weight gain.

4.6 Zotepine

Zotepine is also structurally related to clozapine (Prakash and Lam 1998). Similar to ziprasidone, it inhibits the re-uptake of noradrenaline with a potency comparable to tricyclic antidepressants (Rowley et al. 1998). The recommended dose is 50–450 mg daily.

Again, results of clinical trials resemble those found with other novel antipsychotics (Fleischhacker et al. 1989; Dieterle et al. 1991; Klieser et al. 1991; Petit et al. 1996). It was also used successfully in a small study treating patients with predominantly negative symptoms (Barnas et al. 1992). When patients with stable symptoms were switched to either placebo or zotepine and followed for 6 months, the risk of relapse was found to be significantly higher in the placebo group (Cooper et al. 2000).

Zotepine has fewer side-effects on the extrapyramidal motor system than haloperidol. Dose-dependent adverse effects include sedation, transient elevations of lever enzymes and seizures.

4.7 Methodological Considerations of Treatment Trials with Antipsychotic Drugs

When analysing the clinical trials initiated by various pharmaceutical companies in order to licence new antipsychotics over the last decade, it needs to be considered that all of these studies were performed in a highly selected group of patients. These were usually in their late thirties, and two-thirds were of male gender. The average duration of illness was over 10 years in many cases, and frequent hospitalizations have preceded the clinical trial. Little is known about the pre-treatment of these patients, and wash-out phases were generally short. In almost all studies,

about 20% of the population suffered from schizoaffective disorder.

When looking at response rates of these 6- to 8-week clinical trials, it becomes evident that the scores of the rating scales used to measure psychopathology (the Brief Psychiatric Rating Scale, BPRS; the Positive and Negative Symptom Scale, PANSS) were reduced by 20%–40% at the most. This holds true for both the experimental and the comparator drugs (mostly haloperidol or chlorpromazine).

As stated at the outset, this brief summary of recent clinical trials exemplifies that the investigated sample was a highly selected one. Chronically ill, frequently hospitalized male schizophrenic patients who have probably responded only insufficiently to previous treatment attempts, and were usually far from reaching full remission during the course of the clinical trial, constituted the core group on which the risk–benefit evaluation of the test substance was based.

The attempt to generalize results obtained from studying this group of patients to the overall population of patients with schizophrenia appears problematic, to say the least. The spectrum of schizophrenia includes many patients who hardly ever make it into phase II and III clinical trials. These range from early-onset female patients with suicidal ideation all the way to hostile and agitated treatment-refractory patients. It is therefore not surprising if the results from clinical trials cannot always be translated into clinical practice.

Some clinicians are disappointed with the efficacy of the novel compounds in treating acutely psychotic patients, which may be related to the fact that many of the new drugs have fewer sedative effects than their traditional counterparts. This often calls for additional sedative medication in the early stages of treatment. A lack of sedation must not be confused with a lack of antipsychotic efficacy, which, even with the older drugs, always takes days to weeks to kick in.

Most of the novel antipsychotics have been reported to be more efficacious in treating negative symptoms than traditional neuroleptics. These findings have to be interpreted with caution, as this effect is only documented for patients suffering from positive and negative symptoms concomitantly for most antipsychotics, with the exception of amisulpride (Loo et al. 1997) and, to a lesser extent, zotepine (Barnas et al. 1992). It is accepted knowledge that negative symptoms in this group of patients respond to treatment considerably better than primary enduring negative symptoms, as seen in patients with deficit states of schizophrenia (Carpenter 1996). Post hoc path analyses have been performed for olanzapine (Tollefson et al. 1997c) and risperidone (Möller 1995) to analyse how much the improvement of negative symptoms has been influenced by intervening variables. It has been reported that at least part of the therapeutic efficacy is independent of an improvement in positive symptoms or a reduction of EPS (Möller 1995; Tollefson et al. 1997c). As path analysis is not a confirmatory, inferential statistical method, these results need to be replicated using other study designs. Amisulpride (Loo et al. 1997), ritanserin (Duinkerke et al. 1993) and zotepine (Barnas et al. 1992) have been studied in patients with predominantly negative symptoms and have been documented to lead to an amelioration of deficit states. Ritanserin was used as an add-on to an ongoing neuroleptic treatment; amisulpride and zotepine were shown to be beneficial as monotherapy in low doses.

The evaluation of the tolerability of novel antipsychotics contains some potential pitfalls as well. Side-effect assessment is performed in different ways in different studies: while some trials include specific side-effect rating scales, others rely on spontaneous reporting by patients and/or clinicians. These two methods will understandably yield different incidence rates. A reliable comparison of antipsychotics is only possible if drugs are compared within the same clinical trial.

EPS present a specific problem. All published controlled clinical trials with the novel antipsychotics have found an EPS risk that corresponds to that found in the placebo group. In this context, it needs to be noted that patients in the placebo group generally develop EPS at a rate of about 20%. This seemingly paradox finding can be explained in different ways: as clinical trials tend to have brief wash-out periods, EPS resulting from pre-treatment often carry over into the comparative part of the study. If these effects occur, they may be misinterpreted as induced by whatever drug these patients are being treated with in the clinical trial. Withdrawal dyskinesias, which occur frequently upon cessation of treatment if tardive dyskinesia had been masked by traditional antipsychotics (Gardos and Cole 1995; Schultz et al. 1995), may be an alternative explanation. Furthermore, problems in differential diagnosis, such as distinguishing between acute akathisia and agitation in acutely psychotic patients, can lead to a misdiagnosis of motor symptoms (Miller and Fleischhacker 2000). Kraepelin (1919) already reported that schizophrenia may be accompanied by motor symptoms. Phenomenologically, these are very similar to drug-induced dyskinesias. Several reports in recent years have also found movement disorders in drug-naive patients with schizophrenia (Caligiuri et al. 1993; Chatterjee et al. 1995). Lastly, a true placebo effect, as may be expected in neuroleptic-experienced patients, and the tendency of patients abusing anticholinergics to feign EPS in order to receive the desired prescription may be seen as possible reasons for the surprisingly high EPS incidence in the placebo groups of clinical trials.

It must be borne in mind that the fact that the EPS rate of a new drug corresponds to that found with

placebo should not lead us to the false assumption that the novel compound has no EPS risk at all.

In conclusion, it is emphasized that efficacy and tolerability data from phase III clinical trials should be converted into everyday clinical practice with all due caution and with regard to the potential sources of error discussed above, in order to prevent unrealistic expectations.

5
Special Aspects of the Pharmacotherapy of Schizophrenia

5.1
Negative and Depressive Symptoms, Suicidality

Various methodological issues impede the interpretation of clinical trials in patients suffering from negative and/or depressive symptoms in the course of schizophrenia. Differential diagnosis is one of them: negative symptoms, depression and akinesia share many common features. This makes diagnosis difficult, especially in a cross-sectional evaluation. Negative symptoms can be of secondary nature, e.g. as a result of EPS or as sequelae of positive symptoms or psychosocial deprivation (Carpenter 1996). Depressive syndromes in schizophrenia are commonly seen as an inherent feature of the illness; they can also occur as a psychological reaction to the diagnosis (Liddle et al. 1993; Siris 1995). Problems of clinical trials evaluating negative symptoms have been dealt with in the previous section.

Imipramine has a positive effect against depressive syndromes in schizophrenia patients (Siris 1995). Recently, serotonin re-uptake inhibitors have also been shown to exert beneficial effects in this indication as well as against negative symptoms (Goff et al. 1995). Before pharmacological treatment is initiated, it is important to reach a reliable diagnosis. The fact that negative and depressive syndromes are difficult to manage should not lead to therapeutic nihilism. This is especially important in light of the fact that patients suffering from schizophrenia have a high suicide risk (Roy 1990). Preliminary results on the suicide prophylactic effect of clozapine are very encouraging (Meltzer and Okayli 1995).

5.2
Cognitive Functions

Cognitive disturbances have been described ever since the first systematic research in schizophrenia. They include a general intellectual deficit that has a special emphasis in memory and executive functions (Mortimer 1997; Sharma and Mockler 1998). These dysfunctions hinder the rehabilitation of patients regardless of the other psychopathological symptoms (Goldberg and Gold 1995). They constitute a negative predictor for the course of the illness (Kolakowska et al. 1985; Perlick et al. 1992). Classical neuroleptics have little influence on cognitive dysfunctions, if at all; they may even lead to a deterioration of cognitive abilities (Mortimer 1997). This seems to be different with novel antipsychotics. Clozapine, for instance, has been shown to ameliorate various cognitive functions, especially attention and verbal fluidity (Schall et al. 1998). An improvement of working memory has been found after treatment with risperidone (Green et al. 1997) and olantapine also enhanced cognition in a longer term study (Purdon et al. 2000). These findings and other research indicate that novel antipsychotics might also have advantages over the traditional drugs in this important facet of the schizophrenia syndrome.

5.3
Compliance

As in any other illness in which the long-term intake of medication is necessary, compliance is of high clinical relevance in patients suffering from schizophrenic disorders. This is also documented by the fact that less than 50% of patients with schizophrenia in long-term treatment take their medication according to the physician's recommendations (Fenton et al. 1997). Non-compliance has implications that go beyond mere patient management. It may distort the results of psychopharmacological trials, especially of placebo-controlled studies in which the experimental drug has more side-effects than placebo. In this case, non-compliance can be expected to be more prominent in the active treatment group than in the placebo-treated patients. Consequently, these non-compliant patients will show a lower rate of treatment response, whereby the difference between placebo and active compound may be artificially diminished. This may lead to an underestimation of potential therapeutic advantages of the experimental drug. Similar problems can occur in a comparison of two active drugs that show different side-effect profiles.

Compliance is influenced by a host of different variables. These are usually grouped into patient-related, clinician-related, treatment-related and environment-related factors (Fleischhacker et al. 1994b). Patient-related factors include demographic characteristics such as age (Schwartz et al. 1962), sex (Danion et al. 1987; Swett and Noones 1989) and social status, but also illness-associated characteristics such as type of disorder and psychopathological symptoms (Schou 1997; Drake et al. 1989; Pan and Tantam 1989).

Various aspects of the clinician–patient relationship (Ley and Spelman 1965; Meise et al. 1992) and the information accessible to patients are also part of the influences on compliance (Bäuml et al. 1993).

The psychosocial environment of patients also determines attitudes towards treatment and health belief concepts which, in turn, affect compliance (Blackwell 1973; Hoge et al. 1990).

Adverse effects of drugs are the best studied treatment-related factor. These include EPS, in particular akathisia (Van Putten 1974), as well as weight gain and sexual dysfunctions (Silverstone et al. 1988; Pfeiffer et al. 1991; Buchanan 1992). In contrast, several authors have reported that side-effects do not have a negative effect on compliance (Middelboe 1995; Fleischhacker et al. 1994b; Hummer et al. 1999). These findings can easily be explained by an indirect improvement of the clinician–patient relationship through information about treatment in the case of the advent of drug-induced adverse events which appears to have a positive influence on compliance that outweighs the negative consequences of side-effects. Another important treatment-related factor is the complexity of treatment. Polypharmacy and multiple therapists with poorly defined roles impede the cooperation between the patient and the treatment team.

Interventions to enhance compliance can be targeted to all of these levels (Hummer and Fleischhacker 1999). Ideally, they should be implemented at the beginning of treatment. The success of compliance-enhancing measures is well documented (Eckman et al. 1990; Kemp et al. 1998). On the basis of a working doctor–patient relationship, concise and relevant information provided for patients and their significant others must be an integral part of these efforts. Information must also include illness concepts, as these are often unrealistic. In addition, the prevention and/or rapid management of adverse effects plays a crucial role.

Lastly, compliance by clinicians with rational, scientifically determined treatment guidelines and recommendations will optimize treatment efforts and reduce the uncertainty of patients and/or relatives confronted with different treatment concepts suggested by different doctors.

6 Conclusion

Pharmacotherapy is the basis of the management of patients suffering from schizophrenic disorders. There is no doubt that the novel antipsychotics enhance the spectrum of acute and long-term treatment options. It is to be expected that the better tolerability of these agents, especially in terms of a reduced EPS risk, will also facilitate rehabilitative measures. An improvement in cognitive functions should contribute to that. On the other hand, it must be emphasized that antipsychotics may induce non-extrapyramidal side-effects which can have a negative influence on compliance and psychosocial treatment. The prescriber will therefore still be charged with establishing a sound risk–benefit profile of medication, which is not yet available for the novel antipsychotics to a satisfactory extent. More post-marketing research data are necessary to establish such profiles.

New findings in the field of psychopharmacology, together with advances in psychosocial treatment strategies, have given rise to a new optimism in the treatment of this severe psychiatric disorder. The continuous process of scientific evaluation of new treatment options is a prerequisite for an optimal integration of these into the treatment of schizophrenia.

7 References

Adler LA, Edson R, Lavori P, Peselow E, Duncan E, Rosenthal M, Rotrosen J (1998) Long-term treatment effects of vitamin E for tardive dyskinesia. Biol Psychiatry 43: 868–872

Aitchison K, Kerwin R (1997) Cost-effectiveness of clozapine. Br J Psychiatry 171: 125–130

Allison DB, Mentore JL, Heo M, Chandler LP, Cappelleri JC, Infante MC, Weiden PJ (1999) Antipsychotic-induced weight gain: a comprehensive research synthesis. Am J Psychiatry 156(11): 1686–1696

*Altamura CA, Colacurcio F, Mauri MC, Moro AR, DeNovellis F (1989) Haloperidol decanoate in chronic schizophrenia: a study of 12 months with plasma levels. Prog Neuropsychopharmacol Biol Psychiatry 14: 25–35

Alvir JMJ, Lieberman JA, Safferman AZ, Schwimmer JL, Schaaf JA (1993) Clozapine-induced agranulocytosis: incidence and risk factors in the United States. N Engl J Med 329: 162–167

*American Psychiatric Association (1992) Tardive dyskinesia: a task force report of the American Psychiatric Association. American Psychiatric Press, Washington, DC

*American Psychiatric Association (1997) Practice guideline for the treatment of patients with schizophrenia. Am J Psychiatry 54[Suppl]

Arato M, O'Connor R, Meltzer H, Bradbury J (1997) Ziprasidone: efficacy in the prevention of relapse and in the long-term treatment of negative symptoms of chronic schizophrenia. Eur Neuropsychopharmacol 7[Suppl 2]: 214

*Arvanitis LA, Miller BG, Seroquel Trial 13 Study Group (1997) Multiple fixed doses of "Seroquel" (quetiapine) in patients with acute exacerbation of schizophrenia: a comparison with haloperidol and placebo. Biol Psychiatry 42: 233–246

Baldessarini RJ, Cohen BM, Teicher MH (1988) Significance of neuroleptic dose and plasma level in the pharmacological treatment of psychoses. Arch Gen Psychiatry 45: 79–91

Barnas C, Stuppäck C, Miller C, Haring C, Sperner-Unterweger B, Fleischhacker WW (1992) Zotepine in the treatment of schizophrenic patients with prevailingly negative symptoms:

a double blind trial vs. haloperidol. Int Clin Psychopharmacol 7: 23–27

*Bäuml J, Kissling W, Buttner P, Pitschell-Walz G, Mayer C, Boerner R, Engel R, Peuker I, Welschehold M (1993) Informationszentrierte Patienten- und Angehörigengruppen zur Complianceverbesserung bei schizophrenen Psychosen. Verhaltenstherapie 3[Suppl 1]: 1–96

*Beasley CM, Tollefson G, Tran P, Satterlee W, Sanger T, Hamilton S (1996) Olanzapine versus placebo and haloperidol. Neuropsychopharmacology 14: 111–123

*Bissette G, Nemeroff CB (1995) The neurobiology of neurotensin. In: Bloom FE, Kupfer DJ (eds) Psychopharmacology: the fourth generation of progress. Raven, New York, pp 573–583

Blackwell B (1973) Drug therapy. N Engl J Med 2: 249–252

Bondolfi G, Dufour H, Patris M, May JP, Billeter U, Eap CB, Baumann P, on Behalf of the Risperidone Study Group (1998) Risperidone versus clozapine in treatment-resistant chronic schizophrenia: a randomized double-blind study. Am J Psychiatry 155: 499–504

BorisonRL, Arvanitis LA, Miller BG, Seroquel Study Group (1996) ICI 204.636. An atypical antipsychotic: efficacy and safety in a multicenter, placebo controlled trial in patients with schizophrenia. J Clin Psychopharmacol 16: 158–169

Buchanan A (1992) A two-year prospective study of treatment compliance in patients with schizophrenia. Psychol Med 22: 787–797

*Bunney BG, Bunney WE, Carlsson A (1995) Schizophrenia and glutamate. In: Bloom FE, Kupfer DJ (eds) Psychopharmacology: the fourth generation of progress. Raven, New York, pp 1205–1214

Byerly MJ, DeVane CL (1996) Pharmacokinetics of clozapine and risperidone: a review of recent literature. J Clin Psychopharmacol 16: 177–187

Bymaster FP, Calligaro DO, Falcone JF, Marsh RD, Moore NA, Tye NC, Seeman P, Wong DT (1996) Radioreceptor binding profile of the atypical antipsychotic olanzapine. Neuropsychopharmacology 14: 87–96

Caligiuri MP, Lohr JB, Jeste DV (1993) Parkinsonism in neuroleptic-naive schizophrenic patients. Am J Psychiatry 150: 1343–1348

Carpenter WT (1996) The treatment of negative symptoms: pharmacological and methodological issues. Br J Psychiatry 168[Suppl 29]: 17–22

Carpenter WT, Hanlon TE, Heinrichs DW, Summerfelt AT, Kirkpatrick B, Levine J, Buchanan RW (1990) Continuous versus targeted medication in schizophrenic outpatients: outcome results. Am J Psychiatry 147: 1138–1148

*Casey DE (1996) Extrapyramidal syndromes: epidemiology, pathophysiology and the diagnostic dilemma. CNS Drugs 5[Suppl 1]: 1–12

Chatterjee A, Chakos M, Koreen AR, Geisler S, Sheitman B, Woerner M, Kane JM, Alvir JMJ, Lieberman MA (1995) Prevalence and clinical correlates of extrapyramidal signs and spontaneous dyskinesia in never-medicated schizophrenic patients. Am J Psychiatry 152: 1724 -1729

Claghorn J, Honigfeld G, Abuzzahab FS, Wang R, Steinbook R, Tuason V, Klerman GL (1987) The risks and benefits of clozapine versus chlorpromazine. J Clin Psychopharmacol 7: 377–384

Collins PJ, Larkin EP, Shubsachs APW (1991) Lithium carbonate in chronic schizophrenia – a brief trial of lithium carbonate added to neuroleptics for treatment of resistant schizophrenic patients. Acta Psychiatr Scand 84: 150–154

Cooper SJ, Butler A, Tweed J, Welch C, Raniwalla J (2000) Zotepine in the prevention of recurrence: a double-blind, placebo controlled study in chronic schizophrenics. Psychopharmacology (accepted for publication)

Coryell W, Miller DD, Perry PJ (1998) Haloperidol plasma levels and dose optimization. Am J Psychiatry 155: 48–53

Crow TJ, McMillan JF, Johnson AL, Johnstone EC (1986) The Northwick Park study of first episodes of schizophrenia. II. A randomized controlled trial of prophylactic neuroleptic treatment. Br J Psychiatry 148: 120–127

*Csernansky JG, Newcomer JG (1995) Maintenance drug treatment for schizophrenia. In: Bloom FE, Kupfer DJ (eds) Psychopharmacology: the fourth generation of progress. Raven, New York, pp 1267–1275

Daniel DG, Wozniak P, Mack RJ, McCarthy BG, Sertindole Study Group (1998) Long-term efficacy and safety comparison of sertindole and haloperidol in the treatment of schizophrenia. Psychopharmacol Bull 34: 61–69

Danion JM, Neunreuther C, Krieger-Finance F, Imbs JL, Singer L (1987) Compliance with long-term lithium treatment in major affective disorders. Pharmacopsychiatry 20: 230–231

Davis JM, Kane JM, Marder S, Brauzer B, Gierl B, Schooler NR, Casey DE, Hassan M (1993) Dose response of prophylactic antipsychotics. J Clin Psychiatry 54[Suppl 3]: 24–30

*Davis JM, Matalon L, Watanabe MD, Blake LM (1994) Depot antipsychotic drugs; place in therapy. Drugs 47: 741–773

*Davis R, Markham A (1997) Ziprasidone. CNS Drugs 8: 153–159

Decina P, Mukherjee S, Bocola V, Saraceni F, Hadjichristos C, Scapicchio P (1994) Adjunctive trazodone in the treatment of negative symptoms of schizophrenia. Hosp Comm Psychiatry 45: 1220–1223

Deutch AY, Duman RS (1996) The effects of antipsychotic drugs on Fos protein expression in the prefrontal cortex: cellular localization and pharmacological characterization. Neuroscience 70: 377–389

Dieterle DM, Müller-Spahn F, Ackenheil M (1991) Efficacy and tolerance of zotepine in a double-blind comparison with perazine in schizophrenics. Fortschr Neurol Psychiatr 59[Suppl 4]: 18–22

Drake RE, Osher FC, Wallach MA (1989) Alcohol use and abuse in schizophrenia. J Nerv Ment Dis 177: 408–414

Duinkerke SJ, Botter PA, Jansen AAI, Van Dongen PAM, Van Haaften AJ, Boom AJ, Van Laarhoven JHM, Busard HLSM (1993) Ritanserin, a selective 5-HT2/1C antagonist, and negative symptoms in schizophrenia: a placebo-controlled double-blind trial. Br J Psychiatry 163: 451–455

*Dunn CJ, Fitton A (1996) Sertindole. CNS Drugs 5: 224–230

Eckman TA, Liberman RP, Phipps CC, Blair KE (1990) Teaching medication management skills to schizophrenic patients. J Clin Psychopharmacol 10: 33–38

*Farde L, Nyberg S (1998) Dosing determination for novel antipsychotics – a PET-based approach. Int J Psychiatr Clin Pract 2[Suppl 1]: 39–42

Farde L, Nordström AL, Wiesel FA, Pauli S, Halldin C, Sedvall G (1992) PET analysis of central D, and D2 dopamine receptor occupancy in patients treated with classic neuroleptics and clozapine – relationship to extrapyramidal side effects. Arch Gen Psychiatry 49: 538–544

Fayen M, Goldman MB, Moulthrop MA, Luchins DJ (1988) Differential memory function with dopaminergic versus

anticholinergic treatment of drug induced extrapyramidal symptoms. Am J Psychiatry 145: 483–486

Fenton WS, Blyler CR, Heinssen RK (1997) Determinants of medication compliance in schizophrenia: empirical and clinical findings. Schizophr Res 23: 637–651

Fischer-Cornelssen KA, Ferner VJ (1976) An example of European multicenter trials: multispectral analysis of clozapine. Psychopharmacol Bull 2: 34–39

Fitton A, Heel R (1990) Clozapine. A review of its pharmacological properties, and therapeutic use in schizophrenia. Drugs 40: 722–747

Fleischhacker WW (1999) Clozapine: a comparison with other novel antipsychotics. J Clin Psychiatry 60[Suppl 12]: 30–34

Fleischhacker W, Hummer M (1997) Drug treatment of schizophrenia in the 1990s: achievements and future possibilities in optimising outcomes. Drugs 53: 915–929

Fleischhacker WW, Barnas C, Günther V, Meise U, Stuppäck CH, Unterweger B (1987) Mood-altering effects of biperiden in healthy volunteers. J Affect Disord 12: 153–157

Fleischhacker WW, Barnas C, Stuppäck CH, Unterweger B, Miller CH, Hinterhuber H (1989) Zotepine vs. haloperidol in paranoid schizophrenia: a double-blind trial. Psychopharmacol Bull 25: 97–100

Fleischhacker WW, Roth SD, Kane JM (1990) The pharmacologic treatment of neuroleptic-induced akathisia. J Clin Psychopharmacol 10: 12–21

Fleischhacker WW, Hummer M, Kurz M, Kurzthaler I, Lieberman J, Pollack S, Safferman A, Kane J (1994a) Clozapine dose in the US and Europe: implications for therapeutic and adverse effects. J Clin Psychiatry 55[Suppl B]: 78–81

Fleischhacker WW, Meise U, Günther V, Kurz M (1994b) Compliance with antipsychotic drug treatment: influence of side effects. Acta Psychiatr Scand [Suppl 382]: 11–15

Fulton B, Goa KL (1995) ICI-204,636. An initial appraisal of its pharmacological properties and clinical potential in the treatment of schizophrenia. CNS Drugs 4: 68–78

Gaebel W, Falkai P (1998) Praxisleitlinien in Psychiatrie und Psychotherapie, vol 1: Behandlungsleitlinie Schizophrenie. Steinkopff, Darmstadt

Gardos G, Cole JO (1995) The treatment of tardive dyskinesia. In: Bloom FE, Kupfer DJ (eds) Psychopharmacology: the fourth generation of progress. Raven, New York, 1503–1511

*Gardos G, Casey DE, Cole JO, Perenyi A, Kocsis E, Arato M, Samson J, Conley C (1994) Ten-year outcome of tardive dyskinesia. Am J Psychiatry 151: 836–841

Gerlach I, Koppelhus P, Helweg E, Monrad A (1974) Clozapine and haloperidol in a single-blind cross-over trial: therapeutic and biochemical aspects in the treatment of schizophrenia. Acta Psychiatr Scand 50: 410–424

*Gilbert PL, Harris MJ, McAdams LA, Jeste DV (1995) Neuroleptic withdrawal in schizophrenic patients. Arch Gen Psychiatry 52: 173–188

Glazer WM, Johnstone BM (1997) Pharmacoeconomic evaluation of antipsychotic therapy for schizophrenia. J Clin Psychiatry 58[Suppl 10]: 50–54

Goff DC, Brotman AW, Waites M, McCormick S (1990) Trial of fluoxetine added to neuroleptics for treatment-resistant schizophrenic patients. Am J Psychiatry 47: 492–494

Goff DC, Kamal KM, Sarid-Segal O, Hubbard JW, Amico E (1995) A placebo-controlled trial of fluoxetine added to neuroleptics in patients with schizophrenia. Psychopharmacology 117: 417–423

Goff DC, Posever T, Herz L, Simmons J, Kletti N, Lapierre K, Wilner KD, Law CG, Ko GN (1998) An exploratory haloperidol-controlled dose-finding study of ziprasidone in hospitalized patients with schizophrenia or schizoaffective disorder. J Clin Psychopharmacol 18: 296–304

*Goldberg TE, Gold JM (1995) Neurocognitive deficits in schizophrenia. In: Hirsch SR, Weinberger DR (eds) Schizophrenia. Blackwell, Oxford, pp 146–162

Green M, Marshall B, Wirshing W, Ames D, Marder S, McGurk S, Kern RS, Mintz J (1997) Does risperidone improve verbal working memory in treatment-resistant schizophrenia? Am J Psychiatry 154: 799–804

Growe GA, Crayton JA, Klass DB (1979) Lithium in chronic schizophrenia. Am J Psychiatry 136: 454–455

Hale A, Van der Burght M, Wehnert A, Friberg HH (1996) A European dose-range study comparing the efficacy, tolerability and safety of four doses of sertindole and one dose of haloperidol in schizophrenic patients. Poster presented at the 20th CINP Congress, Melbourne, 23–27 June

Hamilton SH, Revicki DA, Genduso LA, Beasley CM Jr. (1998) Olanzapine versus placebo and haloperidol: quality of life and efficacy results of the North American double-blind trial. Neuropsychopharmacology 18: 41–49

*Haring C, Fleischhacker WW, Schett P, Humpel C, Barnas C, Saria A (1990) Influence of patient-related variables on clozapine plasma levels. Am J Psychiatry 147: 1471–1475

Haring C, Neudorfer C, Schwitzer J, Hummer M, Saria A, Hinterhuber H, Fleischhacker WW (1994) EEG alterations in patients treated with clozapine in relation to plasma levels. Psychopharmacology 114: 97–100

Herz MI, Glazer WM, Mostert MA, Sheard MA, Szymanski HV (1991) Intermittent vs. maintenance medication in schizophrenia: two-year results. Arch Gen Psychiatry 48: 333–339

Hiemke C, Weigmann H, Härtter S, Dahmen N, Wetzel H, Müller H (1994) Elevated serum levels of clozapine after addition of fluvoxamine. Clin Psychopharmacol 14: 279–281

Hirsch S, Link CG, Goldstein JM, Arvanitis LA (1996) ICI 204,636. A new atypical antipsychotic drug. Br J Psychiatry 168[Suppl 29]: 45–56

*Hoge SK, Appelbaum PS, Lawlor T, Beck JC, Litman F, Greer A, Gutheil TG, Kaglun E (1990) A prospective, multicenter study of patients' refusal of antipsychotic medication. Arch Gen Psychiatry 47: 949–956

Hummer M, Fleischhacker WW (1999) Ways of improving compliance. In: Lader M, Naber D (eds) Difficult clinical problems in psychiatry. Martin Dunitz, London, pp 229–238

*Hummer M, Kemmler G, Kurz M, Kurzthaler I, Oberbauer H, Fleischhacker WW (1999) Sexual disturbances during clozapine and haloperidol treatment. Am J Psychiatry 156: 631–633

Janicak PG, Javaid JI, Sharma RP, Leach A, Dowd S, Davis JM (1997) A two-phase, double-blind randomized study of three haloperidol plasma levels for acute psychosis with reassignment of initial non-responders. Acta Psychiatr Scand 95: 343–350

Johnson DAW (1979) Further observations on the duration of depot neuroleptic maintenance therapy in schizophrenia. Br J Psychiatry 135: 524–530

Jolley AG, Hirsch SR, McRink A, Manchanda R (1989) Trial of brief intermittent neuroleptic prophylaxis for selected schizophrenic outpatients: clinical outcome at one year. Br Med J 298: 985–990

*Kane JM, Lieberman JA (1987) Maintenance pharmacotherapy in schizophrenia. In: Meltzer HY (ed) Psychopharma-

cology – the third generation of progress. Raven, New York, pp 1103–1109

Kane JM, Honigfeld G, Singer J, Meltzer HY (1988a) Clozapine for the treatment-resistant schizophrenic. Arch Gen Psychiatry 45: 789–796

*Kane JM, Woerner M, Lieberman J (1988b) Tardive dyskinesia: prevalence, incidence and risk factors. J Clin Psychopharmacol 8: 52S–56S

Kane JM, Aguglia E, Altamura AC et al (1998) Guidelines for depot antipsychotic treatment in schizophrenia. Eur Neuropsychopharmacol 8(1): 55–66

Keck P Jr, Buffenstein A, Ferguson J, Feighner J, Jaffe W, Harrigan EP, Morrissey MR, Ziprasidone Study Group (1998) Ziprasidone 40 and 120 mg/day in the acute exacerbation of schizophrenia and schizoaffective disorder: a 4-week placebo-controlled trial. Psychopharmacology 140: 173–184

*Kemp R, Kirov G, Everitt B, Hayward P, David A (1998) Randomised controlled trial of compliance therapy. Br J Psychiatry 172: 413–419

Kinon BJ, Kane JM, Johns C, Perovich R, Ismi M, Koreen AR, Weiden PJ (1993) Treatment of neuroleptic-resistant schizophrenic relapse. Psychopharmacol Bull 29: 309–314

Kissling W (1991) Duration of neuroleptic maintenance treatment. In: Kissling W (ed) Guidelines for neuroleptic relapse prevention in schizophrenia. Springer, Berlin Heidelberg New York, pp 94–112

*Kissling W, Kane JM, Barnes TRE, Dencker SJ, Fleischhacker WW, Goldstein JM, Johnson DAW, Marder SR, Müller-Spahn F, Tegeler J et al (1991) Guidelines for neuroleptic relapse prevention in schizophrenia: towards a consensus view. In: Kissling W (ed) Guidelines for neuroleptic relapse prevention in schizophrenia. Springer, Berlin Heidelberg New York, pp 155–163

Klieser E, Lehmann E, Tegeler J (1991) Double-blind comparison of 3 × 75 mg zotepine and 3 × 4 mg haloperidol in acute schizophrenics. Fortschr Neurol Psychiatr 59[Suppl 1]: 14–17

Kolakowska T, Williams AO, Ardern M, Reveley MA, Jambor K, Gelder MG, Mandelbrote BM (1985) Schizophrenia with good and poor outcome. I. Early clinical features, response to neuroleptics and signs of organic dysfunction. Br J Psychiatry 146: 229–246

*Kraepelin E (1919) Dementia praecox and paraphrenia. Livingstone, Edinburgh

*Krueger RB, Sackheim HA (1995) Electroconvulsive therapy and schizophrenia. In: Hirsch SR, Weinberger DR (eds) Schizophrenia. Blackwell Science, Oxford, pp 503–545

Kurz M, Hummer M, Oberbauer H, Fleischhacker WW (1995a) Extrapyramidal side effects of clozapine and haloperidol. Psychopharmacology 118: 52–56

Kurz M, Hummer M, Kurzthaler I, Oberbauer H, Fleischhacker WW (1995b) Efficacy of medium-dose clozapine for treatment resistant schizophrenia. Am J Psychiatry 152: 1690–1691

*Lehmann AF, Steinwachs DM, Co-investigators of the PORT project (1998) At issue: translating research into practice: the schizophrenia patient outcomes research team (PORT) treatment recommendations. Schizophr Bull 24: 1–10

Lemmens P, Brecher M, Van Baelen B (1999) A combined analysis of double-blind studies with risperidone vs. placebo and other antipsychotic agents: factors associated with extrapyramidal symptoms. Acta Psychiatr Scand 99(3): 160–170

Levinson DF, Simpson GM, Singh H, Yadalam K, Jain A, Stephanos MJ, Silver P (1990) Fluphenazine dose, clinical response, and extrapyramidal symptoms during acute treatment. Arch Gen Psychiatry 47: 761–768

Levinson DF, Singh H, Simpson GM (1992) Timing of acute clinical response to fluphenazine. Br J Psychiatry 160: 365–371

Levinson DF, Simpson GM, Lo ES, Cooper TB, Singh H, Yadalam K, Stephanos MJ (1995) Fluphenazine plasma levels, dosage, efficacy, and side effects. Am J Psychiatry 152: 765–771

Ley P, Spelman MS (1965) Communication in an outpatient setting. Br J Soc Clin Psychol 4: 114–116

Liddle PF, Barnes TRE, Curson DA, Patel M (1993) Depression and the experience of psychological deficits in schizophrenia. Acta Psychiatr Scand 88: 243–247

Loebel A, Lieberman J, Alvir J, Geisler J, Koreen A, Chakos M (1995) Time to treatment response in successive episodes of early onset schizophrenia. Schizophr Res 15: 158

*Loo H, Poirier-Littre M-F, Theron M, Rein W, Fleurot O (1997) Amisulpride versus placebo in the medium-term treatment of the negative symptoms of schizophrenia. Br J Psychiatry 170: 18–22

*Maj M, Sartorius N (eds) (1999) Schizophrenia. Wiley, New York

*Mandoki MW (1995) Risperidone treatment of children and adolescents: increased risk of extrapyramidal side effects? J Child Adolesc Psychopharmacol 5: 49–67

Marder SR (1998) Facilitating compliance with antipsychotic medication. J Clin Psychiatry 59[Suppl 3]: 21–25

Marder SR, Meibach RC (1994) Risperidone in the treatment of schizophrenia. Am J Psychiatry 151: 825–835

McEvoy JP, Hogarty GE, Steingard S (1991) Optimal dose of neuroleptics in acute schizophrenia. Arch Gen Psychiatry 48: 739–745

Meise U, Günther V, Gritsch S (1992) Die Bedeutung der Arzt-Patienten-Beziehung für die Patientencompliance. Wien Klin Wochenschr 104: 267–271

Meltzer HY (1989) Duration of a clozapine trial in neuroleptic-resistant schizophrenia. Arch Gen Psychiatry 46: 672

Meltzer HY, Okayli G (1995) The reduction of suicidality during clozapine treatment in neuroleptic-resistant schizophrenia: impact on risk-benefit assessment. Am J Psychiatry 152: 183–190

Meltzer HY, Burnett S, Bastani B, Ramirez LF (1990) Effects of 6 months of clozapine treatment on the quality of life of chronic schizophrenic patients. Hosp Comm Psychiatry 41: 892–897

*Middelboe T (1995) Predictors of treatment compliance in long-term mentally ill. Eur Neuropsychopharmacol 5(3): 318

*Miller CH, Fleischhacker WW (2000) Managing antipsychotic-induced acute and chronic akathisia. Drug Safety 22(1): 73–81

Miller DD, Fleming F, Holman TL, Perry PJ (1994) Plasma clozapine concentrations as a predictor of clinical response: a follow-up study. J Clin Psychiatry 55[Suppl 13]: 117–121

Möller HJ (1995) The negative component in schizophrenia. Acta Psychiatr Scand 91[Suppl 388]: 11–14

*Möller HJ, van Zerssen D (1995) Course and outcome of schizophrenia. In: Hirsch SR, Weinberger DR (eds) Schizophrenia. Blackwell Science, Oxford, pp 106–127

Mortimer A (1997) Cognitive function in schizophrenia – do neuroleptics make a difference? Pharmacol Biochem Behav 56: 789–795

Nordström A, Farde L, Nyberg S, Karlsson P, Halldin C, Sedvall G (1995) D1, D2, and 5-HT2 receptor occupancy in relation to clozapine serum concentration: a PET study of schizophrenic patients. Am J Psychiatry 152: 1444–1449

Nyberg S, Farde L, Halldin C (1997) A PET study of 5-HT2 and D2 clozapine receptor occupancy induced by olanzapine in healthy subjects. Neuropsychopharmacology 16: 1–7

Pan PC, Tantam D (1989) Clinical characteristics, health beliefs and compliance with maintenance treatment: a comparison between regular and irregular attenders at a depot clinic. Acta Psychiatr Scand 79: 564–570

Perlick D, Mattis S, Stasny P, Teresi J (1992) Neuropsychological discriminators of long-term inpatient or outpatient status in chronic schizophrenia. J Neuropsychiatr Clin Neurosci 4: 428–434

Petit M, Raniwalla J, Tweed J, Leutenegger E, Dollfus S, Kelly F (1996) A comparison of an atypical and typical antipsychotic, zotepine versus haloperidol in patients with acute exacerbation of schizophrenia: a parallel group double-blind trial. Psychopharmacol Bull 32: 81–87

*Peuskens J (1995) Risperidone in the treatment of patients with chronic schizophrenia: a multi-national, multi-centre, double-blind, parallel-group study versus haloperidol. Br J Psychiatry 166: 712–726

Pfeiffer W, Kockott G, Fischl B, Schleuning G (1991) Unerwünschte Wirkungen psychopharmakologischer Langzeittherapie auf die sexuellen Funktionen. Psychiatr Prax 18: 92–98

Pietzcker A, Gaebel W, Kopcke W, Linden M, Müller P, Müller-Spahn F, Tegeler J (1993) Intermittent versus maintenance neuroleptic long-term treatment in schizophrenia – 2-year results of a German multicenter study. J Psychiatr Res 27: 321–339

Pilowsky LS, O'Connell P, Davies N, Busatto GF, Costa DC, Murray PJ, Kerwin RW (1997) In vivo effects on striatal dopamine D2 receptor binding by the novel atypical antipsychotic drug sertindole – a 123 I IBZM single photon emission tomography (SPET) study. Psychopharmacology 130: 152–158

Potkin SG, Bera R, Gulasekaram B, Jin Y, Costa J, Gerber B, Richmond G, Ploszaj D, Carreon D, Cooper T, Sitanggan K (1994) Plasma clozapine concentrations predict clinical response in treatment-resistant schizophrenia. J Clin Psychiatry 55[Suppl B]: 133–136

*Prakash A, Lamb HM (1998) Zotepine: a review of its pharmacodynamic and pharmacokinetic properties and therapeutic efficacy in the management of schizophrenia. CNS Drugs 9: 154–175

Purdon SE, Jones BD, Stip E, Labelle A, Addington D, David SR, Breier A, Tollefson GD (2000) Neuropsychological change in early phase schizophrenia during 12 month treatment with olanzapine, risperidone, or haloperidol. The Canadian Collaborative Group for research in schizophrenia. Arch Gen Psychiatry 57(3): 249–258

*Remington G, Bezchlibnyk-Butier K (1996) Management of acute antipsychotic-induced extrapyramidal syndromes. CNS Drugs 5[Suppl 1]: 21–35

Rifkin A, Doddi S, Karagi B (1991) Dosage of haloperidol for schizophrenia. Arch Gen Psychiatry 48: 166–170

Rosenheck R, Cramer J, Xu W, Thomas J, Henderson W, Frisman L, Fye C, Charney D (1997) A comparison of clozapine and haloperidol in hospitalized patients with refractory schizophrenia. N Engl J Med 337: 809–815

*Roth BL, Meltzer HY (1995) The role of serotonin in schizophrenia. In: Bloom FE, Kupfer DJ (eds) Psychopharmacology: the fourth generation of progress. Raven, New York, pp 1215–1227

Rowley H, Kilpatrick L, Needham P, Heat D (1998) Elevation of extracellular cortical noradrenaline may contribute to the antidepressant activity of zotepine: an in vivo microdialysis study in freely moving rats. Neuropharmacology 37: 937–944

Roy A (1990) Relationship between depression and suicidal behaviour in schizophrenia. In: DeLisi LE (ed) Depression in schizophrenia. American Psychiatric Press, Washington, DC

Saller FC, Salama AJ (1993) Seroquel: biochemical profile of a potential atypical antipsychotic. Psychopharmacology 112: 285–292

Schall U, Catts S, Chaturvedi S, Liebert B, Redenbach J, Karayanidis F, Ward P (1998) The effect of clozapine therapy on frontal lobe dysfunction in schizophrenia: neuropsychology and event-related potential measures. Int J Neuropsychopharmacol 1: 19–29

*Schmutz J, Hunziker F, Stille G, Lauener H (1967) Constitution chimique et action pharmacologique d'un nouveau groupe de neuroleptiques tricycliques. Bull Chim Therapeut 424

Schooler NR, Keith SJ, Severe JB, Matthews SM, Bellack AS, Glick ID, Hargreaves WA, Kane JM, Ninan PT, Frances A et al (1997) Relapse and rehospitalization during maintenance treatment of schizophrenia – the effects of dose reduction and family treatment. Arch Gen Psychiatry 54: 453–463

Schou M (1997) The combat of non-compliance during prophylactic lithium treatment. Acta Psychiatr Scand 95: 361–363

Schultz SK, Miller DD, Arndt S, Ziebell S, Gupta S, Andreasen NC (1995) Withdrawal-emergent dyskinesia in patients with schizophrenia during antipsychotic discontinuation. Biol Psychiatry 38: 713–719

Schulz SC, Kahn EM, Baker RW, Conley RR (1990) Lithium and carbamazepine augmentation in treatment refractory schizophrenia. In: Angrist B, Schulz SC (eds) The neuroleptic nonresponsive patient: characterization and treatment. American Psychiatric Association, Washington DC, pp 109–136

Schwartz D, Wang W, Zeitz L, Goss ME (1962) Medication errors made by elderly, chronically ill patients. Am J Public Health 52: 2018–2029

Shalev A, Hermesh H, Rothberg J, Munitz H (1993) Poor neuroleptic response in acutely exacerbated schizophrenic patients. Acta Psychiatr Scand 87: 86–91

Sharma T, Mockler D (1998) The cognitive efficacy of atypical antipsychotics in schizophrenia. J Clin Psychopharmacol 18[Suppl 1]: 12S–19S

Shiloh R, Zemishlany Z, Aizenberg D, Radwan M, Schwartz B, Dorfman-Etrog P, Modai L, Khaikin M, Weitman A (1997) Sulpiride augmentation in people with schizophrenia partially responsive to clozapine: a double-blind, placebo-controlled study. Br J Psychiatry 171: 569–573

Silver H, Kushnir M, Kaplan A (1996) Fluvoxamine augmentation in clozapine-resistant schizophrenia: an open pilot study. Biol Psychiatry 40: 671–674

Silverstone T, Smith G, Goodall E (1988) Prevalence of obesity in patients receiving depot antipsychotics. Br J Psychiatry 153: 214–217

*Siris SG (1995) Depression and schizophrenia. In: Hirsch SR, Weinberger DR (eds) Schizophrenia. Blackwell Science, Oxford, pp 128–145

Skarsfeldt T (1995) Differential effects of repeated administration of novel antipsychotic drugs on the activity of midbrain dopamine neurons in the rat. Eur J Pharmacol 281: 289–294

Small JG, Hirsch SR, Arvanitis LA, Miller BG, Link CGG, Seroquel Study Group (1997) Quetiapine in patients with schizophrenia: a high- and low-dose double-blind comparison with placebo. Arch Gen Psychiatry 54: 549–557

Smith JM (1980) Abuse of antiparkinsonian drugs – a review of the literature. J Clin Psychiatry 41: 351–354

Steingard S, Allen M, Schooler NR (1994) A study of the pharmacologic treatment of medication-compliant schizophrenics who relapse. J Clin Psychiatry 55: 470–472

Sternlicht HC, Wells SR (1995) Risperidone in childhood schizophrenia. J Am Acad Child Adolesc Psychiatry 34: 5

*Stille G, Hippius H (1971) Kritische Stellungnahme zum Begriff der Neuroleptika. Pharmakopsychiatr Neuropharmakol 4: 182–191

Swett J, Noones J (1989) Factors associated with premature termination from outpatient treatment. Hosp Comm Psychiatr 40: 947–951

Tandon R, Mann NA, Eisner WH, Coppard N (1990) Effect of anticholinergic medication on positive and negative symptoms in medication free schizophrenic patients. Psychiatry Res 31: 235–241

Tandon R, Harrigan E, Zorn SH (1997) Ziprasidone: a novel antipsychotic with unique pharmacology and therapeutic potential. J Serotonin Res 4: 159–177

*Tollefson GD, Beasley CM, Tamura RN, Tran PV, Potvin JH (1997a) Blind, controlled, long-term study of the comparative incidence of treatment-emergent tardive dyskinesia with olanzapine or haloperidol. Am J Psychiatry 154: 1248–1254

Tollefson GD, Beasley CM, Tran PV, Street JS, Krueger JA, Tamura RN, Graffeo KA, Thierne ME (1997b) Olanzapine versus haloperidol in the treatment of schizophrenia and schizoaffective and schizophreniform disorders: results of an international collaborative trial. Am J Psychiatry 154: 457–465

Tollefson GD, Sanger TM, Beasley CM (1997c) Negative symptoms: a path analytic approach to a double-blind, placebo- and haloperidol-controlled clinical trial with olanzapine. Am J Psychiatry 154: 466–474

Tollefson GD, Sanger TM, Thierne ME (1998) Depressive signs and symptoms in schizophrenia: a prospective blinded trial of olanzapine and haloperidol. Arch Gen Psychiatry 55: 250–258

Tran PV, Dellva MA, Tollefson GD, Beasley CM, Potvin JH, Kiesler GM (1997) Extrapyramidal symptoms and tolerability of olanzapine versus haloperidol in the acute treatment of schizophrenia. J Clin Psychiatry 58: 205–211

Tran PV, Dellva MA, Tollefson GD, Wentley AL, Beasley CM (1998) Oral olanzapine versus oral haloperidol in the maintenance treatment of schizophrenia and related psychoses. Br J Psychiatry 172: 499–505

Travis MJ, Busatto GF, Pilowsky LS, Mulligan R, Acton PD, Gacinovic S, Mertens J, Terrière D, Costa DC, Ell PJ et al (1998) 5-HT2A receptor blockade in patients with schizophrenia treated with risperidone or clozapine: a SPECT study using the novel 5-HT2A ligand 123 I-5-I-R-91150. Br J Psychiatry 173: 236–241

*VanderZwaag C, McGee M, McEvoy JP, Freudenreich O, Wilson WH, Cooper TB (1996) Response of patients with treatment-refractory schizophrenia to clozapine within three serum level ranges. Am J Psychiatry 153: 1579–1584

Van Harten PN, Hoek HW, Matroos GE, Koeter M, Kahn RS (1998) Intermittent neuroleptic treatment and risk for tardive dyskinesia: Curacao extrapyramidal syndromes study III. Am J Psychiatry 155: 565–567

Van Kammen DP, McEvoy JP, Targum SD, Kardatzke D, Sebree T, Sertindole Study Group (1996) A randomized, controlled, dose-ranging trial of sertindole in patients with schizophrenia. Psychopharmacology 124: 168–175

Van Putten T (1974) Why do schizophrenic patients refuse to take their drugs? Arch Gen Psychiatry 31: 67–72

Van Putten T, Marder SR, Mintz J, Poland RE (1992) Haloperidol plasma levels and clinical response: a therapeutic window relationship. Am J Psychiatry 149: 500–505

*Volavka J, Cooper T, Czobor P, Bitter I, Meisner M, Laska E, Gastanaga P, Krakowski M, Chou SCY, Crowner M et al (1992) Haloperidol blood levels and clinical effects. Arch Gen Psychiatry 49: 354–361

Weiss F, Danzl C, Hummer M, Kemmler G, Lindner C, Reinstadler K, Fleischhacker WW (1998) Weight gain induced by olanzapine. Schizophr Res 29 (special issue 1–2): 179

Whitworth AB, Fleischhacker WW (1995) Adverse effects of antipsychotic drugs. Int Clin Psychopharmacol 9[Suppl 5]: 21–27

Wilson WH (1993) Addition of lithium to haloperidol in non-affective, antipsychotic non-responsive schizophrenia: a double blind, placebo controlled, parallel design clinical trial. Psychopharmacology 111: 359–366

*Wirshing WC, Marder SR, Van Putten T, Ames D (1995) Acute treatment of schizophrenia. In: Bloom FE, Kupfer DJ (eds) Psychopharmacology: the fourth generation of progress. Raven, New York, pp 1259–1266

*Wolkowitz OM, Rapoport MH, Pickar D (1990) Benzodiazepine augmentation of neuroleptics. In: Angrist B, Schulz SC (eds) The neuroleptic nonresponsive patient: characterization and treatment. American Psychiatric Association, Washington DC, pp 87–108

World Health Organization (1990) Prophylactic use of anticholinergics in patients on long-term neuroleptic treatment. Br J Psychiatry 156: 412–414

Wyatt RJ (1992) Neuroleptics and the natural course of schizophrenia. Schizophr Bull 17: 325–351

Zimbroff DL, Kane JM, Tamminga CA, Daniel DG, Mack RJ, Wozniak TJ, Sebree TB, Wallin BA, Kashkin KB, Sertindole Study Group (1997) Controlled, dose-response study of sertindole and haloperidol in the treatment of schizophrenia. Am J Psychiatry 154: 782–791

CHAPTER 12

H.D. Brenner, H. Hoffmann, H. Heise

Sociotherapy and Psychotherapy of Schizophrenic Disorders

Translator: E. Taub

1 Introduction

Although the treatment of schizophrenic patients without sociotherapeutic interventions would be unthinkable today, the importance of psychotherapy in the narrower sense for these patients remains as controversial today as in the past. These two forms of therapy have come closer to one another in recent years, and it therefore hardly seems reasonable now to discuss them separately; however, each of them emerged from, and must be understood as the product of, a particular historical development.

If we take sociotherapy to include all types of treatment dealing with the social environment of the patient, then sociotherapy is certainly as old as psychiatry itself. In this sense, the unchaining of the patients of the Salpêtrière in 1772 on the order of Pinel must count as an early and much-applauded example of sociotherapeutic intervention. In the English-speaking countries, too, various reform movements devoted to a general improvement in the living conditions of the mentally ill, and advocating specific socio- and psychotherapeutic interventions, arose as early as the beginning of the nineteenth century. Especially worth mentioning in this context are the concepts of "moral treatment," based on the "open door" principle and the special training of nursing staff, and "no-restraint therapy."

In Germany, the social element was, at first, a strong emphasis of romantic psychiatry but was soon replaced, at least on the level of the construction of theories, by other focuses of interest. In practice, however, continuing efforts were made to reform psychiatric institutions and integrate mental patients into society. Griesinger's conception of the *Stadtasyl* (city asylum) (Griesinger 1868/1869) was, indeed, an early forerunner of many of the current principles of community-based psychiatric care.

Further developments in the second half of the nineteenth century seem, at first glance, to be characterized largely by an emphasis on somatic medical ideas and by the development of medically oriented university psychiatry. Nevertheless, even in this period, which has been called the age of "brain psychiatry," it was precisely the lack of effective somatic therapies in the everyday practice of the then dominant form of institutional psychiatry that forced even "somatically oriented" psychiatrists to apply sociotherapeutic treatment strategies and to develop them further. Indeed, many of these psychiatrists were thoroughly progressive with respect to their sociopsychiatric approach.

This complex interaction of multiple currents and forces is illustrated particularly vividly by the early history of work therapy. In the age before psychiatric institutions began to arise, the mobilization of the mentally ill into the workforce was more the rule than the exception, because exemption from labor was seen as a privilege of the aristocracy, for both the healthy and the mentally ill alike. The idea that the mentally ill had a just claim to medical treatment, which came about when psychiatric institutions were created, was followed at first by their large-scale exclusion from the workforce and then, shortly afterward, by a countermovement. While the proponents of "physical" treatment, by direct analogy to somatic medicine, advocated putting the brain at rest to promote healing, as would be done with any other diseased organ, the proponents of "moral" treatment thought that work had a major, beneficial therapeutic effect.

By 1850, patients had been put to work under medical supervision across a broad front in German psychiatric institutions, albeit mainly in helping in the running of the clinics themselves at first. Both therapeutic and economic considerations favored such a development. An early indication of the potential for conflict inherent to this situation is provided by the directive issued under Pienitz in 1828 at the *Königlich sächsische Heil- und Verpflegungsanstalt Sonnenstein* (Royal Saxon Healing and Boarding Institution Sonnenstein), to the effect that the work performed must be in accordance with the goal of healing (Willis and Reimer 1988).

In the second half of the nineteenth century, many institutions created special work opportunities for their patients, mainly in agriculture at first and then later also in industry (e.g. in Göppingen and Hildesheim). From 1880 onward, Koeppe elevated the concept of regular work to the dominating organizational principle of the Alt-Scherbitz Institution and, at the turn of the century, private sanatoria and healing institutions took the first steps toward the individualization of work therapy for middle-class patients. An emphasis was placed on personal motivation, and being able to carry out one's own occupation again was declared as the goal. This was the context in which two nearly synonymous terms were coined, *Arbeitstherapie* ("work therapy") and *Beschäftigungstherapie* ("keeping-busy therapy"), which were later adopted by institutional psychiatry to describe its own proffered forms of treatment.

During and after the First World War, an extreme shortage of staff and financial means made itself felt. It was in this setting that Hermann Simon, in 1919, assumed the direction of the not yet completed Gütersloh Institution and put the patients to work in finishing its construction. Once this work was over, he organized the clinic, in analogy to educational and military patterns, into classes, in each of which the patients were required to perform a different type of work. There was also a differential schedule of

disciplinary measures for refusal to work (H. Simon 1929). This model, tailored primarily to the needs of chronic patients, earned Simon the reputation of being the founder of work therapy.

The lack of effective somatic interventions that was still evident in the remainder of the first half of the twentieth century was also, in part, responsible for the rapid assumption of a central role in psychiatry by psychoanalysis, still a very young discipline at the time, and the consequent introduction of psychoanalytically inspired concepts into sociotherapy. Thus Sullivan (1931) was concerned with the creation of a "therapeutic milieu" and coined phrases such as "patient participation" or "patient government." These ideas paved the way for still more influential developments. While the understanding of mental illness had previously been based on medical or psychological conceptions relating to the individual patient, the findings of empirical social research began to have an impact on psychiatry in the 1940s in the English-speaking countries. Finally, against the background of sociopsychological theories such as those of Erikson, Parsons, and Lewin, the organizational structures of psychiatric institutions themselves and their effects on the course of mental illnesses came to be a topic of research and a point of departure for new conceptions of therapeutic communities.

The roots of sociotherapy can thus be traced back quite far into the past. Psychotherapy proper, in contrast, first arose in the latter years of the nineteenth century with the development of psychoanalysis. While Freud himself had low expectations for the success of psychoanalysis in schizophrenics, many of his pupils nonetheless made such attempts, mainly because any real therapeutic alternative was lacking until many years later.

An initial phase saw the development of an astounding variety of theoretical constructs, such as confrontation with a tyrannical superego (Hinsie 1930), strengthening of ego boundaries (Federn 1952), or unconditional motherly love (Schwing 1954). Klein (1930), however, advocated the nearly unchanged application of classical analytical techniques, as was done thereafter for many years, particularly in South America. Sullivan (1931) and Fromm-Reichmann (1950), during and after the Second World War, created an approach that was relatively far removed from its psychoanalytic roots and, in its practical application, already incorporated elements reminiscent of modern supportive and cognitive-behavioral techniques. The psychoanalytically oriented treatment of schizophrenia reached its zenith between 1945 and about 1965, after which the emphasis shifted to family therapy and, later, to cognitive-behavioral approaches. Nonetheless, the psychoanalytic therapy of schizophrenics is still alive today.

The different therapeutic orientations that emerged from these processes of historical development will be discussed individually below. Special attention will be paid to the present status of empirical research in these areas, in view of the current development of "guidelines" and the demand for quality assurance by means of "evidence-based medicine." Because, as already mentioned, sociotherapy and psychotherapy of schizophrenic patients are closely interconnected, the subdivision of topics below will not follow these two conceptual categories, but will rather be oriented toward the particular uses of, and differences between, the various therapeutic approaches in clinical practice.

2 Supportive Psychotherapy

The greatest number of therapeutic efforts in schizophrenic patients aside from pharmacological treatment fall into the category of supportive psychotherapy, earlier known in German as *Stütztherapie* ("support therapy"). For many years, the degree of theoretical interest in such measures was exceedingly low, despite their widespread application. This situation led Heim (1980) to remark:

> All of us practice these techniques to a greater or lesser degree; but we hardly ever talk about them with our colleagues, we organize no meetings on the subject, we scarcely find it mentioned in textbooks – we remain silent, as we are somewhat ashamed to use such straightforward methods to help those of our patients who, in fact, require them (p. 262).

Supportive psychotherapy was evidently held to stand in contradiction to the fundamental goal of psychotherapy, i.e. that of setting processes of deep-seated change in motion. It is primarily because of the emergence of a more sophisticated awareness of the possibilities and limitations of psychotherapy that supportive psychotherapy is now accepted as an independent treatment approach. Nonetheless, there is still no generally recognized definition of supportive psychotherapy and no truly satisfactory theoretical foundation for it. The conceptions that have been proposed are extremely diverse. Some think that supportive psychotherapy is based on a medical model of illness, others that it is a logical application of psychoanalytic concepts in a certain group of patients. According to other views, in accordance with the concepts of humanistic psychology, supportive psychotherapy is a practical collective term for the therapeutic techniques used in patients in which theoretical conceptions no longer afford any help. The relationship

between the therapist and the patient is generally identified as the major factor contributing to its efficacy.

Relatively elaborate conceptions of supportive psychotherapy were developed recently by Werman (1984) and by Hartland (1991). Both of these conceptions focused on the specific problems that were thought to be most amenable to supportive therapy: inadequately developed mental functions, lack of justified expectation of regaining the premorbid functional level, protracted psychiatric history, and repeated need for institutionalization and assistance in coping with everyday life. According to the conceptions of Werman and Hartland, the diagnostic categories in which supportive psychotherapy is most relevant include chronic psychoses, anxiety disorders that have become chronic, somatization disorders, depressive syndromes, and severe personality disorders.

The lack of any therapeutic progress worthy of the name in many patients in these diagnostic groups, and the fact that they frequently live on the margins of society, place especially high demands on the therapist (despite the prevailing opinion to the contrary). In addition to experience, absolute mastery of "technical" matters, and knowledge of one's own personality, including weaknesses and specific countertransference reactions, what is required most of all is a high frustration threshold and a lack of dependence on narcissistic confirmation.

With regard to the specific techniques to be used in supportive psychotherapy, only a few general remarks can be made. Rogers (1957) worked out a set of essential therapeutic variables that he considered necessary for successful treatment. We shall state these requirements in somewhat modified fashion. The therapist's behavior must be marked by genuineness, predictability, and preservation of the necessary emotional distance. Furthermore, an especially large measure of patience and flexibility is needed. According to many who have written on the subject, a durable therapeutic relationship for supportive psychotherapy is established less through conversations with the patient than through common activities such as shared meals or excursions, and by the communication of life experiences not only from patient to therapist, but also from therapist to patient.

In terms of content, the most important objective is coping with the demands of everyday life. The following elements of therapy are the most important ones contributing to the achievement of this goal: empathic listening, emotional support, clarification of acute problem situations, encouragement and active guidance toward problem-solving, strengthening of preexisting defense strategies and adaptive behavior, prevention of relapse by psychoeducation, and clear boundary-setting in case of regressive behavior. The duration and frequency of therapy sessions may vary greatly according to need, from a few minutes to an hour, and from once every 3–6 months to once a day in crisis situations.

The efficacy of supportive psychotherapy in schizophrenia has scarcely been considered an independent subject for empirical research until the present, but it has nevertheless been repeatedly measured as a control condition in group-comparison studies on schizophrenic patients, such as the widely known Boston Psychotherapy Study, in which supportive psychotherapy was found to be highly effective in comparison to psychoanalytic therapy (Stanton et al. 1984; Gunderson et al. 1984). Nonetheless, supportive psychotherapy in schizophrenic patients is not without a certain danger. If other types of psychotherapeutic intervention are rejected a priori, the appearance or progression of disease chronification may be promoted. It is perhaps a recognition of this possibility that underlies the current trend toward a fusion of supportive therapy with certain elements of psychoeducative and cognitive therapy, through which supportive therapy may become transformed into a special type of diagnosis-specific psychotherapy. Such a development would surely heighten scientific interest in supportive psychotherapy to a level more in keeping with its immense practical importance.

3 Psychoanalytic Therapy

Freud developed psychoanalysis both as a depth-psychological theory and as a therapeutic method on the basis of knowledge he had gained from treating patients with hysteria and other neurotic illnesses. As mentioned above, he was sceptical about the possibility of treating schizophrenics by psychoanalysis. He believed psychoanalytical healing to require the presence of a so-called transference neurosis, and he considered schizophrenic patients to be incapable of this because of their disturbed self- and object differentiation and the regression of their libido into the ego (Freud 1963).

As already mentioned in the brief historical survey at the beginning of this chapter, other early psychoanalysts were indeed very much occupied with the treatment of schizophrenic patients, including Abraham, Federn, and Jung (see Benedetti 1987; Müller 1972). Federn (1952) was the first to describe the loss of ego boundaries. Klein (1930) developed her own theoretical conception, in which she described the paranoid schizoid position, with its severe persecution anxieties, as a transitional syndrome of normal ego development. According to this view, projection and identification serve as defenses against anxiety in this stage. Schizophrenic patients suffer from these early

childhood anxieties and react to them with the corresponding defense mechanisms.

The psychoanalytic treatment of schizophrenia gained considerable ground, primarily in the United States, in the mid-twentieth century; the heyday of this form of therapy is connected with such names as Sullivan, Fromm-Reichmann, and Searls. The therapists of this period built up a relationship with the schizophrenic patient through enormous patience and endurance and then used the relationship as a bridge across which patients could come to terms with the outside world. The libido theory receded in importance for Sullivan (1962), whose conception instead stressed the development of interpersonal relationships leading progressively, through infantile, early, and late childhood stages, and several stages of adolescence, to maturity. Fromm-Reichmann's (1959) view of the influence of the mother on her schizophrenic child was widely transmitted in an oversimplified form. In its guise as the concept of the "schizophrenogenic mother," it led to the raising of completely unjustified accusations against the families of schizophrenics.

In the German-speaking countries, psychoanalysis came into widespread use again only after the end of the Second World War, because the National Socialist dictatorship had vilified its best-known proponents and forced them into exile. In 1949, Schindler (1980) introduced so-called bifocal group therapy in Vienna – two analytic groups led by the same therapist, one for the patients and another for their families. Individual psychotherapy on an analytical basis was initially practiced mainly in Switzerland (Sechehaye 1986; Benedetti 1987). Benedetti, in particular, made many contributions to understanding schizophrenics, proceeding from his own therapeutic observations and his wide experience with supervision. Among his other contributions, he worked out specific features of psychotic transference and special disturbances of symbol construction and also described levels of helpful relationships and the significance of the therapist's unconscious fantasies. In *Alienazione e Personazione nella Psicoterapia della Malattia Mentale* (Benedetti 1980), he gave a comprehensive account of the psychopathology, psychodynamics, and psychotherapeutic treatment of schizophrenic patients.

The task of demonstrating the effectiveness of psychoanalytic therapy for schizophrenic disorders is certainly fraught with major methodological difficulties, not least because of its special setting and long duration, but this fact cannot free its proponents of the responsibility of providing empirical support for it. In the Boston Psychotherapy Study, already mentioned above (Stanton et al. 1984; Gunderson et al. 1984), which is the most comprehensive of the few group comparison studies performed to date, no significant difference was found between the results of supportive and psychoanalytic therapy. Despite these rather disappointing results, however, this study also yielded indications that some patients, particularly the highly educated but poorly socially adapted, do indeed stand to benefit from a psychodynamically oriented treatment. The so-called Menninger Study of long-term analysis in 42 patients (Wallerstein 1989) further revealed that elements of supportive therapy are often applied in addition to psychoanalytical methods in the treatment of severely disturbed patients. Accordingly, further attempts have been made recently to develop specific psychodynamic approaches for schizophrenic patients (see Munich 1987; Rosberg and Stunden 1990; Mentzos 1992; Lotterman 1996).

The use of neuroleptics has brought about a major change in psychoanalytic therapy in the last 40 years. The enormously difficult and demanding therapies of the pre-neuroleptic age have given way to newer psychotherapeutic treatments; these are certainly still demanding, but are somewhat more "normal." Even psychoanalysts today generally carry out psychoanalytic therapy in only one or two sessions per week, instead of the classical psychoanalysis that was previously usual. An essential feature of psychoanalytic therapy that has remained constant is the endeavor to gain an understanding of symptoms and reactions, on the basis of which changes can then be attempted. Alongside the analytical process in the narrower sense, however, it is of the highest importance when treating schizophrenic patients to build a durable therapeutic relationship and to use elements of supportive treatment to help achieve this objective.

4 Cognitive-Behavioral Therapy

The development of effective antipsychotic medications, the process of deinstitutionalization, and, above all, the formulation and wide acceptance of the vulnerability–stress–coping model have led to a major expansion of the role of cognitive-behavioral therapy in the treatment of schizophrenic illnesses. According to the vulnerability–stress–coping model, an effective therapy for schizophrenia must favorably influence at least one of the following three pathogenetically interacting factors: (1) biologically determined neuropsychological vulnerability, (2) psychosocial stressors, and/or (3) the affected individual's coping ability. As Smith and colleagues discuss in their chapter on rehabilitation in schizophrenic illnesses (Chap. 13, Vol. 3, Part 1), this heuristic framework is also useful for the evaluation of rehabilitative efforts.

A great number of studies over the last two decades have shown that cognitive-behavioral therapy

processes counteract neuropsychological vulnerability both indirectly, by means of an improved compliance with medications, and directly, by means of compensation strategies; that they reduce psychosocial stressors in the patient's environment; and that they can strengthen the patient's coping resources (Bellack and Mueser 1993; Kopelowicz and Liberman 1995; Penn and Mueser 1996). They are thus used not only at the end of the acute phase, but also in the stabilization phase that follows, and in the remission phase of schizophrenic illnesses.

The boundaries between cognitive-behavioral therapy on the one hand and rehabilitative therapy on the other are fluid, and the same techniques may be used in both (see Chap. 13, Vol. 3, Part 1). Differences are evident with respect to the problem areas on which these two forms of therapy are focused and the content of therapy in the individual case. As for the results of treatment, it should be observed that the different forms of therapy are always applied in combination with neuroleptic therapy in the relevant studies. According to the most comprehensive meta-analysis performed to date (Mojtabai et al. 1998), a combination of this type leads to a more favorable outcome than any of these forms of therapy used alone.

4.1 Social Skills Training

Schizophrenic patients suffer from numerous instrumental and interpersonal deficits because of neuropsychological dysfunction interfering with their schooling, because of the manifestations of the illness itself, and (above all in chronic cases) because of a lack of social stimulation. All of these factors increase the likelihood of tense and stressful interactions and further social isolation and can thus contribute to a worsening of the course of the illness. Good instrumental and interpersonal competence, on the other hand, helps the patient obtain access to, and make good use of, the available social resources. The training of social skills is intended to promote these forms of competence. Relevant social situations and instrumental skill areas that are important for coping with illness are broken down into their elements in suitably designed training programs and then practiced through instruction, model learning, role playing, homework, and positive social reinforcement (see Chap. 13, this volume). The efficacy of such training programs in schizophrenic illnesses has been intensively investigated in a series of controlled treatment comparison studies reaching back to the early 1980s (Bellack et al. 1984; Wallace and Liberman 1985; Hogarty et al. 1986, 1991; Dobson et al. 1995; Hayes et al. 1995).

These studies and several meta-analyses (e.g. Benton and Schröder 1990) have shown that schizophrenic patients are able, in the long term, to acquire the originally deficient instrumental and social skills and to carry these over, at least in part, to their everyday social environment. Focused approaches aimed at an improvement of skills in narrowly defined instrumental skill areas that are especially important for dealing with the disease (e.g. medication and symptom management) have more strongly generalizable results than approaches aimed at general social competence. Many patients feel more self-assured after participating in social skills training, have fewer social anxieties, and, moreover, require shorter periods of hospitalization. Overall social adaptation, however, is only mildly and transiently improved by the training of social skills. Such training programs must apparently be applied for at least 6 months to 1 year, with refresher sessions afterwards, to maximize their therapeutic potential with respect to clinical indicators such as disease manifestations, level of social functioning, or recurrence rate. Severely disturbed patients may require even more time for the acquisition of individual skills.

Despite the large number of pertinent studies now available, a number of important questions remain unanswered. It is not yet known precisely how psychotic manifestations and neuropsychological deficits interfere with the acquisition of social skills. Open questions with respect to the efficacy of social skills training include its optimal frequency and duration, its specific effective components, and the specific techniques and specific intensity of treatment that provide the greatest therapeutic benefit at each stage of the disease. Not least, procedures must be developed for improving the transfer of the learned skills into the patient's natural living environment and the fulfillment of social roles, which is, of course, the ultimate goal of social skills training. Its potential additive and interactive effects with newer atypical neuroleptics and with other psychological treatment approaches are a particular focus of current interest. Finally, more attention must be paid in future to the integration of social skills training with work rehabilitation measures and with care strategies such as case management.

4.2 Cognitive Deficits

Another current application of cognitive-behavioral therapy is in the treatment of cognitive deficits. A distinction can be drawn between procedure-oriented and content-oriented techniques. Procedure-oriented techniques are aimed at deficits of information processing, i.e. deficits of attention, perception, concept formation, memory, and so forth, and are known in the

current literature as cognitive remediation. In recent years, such disturbances of information processing have been held to be important factors in the emergence of overt schizophrenic illness, and thus cognitive remediation may, in fact, be a means of rapid and effective therapeutic intervention. Both elementary and complex deficits of information processing may disrupt the acquisition, generalization, transfer, maintenance, and application of social skills and raise the patient's vulnerability to stress. Smith and colleagues (Chap. 13, Vol. 3, Part 1) point out that the disturbances of information processing typical of schizophrenia have a greater predictive value than the psychological manifestations of the illness with regard to the successful learning of treatment content. They therefore advocate the application of cognitive remediation techniques in psychiatric rehabilitation as well.

Content-oriented techniques, in contrast, are intended to reduce or eliminate persistent positive manifestations such as delusions and hallucinations. The latter are often accompanied by anxiety and depressive manifestations and usually lead to far-reaching difficulties in the social and occupational spheres and to a significant impairment of the overall quality of life of patients and their families.

A number of neuropsychological tests have been applied and tested for use in cognitive remediation in recent years, including the Continuous Performance Test (CPT), the Span of Apprehension Test (SAT), and especially the Wisconsin Card Sorting Test (WCST). It has been shown that the use of these tests can improve poor performance in specific areas. The greatest improvements in performance are apparently obtained through a combination of reinforcement and instruction techniques or through special procedural training techniques (e.g. support in isolated aspects of a task in which the patient is prone to error) (Green 1993; Penn and Mueser 1996).

The Integrated Psychological Therapy (IPT) Program for Schizophrenic Patients is an extended approach in the sense that it combines the remediation of procedural cognitive deficits with the training of social skills (Hodel and Brenner 1994; Brenner et al. 1994). It is assumed that defects in information processing and in instrumental and interpersonal skills reinforce each other in a vicious circle, and a set of subprograms is therefore used to redirect the therapeutic emphasis from the cognitive to the social aspects of behavior, while the interdependence of the two is always borne in mind.

Most studies on the efficacy of cognitive remediation techniques are based on very small samples. Much more extensive, controlled group studies are needed to document efficacy unequivocally. It is not yet definitively known how long the therapeutic benefit of such techniques lasts or whether it is transferred to neuropsychological vulnerability in general or to the acquisition of social skills and social competence in particular. A problem in the design of training programs is that it remains unclear which specific disturbances of information processing impede or prevent the acquisition of social skills and thus ought to be the focus of the therapeutic effort. Current developments are aimed at clarifying the relationship between cognitive dysfunction and social learning, so that comprehensive rehabilitation programs may be targeted at those particular types of cognitive dysfunction ("rate-limiting factors") that most severely limit the success of other therapeutic interventions in schizophrenic patients.

Two approaches to the treatment of content-related cognitive dysfunction in schizophrenic patients, in the sense of positive (residual) disease manifestations, have received much attention in the past 5–10 years: cognitive techniques and self-management techniques. Cognitive techniques are based on the assumption that schizophrenic delusions are based on irrational distortions of functional thought processes. These irrational cognitive styles are, accordingly, called into question; alternative, rational explanations are offered; and the truth content of delusional ideas is checked against reality. Several case studies have shown that this type of cognitive technique can bring about a major reduction in schizophrenic patients' subjective certainty about, and preoccupation with, their delusions and thereby alleviate the accompanying anxieties and depressive feelings (Chadwick and Lowe 1994). Furthermore, a change in patients' ideas concerning the origin, identity, power, and authority of auditory hallucinations can reduce their frequency of occurrence and mitigate their effects on the patients' behavior and subjective well-being (Chadwick and Birchwood 1994; Bentall et al. 1994).

Self-management techniques proceed in a different direction. Their objective is not the direct therapeutic amelioration of positive (residual) manifestations of disease, but rather an improvement of the patient's ability to cope with them. Therapy programs of this type begin with a detailed analysis of disease manifestations, the conditions that provoke them, and their consequences. Suitable coping strategies, making use of the coping mechanisms that each patient already has, are then selected for the individual manifestations. Finally, these strategies are practiced in simulated situations, or in vivo when positive manifestations are present. In a group comparison study (Tarrier et al. 1993), a self-management technique of this type was found to result in a significant reduction of the number and severity of positive manifestations and the anxiety accompanying them; nonetheless, the effect of a self-management technique on negative manifestations, depressive manifestations, and indicators of

the level of social functioning was no better than the effect of training in problem-solving or that of being in a waiting-list control group.

These findings give rise to hope that cognitive-behavioral therapy techniques might, in future, play a major role in the treatment of patients with persistent, neuroleptic-resistant positive manifestations. The empirical support for this hope is not yet solid, as controlled group-comparison studies are still lacking, and the list of open questions is long. There is an urgent need to determine how long the desired therapeutic benefit lasts, for example, and what effect it might have on other clinical parameters, such as recurrence rates or social adaptation. The data available on these questions to date are inconsistent. The individual techniques can be used selectively to best advantage only if their specific mechanisms of effectiveness are identified and distinguished from one another and their effects when used in combination with other therapeutic measures are determined.

4.3 Family Therapy

No other class of psychological therapies has been studied as intensively as cognitive-behavioral family therapy in the last 15–20 years. The motivation for research in this area came, on the one hand, from study findings that a family climate characterized by overprotective, critical, or hostile expressions ("high expressed emotion") affects the course of schizophrenic illnesses and, on the other hand, from the recognition that the families of schizophrenic patients carry an enormous psychological burden. Thus treatment concepts were developed that were intended to produce a change in types of family interaction that affect the disease course unfavorably (reduction of the level of expressed emotion) and/or amplify the family's coping resources.

Despite differences in content and therapeutic procedure, these concepts share essential core features. They all involve a structured procedure oriented toward concrete problem-solving, beginning with an introductory provision of information about the illness (psychoeducation) and proceeding to a more or less broad range of behavior-therapeutic interventions aimed at improving the interpersonal communication skills of the patient's family members, as well as their problem-solving ability and their ability to cope with stress. A brief list of the major common features of the different approaches with respect to content and procedure is found in Chap. 13 (Vol. 3, Part 1).

A number of controlled studies have been performed since the early 1980s to assess these types of intervention (Falloon et al. 1985; Leff et al. 1985; Tarrier et al. 1989; Hogarty et al. 1991; Randolph et al. 1994; Xiong et al. 1994; Zhang et al. 1994). These studies reveal that cognitive-behavioral family therapy can result in a significant reduction of the recurrence rate in the first 2 years after therapy is begun. In studies including an assessment of the length of hospitalization, the level of social functioning, and family stress as additional criteria for success, cognitive-behavioral family therapy was found to reduce the number of hospitalizations, improve social adaptation, and reduce the stress on family members (Falloon et al. 1985; Barrowclough and Tarrier 1990; Hogarty et al. 1991). Comparisons between therapies carried out with individual families and those carried out with groups of families, or groups of patients' relatives, have shown, in part, that the group format confers a greater benefit (Leff et al. 1990; McFarlane et al. 1995; Schooler et al. 1997). Studies have also shown that cognitive-behavioral family therapy is cost-effective, at least in the intermediate term (Cardin et al. 1985; Tarrier et al. 1991; Held et al. 1992; Rund et al. 1994).

The specific mechanisms by which these therapeutic techniques exert their beneficial effect, their optimal frequency and duration, and the therapeutic processes of change that they set in motion are still largely unknown. It is therefore difficult to plan therapy in a goal-directed manner with specific attention to the initial conditions, socioeconomic situation, and specific needs and resources of the individual patient's family. This may be one of the reasons why many families cannot be persuaded to participate in therapy at all or terminate it after a short time. It is also unclear how long the effect of therapy lasts after the therapy itself has come to an end. All of the studies have shown that the recurrence rate rises as the follow-up interval increases, which can be interpreted to imply that cognitive-behavioral family therapy mainly postpones recurrence, rather than preventing it. If so, it should be used in the long term, like pharmacotherapy, and perhaps during the entire period of the patient's life in which he or she is at risk. The follow-up study by Tarrier et al. (1994), however, showed that patients treated in this way still had a significantly lower rate of relapse than patients in the control group as long as 8 years after the end of therapy.

5 Systemic Therapy

Like cognitive-behavioral family therapy, systemic therapy today has a secure place in the outpatient and inpatient treatment of schizophrenia. The earliest, ground-breaking studies on the systemic therapy of schizophrenic disorders were performed in the United

States in the 1950s by three teams of researchers and therapists. These were the groups led by Lidz and by Wynne and the Palo Alto group of Bateson and colleagues, whose writings (e.g. Bateson et al. 1956; Bateson 1987) remain the most important sources for the theory of systemic therapy today. The further development of the systemic approach to the therapy of schizophrenia was carried out by the Milan group of Selvini Palazzoli and colleagues (Selvini Palazzoli 1985; Selvini Palazzoli et al. 1980) and, more recently, by the Heidelberg group of Stierlin and colleagues (Stierlin 1975; F.B. Simon 1988, 1990; Retzer 1994). Specific patterns of interaction of families with a schizophrenic member, and of families with a schizoaffective or manic-depressive member, have been identified and made the target of specifically designed therapeutic interventions (F.B. Simon et al. 1989; Retzer 1994).

The goal of systemic therapy is to set problem-stabilizing patterns of family interaction in motion and to design constructions of reality that facilitate a change toward an assumption of personal responsibility and autonomy. A further goal is to lessen the degree of importance attached to psychiatric diagnoses. Thus system therapists tend to take a sceptical view of the vulnerability–stress–coping model of schizophrenia, as well as other models. The systemic approach seems to be similarly incompatible with cognitive-behavioral approaches, based as they are on the concept of expressed emotion and on the vulnerability–stress–coping model. System therapists acknowledge that cognitive-behavioral therapy lessens the rate of relapse, but they also suspect that it promotes chronification – a hypothesis that is difficult to test empirically.

There are practically no empirical studies available on the efficacy of systemic therapy. Only Retzer (1994), in an uncontrolled follow-up study involving patients with schizophrenia according to the DSM-III-R criteria, was able to show that an improvement had occurred in 65% of patients and the relapse rate had fallen by 60% at an average of 3.2 years after the end of family therapy. The question of whether systemic therapy is as effective as other methods of treating families with a schizophrenic member thus remains to be definitively answered. Because even system therapists must acknowledge that they can do little to help once chronification has occurred, the Heidelberg group has recently devoted itself intensively to the theme of chronification (Simon and Weber 1987; F.B. Simon 1993).

6 Milieu Therapy

Although Sullivan (1931) had already pointed out the importance of the patient's social environment for the recovery process in schizophrenia, Cumming and Cumming are considered to have founded modern milieu therapy with their book *Ego and Milieu* (1962). Their basic conception was that the patient should do everything that he or she is capable of; they considered action, rather than the passive receipt of treatment, to be the essential therapeutic principle. Especially in psychotic patients, they maintained, such activity leads to an improvement of identity and to "ego growth."

Gunderson (1980) distinguished two fundamentally different types of milieu therapy: (1) the therapeutic community in the sense of Jones (1953) and (2) behavioral milieu therapy (e.g. the "token economy" approach; Ayllon and Azrin 1968). He stated that three qualitative factors were determinative of the success of milieu therapy: (1) the sharing of responsibility and decision-making authority, (2) the clear definition of treatment programs, roles, and leadership, and (3) a high degree of caregiver–patient interaction. Tucker (1983) understood the milieu as an expression both of the treating institution's particular attitude toward treatment and of its organizational structure. One of the most important aspects of the therapeutic community is the attempt to give patients the chance to assume responsibility for their own treatment and that of other patients. The distinguishing feature of behavioral milieu therapy is the consistent application of principles of learning theory in the framework of the individual patient's situation and treatment plan.

Therapeutic milieus differ considerably from one another with respect to the degree of control, support, structure, engagement, and evaluation. There are thus many different types of "milieu," and all psychiatric institutions today claim to offer milieu therapy. The concept of milieu therapy has thereby become diffuse and lost much of its original meaning. The concept of the therapeutic community has also undergone much change. Jones himself, in his later years, considered the distinction between the therapeutic community and milieu therapy superfluous and proposed the overarching concept of learning as a social process (Jones 1983).

Van Putten and May (1976) expressed the pessimistic view that milieu therapy has only a small additional effect beyond that of pharmacotherapy in an adequate dose. Gunderson (1980) and Ellsworth (1983), however, in their reviews of the literature, concluded that milieu therapy does indeed lead to a significant improvement of disease manifestations and of the level of social functioning in patients with both acute and chronic schizophrenic illnesses. Paul and Lentz (1977), in their pioneering study, were able to show that milieu therapy and social learning for the chronically mentally ill produce better results than treatment in a custodial environment, independently of the pharmacologic treatment given. This finding had a strong and far-reaching influence on the further

development of psychotherapeutic and psychosocial treatment programs for schizophrenic patients.

An exceptionally consistent application of the milieu-therapeutic approach to the treatment of schizophrenia was carried out in the San Francisco Soteria House, founded by Mosher et al. (1975), in which a small group of acutely affected schizophrenics were treated by lay caregivers in a community setting practically without the use of neuroleptic medications. After 2 years of clinical follow-up, Mosher and Menn (1978) found that the Soteria patients had done no better with respect to disease manifestations and recurrence rate than a control group of patients managed in a traditional inpatient setting, but they did have a higher rate of employment and were living more independently than the patients in the control group.

Ciompi, in Bern, Switzerland, replicated this concept and developed it further. On the basis of his theory of affect logic (Ciompi 1982), he laid down the following eight therapeutic principles for Soteria Bern (Ciompi et al. 1991):

1. A therapeutic milieu that is small, as "normal" as possible, transparent, relaxing, and protected from excessive stimulation
2. Attentive and continual emotional support during the psychotic crisis, provided by a small number of selected individuals
3. Continuity of concept and of treating staff, lasting from the phase of acute treatment all the way to reintegration in the community
4. Provision of clear and consistent information to patients, relatives, and caregivers with respect to the illness, its prognosis, and its treatment
5. Continual close cooperation with family members and significant others
6. Development of common concrete objectives and priorities on the living and work axes, with the generation of realistic, cautiously optimistic expectations
7. Use of neuroleptics only in the presence of not otherwise treatable danger to the patient or others, in the absence of signs of improvement after 4–5 weeks, or when the risk of relapse cannot be otherwise reduced in the post-acute treatment phase
8. Systematic post-acute treatment and relapse prophylaxis for 2 years on the basis of a preceding analysis of individual prodromal symptoms, stressful situations, and possible coping strategies, carried out with the cooperation of patients, their relatives, and caregivers

These principles of treatment, with the exception of (7), can be profitably carried over to other outpatient, part-outpatient, and inpatient therapeutic milieus as well. With regard to medication, however, the practice has changed, even at Soteria Bern, to the administration of a low basal dose of medication for relapse prevention, with an increase in dose if prodromal symptoms appear. In a controlled, prospective study involving follow-up at 2 years, the results of Soteria treatment with respect to social functioning level and recurrence rates were found to be comparable to those of treatment in traditional inpatient settings. While Soteria treatment reduced the use of neuroleptics by half, the patients stayed in the treating institution twice as long, which increased the cost of treatment (Ciompi et al. 1993). It was later possible to reduce the length of stay. These results are in accordance with those of Mosher and Menn (1978). Soteria treatment was also felt by many patients to be less traumatizing and stigmatizing (because less isolating) than inpatient treatment in a psychiatric ward.

The milieu-therapeutic Soteria approach has a number of evident limitations. This form of treatment is necessarily voluntary and is thus eminently unsuitable for use among young, chronically schizophrenic patients who have no insight into their illness, behave aggressively, and abuse addictive substances. Such difficult patients require a clearly structured milieu, and hospital wards of the traditional type often hold more than their share of them. A further problem is that the success of milieu therapy depends, in part, on a high degree of involvement on the part of the caregiver, who is, in turn, motivated by ideals and values propounded by a charismatic leader. This can easily lead to a romantic idealization of the proffered form of therapy.

7 Long-Term Care and Case Management

In the last 20 years, the deinstitutionalization process and the establishment of community-integrated structures for the delivery of social psychiatric care have led to the relocation of the focus of long-term care for schizophrenic patients from the hospital into the community. In line with this development, a broad range of facilities has developed, including outpatient and inpatient crisis centers, day clinics, low-threshold outpatient services serving as contact points, outpatient follow-up clinics, and visiting services, all of which offer diverse types of socio- and psychotherapeutic assistance in a community-based setting.

Regionalized or sectorized care structures, and the concepts of the long-term care team (Test 1979; Torrey 1986; Johnson et al. 1997) and case management (Sledge et al. 1995), were developed in order to enable patient-centered coordination of these diverse offerings and to ensure continuity of care. The long-term

"care team" is a multidisciplinary team of professionals that assumes responsibility for a defined group of patients, ideally in both the outpatient and the inpatient settings.

The concept of case management is represented by several considerably different approaches, ranging from the purely organizational concept of "brokering" (i.e. mediation of treatment and rehabilitation offerings) to the so-called clinical case management model, which bears a strong resemblance to the long-term caregiver model (Mueser et al. 1998). Studies by Bond et al. (1988) and by Borland et al. (1989) showed that intensive case management shortens hospital stays and may thereby result in a saving of costs, but it does not improve the patients' functional level or quality of life. These findings were not supported by Rössler et al. (1993), who found that care by a sociopsychiatric service reduced neither the rate of hospitalization nor the length of hospital stay. Similarly, a controlled study by Marshall et al. (1995) revealed no advantage to case management in the care of chronic patients.

On the other hand, Quinlivan et al. (1995) found that case management can indeed reduce costs significantly, and precisely in the group of chronic patients. Rosenheck et al. (1995) came to the same conclusion in a controlled multicenter study of long-term care; the most important factor in cost reduction was a reduced need for inpatient treatment. The limited efficacy of intensive approaches to long-term care may be due to the fact that such approaches still usually do not include any specific rehabilitation techniques that improve the level of cognitive or psychosocial functioning, e.g. techniques of cognitive remediation or social skills training.

Thus, although most of the findings have been encouraging, the present state of research on long-term care and case management does not yet allow any reliable conclusion about their efficacy. The results of different studies may conflict at least in part because long-term care and case management function optimally only when patients have a suitably comprehensive mental health care system and diverse therapeutic offerings at their disposal. The conditions under which the studies mentioned above were carried out varied widely in this respect. Related considerations of mental health care policy are discussed in greater detail in Chap. 13 (Vol. 3, Part 1).

Family care, which was practiced in Germany as early as the nineteenth century but was forgotten thereafter, is now assuming an importance that is not to be underestimated in the long-term care of chronically schizophrenic patients. In this form of care, a family framework with the necessary social support is provided by a family other than the patient's own, when the patient no longer needs inpatient treatment but cannot live alone. Recent studies have shown that the results obtained with family care for patients of this type are generally favorable (e.g. Schmidt-Michel et al. 1992).

8 Concluding Remarks

In the last two decades, major progress has been made in the socio- and psychotherapy of schizophrenic disorders. This is most evident in the larger role now played by cognitive-behavioral therapy approaches and in the increased attention paid to milieu-therapeutic aspects. Psychoanalytic therapy and, to a lesser degree, systemic therapy for schizophrenic disorders are less influential today than they once were, but innovative developments are still taking place in these fields.

All modern approaches to the socio- and psychotherapy of schizophrenic patients have in common a greater emphasis on the active participation of patients and family members, so that the result of treatment will be socially relevant, as well as on the optimal embedding in a coordinated overall treatment plan. As the scientific understanding of these forms of therapy has improved, questions have been raised about their differential indications, specific mechanisms of action, and interactions with other psychological interventions and various pharmacological treatment strategies – questions which must now be investigated more systematically than before. It also remains to be seen whether these various approaches can produce results in routine care that are as good as those obtained under experimental testing conditions, and whether their application is cost-effective.

Finally, a further major deficiency of the current state of scientific knowledge is that the effectiveness of different socio- or psychotherapeutic interventions and structures of psychiatric care has almost always been studied separately for each intervention or structure. This artificial separation will have to be overcome. A second generation of efficacy studies is needed in which an integrated evaluation of socio- and psychotherapy and of care structures is performed.

9 References

Ayllon T, Azrin NH (1968) The token economy. Appleton-Century-Crofts, New York

Barrowclough C, Tarrier N (1990) Social functioning in schizophrenic patients. 1. The effects of expressed emotion and family intervention. Soc Psychiatry Psychiatr Epidemiol 25: 125–129

*Bateson G (1987) Steps to an ecology of mind. Collected essays in anthropology, psychiatry, evolution, and epistemology. Aronson, Northvale

**Bateson G, Jackson DD, Hayley J, Weakland J (1956) Towards a theory of schizophrenia. Behav Sci 191: 251–264

Bellack AS, Mueser KT (1993) Psychosocial treatment for schizophrenia. Schizophr Bull 19(2): 317–336

Bellack AE, Turner SM, Hersen M, Luber RF (1984) An examination of the efficacy of social skills training for chronic schizophrenic patients. Hosp Community Psychiatry 35: 1023–1028

*Benedetti G (1980) Alienazione e personazione nella psicoterapia della malattia mentale. Einaudi, Torino

Benedetti G (1987) Psychotherapeutische Behandlungsmethoden. In: Kisker KP, Lauter H, Meyer JE, Müller C, Strömgren E (eds) Psychiatrie der Gegenwart, vol 4, 3rd edn. Springer, Berlin Heidelberg New York, pp 285–323

Bentall RP, Haddock G, Slade PD (1994) Cognitive therapy for persistent auditory hallucinations: from theory to therapy. Behav Ther 25: 51–66

Benton MK, Schröder HE (1990) Social skills training with schizophrenics: a meta-analytic evaluation. J Consult Clin Psychol 58: 741–747

Bond GR, Miller LD, Krumwied RD, Ward RS (1988) Assertive case management in three CMHCs: a controlled study. Hosp Community Psychiatry 39: 411–418

Borland A, McRae J, Lycan C (1989) Outcomes of five years of continuous intensive case management. Hosp Community Psychiatry 40: 369–376

Brenner HD, Roder V, Hodel B, Kienzle N, Reed D, Liberman RP (1994) Integrated psychological therapy for schizophrenic patients (IPT). Hogrefe and Huber, Seattle

Cardin VA, McGill CW, Falloon IRH (1985) An economic analysis: costs, benefits and effectiveness. In: Falloon IRH (ed) Family management of schizophrenia. John Hopkins University Press, Baltimore

Chadwick PDJ, Birchwood M (1994) The omnipotence of voices. A cognitive approach to auditory hallucinations. Br J Psychiatry 164: 190–201

Chadwick PDJ, Lowe CF (1994) A cognitive approach to measuring and modifying delusions. Behav Res Ther 32: 355–367

Ciompi L (1982) Affektlogik. Über die Struktur der Psyche und ihre Entwicklung. Ein Beitrag zur Schizophrenieforschung. Klett-Cotta, Stuttgart

Ciompi L, Dauwalder H-P, Maier Ch, Aebi E (1991) Das Pilotprojekt "Soteria Bern" zur Behandlung akut Schizophrener. I. Konzeptuelle Grundlagen, praktische Realisierung, kliniksche Erfahrungen. Nervenarzt 62: 428–435

Ciompi L, Kupper Z, Aebi E, Dauwalder HP, Hubschmid T, Trütsch K, Rutishauser C (1993) Das Pilotprojekt "Soteria Bern" zur Behandlung akut Schizophrener. II. Ergebnisse einer vergleichenden prospektiven Verlaufsstudie über 2 Jahre. Nervenarzt 64: 440–450

*Cumming J, Cumming E (1962) Ego and milieu; theory and practice of environmental therapy. Atherton, New York

Dobson DJG, McDougall G, Busheikin J, Aldous J (1995) Effects of social skills training and social milieu treatment on symptoms of schizophrenia. Hosp Community Psychiatry 46: 376–380

Ellsworth RB (1983) Characteristics of effective treatment milieus. In: Gunderson JG, Will OA, Mosher LR (eds) Principles and practice of milieu therapy. Aronson, New York, pp 87–123

Falloon IRH, Boyd JL, McGill CW et al (1985) Family management in the prevention of morbidity of schizophrenia. Clinical outcome of a two-year longitudinal study. Arch Gen Psychiatry 42: 887–896

Federn P (1952) Ego psychology and the psychoses. Basic Books, New York

Freud S (1963) Neurosis and psychosis (collected papers). Collier, New York

Fromm-Reichmann F (1950) Principles of intensive psychotherapy. University of Chicago Press, Chicago

Fromm-Reichmann F (1959) Psychoanalysis and psychotherapy. University of Chicago Press, Chicago

Green MF (1993) Cognitive remediation in schizophrenia: is it time yet? Am J Psychiatry 150: 178–187

Griesinger W (1868/1869) Über Irrenanstalten und deren Weiterentwicklung in Deutschland. Arch Psychiatr Nervenheilkd 1: 8–43

Gunderson JG (1980) A reevaluation of milieu therapy for nonchronic schizophrenic patients. Schizophr Bull 6: 64–69

*Gunderson JG, Frank AF, Katz HM et al (1984) Effects of psychotherapy in schizophrenia. II. Comparative outcome of two forms of treatment. Schizophr Bull 10: 564–598

Hartland S (1991) Supportive psychotherapy. In: Holmes J (ed) Textbook of psychotherapy in psychiatric practice. Churchill Livingstone, London

Hayes RL, Halford WK, Varghese FT (1995) Social skills training with chronic schizophrenic patients: effects on negative symptoms and community functioning. Behav Ther 26: 433–449

Heim E (1980) "Stütztherapie" – neu entdeckt? Plädoyer für adaptive Psychotherapien. Psychother Med Psychol 30: 261–273

Held T, Bockhorn A, Schoonen B (1992) Schizophreniebehandlung in der Familie (Modellbericht). Modellverbund Psychiatrie, Bonn

Hinsie LE (1930) The treatment of schizophrenia. Williams and Wilkins, Baltimore

Hodel B, Brenner HD (1994) Cognitive therapy with schizophrenic patients: conceptual basis, present state, future directions. Acta Psychiatr Scand 90[Suppl 384]: 108–115

Hogarty GE, Anderson CM, Reiss DJ, Kornblith SJ, Greenwald DP, Javna CD, Madonia MJ (1986) Family psychoeducation, social skills training, and maintenance chemotherapy in the aftercare treatment of schizophrenia. I. One-year effects of a controlled study on relapse and expressed emotion. Arch Gen Psychiatry 43: 633–642

*Hogarty GE, Anderson CM, Reiss DJ et al (1991) Family psychoeducation, social skills training, and maintenance chemotherapy in the aftercare treatment of schizophrenia. II. Two-year effects of a controlled study on relapse and adjustment. Arch Gen Psychiatry 48: 340–347

Johnson S, Prosser D, Bindman J (1997) Continuity of care for the severely mentally ill: concepts and measures. Soc Psychiatry Psychiatr Epidemiol 32: 137–142

Jones M (1953) The therapeutic community. Aronson, New York

Jones M (1983) Therapeutic community as a system for change. In: Gunderson JG, Will OA, Mosher LR (eds) Principles and practice of milieu therapy. Aronson, New York, pp 177–184

Klein M (1930) The psychotherapy of the psychoses. Br J Med Psychol 10: 242–244

Kopelowicz A, Liberman RP (1995) Biobehavioral treatment and rehabilitation of schizophrenia. Harv Rev Psychiatry 3: 55–64

Leff J, Kuipers L, Berkowitz R, Sturgeon D (1985) A controlled trial of social intervention in the families of schizophrenic patients: two year follow-up. Br J Psychiatry 146: 594–600

Leff J, Berkowitz R, Shavit N, Strachan A, Glass I, Vaughn C (1990) A trial of family therapy versus a relatives' group for schizophrenia: two-year follow-up. Br J Psychiatry 157: 571–577

Lotterman A (1996) Specific techniques for the psychotherapy of schizophrenic patients. International University Press, Madison

Marshall M, Lockwood A, Gath D (1995) Social services case-management for long-term mental disorders: a randomised controlled trial. Lancet 345: 409–412

McFarlane WR, Lukens E, Link B et al (1995) Multiple-family groups and psychoeducation in the treatment of schizophrenia. Arch Gen Psychiatry 52: 679–687

*Mentzos S (1992) Psychose und Konflikt. Vandenhoeck and Ruprecht, Göttingen

Mojtabai R, Nicholson RA, Carpenter BN (1998) Role of psychosocial treatments in management of schizophrenia: a meta-analytic review of controlled outcome studies. Schizophr Bull 24(4): 569–586

Mosher LR, Menn AZ (1978) Community residential treatment for schizophrenia: two year follow-up. Hosp Community Psychiatry 29: 715–723

Mosher LR, Menn A, Matthews SM (1975) Soteria: evaluation of a home-based treatment for schizophrenia. Am J Orthopsychiatry 45: 455–467

**Mueser KT, Bond GR, Drake RE, Resnick SG (1998) Models of community care for severe mental illness: a review of research on case management. Schizophr Bull 24: 37–74

Müller C (1972) Psychotherapie und Soziotherapie der endogenen Psychosen. In: Kisker KP, Meyer JE, Müller M, Stömgren E (eds) Psychiatrie der Gegenwart, vol 3, 2nd edn. Springer, Berlin Heidelberg New York

Munich RL (1987) Conceptual trends and issues in the psychotherapy of schizophrenia. Am J Psychother 41: 23–37

Paul GL, Lentz RJ (1977) Psychosocial treatment of chronic mental patient. Milieu versus social-learning programs. Harvard University Press, Cambridge

**Penn DL, Mueser KT (1996) Research update on the psychosocial treatment of schizophrenia. Am J Psychiatry 153: 607–617

Quinlivan R, Hough R, Crowell A, Beach C, Hofstetter R, Kenworthy K (1995) Service utilization and costs of care for severely mentally ill clients in an intensive case management program. Psychiatr Serv 46: 365–371

Randolph ET, Eth S, Glynn S et al (1994) Behavioural family management in schizophrenia: outcome of a clinic-based intervention. Br J Psychiatry 164: 501–506

Retzer A (1994) Familie und Psychose. Fischer, Stuttgart

Rogers CR (1957) The necessary and sufficient conditions of therapeutic personality change. J Consult Psychol 21: 95

Rosberg J, Stunden AA (1990) The use of direct confrontation: the treatment-resistant schizophrenic patient. Acta Psychiatr Scand 81: 352–358

Rosenheck R, Neale M, Leaf P, Milstein R, Frisman L (1995) Multisite experimental cost study of intensive psychiatric community care. Schizophr Bull 21: 129–140

Rössler W, Löffler W, Fätkenheuer B, Riecher-Rössler A (1993) Does case management reduce the rehospitalization? Acta Psychiatr Scand 86: 445–449

Rund BR, Moe LC, Sollien T et al (1994) An efficiency study of a psychoeducational treatment programme for schizophrenic adolescents. Acta Psychiatr Scand 89: 211–218

Schindler R (1980) Die Veränderung psychotischer Langzeitverläufe nach Psychotherapie. Psychiatr Clin 13: 206–216

Schmidt-Michel PO, Ostroga G, Kenntner S, Konrad M, Krüger M, Hoffman M (1992) Rehabilitationsverläufe in der psychiatrischen Familienpflege. Nervenarzt 63: 34–41

*Schooler NR, Keith SJ, Severe JB et al (1997) Relapse and rehospitalization during maintenance treatment of schizophrenia: the effects of dose reduction and family treatment. Arch Gen Psychiatry 54: 453–463

Schwing G (1954) A way to the soul of the mentally ill. International University Press, New York

Sechehaye MA (1986) Eine Psychotherapie der Schizophrenen. Klett-Cotta, Stuttgart

*Selvini Palazzoli M (1985) Paradox and counterparadox: a new model in the therapy of the family in schizophrenic transaction. Aronson, New York

Selvini Palazzoli M, Boscolo L, Cecchin F, Prata G (1980) Hypothesizing – circularity – neutrality: three guidelines for the conductor of the sessions. Fam Proc 19: 3–12

Simon FB (1988) Unterschiede, die Unterschiede machen. Klinische Epistemologie: Grundlagen einer systemischen Psychiatrie und Psychosomatik. Springer, Berlin Heidelberg New York

Simon FB (1990) Meine Psychose, mein Fahrrad und ich. Zur Selbstorganisation der Verrücktheit. Auer, Heidelberg

Simon FB (1993) Die Kunst der Chronifizierung. System Familie 6: 139–150

Simon FB, Weber G (1987) "Wie chronifiziere ich meine Patienten am besten?" Ein Sience-fiction-Märchen. In: Stierlin H, Simon FB, Schmidt G (eds) Familiäre Wirklichkeiten. Klett-Cotta, Stuttgart, pp 157–163

Simon FB, Weber G, Stierlin H, Retzer A, Schmidt G (1989) "Schizo-affektive" Muster: Eine systemische Beschreibung. Familiendynamik 14: 190–213

Simon H (1929) Die aktivere Krankenbehandlung in der Irrenanstalt. de Gruyter, Berlin Leipzig

Sledge WH, Astrachan B, Thompson K et al (1995) Case management in psychiatry: an analysis of tasks. Am J Psychiatry 152: 1259–1265

Stanton AH, Gunderson JG, Knapp PH et al (1984) Effects of psychotherapy in schizophrenia. I. Design and implementation of a controlled study. Schizophr Bull 10: 520–563

Stierlin H (1975) Von der Psychoanalyse zur Familientherapie. Klett, Stuttgart

Sullivan HS (1931) Socio-psychiatric research: its implications for the schizophrenia problem and for mental hygiene. Am J Psychiatry 10: 977–991

Sullivan HS (1962) Schizophrenia as a human process. Norton, New York

Tarrier N, Barrowclough C, Vaughn C, Bamrah JS, Porceddu K, Watts S, Freeman H (1989) Community management of schizophrenia. A two-year follow-up of a behavioural intervention with families. Br J Psychiatry 154: 625–628

Tarrier N, Lowson K, Barrowclough C (1991) Some aspects of family interventions in schizophrenia. II. Financial considerations. Br J Psychiatry 159: 481–484

Tarrier N, Beckett R, Harwood S, Baker A, Yusupoff L, Ugarteburu I (1993) A trial of two cognitive-behavioural methods of treating drug-resistant residual psychotic symptoms in schizophrenia patients. I. Outcome. Br J Psychiatry 162: 524–532

Tarrier N, Barrowclough C, Porceddu K, Fitzpatrick E (1994) The Salford Family Intervention Project: relapse rates of schizophrenia at five and eight years. Br J Psychiatry 165: 829–832

Test MA (1979) Continuity of care in community treatment. N Direct Ment Health Serv 2: 15–23

Torrey EF (1986) Continuous treatment teams in the care of the chronic mentally ill. Hosp Community Psychiatry 37: 1243–1247

Tucker GJ (1983) Therapeutic communities. In: Gunderson JG, Will OA, Mosher LR (eds) Principles and practice of milieu therapy. Aronson, New York

Van Putten T, May PRA (1976) Milieu therapies of the schizophrenias. In: West LJ, Flinn DE (eds) Treatment of schizophrenia. Progress and prospects. Grune and Stratton, New York, pp 217–243

Wallace CJ, Liberman RP (1985) Social skills training for patients with schizophrenia: a controlled clinical trial. Psychiatry Res 15: 239–247

Wallerstein RS (1989) Follow-up in psychoanalysis clinical and research values. J Am Psychoanal Assoc 37(4): 921–941

*Werman DS (1984) The practice of supportive psychotherapy. Brunner Mazel, New York

Willis E, Reimer F (1988) Arbeitstherapie und berufliche Rehabilitation aus der Sicht des Psychiatrischen Landeskrankenhauses. In: Schubert A, Reihl D, Bungard W (eds) Chancen im Arbeitsleben für psychisch Kranke. Ehrenhof, Mannheim

Xiong W, Phillips MR, Hu X, Ruiwen W, Dai Q, Kleinman J, Kleinman A (1994) Family-based intervention for schizophrenic patients in China: a randomised trial. Br J Psychiatry 165: 239–247

Zhang M, Wang M, Li J, Phillips MR (1994) Randomised-controlled trial of family intervention for 78 first-episode male schizophrenic patients: an 18-month study in Suzhou, Jiangsu. Br J Psychiatry 165: 96–102

CHAPTER

13

T.E. Smith, R.P. Liberman, A. Kopelowicz

Schizophrenic Disorders: Rehabilitation

1 Importance of Rehabilitation Treatments for Schizophrenia

Pharmacotherapy with antipsychotic, mood-stabilizing, and antidepressant drugs forms the basis of treatment for the major psychotic and mood disorders, which include functional disability as a diagnostic criterion. While the evidence for the efficacy of these agents for the acute and maintenance treatment of these disorders is quite robust, limitations in their impact have been noted for primary negative symptoms, cognitive deficits, psychosocial functioning, and quality of life (Liberman et al. 1995). Side effects of maintenance pharmacotherapy diminish the value of these treatments, with examples including the sedation, autonomic, and neurological side effects that can impair social and vocational functioning (Mintz et al. 1992).

In addition, the subjective distress caused by side effects can trigger noncompliance with the drug regimen, with high risk for relapse. While the introduction of atypical antipsychotic drugs appears to confer additional benefits in terms of improvement in neurocognitive deficits (Green et al. 1997) and reductions in negative symptoms and side effects (Marder and Meibach 1994), the degree to which these new drugs will favorably alter the long-term course of psychotic disorders remains to be seen.

During the past two decades, substantial evidence has accumulated for the effectiveness of psychosocial interventions to supplement drug therapy for disabling mental disorders (Dilk and Bond 1996; Penn and Mueser 1996; Smith et al. 1996; Scott and Dixon 1995; Marder et al. 1996; Liberman and Kopelowicz 1995). Psychosocial interventions such as behavioral family management, social skills training, assertive case management, and supported employment have been designed to (a) reduce the stress experienced by individuals who are vulnerable to relapse – especially stressors from the family emotional climate, (b) strengthen the individual's coping capacities to improve social functioning, and (c) provide social supports that compensate for the deficits in community functioning and buffer stressors that afflict the seriously mentally ill. The advent of a combined biobehavioral approach to treatment of individuals with disabling forms of mental disorders has led to the development of the field of psychiatric rehabilitation.

Psychiatric rehabilitation can be defined as comprising those biobehavioral interventions that are aimed at enabling an individual with a disabling mental disorder to (a) build skills that improve adaptive functioning, (b) achieve personally relevant goals that are consistent with as high a level of independence and quality of life as is feasible, and (c) live in a supportive environment that enables the person to enjoy a higher quality of life when symptoms or deficiencies in life skills persist despite the best efforts at rehabilitation. Interventions are considered rehabilitative if they remove obstacles or impediments to these goals (e.g. symptoms, bizarre behavior, deficits in social functioning).

The conceptual framework underlying psychiatric rehabilitation is the vulnerability–stress–protective factors model of serious mental disorder. Vulnerability is presumed to be biologically and genetically mediated and to persist even during periods of symptomatic remission. Thus the aim of psychiatric rehabilitation is to reduce stressors and provide protection against vulnerability through strengthening the individual and modifying the environment.

Several conclusions can be articulated from the research literature on psychiatric rehabilitation. Psychosocial treatments are treatment specific; in other words, to achieve favorable vocational outcomes, vocational rehabilitation has to be highly structured and well-organized (e.g. supported employment), and to achieve improvements in social competence, structured methods of social skills training must be employed (Kopelowicz and Liberman 1998). A second conclusion is the importance of long-term deployment of psychosocial interventions for improving community tenure of patients. This has been well demonstrated in the studies of assertive community treatment where rehospitalization rates remain low as long as the case management continues to provide interventions with outreach and mobility. Finally, schizophrenia is a stress-related disorder; thus interventions aimed at reducing stress (e.g. high expressed emotion within families) are likely to succeed in reducing relapse rates (Scott and Dixon 1995).

Psychosocial treatment should be combined with optimal types and doses of antipsychotic medications, so that symptoms and side effects do not interfere with compliance and instrumental role functioning. Negative symptoms and conceptual disorganization/thought disorder, in particular, have been associated with poor outcomes from psychosocial treatment (Mueser et al. 1991; Kopelowicz et al. 1997). The new antipsychotic drugs may offer special promise in terms of reducing the neurocognitive and learning disabilities that can compromise the psychosocial rehabilitation of individuals with schizophrenia (Green 1996; Green et al. 1997; Mueser et al. 1991).

Finally, treatment and rehabilitation should be linked to the phase of the patient's disorder; thus the type of psychosocial intervention appropriate for the acute and florid phase of illness (e.g. involving the

family in a therapeutic alliance as through psychoeducation) will not necessarily be the same intervention used in the stable or recovery phase of the disorder (e.g. supported employment, intensive social skills training for conversation and friendship skills).

Psychosocial treatment and rehabilitation can be categorized as primarily remedial or primarily compensatory in terms of the psychosocial and functional deficits of individuals with schizophrenia. In remedial strategies, attempts are made to improve the individual's functional deficits as by teaching skills and broadening the individual's behavioral repertoire, coping capacities, and resilience to stress. Social skills training is an example of this strategy. In compensatory approaches, treatment aims to create environmental supports for the individual that reduce the burden, stress, and requirements for functioning, while at the same time providing for the individual's needs and quality of life. Assertive case management or training in community living, as well as supported employment and supported living techniques, reflect successful compensatory strategies of intervention (Test 1992; Burns and Santos 1995; Hromco et al. 1997; Bond et al. 1997).

Compensatory and remedial strategies for psychosocial rehabilitation can be viewed as means of overcoming obstacles to normative social and role functioning that are found in the environment, the individual, and in the roles that individuals must perform to function within the community. These various obstacles, or variables, are depicted in Fig. 1. Remedial interventions aim to strengthen an individual's coping capacity by teaching skills, while compensatory interventions are primarily imbedded in environmental supports. The following two sections of this chapter will review these strategies in detail.

2 Rehabilitation Services

Specific rehabilitation services include social skills training, family psychoeducation, cognitive remediation, vocational rehabilitation, and self-help programs. The utility and effectiveness of these services have been recognized, and they are critical elements of rehabilitation programs around the world.

2.1 Social Skills Training

Social skills training is defined by behavioral techniques or learning activities that enable patients to acquire instrumental and affiliative skills in domains required to meet the interpersonal, self-care, and coping demands of community living. Skills training can be done with individuals, patient groups, or families and may continue for years as the person's

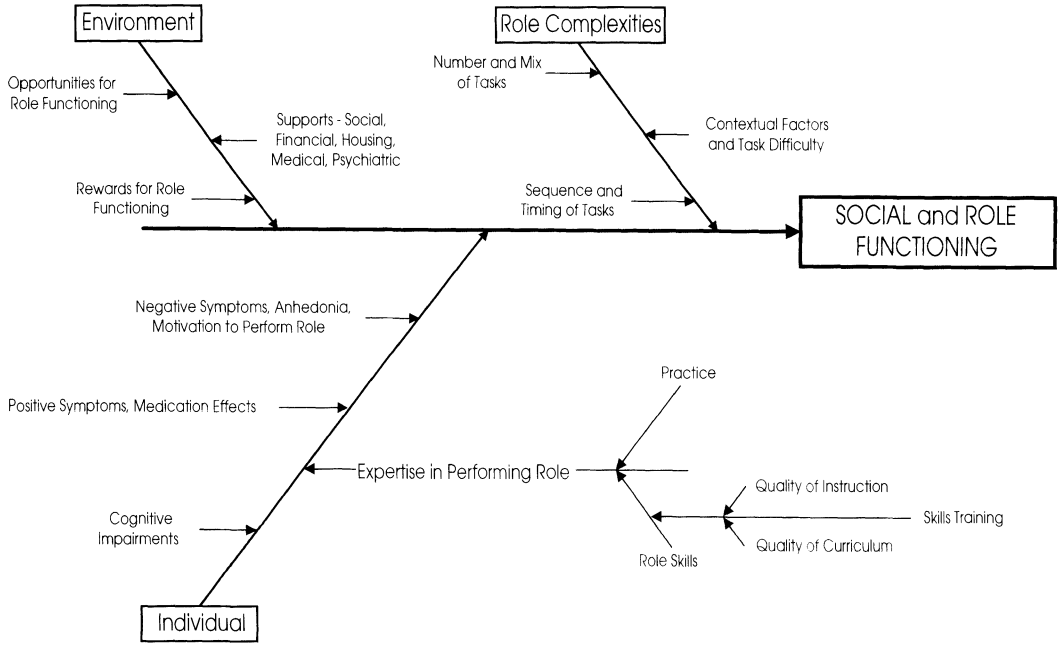


Fig. 1. Factors that influence social and role functioning

Table 1. Framework for social problem solving

	Problem-solving sequence	Cognitive-behavioral requirements
Step 1	Identifying the problem	Social and self-perception
Step 2	Generating list of potential solutions	Information-processing capacities
Step 3	Reviewing pros and cons of potential solutions	Information-processing capacities
Step 4	Choosing a solution	Decision-making abilities
Step 5	Implementing the plan	Behavioral skills
Step 6	Reviewing the outcome	Social and self-perception

abilities, goals, and values ascend a hierarchy of community adaptation.

Skills training approaches draw upon theoretical models of social problem-solving, which involve variations on a stepwise process of social perception, information processing, and behavioral response (Bellack et al. 1994). Patients with chronic psychotic disorders often have difficulty in accurately perceiving and interpreting affective and cognitive cues that are critical elements of communication. Training skills in social perception address these deficits and provide a foundation upon which more specific social and coping skills can be developed. As a next step, skills training uses cognitive techniques to teach strategies for identifying social problems, generating and evaluating alternative solutions, and choosing a plan of action. As a final step, skills training addresses discrete verbal, paralinguistic, and nonverbal skills which are used to effect a competent social response. Table 1 lists these social problem-solving strategies in sequential order.

Based on the social problem-solving model of social skills training, a set of psychoeducational modules has been developed at the University of California at Los Angeles (UCLA) Clinical Research Center for Schizophrenia and Psychiatric Rehabilitation (Liberman et al. 1993). Specific training modules target skills for self-management of antipsychotic medication, symptom management, grooming and personal hygiene, recreation for leisure, interpersonal problem-solving, job finding, community reintegration, safe and satisfying sex, family coping, and engaging in friendly conversations. These modules use instructional techniques including didactic lectures, videotape demonstrations, role-playing, and in vivo homework assignments to help patients master specific skill areas. Each module is a self-contained package that can be adopted alone or in combination with other modules in comprehensive rehabilitation programs.

Meta-analyses and reviews of the more than 50 controlled studies of social skills training have shown that individuals with schizophrenia can acquire and retain skills and that training is associated with significant favorable effects on social adjustment, symptoms, relapse, and rehospitalization rates (Benton and Schroeder 1990; Dilk and Bond 1996; Smith et al. 1996; Hogarty et al. 1997). Skills training is especially effective if it is intensive (more than two sessions per week) and of sufficient duration (at least 6 months). While schizophrenic patients with even high levels of hallucinations and delusions can acquire skills through systematic training, cognitive disorganization (e.g. severe distractibility and thought disorder) and the deficit syndrome (i.e. primary negative symptoms) are likely to interfere with the training process.

2.2 Family Psychoeducation

Many studies have replicated the finding that family stress, often reflected in high expressed emotion attitudes of criticism and emotional overinvolvement toward the mentally ill relative, is a powerful predictor of relapse in schizophrenia (De Jesus Mari and Streiner 1994). Several modes of family intervention have therefore been designed and empirically validated for the ability to equip relatives with coping skills and thereby change the emotional climate of the family and reduce the incidence of relapses and rehospitalizations. Patients who have contact with their family members will benefit from participating in a psychoeducational and skill-building program aimed at improving communication and problem-solving.

Approaches to family psychoeducation share several features. First, it is essential that the treatment team develop collegial relationships with family members and other support persons. In this context of collaboration, specific psychoeducational techniques are used with the aim of teaching what is known scientifically about the patient's mental disorder and directing the family and other caregivers to locally available treatment and rehabilitation services. Another feature involves the assumption of least pathology. The attitude is taken that family members are always doing the best they can and acting in the interests of the

patient and family, given their coping capacities and biobehavioral vulnerabilities. This approach sets the stage for the development of communication and problem-solving abilities. To produce durable clinical effects, family interventions must go beyond education to train family members in necessary coping skills, including basic communication and contingency management. Examples of communication skills that form the basis for effective problem-solving include active listening, giving positive feedback, and making positive requests.

Over two dozen well-controlled studies in the past decade have examined the effects of family psychoeducation on patients and their families (Scott and Dixon 1995; Solomon et al. 1996; Anderson et al. 1986; McFarlane et al. 1995). Results showed that the relapse and hospitalization rates for patients who participated in family psychoeducation were significantly less than for patients who completed various comparison treatments. Studies have also shown that family psychoeducation significantly lowers family burden and improves self-efficacy and esteem.

2.3 Cognitive Remediation

In recent years, there has been interest in cognitive remediation treatments for schizophrenia. The cognitive and neuropsychological deficits that are core features of the disorder play a role in determining the success of rehabilitation strategies, with studies showing that enduring thought disorder and short-term memory and verbal learning deficits are more predictive of skill acquisition than psychotic symptoms (Mueser et al. 1991; Kern et al. 1992; McKee et al. 1997). In addition, deficits in vigilance, memory, and executive functioning have also repeatedly been associated with social skills and overall social adjustment (Green 1996). While many clinicians have assumed that the cognitive deficits of schizophrenia represent an irreversible and enduring form of dementia and hence cannot be mitigated by rehabilitation, evidence is accumulating to justify strategies for remediation of these basic cognitive deficits.

Two types of cognitive remediation strategies have been developed. The first involves direct remediation of basic cognitive deficits. Demonstrations have shown that laboratory-based measures of cognitive dysfunctions such as vigilance and card sorting can improve significantly with behavioral training (Benedict et al. 1994; Stratta et al. 1994). Remediation strategies include repeated practice, instructional modification, positive reinforcement (e.g. money), and errorless learning that teaches discrimination and problem-solving in small steps where success is maximized and trial-and-error learning is minimized. Empirical evaluation of the efficacy of these direct approaches to cognitive training suggests that striking improvements are possible in the cognitive tasks, but the links between improvements in laboratory-based molecular levels and in molar social and clinical status have yet to be satisfactorily documented (Penn and Mueser 1996). Current research aims to promote clinical generalization of direct cognitive remediation through identifying and strengthening the cognitive, behavioral, and social processes that mediate learning of adaptive skills.

A different remediation strategy targets amelioration of psychotic symptoms through cognitive restructuring and behavioral learning principles. To date, strategies have been developed for treating delusions, hallucinations, and negative symptoms. Sometimes referred to as cognitive strategy enhancement, these approaches involve the identification of specific symptoms with subsequent training in the use of cognitive coping strategies including distraction, reframing, self-reinforcement, reality testing, and verbal challenging. Several reports have shown these strategies to be effective, at least during short-term follow-up in hospital and clinic settings (Tarrier et al. 1993).

2.4 Vocational Rehabilitation

Vocational rehabilitation techniques for schizophrenia have improved dramatically since the era of institution-bound, sheltered work programs. In the 1970s and 1980s, transitional employment programs were developed largely within psychosocial rehabilitation clubs and often without input from mental health professionals. Transitional employment comprised prevocational work activities or work enclaves in industrial settings that had the philosophy of "train then place" in a real job. A job in the competitive work sector was considered stressful, requiring gradual work hardening. Evaluations of the transitional employment model have produced mixed results, and there are concerns that this approach may be less effective due to the disparity between the controlled environments of prevocational activities and the competitive workplace (Wallace 1993).

The supported employment model of rehabilitation evolved in an effort to improve these vocational outcomes. This approach arises from the understanding that individuals with psychiatric disabilities require ongoing rehabilitation and support after they secure competitive employment. It de-emphasizes the importance of prevocational training, advocating instead a "place then train" approach. Clients are placed in employment settings based on their interests and

abilities and then offered the training and supports necessary to maintain their positions. In its fully applied form, individuals are offered services indefinitely, with job coaches visiting workplaces to help with learning and retention of technical, interpersonal, and problem-solving skills required to sustain employment.

Evaluations of supported employment have revealed successful job placements in competitive, community-based work in over 50% of participants (Bond et al. 1997; Drake and Becker 1996). Placement rates are substantially higher when the rehabilitation work is part of an integrated mental health treatment plan. Supported employment requires close collaboration and communication between the client, the vocational specialist or job coach, and the interdisciplinary treatment team. This is because many of the factors that negatively influence vocational rehabilitation are those commonly addressed by the clinical treatment team: severity of psychopathology (especially conceptual disorganization and negative symptoms), stress in the family (expressed emotion), neurocognitive deficits, and poor premorbid social and work adjustment.

3 Social and Community Support Programs

Over the past three decades, many countries have transferred the care of the seriously mentally ill from psychiatric hospitals to less restrictive community settings. This shift in locus of care required more aggressive outreach on the part of treatment providers, which is conceptualized in the community support/case management model. In this approach, interventions are delivered in a coordinated fashion by multidisciplinary teams of clinicians who assume long-term responsibility across the spectrum of mental health services, including inpatient units, outpatient programs, and psychosocial rehabilitation centers. Several key principles guide effective case management (Ellison et al. 1995). These include the commitment to treating clients with dignity and confidentiality and the requirement that services be adapted to the changing needs and preferences of each client on the basis of self-determined goals. Most importantly, the service delivery system must provide comprehensive, accessible services for as long as and whenever the client needs them, in settings that are the least restrictive and most normalized.

The many functions provided in case management/community support models of care delivery include the following:

- Outreach and engagement
- Support of basic needs
- Mental health care and treatment
- Crisis/emergency services
- Comprehensive psychosocial services
- Range of housing options
- Support and education for providers and families
- Development of natural supports
- Advocacy and protection
- Case management

Case management programs have been shown to increase community tenure of previously institutionalized patients (Solomon 1992).

Among the most effective adaptations of this approach is the Training in Community Living or Program of Assertive Community Treatment (PACT) developed in Madison, Wisconsin (Burns and Santos 1995; Test 1992). The PACT model uses broad-spectrum case management organized in round-the-clock continuous treatment teams. This program is effective in both rural and urban settings and has been replicated throughout the United States and Europe.

Although it is generally agreed that case management is a desired service for individuals with serious and persistent mental illness, there is little consensus as to which elements of the case management process are most clinically useful.

A case manager in a psychiatric rehabilitation setting has several specific tasks to perform. These include the following: assisting clients in building social networks; facilitating access to housing and employment opportunities; helping clients interact with various service organizations; teaching clients skills for illness management; monitoring clinical progress; and, when necessary, undertaking timely clinical interventions. By acting as the fixed point of responsibility within a continuum of care, case managers contribute to improved vocational functioning, less social isolation, and more independent living.

Another outgrowth of the deinstitutionalization era was the development of fellowship clubs of formerly hospitalized patients. Psychosocial clubhouses now exist in many large cities in Europe and the United States, including the Fountain House in New York City and Thresholds in Chicago. Clubhouses provide basic opportunities for acceptance, friendship, advocacy, housing, destigmatization, and social and recreational activities. Although the psychosocial clubhouse movement developed separately from the medically oriented community mental health approach, the past few years have witnessed an increase in the number of community mental health centers that have converted their traditional day treatment programs to the clubhouse model.

Central to the psychosocial self-help philosophy is the belief that individuals with mental disabilities have a fundamental right to work, socialization, and a home and that those basic needs, when satisfied, generate self-esteem and a positive identity necessary for community adjustment. Thus psychosocial clubhouses focus on developing employment opportunities, peer support, and housing programs tailored to the capabilities of their members. The success of these programs, including those operated by patients themselves (consumer-run agencies), is reflected by reimbursement of the services by government insurance programs, by their certification by accreditation agencies, and by the vitality of the International Association of Psychosocial Rehabilitation Services, an organization spawned by the clubhouse network. Increasingly, psychosocial self-help programs are evolving toward full service enterprises with assertive community management, pharmacotherapy, and community support.

4 References

Anderson CM, Reiss DJ, Hogarty GE (1986) Schizophrenia and the Family. Guilford, New York

Bellack AS, Sayers MD, Mueser KT, Bennett M (1994) Evaluation of social problem solving in schizophrenia. J Abnorm Psychol 103: 371–378

Benedict RHB, Harris AE, Markow T, McCormick JA, Nuechterlein KH, Asarnow RF (1994) Effects of attention training on information processing in schizophrenia. Schizophr Bull 20: 537–546

Benton MK, Schroeder HE (1990) Social skills training with schizophrenics: a meta-analytic evaluation. J Consult Clin Psychol 58: 741–747

Bond GR, Drake RE, Mueser KT, Becker DR (1997) An update on supported employment for people with severe mental illness. Psych Serv 48: 335–346

Burns BJ, Santos AB (1995) Assertive community treatment: an update of randomized trials. Psych Serv 46: 669–675

De Jesus Mari J, Streiner DL (1994) An overview of family interventions and relapse on schizophrenia: meta-analysis of research findings. Psychol Med 24: 565–578

Dilk MN, Bond GR (1996) Meta-analytic evaluation of skills training research for individuals with severe mental illness. J Consult Clin Psychol 64: 1337–1346

Drake RE, Becker DR (1996) The individual placement and support model of supported employment. Psych Serv 47: 472–475

Ellison ML, Rogers ES, Sciarappa K, Cohen M, Forbess R (1995) Characteristics of mental health case management: results of a national survey. J Ment Health Admin 22: 101–112

**Green MF (1996) What are the functional consequences of neurocognitive deficits in schizophrenia? Am J Psychiatry 153: 321–330

Green MF, Marshall BD, Wirshing WC, Ames D, Marder SR, McGurk S, Kern RS, Mintz J (1997) Does risperidone improve verbal working memory in treatment-resistant schizophrenia? Am J Psychiatry 154: 799–804

Hogarty GE, Kornblith SJ, Greenwald DP, DiBarry AL, Cooley S, Ulrich RF, Carter M, Flesher S (1997) Three-year trials of personal therapy among schizophrenic patients living with or independent of family. I. Description of study and effects on relapse rates. Am J Psychiatry 154: 1504–1513

Hromco JG, Lyons JS, Nikkel RE (1997) Styles of case management: the philosophy and practice of case managers. Comm Ment Health J 33: 415–428

Kern RS, Green MF, Satz P (1992) Neuropsychological predictors of skills training for chronic psychiatric patients. Psychiatry Res 43: 223–230

Kopelowicz A, Liberman RP (1998) Psychological and behavioral treatments for schizophrenia. In: Nathan PE, Gorman JM (eds) Treatments that work. Oxford University Press, London, pp 190–211

Kopelowicz A, Liberman RP, Mintz J, Zarate R (1997) Comparison of efficacy of social skills training for deficit and nondeficit negative symptoms in schizophrenia. Am J Psychiatry 154: 424–425

Liberman RP, Kopelowicz A (1995) Basic elements in biobehavioral treatment and rehabilitation of schizophrenia. Int Clin Psychopharmacol 9[Suppl 5]: 51–58

Liberman RP, Wallace CJ, Blackwell G, Eckman TA, Vaccaro JV, Kuehnel TG (1993) Innovations in skills training for the seriously mentally ill: the UCLA social and independent living skills modules. Innovations Res 2: 43–60

Liberman RP, Vaccaro JV, Corrigan PW (1995) Psychiatric Rehabilitation. In: Kaplan HI, Sadock BJ (eds) Comprehensive textbook of psychiatry, vol VI. Williams and Wilkins, New York, pp 2696–2717

Marder SR, Meibach RC (1994) Risperidone in the treatment of schizophrenia. Am J Psychiatry 151: 825–835

Marder SR, Wirshing WC, Mintz J, McKenzie J, Johnston K, Eckman TA, Lebell M, Zimmerman K, Liberman RP (1996) Two-year outcome of social skills training and group psychotherapy for outpatients with schizophrenia. Am J Psychiatry 153: 1585–1592

McFarlane WR, Lukens E, Link B, Dushay R, Deakins SA, Newmark M, Dunne EJ, Horen B, Toran J (1995) Multiple-family groups and psychoeducation in the treatment of schizophrenia. Arch Gen Psychiatry 52: 679–687

McKee M, Hull JW, Smith TE (1997) Cognitive and symptom correlates of participation in social skills training groups. Schizophr Res 23: 223–229

Mintz J, Mintz LI, Phipps CC (1992) Treatments of mental disorders and the functional capacity to work. In: Liberman RP (ed) Handbook of psychiatric rehabilitation. Allyn and Bacon, Boston, pp 290–316

Mueser KT, Bellack AS, Douglas MS, Wade JH (1991) Prediction of social skill acquisition in schizophrenic and major affective disorder patients from memory and symptomatology. Psychiatry Res 37: 281–296

**Penn DL, Mueser KT (1996) Research update on the psychosocial treatment of schizophrenia. Am J Psychiatry 153: 607–617

**Scott JE, Dixon LB (1995) Psychological interventions for schizophrenia. Schizophr Bull 21: 621–630

Smith TE, Liberman RP, Bellack AS (1996) Social skills training for schizophrenia: review and future directions. Clin Psychol Rev 16: 599–617

Solomon P (1992) The efficacy of case management services for severely mentally disabled clients. Community Ment Health J 28: 163–180

Solomon P, Draine J, Mannion E, Meisel M (1996) Impact of brief family psychoeducation on self-efficacy. Schizophr Bull 22: 41–50

Stratta P, Mancinia F, Mattei P, Casacchia M, Rossi A (1994) Information processing strategy to remediate Wisconsin Card Sorting Test performance in schizophrenia: a pilot study. Am J Psychiatry 151: 915–918

Tarrier N, Beckett R, Harwood S, Baker A, Yusopoff L, Ugareburu I (1993) A trial of two cognitive-behavioral methods of treating drug-resistant residual psychotic symptoms in schizophrenic patients. I. Outcome. Br J Psychiatry 162: 524–532

Test MA (1992) Training in community living. In: Liberman RP (ed) Handbook of psychiatric rehabilitation. Macmillan, New York

Wallace CJ (1993) Psychiatric rehabilitation. Psychopharmacol Bull 29: 537–548

CHAPTER

14

K.-T. Kronmüller, C. Mundt

Depressive Episodes

Translator: E. Taub

1 Introduction

Depressive disorders are among the most common mental illnesses. The prevailing concepts of their frequency, classification, course, and treatment have changed markedly in the last two decades. Epidemiological studies have shown that depressive disorders are more common, and have a less favorable course, than previously assumed. Concepts of diagnosis have also changed in the same period. The entity of endogenous depression, a part of the systematic, triadic nosology of Kraepelin (1913), was abandoned in favor of depressive episodes, as in ICD-10 (WHO 1992a), and major depression, as in DSM-IV (APA 1994), which became the new major diagnostic categories for depressive disorders. Meanwhile, many related diagnostic categories were developed to provide a more detailed description of the entire psychopathological spectrum of depressive disorders.

In this chapter, we describe the depressive episode and recurrent depressive disorder and present the current state of research into this disorder. This illness essentially corresponds to the major depression of DSM-IV (APA 1994) and is one of the main categories of the current diagnostic scheme for depression. We discuss not only its clinical manifestations, but also instruments that may be used for diagnostic assessment and differential diagnostic viewpoints. Finally, we consider research on the course of depressive disorders and empirical findings about prognostic factors.

2 Manifestations of Illness

2.1 Clinical Features

The term "depression" can be traced back to Cullen (1800). His use of the term reflected a pathophysiological view of this illness as being due to a loss of central vascular tone. This conception was in accordance with the tradition of British sensualism and with the "iatrophysical" characterization of this disease as secondary to a reduction of humoral tone in the brain (Schmidt-Degenhard 1983). The descriptive phenomenological content of the term was first worked out in our modern sense in the earlier decades of the nineteenth century, most notably by Heinroth (1818) (see Schmidt-Degenhard 1991). Nonetheless, interference from the competing concept of melancholia led to difficulties in distinguishing illnesses of relatively autonomous course, with or without psychotic content, from states similar to the familiar mood changes of everyday life. The term "endogenous depression," recently abandoned, also reflected this problem.

There was little movement in the descriptive phenomenology of depression from this early period until recent decades, when questions of symptomatology came under discussion with respect to classification. Nonetheless, many current textbooks and manuals of depression no longer contain a chapter devoted to the descriptive phenomenology of depressive illnesses (Beckham and Leber 1995). Instead, information about depressive manifestations is subsumed under the chapters on epidemiology or classification (Andreasen and Black 1993). There is a tacit assumption that the psychopathological phenomena may, in principle, be accepted as already defined and understood. Detailed descriptions of the manifestations of illness are today more likely to be found in the numerous books that have appeared for the benefit of affected individuals and their families (Luderer 1994; Wittchen et al. 1995; Wolfersdorf 1994).

Nonetheless, the traditional European conception of psychopathology differs from the American conception reflected in DSM, and it is precisely this discrepancy that necessitates a clear understanding of the basic elements of each. Until the 1970s, for example, German psychopathology customarily distinguished between subjectively experienced vital disturbances and objectifiable vegetative disturbances, depressive coenesthesiae in the sense of disturbances of body sensation and body image, and somatoform disorders, which were similar to conversion disorders. This framework of classification no longer exists. The collective term "somatic complaints" (DSM-IV) obliterates the distinctions between these categories, even though there is clearly no theoretical difficulty in separating at least the subjective vital disturbances from the objectifiable, physical vegetative disturbances.

Furthermore, the international English-language literature on anthropological-phenomenological psychopathology (see Chap. 21, Vol. 1, Part 1) is scarcely taken into account, even though this field has yielded important contributions to depression research since its most productive period in Europe, which lasted from the 1950s to the 1970s. These contributions include the descriptions of disturbances of temporal and spatial representation by Erwin Strauss, Viktor von Gebsattel, Ludwig Binswanger, and others (see also Mundt 1998), as well as the existential personality traits that result from these disturbances in depressive patients, including altered relationships to themselves, the world, and the future, as was later discussed by Beck et al. (1979).

In addition to anthropological phenomenology – a holistic ontological approach focusing on personality traits as ideal types – two further theoretical approaches

have taken root in contemporary depression research, which we may call the nosological medical approach and the functional pathopsychological approach. Kuhs and Tölle (1987), for example, classify manifestations of depression in a manner derived from "classical" psychopathology and Kraepelin's medical model, breaking them down into affective disturbances and disturbances of initiative, cognition, perception, and memory. Further types of manifestations include disturbances of body sensation, vital disturbances, daily fluctuations, sleep disturbances, and accessory manifestations that are universal in psychiatry but take on a special coloration when they arise in the setting of depression (e.g. compulsion, alienation, and the development of histrionic manifestations).

Wolfersdorf (1995) attempted to classify the manifestations of depression with reference to basic mental functions, distinguishing disorders of cognition and affect regulation, of initiative, and of vegetative-symptomatic functions. Hautzinger (1997) proposed a similar classification scheme, dividing the manifestations of depression into emotional, motivational, cognitive-imaginative, vegetative-physiological, and behavioral planes.

The reader is referred to chapters in earlier editions of this textbook for consideration of individual manifestations and for the phenomenological conception of melancholia (see Kuhs and Tölle 1987; Kraus 1980). There have been no fundamentally new findings in symptomatology since these chapters appeared, and only a few new findings in phenomenology, notably the anthropological conception of depression by Fuch (1994) as secondary to "uprooting" – an explanatory model that goes beyond the idea of experiences of loss as depression-releasing "life events" and that is of major importance for the psychiatric care of immigrant patients.

2.2 Empirical/Statistical Description

None of the manifestations listed above are found exclusively in patients with depression; they occur with variable frequency, and they may occur in individual patients in highly variable combinations. Thus, to facilitate clinical diagnosis, many attempts have been made to group the manifestations of depression into patterns. This was, at first, a primarily phenomenological process, but the last two decades have seen the increasing use of empirical and statistical techniques for help in finding patterns and establishing constructs. Such techniques play a role complementary to that of phenomenological research (Mundt 1991).

A simple, early approach to depression research consisted of determining the frequency of occurrence of certain manifestations in depressed patients (Winokur et al. 1969). It was found that some manifestations, such as disturbances of sleep, mood, and concentration, occur almost universally. Others, such as loss of appetite, daily fluctuations, and hopelessness, are found in a bare majority of patients, and yet others, such as delusions, suicide attempts, and auditory hallucinations, are much rarer. Hamilton (1989) arrives at similar conclusions, but also states that manifestations of anxiety are almost always present in depressed patients.

2.3 Core and Peripheral Manifestations

Empirical frequency analyses are the point of departure for the construction of further patterns by means of multivariate statistical processes. The overall concept of this research strategy is that the multivariate description of patterns of disease manifestations, rather than the consideration of individual manifestations, ought to lead to the identification of more characteristic subgroups. This approach has been discussed by Nelson and Charney (1981) and by Steck (1988).

Such multivariate analyses may be classified with respect to three main features: the type of patients studied (depressed or non-depressed patients), the statistical techniques used (factor analysis, cluster analysis, or discriminance analysis; Bortz 1985), and, finally, whether individuals or traits are grouped. Nelson and Charney (1981) reviewed a number of multivariate studies and concluded that most of them clearly reveal the presence of a melancholic core group of manifestations, including severe mood disturbance and psychomotor abnormalities such as inhibition, depressive delusions, self-reproach, loss of interest, and inability to be influenced by the environment. In contrast, Steck (1988) reviewed 54 original publications and concluded that a reliable differentiation of endogenous from neurotic clinical pictures was not possible. The study by Matussek (1983) was the only one in which separate endogenous and neurotic syndromes were identified. The findings of most studies, however, support the dichotomous concept of depression proposed by Kendell (1976) and by Blashfield and Morey (1979), in which two types of depression differ with regard to both the quality and the severity of manifestations. The first type is characterized by guilt feelings, sleep disturbances, weight loss, and depression of mood, while the second type is characterized by fluctuating manifestations and lesser severity.

Paykel (1971) performed a study in which four different clusters with varying manifestations of depression were found: a psychotic cluster with the

typical manifestations of endogenous depression, an anxious cluster, a hostile cluster, and a fourth cluster comprising younger patients with personality disorders. These results were replicated in a further study, which also revealed that the clusters respond differently to psychopharmacological treatment (Paykel and Henderson 1977). Andreasen et al. (1980), Andreasen and Grove (1982), and Steinmeyer (1980) all identified three clusters largely corresponding to those of Paykel (1971).

Another approach of classification research consists of multivariate analysis and grouping of the clinical features of these disorders. Philipp and Mayer (1987) empirically studied a number of competing operationalized classification schemes for depressive disorders and endogenous depression, first comparing them with one another on a descriptive level and then determining their common features with multivariate statistical techniques. The 19 classification systems they studied were based on 72 different manifestations of anxiety and depression. Factor analysis and hierarchical cluster analysis were performed in a study group of 1000 patients with mental disturbances not due to any known physical cause. Using these techniques, Philipp et al. (1991) were able to derive a core syndrome of depression, which consists largely, if not entirely, of the same manifestations as the DSM-defined major depression.

Since then, empirically and statistically oriented symptom and classification research has yielded a large amount of further data, from which a few convergent findings emerge. It has not been found possible to separate depressions clearly into endogenous and neurotic syndromes. It has been possible, however, to characterize a core syndrome of depression and to distinguish it from less stable peripheral syndromes. This core syndrome essentially corresponds to the concept of major depression. The existence of melancholic, somatic, and endogenous subtypes of the core syndrome is also a replicable finding.

This classification of depressive syndromes is likely to be provisional. The inconstancy of the findings of multivariate analyses of disease manifestations is reflected in the multiplicity of diagnostic concepts in the area of the depressive disorders. Nevertheless, most of the empirical results can be integrated into a differential continuity hypothesis, by which typical clinical syndromes are increasingly formed as depression becomes more severe (Kendell 1976).

2.4
Transcultural Findings

In recent years, there has been an increasing amount of research, and of controversy, on the question of whether this pattern of manifestations of depression is uniformly valid across cultures or is subject to the influence of cultural factors and, if so, to what extent (Jenkins et al. 1990; Kleinman and Good 1985; Mezzich et al. 1996; Pfeiffer 1994; see also Chap. 14, Vol. 2, Part 1). Most authors agree that there is a culture-independent core syndrome consisting of sad or anxious mood, loss of initiative and interests, diminution of sleep, and loss of appetite (Pfeiffer 1971; Sartorius et al. 1980). There are cultural differences, however, particularly in the degree of severity of feelings of guilt and shame and in the tendency toward somatization (Jenkins et al. 1990).

The dependence of depression inventories on culture has been documented repeatedly (Manson 1996). The Center for Epidemiologic Studies Depression Scale (CES-D; Radloff 1977) is one of the more thoroughly investigated inventories. Studies of construct validation in different cultures reveal differential factor structures of the CES-D and indicate that instruments of measurement for depression are strongly sample dependent, and thus that the experience of depression has a different "internal structure" in different cultures (Manson 1996). In particular, the distinction between affective and somatic patterns of depressive manifestations could not be discerned in societies that do not possess a mind–body dichotomy of this type. These findings also cast doubt on the notion that the core syndrome of depression is culturally independent (Pfeiffer 1990).

3
Classification

3.1
Diagnostic Criteria and Classification Systems

Depression is not only one of the more common psychiatric disorders, but also that for which the greatest number of competing definitions, classifications, and operationalizations have been developed (Philipp et al. 1991). Linden (1979) explains this phenomenon as a consequence of the heterogeneity of depressive mood disorders, as manifested in the variable intensity of simultaneously appearing manifestations, the multiplicity of situational framework conditions, and the variability of disease course. This heterogeneity of the depressive syndrome results in a continuous demand for new attempts at classification.

The two leading definitions of the depressive disorders today are those of DSM-IV (APA 1994) and of ICD-10 (WHO 1992a, 1993; Dilling et al. 1991, 1994). They are both products of much debate and consensus formation (see Chaps. 2 and 3, Vol. 1, Part 2). The multiplicity of manifestations of depression has led to

numerous suggestions of criteria for the definition of this syndrome. An early attempt at operational criteria, the St. Louis Criteria of Feighner et al. (1972), opened the way to the Research Diagnostic Criteria (RDC) of Spitzer et al. (1978), which are the basis of the classification criteria in use today. In contrast to the "classical" concept of endogenous depression, as represented by Weitbrecht (1972), for example, the two classification systems in widest use today, ICD-10 and DSM-IV, postulate the existence either of a broadly conceived main syndrome of depression or of depressive episodes. Individual subgroups of depression are, in turn, differentiated from the main syndrome by means of further classifying criteria such as the psychopathological manifestations and course of the disorder.

The consensus definitions are the product of a long history of controversy. Kraepelin (1913) classified depression under the broad category of "manic-depressive insanity" (*Irresein*), a unitary conception of the affective disorders. Bleuler (1916) and Schneider (1932) perpetuated this conception. In the years that followed, many classification systems were proposed for subtypes of depressive disorders (Paykel 1992). Neurotic depression was defined in distinction to psychotic depression. It was, however, the division of depression into reactive and endogenous types that long dominated the diagnostic conception of this disorder.

Further subdivisions were proposed, such as that of primary versus secondary depression, and that of unipolar and bipolar disorder (Angst 1966; Perris 1966). Among classifications at the syndrome level, the division into an agitated–anxious syndrome and an inhibited–depressed syndrome was important for pharmacological treatment (Kielholz 1971). The occult vegetative depressive syndrome (*vegetativ larviertes depressives Syndrom*) was held to be distinct from these syndromes. A further distinction at the syndrome level was the division of delusional psychotic depression (Parker et al. 1991; Schatzberg and Rothschild 1992) into forms with mood-congruent and mood-incongruent delusional manifestations (Bellini et al. 1992; Burch et al. 1994; Kendler 1991).

Other classification schemes used the course of the disorder as a classifying feature. One such distinction was that between early-onset and late-onset, or involutional, depression. Further possibilities of classification on the basis of types of disease course were proposed by Merikangas et al. (1994) and Angst (1990). The criterion for classification is the course of disease, more specifically, the extent and duration of depressive manifestations and the frequency of depressive phases. Consideration of duration yields a division into brief depressive episodes, and chronic courses. Consideration of the frequency of episodes yields a division into courses with a single depressive episode and recurrent courses. Consideration of intensity yields a division into minor and major depression. These classifying principles underlie DSM-IV (APA 1994). A further entity is that of "double depression," i.e. the comorbid occurrence of a dysthymic disorder with a "superimposed" major depression.

While ICD-9 (WHO 1978; Degwitz et al. 1980) was still oriented toward the classical concept of endogenous depression, ICD-10 has moved toward the position of DSM-IV with its concept of the depressive episode. Depressive disorder as described in ICD-10 is thus nearly identical to major depression as described in DSM-IV. The definition of the depressive episode is broader in DSM-IV than in ICD-10, however, because the former requires the presence of only one of the core manifestations of affective mood disturbance and loss of interest, while ICD-10 requires the simultaneous presence of both.

In DSM-IV, the individual types of disorder of the depressive episode can be characterized more specifically with so-called specifiers, which take into account aspects of the severity of the disorder, the presence of psychotic manifestations, and the course of illness (on a scale ranging from full remission to chronic persistence). Other factors that are taken into account include the appearance of melancholic, catatonic, and atypical features, the presence of seasonal dependence or of a "rapid cycling" course, and the onset of the depressive episode. In contrast, ICD-10 bases its subdivision of depressive syndromes mainly on their duration and the presence or absence of relapses, as well as on severity and the presence of somatic or psychotic manifestations. In principle, both systems strive to be as descriptive and atheoretical as possible and to avoid the incorporation of etiological, genetic, or therapeutic considerations in the classifying criteria.

According to the DSM-IV diagnostic criteria for major depression, the principal manifestations of this disorder are sad mood disturbance and loss of interest. One of these manifestations must be present. Furthermore, at least five of the following must be present: loss of weight, insomnia, psychomotor agitation, inhibition, fatigue or loss of energy, a feeling of worthlessness, and disturbance of concentration. Recurring thoughts of death count as a criterion only if present over a 2-week period. Exclusion criteria include mixed episodes with manic features, and toxic effects of substance ingestion. The syndrome must also be distinguished from a grief reaction. A positive diagnosis generally requires a clinically significant degree of suffering or an impairment of performance ability.

The principal manifestations of depression according to ICD-10 are mood disturbance and diminution of initiative. Other manifestations of diagnostic impor-

tance include diminished concentration and attention, diminished self-confidence and sense of worth, feelings of guilt and worthlessness, negative and pessimistic perspectives on the future, contemplated or actual self-destructive or suicidal behavior, disturbance of sleep, and diminished appetite (Dilling et al. 1991). The subtype "depressive episode with somatic syndrome" is present when at least four of the following manifestations are present: loss of interest or joylessness, lack of ability to react emotionally, early morning awakening, morning low, psychomotor inhibition or agitation, loss of appetite, weight loss, and loss of libido.

When the current DSM-IV and ICD-10 systems are considered from the perspective of the historical controversies about the classification of depressive disorders, it becomes clear that the distinction between unipolar and bipolar disorders has prevailed and that classification on the subsyndrome level has been taken up again, in somewhat modified form. Subclassification with respect to psychotic features is also reflected in the current classification systems. On the other hand, the distinction between primary and secondary depressive disorders has been abandoned, because it was incompatible with the concept of comorbidity (see Chap. 3, Vol. 1, Part 2). The division of depression into disorders of early and late onset has also been abandoned because it was found to be unreliable: the precise time of onset of the illness turned out to be extremely difficult to define. The main change that occurred in the transition from ICD-9 to ICD-10 was the abandonment of the distinction between endogenous and neurotic forms of depression: this distinction was considered to be too etiologically oriented, and it could not be adequately validated on an empirical basis.

3.2 Empirical Validation of Criteria and Systems

The question arises as to the extent to which these new diagnostic conceptions actually lead to their desired result, i.e. the ability to make more reliable and valid diagnoses of the depressive disorders (Spitzer and Fleiss 1974). The World Health Organization (WHO) sponsored a multicenter, multinational field study for the evaluation of the ICD diagnostic criteria. Sartorius et al. (1993) reported that the WHO field study yielded satisfactory reliability coefficients of 0.66 and 0.69 for the diagnoses of depressive episode and recurrent depressive disorder, respectively. Only a few subcategories seemed to have unacceptable reliability coefficients. In the portion of the WHO study carried out in the German-speaking countries (Freyberger et al. 1990; Dilling et al. 1990), similarly high values were obtained for the overall category of the affective disorders, but there was a low degree of agreement on the definition of a depressive episode, for which a κ-value of 0.19 was obtained.

Substantially better reliability coefficients were reported for the ICD-10 research criteria (Thiel et al. 1996). These studies, too, revealed that ICD-10 has moved very close to DSM-III-R: a κ-value of 0.97 was obtained for agreement between the two systems (Freyberger et al. 1990). Siebel et al. (1997), in their multicenter reliability study for depressive disorders, found a κ-coefficient of only 0.40, which corresponds to a low degree of agreement. This value was lower than the value for all disorders taken together, as well as the value for schizophrenic psychoses.

Hiller et al. (1993) obtained similar findings. The ICD-10 diagnoses for all diagnostic categories taken together were satisfactorily reliable, with a κ-value of 0.80, but the κ-value for depressive disorders was only 0.40. Similarly low values were obtained for subtyping classifications, in particular for the determination of the severity of the depressive disorder. The question remains open, at least as far as the reliability studies in the German-speaking countries are concerned, as to whether the new ICD-10 criteria are any more reliable than the older criteria, which were criticized by Spitzer and Fleiss (1974) because of their low κ-coefficient of 0.24 for endogenous depression. Despite the substantial overall improvement of reliability values, no study has yet yielded entirely satisfactory results.

The evaluation of the criteria for major depression in a DSM-IV field study yielded reliability values between 0.52 and 0.72; only the retest reliability was somewhat worse (0.43) (M.B. Keller et al. 1995). The low stability of the diagnosis of a depressive disorder was pointed out by Clayton et al. (1992). After an interval of 6 years, fewer than half of the patients received the same diagnosis, despite the application of a standardized interview. Even though they are operationally defined, depressive disorders cannot be assessed as reliably as most other psychiatric syndromes. This is particularly true of the assessment of severity.

3.3 Critique of the Operational Diagnosis of Depression and Perspectives on Classification

Although a few of the international field studies did yield high reliability values, several studies revealed a low interrater reliability for the diagnosis of depressive disorder. An adequate interrater reliability for the main categories of the disorder seems to be possible only when standardized interviews are administered by trained examiners (Wittchen and Unland 1991). While this procedure is now accepted as standard in scientific research, it is only rarely applied in the clinical field. In

addition to the problem of reliability, the question of validity arises. There are diverse opinions regarding the usefulness of diagnostic classification schemes for psychotherapeutic treatment. According to Schulte and Wittchen (1988) and D. Schulte (1994), diagnosis according to classification schemes is increasingly important in clinical practice, above all because it enables, and is indeed a prerequisite for, the development of more disorder-specific forms of treatment (Fiedler 1997).

A negative feature of the currently practiced classification of depressive disorders is, according to Saß (1987), the polydiagnostic combination of different classification systems that are only partly compatible with one another and, in addition, reappear continually in modified form at intervals of only a few years. Any advantage to be gained from the use of these systems is offset by the uncertainty over whether scientific findings obtained with the use of an older version of a system remain applicable when a newer version is used. A further point concerns the restriction of the concept of disease features to easily observable phenomena. According to Saß (1987), the development of diagnostic instruments has now reached a boundary beyond which a consideration of the psychopathological basis of these disorders is required.

An entirely different attempt to solve this problem consists of the development of diagnostic criteria on the basis of biological or psychosocial processes. The "learned helplessness" described by Abramson et al. (1978) may be cited here as an example: evidence was found for a pathogenetic relationship of the findings in this disorder (though nonspecific) to hypercortisolism and a disturbance of regulation of the hypothalamic–pituitary–adrenal endocrine axis. Another prototypic classification scheme of this type was proposed by Klerman et al. (1984), who found four types of interpersonal problems in connection with depression and formulated a specific form of treatment for each. The advantage of an etiologically oriented classification lies in the differential implications for treatment. It remains unclear whether descriptively oriented classification will continue to hold its ground, or whether more attention will be paid once again to psychopathological and etiological concepts. The multiaxial extension of the classification systems is an attempt to integrate such perspectives more intensively.

4 Instruments of Assessment

There are now many instruments of assessment that may be used to evaluate the nature and severity of depressive disorders and their manifestations in a standardized, reliable, and valid way. When any such technique is used, it must be borne in mind whether its purpose is an assessment on the level of disease manifestations, the determination of the presence of a particular syndrome, or the assignment of a diagnosis of a particular psychiatric disorder. The two most important techniques of this type are the expert interview and the self-assessment scale. Among the available diagnostic instruments for depression, some are intended to assess the manifestations of depression only, while others aim at a multidimensional psychopathological description of the patient's condition; some are intended to determine specifically whether the overall diagnosis of "depression" is applicable, while others also measure the individual manifestations (see Chap. 6, Vol. 1, Part 2). Several reviews are available that discuss diagnostic instruments for depression and the current state of psychometric assessment (Bech 1992, 1993; Erdberg 1990; Hautzinger 1994; Katz et al. 1995; Röhrle 1988; Sartorius and Ban 1986; Stieglitz 1997; Westhoff 1993; see also Chap. 6, Vol. 1, Part 2).

The instruments of assessment that have been developed in recent years constitute a foundation for empirical statistical research on the depressive disorders. The use of a standardized interview for the assignment of a diagnosis is now a standard procedure in depression research. Thus, to understand and evaluate the findings of research in this field, one must first be acquainted with the instruments used and know the theoretical assumptions that underlie them and the psychometric criteria for their validity.

4.1 Assessment of Depression by Interview

Several structured interviews are currently available for the diagnostic assessment and differential diagnosis of depression (see Chap. 6, Vol. 1, Part 2), including the *Strukturiertes Klinisches Interview für DSM-IV* (SKID-I, in German; Wittchen et al. 1997), the Composite International Diagnostic Interview (CIDI; Wittchen and Semler 1991), and the Schedule for Clinical Assessment in Neuropsychiatry (SCAN; Wing et al. 1990); the last is the most widely used. An appropriately trained interviewer may use these techniques to carry out a highly reliable psychiatric assessment and thereby significantly raise the rate of diagnostic agreement beyond that obtainable with standard clinical diagnosis. Thus Wittchen et al. (1991) report a κ-coefficient of 0.70 for the presence of major depression as assessed by the SKID-I. The problem of

reference remains, i.e. an interviewer can only be reliable in relation to his or her particular reference group (Möller and von Zerssen 1983). In addition, the problem of the assessment of the severity of a depressive disorder is still inadequately addressed by these interviews. The use of a standardized interview for the generation of homogeneous samples is now a standard research technique.

The abandonment of the concept of stratification in these diagnostic systems led to the creation of the now extensively developed field of comorbidity research (Feinstein 1970). It is not yet entirely clear whether comorbidity research actually opens substantially new conceptual perspectives or simply reflects methodological problems inherent in the diagnostic instruments used. High reliability rates have been reported only for the application of a particular diagnostic concept, but not for diagnostic agreement in relation to individual patients. The pathogenetic, etiological, and psychodynamic significance of the determined comorbidities thereby remains unexplained.

The most commonly applied procedure for the assessment of the depressive syndrome with a semi-structured interview is the Hamilton Depression Scale (HAM, or HAM-D in its German version; Hamilton 1960; Baumann 1976; CIPS 1986). The HAM was the point of departure for a number of abbreviated scales and adaptations such as the Bech-Rafaelson Melancholia Scale (BRMS; Bech and Rafaelsen 1986) and the Montgomery-Asberg Depression Scale (MADRS; Montgomery and Asberg 1979). In a study by Maier and Philipp (1993), the HAM-D, BRMS, and MADRS were compared with respect to their psychometric properties. The results suggest that the HAM-D should not be used as the single standard for depression research. The AMDP System (AMDP 1995; Woggon 1986) is also frequently used for evaluation on the level of manifestations and on the syndrome level. A newly developed German-language scale, the *Inventar Depressiver Symptome* (IDS), is oriented toward the concepts of DSM-III-R (Hautzinger and Bailer 1994). The IDS is particularly effective in the measurement of changes (van Gülick-Bailer and Hautzinger 1990).

The severity of depression is gauged by the magnitude of the raw numerical values obtained by these instruments. Such conclusions can be drawn only with certain limitations, because of the unclear factorial structure of these instruments, their differential weighting of individual types of depressive manifestations, and their differential validity for individual subgroups of depression. The severity of illness is more reliably assessed by means of these scales than by evaluation in the sense of ICD-10 (Freyberger and Dilling 1996), but these instruments cannot be used for the assignment of a psychiatric diagnosis.

4.2 Assessment of Depression by Questionnaire

According to Bouman (1993), there are more than 100 different questionnaires for depression, most of them in the English language. Depression scales may be divided into instruments that measure psychopathological manifestations multidimensionally and specific questionnaires for depression (see Chap. 6, Vol. 1, Part 2). A further group consists of questionnaires that attempt only to assess specific aspects of the depressive disorders (Hautzinger 1994; Röhrle 1988).

The best established and most commonly used questionnaires in the assessment of depression on the syndrome level are certainly the Beck Depression Inventory (BDI; Beck et al. 1961; Beck 1978; Hautzinger et al. 1993) and the Self-Rating Depression Scale (SDS; Zung 1965). In the German-speaking countries, the *Paranoid-Depressivitäts-Skala* (PD-S; von Zerssen 1976a) and the *Allgemeine Depressionsskala* (ADS; Hautzinger and Bailer 1993) have been used increasingly alongside these. In addition to questionnaires that attempt to assess depression unidimensionally as a symptom, there are a large number of scales that characterize specific aspects of depression on the levels of mood, behavior, and cognition (Hautzinger 1994; Röhrle 1988).

In summary, questionnaires are an important extension of expert assessment by means of structured and semi-structured interviews. Despite their low diagnostic potential and the difficulty of using them to distinguish depression from related constructs, such as anxiety, they have become indispensable for scientific research. If we consider how often depressive disorders remain unrecognized even today, we realize that these scales offer an economic means of screening. Progress is to be expected primarily in the development of new and more valid depression scales and of scales for specific patient groups, as well as in the assessment of specific individual aspects of depression.

4.3 Comparison of Different Techniques and Instruments

The question arises as to whether different depression questionnaires all measure the same thing, and whether assessment by means of interviews and questionnaires may lead to different conclusions. The results of interviews and questionnaires for the assessment of depression are, in general, moderately highly correlated. In a review of the subject, Bouman and Kok (1987) found an average correlation of 0.53 between the assessment modalities, although approximately 30% of the correlations lay between 0.20 and

0.40. Considerably higher values were reported by Bech (1992), who found an average correlation of 0.73 (0.61–0.86) between the HAM-D and the BDI and similar values for other depression inventories. A more recent meta-analysis by Richter et al. (1998) revealed lower values for this correlation – a mean of 0.56 and a range of 0.34 to 0.86.

The correlation of depression questionnaires with one another was somewhat higher, with a value of 0.69 (Bouman 1993). It was also pointed out, however, that depression questionnaires were just as highly correlated with anxiety questionnaires and neuroticism scales. This implies that depression scales actually measure depression to an average extent of about 30% (Paykel and Norton 1986). While interviews have the highest degree of external validity, questionnaires for the assessment of depression are typified by higher reliability. A degree of external validity and of reliability as high as this is obtainable with interview assessment techniques only after intensive training of the interviewer. Although questionnaire techniques have a high degree of apparent validity, the development of psychometric test procedures has still not yielded any single one that can be recommended unreservedly and has established itself as a standard.

4.4
Discussion of a Standard Test Battery for Depression

More than 100 diagnostic questionnaires for depression are now available (Bouman 1993; Katz et al. 1995; see also Chap. 6, Vol. 1, Part 2). Many diagnostic instruments for depression have not been adequately evaluated. The most commonly used techniques suffer from major defects. In recent years, therefore, there has been a discussion of which test procedures and instruments of assessment should be incorporated into a standard test battery for the diagnostic assessment of depression (Fydrich et al. 1996a; Katz et al. 1995). In view of the current state of research, this question can only be answered provisionally. The most important considerations for the development of a standard test battery are the quality of the instruments as psychometric tools, their clinical usefulness, and the confirmation of their value by experience. The underlying idea of using multiple measurements in depression is to capitalize on the advantages of each technique while lessening the effect of their individual disadvantages.

Katz et al. (1995) recommended a test battery that we, too, recommend here, albeit in somewhat modified form; we further suggest alternative instruments of measurement for each of those used in this battery. Either the SKID-I or the HAM-D can be used as an expert interview in the initial diagnostic phase. If the diagnosis is to be made according to ICD-10, the SCAN interview should be used. The IDS may be used as an alternative to the HAM-D. Among self-assessment instruments, the Symptom Checklist (SCL-90-R) and the BDI can be used at the beginning of treatment. For the measurement of changes, the BDI or, alternatively, the ADS, can be used at intervals of 2–4 weeks. The German-language *Depressivitäts-Skala* ("Depressivity Scale," D-S; von Zerssen 1976a) may also be used for this purpose; norm values of the D-S are available for a representative sample group. Assessment at the end of a period of treatment can be performed by renewed application of the SKID-I or the SCL-90-R. Furthermore, the visual analogue scale of Aitken (1969) can be used for daily assessment of depressivity, but, if this is done, circadian fluctuations must be taken into account by testing both in the morning and in the evening.

These recommendations seem not to be reflected in the customary manner of use of instruments for the assessment of depression, at least as it was until the mid-1980s. Thus Fähndrich et al. (1986) showed that the three instruments for depression most commonly used in German-language studies were the *Befindlichkeits-Skala* ("Well-Being Scale," Bf-S) of von Zerssen (1976b), the HAM-D, and the AMDP-System (AMDP 1995). The BDI and the visual analogue scales were applied in fewer than 5% of the studies. Diagnostic habits may well have changed in the last few years, although there are no more recent empirical studies to document this possibility. In larger research studies, the problem of the absence of a diagnostic standard is usually solved by the parallel application of several instruments on a single level of assessment, which, in turn, leads to new methodological problems.

5
Diagnostic Boundaries, Differential Diagnosis, and Comorbidity

It has been found that a high percentage of patients in whom a depressive disorder has been diagnosed also have other mental and physical illnesses. The percentage of depressed patients with comorbid illnesses is higher than that among patients with psychiatric disorders of other types (Moldin et al. 1993; Mezzich et al. 1990). Rohde et al. (1991) showed that 42% of unipolar depressives fulfill the criteria for one or more further DSM-III diagnoses. The most common of these were anxiety disorders (18%–20%) and substance dependencies (14%–20%). The frequency of physical illness in depressed patients is comparably high, though also relatively strongly influenced by the age of the group under study.

These comorbid disturbances are of great importance not only for diagnostic assessment, differential

diagnosis, and treatment planning, but also for the scientific understanding of mental and physical illnesses. Beyond classical psychopathological diagnosis, comorbidity research, which has by now established itself as a field of study in its own right (Feinstein 1970), has yielded findings of importance to differential diagnosis. These findings have been made in both epidemiological and clinical studies (Maser and Cloninger 1990; Robertson and Katona 1997).

Under the influence of descriptively conceived diagnostic systems such as ICD-10 and DSM-IV, the prevailing concept of diagnosis has evolved from a stratified concept to a concept of comorbidity. Several flowcharts and decision schemes for syndromal differential diagnosis have been developed (Saß et al. 1996; Linden 1979). Nonetheless, a complete clinical diagnostic assessment requires not only the identification of a syndrome, but also the development of an etiopathogenetic hypothesis (Seidenstücker and Baumann 1978) – and this is all the more so when several disorders are present simultaneously.

One of the most important problems regarding the drawing of diagnostic boundaries is the differentiation of affective disorders from each other, but their differentiation from other axis I disorders, such as schizoaffective and dementing disorders, may also be difficult. An increasing amount of attention is being devoted to the differential diagnosis of depressive disorders from personality disorders and organic illnesses. Alongside descriptive comorbidity research, more complex etiology-theoretic models have been developed in recent years to explain the coexistence of these disorders and their overlapping features (Klein et al. 1993; Mayou 1997). As yet, however, there is little empirical confirmation of these models, and little evidence to show in what sphere each of them is valid.

5.1 Differentiation Within the Affective Disorders

Reviews of this subject have been written by Maser et al. (1995), Maser and Cloninger (1990), and Clayton (Chap. 15, Vol. 3, Part 1). The dichotomy between unipolar and bipolar disorders, which was established in the wake of the studies by Leonhard et al. (1962), Winokur et al. (1969), Perris (1966), and Angst (1966), is now coming under criticism again, for several reasons. One problem is that it is not known which patients will later develop a bipolar disorder. The literature documents rates between 4% and 33% (Clayton 1981) and between 0% and 41% (Coryell and Winokur 1992), depending on the duration of observation. The risk of developing a bipolar disorder does not decrease even after numerous exclusively depressive phases have occurred (Angst 1978).

In contrast, Marneros et al. (1991) found that a syndrome change rarely occurred more than 5 years into the course of the illness. A further difficulty is that of retrospectively diagnosing manic and, especially, hypomanic phases, not least because these often do not cause the patients any suffering and may not be perceived at all as being pathological. The same may be said for hypomanic fluctuations after the termination of a depressive phase. Yet another differential diagnostic problem is posed by the existence of mixed states and mixed affects lying in between the depressive and manic syndromes (Kuhs and Tölle 1987; Cassidy et al. 1997).

One of the most difficult differential diagnoses is that of a depressive disorder from a dysthymic disorder. The question of whether we are dealing here with a bimodal distribution of two separable syndromes or, alternatively, with a continuum (Kendell and Gourlay 1970; Judd 1997) has not yet been definitively answered. While several studies on the subtyping of depression have revealed that a melancholic symptom pattern is characteristic of endogenous depression, syndromes of peripheral types are considerably less stable and less typical. Differential diagnosis is further complicated by the concept of depressive personality disorder (Hirschfeld 1994).

In addition to the psychopathological findings, the course of the disorder is a further, crucial differential diagnostic criterion. Here, too, consideration must be given to aspects of comorbidity. The concept of "double depression" refers to a course in which one or more episodes of major depression are superimposed on a dysthymic disorder (Keller and Shapiro 1982). In the latter study, dysthymia was diagnosed in 26% of patients who underwent outpatient and then inpatient treatment for major depression. In the Epidemiological Catchment Area Study, patients with dysthymia were found to have a comorbidity with major depression in almost 40% of cases. Longitudinal studies have revealed that 90% of individuals with a dysthymic disorder eventually develop major depression (Akiskal et al. 1981; Lewinsohn et al. 1991). The opposite sequence, i.e. the development of dysthymia after major depression, is rare.

A further topic of increasing importance is the differentiation of a depressive episode from subsyndromal or subdiagnostic depressive disorders, primarily as they are conceived of in DSM-IV (APA 1994; see Chap. 16, Vol. 3, Part 1). According to Angst and Merikangas (1997), these include recurrent brief depression, minor depression, and "subsyndromal depressive symptomatology." It was shown that subsyndromal forms of depression are common and cause significant impairment. They are a risk factor for the development of major depression as well as a phase in the disease course of patients with diagnosed major depression (Helmchen et al. 1996; Judd 1997). There is

now debate over whether these findings should be interpreted as reflecting the existence of a spectrum of depressive disorders (Winokur 1979). There is also a question as to whether depression with psychotic features must be granted a special position within this spectrum (Coryell 1997).

Symptoms of anxiety are very common in depressive disorders and are not uncommonly severe enough to merit designation as a disorder in their own right (Stavrakaki and Vargo 1986). The odds ratio is a measure of the degree to which anxiety disorder and depressive disorder occur together more often than would be predicted by chance. The Zurich study yielded the following odds ratios for disorders in association with major depression: for agoraphobia, 2.8; for panic disorders, 1.9; and for generalized anxiety disorder, 4.2 (Angst 1993). The values reported by Boyd et al. (1984) were considerably higher. Rohde et al. (1991) found an anxiety disorder in 20% of the depressed patients they studied, while Sanderson et al. (1990) arrived at the figure of 42%. From the point of view of temporal sequence, it seems to be more common for a patient with an anxiety disorder to develop superimposed major depression than for the reverse to occur (Angst et al. 1990). Nonetheless, the latter occurrence was reported by Sanderson et al. (1990). It should be recalled that instruments of assessment never measure depressivity or manifestations of anxiety alone; rather, such instruments are confounded with one another and have a high degree of statistical correlation.

Kuhs (1990) found that melancholic and neurotic-depressive patients experience anxiety differently. Qualitative analysis revealed that patients with endogenous depression suffered more from everyday anxieties, while situational anxieties and anxieties related to interpersonal problems predominated in neurotic depressives, with a broader spectrum of content.

A mixed disorder composed of both anxiety of depression was introduced as a new diagnostic category in ICD-10. This disorder may be diagnosed in the presence of manifestations belonging to both groups, even when the diagnostic criteria for a depressive episode or for an anxiety disorder are not fulfilled (Sartorius and Üstün 1995). The practical value of this additional diagnostic category, alongside other types of subsyndromal disorder, is a topic of current research (Wittchen and Essau 1993; Boulenger and Lavallée 1993; Liebowitz 1993).

5.2 Differentiation from Other Psychiatric Disorders

The differentiation of depressive disorder from the schizoaffective psychoses and schizophrenias is very important. The differential diagnosis of depressive disorder from the dementing diseases is also of great practical relevance. In addition, depressive disorders often appear together with alcohol and drug dependency and with Parkinson's disease (Maser et al. 1995).

Many patients (15%–57%) with schizophrenic disorders also suffer from depressive disorders (Maser et al. 1995). The risk of developing major depression is elevated by a factor of almost 30 in schizophrenic patients. Not uncommonly, post-remission depression develops after the end of the schizophrenic phase. According to the findings of Marneros et al. (1991), 6% of patients who initially present with major depression of melancholic type go on to develop a schizoaffective psychosis.

As for the differential diagnosis of depression and the dementing diseases (see also Chap. 8, Vol. 2, Part 1), epidemiological studies have shown that depression is the second most common mental illness in old age, after dementia. A problem of differential diagnosis regularly arises in clinical practice because of the similarity of the manifestations of depression and dementia. A major difference between them is that the cognitive impairment associated with depression, which has been given the designation "pseudodementia," is not associated with morphological changes and regresses after the depressive manifestations have subsided. Pseudodementia remains a controversial concept (Zimmer and Lauter 1984) despite a number of suggestions for the definition of diagnostic criteria (Rabins et al. 1984). Clinical diagnosis must take multiple criteria of prior history, psychiatric exploration, neuropsychological findings, and the results of functional imaging studies into account (Kurz 1997).

It has been found repeatedly that incipient dementia imparts an elevated risk for depression. There are conflicting findings on the question of whether, conversely, a preexisting depression predisposes to the development of a dementing illness. Devanand et al. (1996) were able to show such an effect, but Ernst and Angst (1995) found no elevation of the risk of dementia in elderly depressed patients. Further longitudinal studies taking account of specific etiological and risk factors are needed to determine whether depression and dementia are comorbid disorders in the strict sense, or whether a more complex relationship exists between them.

5.3 Depression and Personality Disorders

The differential diagnosis of the depressive disorders must also take personality disorders into account (Millon and Kotik-Harper 1995). Comorbidity studies have revealed the presence of a personality disorder in

30%–70% of patients with major depression (Farmer and Nelson-Gray 1990). Fydrich et al. (1996b) cited studies with comorbidity rates of 35%–88%. Figures obtained by the use of questionnaires were somewhat higher than those obtained by diagnostic interviews. At least half of all patients with major depression have an accompanying personality disorder. The most commonly diagnosed personality disorders are the insecure (anxious-avoidant), dependent, and compulsive personality disorders. A more detailed discussion of the relationship of depression to individual personality disorders is found in Millon and Kotik-Harper (1995).

The reintroduction of the concept of depressive personality disorder in the research supplement to DSM-IV led to new difficulties for differential diagnosis (Philipps et al. 1990; Sherman 1995). Although theoretical explanations have been given, the relationships between depressive disorder, dysthymia, and depressive personality disorder have not yet been adequately clarified (Hirschfeld 1994). Differential diagnosis is particularly difficult in chronic cases (Shea and Hirschfeld 1996). There are extremely divergent explanatory models for the relationship between personality and depression. The most important models, including the predisposition, complication, coeffect, spectrum, and overlapping-symptom models, and the empirical support for each, are discussed by Klein et al. (1993).

The comorbidity of subaffective personality disorders with depressive manifestations and disorders has been intensively studied in recent years (Herpertz et al. 1996). This spectrum also includes studies of "typus melancholicus," a personality configuration often found in patients with depressive disorder (Mundt et al. 1997). Pfohl et al. (1991) found that depressive patients with personality disorders differed from those without by their earlier age of disease onset, lesser degree of social support, greater psychosocial stress, and worse response to treatment. These findings led to the development of initial attempts at differential psychotherapeutic treatment strategies (Mundt 1996).

5.4 Differentiation from Organic Disorders

Physically ill individuals suffer from depressive disorders at an elevated frequency (Robertson and Katona 1997; Lang 1991). Conversely, depressive patients are more likely to have physical illnesses (Stevens et al. 1995). Lobo and Campos (1997), in their study of patients in the primary medical care system, reported on rates of major depression ranging from 4.8% to 13.5%. Rodin et al. (1991) found an overall rate of 22% and a still higher rate for patients with neurological disorders. When the frequency of depressive manifestations, rather than of depression, is measured, it is found to be even higher, with reported rates between 20% and 83% (Stevens et al. 1995). Similar results on the frequency of depressive disorders in hospitalized patients were reported by Arolt (1997). Arolt et al. (1995) found that 4.1% of surgical and internal medical patients had at least one depressive episode.

Despite their common occurrence, depressive disorders in the physically ill are often overlooked (Lobo and Campos 1997). Thus Ormel et al. (1991) found that general practitioners recognized the presence of depressive disorders and anxiety disorders in their physically ill patients in only 47% of cases. Coyne et al. (1995) reported similar findings.

Comorbidity research regarding depression and physical illness must be seen in the context of the traditional conceptions of primary and secondary depression (Feighner et al. 1972). In recent years, this conceptual area has been extended by the proposal of increasingly complex models of the relationship between these two types of disorders (Moffic and Paykel 1975; Lang 1991). The presence of a physical illness is held to be an important predictor of an unfavorable course of the depressive disorder and of its chronification (Angst 1988b; Zimmer 1991). In this context, F. Post (1962) formulated the hypothesis of age-dependent vulnerability, according to which younger individuals develop depression because of their genetic predisposition and personality factors, while elderly persons require an additional physical illness, or a disorder of the brain, before a depressive phase can be triggered, often on a subsyndromal level.

It can also be difficult to differentiate an organically caused depressive disorder from another form of comorbidity. Such relationships may be quite complex in individual cases. Several studies have demonstrated that depressive patients with and without accompanying physical illnesses have different patterns of disease manifestations and clinical features (Berrios and Samuel 1987; Clarke et al. 1983).

Somatized or occult depression poses a special diagnostic problem. Bridges and Goldberg (1985) offered a definition of somatized depression. Lobo et al. (1996) found this type of depressive disorder in 15% of patients with a depressive episode. Nonetheless, the conceptual problems of differential diagnosis in this area have not yet been satisfactorily solved.

Somatic illnesses are often associated with a depressive disorder, but subsyndromal depression occurs even more frequently, mostly in the elderly. Several studies have revealed that depression is correlated not with the presence of illness per se, but with the resulting limitation of everyday activities. Likewise, it is not old age itself that is a risk factor for the development of depression, but rather the age-associated

risk of having one's everyday activities limited by illness (Helmchen et al. 1996; Zeiss et al. 1996).

Such problems regarding the drawing of diagnostic boundaries are highly relevant not only to differential diagnosis in clinical practice, but also to the further refinement of scientific concepts of classification and etiology. Research findings concerning the frequency of disorders occurring together are an important contribution in this direction and also serve as a point of departure for the further development of models concerning the relationships among these disorders and for the empirical study of these models' applicability. An appreciation of the significance of comorbid disorders in depression is very important for the development of new therapeutic strategies.

6 Disease Course and Prognostic Factors

6.1 Historic and Terminological Aspects of Research on Disease Course

Kraepelin (1913) developed the model of manic-depressive illness as a disorder with a phasic and remitting course. Bleuler (1916) and Schneider (1932) adopted Kraepelin's conception that depressive disorder had a relatively favorable prognosis. According to Paykel (1994), research on the course of depression was dominated by this idea and, at first, by therapeutic optimism, from the 1950s until the 1970s. Non-remitting depressive manifestations were viewed during this period as characterologic deviations of personality. It was only in the late 1970s that the findings of longitudinal studies of disease course led to the realization that depressive disorders had a less favorable course than previously assumed (Angst 1987). The prospective clinical and epidemiologic studies of disease course carried out in recent years, some of them on a large scale, confirmed and refined this new appraisal. Research on disease course has concentrated in recent years on the performance of prospective studies, the definition of uniform standards for use in research, and the study of the problem of unfavorable courses, chronic depressive states, prodromal manifestations, and milder forms of depression.

The first large-scale studies of disease course were carried out in Scandinavia in the 1960s (Astrup et al. 1959; Lundquist 1945; Perris 1966; Stenstedt 1952), in Switzerland by Kinkelin (1954) and Angst (1966), and in the United States by Winokur et al. (1969). The main finding of these studies was that unipolar and bipolar affective illnesses have different course dynamics. This finding, along with the results of genetic studies of the affective disorders (Angst 1966; Perris 1966; Zerbin-Rüdin 1969), led to the abandonment of Kraepelin's unitary conception of manic-depressive illness. This found its expression in the St. Louis Criteria (Feighner et al. 1972) and the Research Diagnostic Criteria (RDC; Spitzer et al. 1978). The second important consequence of these studies was a revision of the previously optimistic view of the course of the affective disorders, which turned out to be marked by frequent relapses, chronification, and suicidal outcomes much more commonly than had been assumed.

Although a large number of studies have been performed on the course of the depressive disorders, it is difficult to judge their results collectively, because of the lack of uniformity in the definition of terms used to characterize disease course (Angst 1987). For this reason, efforts have been made in recent years to establish uniform, operationally defined terms for use in research on the course of depressive disorders. Prien et al. (1991) discuss several different operational criteria for the definition of terms relating to disease course and treatment outcome in major depression.

Frank et al. (1991) recommended specific operationalized definitions of the most important terms used to describe disease course. They drew distinctions between remission, recovery, relapse, and recurrence of illness. These terms correspond to the division of forms of treatment into acute, maintenance, and chronic treatment (Thase 1992). Remission was defined as a state of partial or complete improvement of the depressive manifestations. Recovery was taken to imply a complete remission over a longer period of time (2–6 months, depending on the particular criterion used). Frank et al. (1991) criticized the absence of a distinction between remission and recovery in most studies. Relapse was defined as the appearance of depressive manifestations during a period of remission. Recurrence of illness, finally, can occur only after a patient has recovered. Chronic depression is a depressive episode that persists over the entire relevant period of observation (generally at least 2 years).

It is impossible to draw such distinctions from the reported results of most of the earlier studies of disease course. Von Riso et al. (1997) proposed operational definitions of these terms used to describe disease course based on the use of clinical rating scales and well-defined temporal intervals. This study yielded evidence for the predictive validity of such criteria. The definitions presented by Frank et al. (1991) were an important contribution to the unification of terminology in research on the course of depressive disorders and made it possible to compare results obtained across studies.

6.2 Long-Term Outcome of Depression

Piccinelli and Wilkinson (1994) performed a meta-analysis of the then available studies on the course of depression on the basis of the definitions provided by Frank et al. (1991). They analyzed 51 studies published from 1979 to 1992 and classified them by the length of clinical observation in each: 6 months, 1 year, 2–5 years, and more than 10 years. In data obtained as a weighted average from the studies involving more than 10 years of observation, recovery was found to have occurred during the period of observation in 24% of patients, while a course with recurrences occurred in 76%. Chronic depression was found in an average of 12% of patients studied (*note:* the use of weighted averages may result in percentages that sum to more than 100%). There was a high correlation between the length of the period of observation and the number of disease recurrences. On the other hand, there was no statistically significant relationship between the length of the period of observation and the number of patients developing chronic depression. These calculated frequencies of the individual types of disease course in depressive disorder are largely in agreement with the results reported by Angst (1987), M.B. Keller (1994), and Judd (1997).

Data concerning the frequency of unipolar depressive disorders consisting of a single episode are inconsistent and heavily dependent on the sample populations and periods of observation used. Angst (1990) concluded from his review of the literature on epidemiologic studies that an average of 50% of all individuals with depression have only a single depressive episode. Coryell and Winokur (1992) found a monophasic course in 5%–60% of patients in the studies of long-term course reviewed by them; however, in an epidemiologic sample consisting of very young individuals (age 20), Angst (1990) found a rate of only 22% of monophasic courses over a period of 10 years. Wittchen and von Zerssen (1987) reported a rate of 28% of monophasic courses in depressive patients who had undergone inpatient treatment.

A further question concerns the frequency of courses with relapse, as well as the chronification of depressive disorders. According to Angst (1990), 20% of cases of depression in the general population have a relapsing course. This is in marked distinction to the 80% frequency of relapsing courses reported in patients who had undergone inpatient treatment for endogenous depression, after periods of observation of 13–17 years. A total of 20% of patients had two phases, and 60% had three or more (Angst and Frey 1977). The average number of episodes that a patient with depression suffers during his or her lifetime is four (Judd 1997). With regard to relapse dynamics, it was found that the period after discharge from inpatient psychiatric treatment is particularly fraught with risk. A total of 25% of these patients have a relapse in the first 3 months after discharge, 33% within 1 year, and 73% over the next 8 years (M.B. Keller et al. 1983). The National Institute of Mental Health (NIMH-NIH 1985) estimates, on the basis of the available studies, that 50% of patients suffer a relapse within 2 years. Similar results on the prediction of relapses in depression were obtained in the Heidelberg Study (Mundt et al. 1998a).

The descriptive parameters used in longitudinal studies of the depressive disorders to characterize the course of the disease include the age of onset, the number and duration of phases, the duration of the asymptomatic interval, and the duration of the cycle (i.e. the interval between the onset of two consecutive phases).

A depressive disorder may appear at any age, even in childhood (Speier et al. 1995). Although a median age of onset of 30–40 years was found in earlier studies (Lewinsohn et al. 1986; Angst 1987), more recent studies have moved the apparent peak of the age distribution of disease onset to between 20 and 30 years (Kessler et al. 1994). The distribution of ages of onset is skewed to the left, i.e. the risk of developing a depressive disorder declines steadily with age. Knäuper and Wittchen (1995) also found that younger age cohorts have a significantly elevated risk of developing depressive disorders, and a less pronounced difference of incidence between the sexes, than older age cohorts. The age of earliest manifestations of the disease does not differ between the sexes (Lewinsohn et al. 1986). The bimodal distribution of the age of onset of disease postulated by Angst (1987) was confirmed by the Baltimore Epidemiologic Catchment Area Follow-up Study (Eaton et al. 1997), which revealed a smaller peak between 50 and 60 years of age. The hypothesis that depressive disorders of later onset have a worse prognosis (Angst 1987) was not confirmed in more recent studies (Hinrichsen 1992). Nonetheless, the interpretation of findings is complicated here by the problem of incomparability of sample groups (Maj 1994).

Earlier authors estimated the average duration of depressive episodes as between 3 and 6 months (Pilcz 1901; Ziehen 1896); Kraepelin's figure (1913) was between 6 and 8 months. Recent studies, practically all of which reveal a median episode duration of approximately 5 months (Angst 1987; Solomon et al. 1997), are in accordance with these estimates. This means that modern pharmacotherapy has not led to a shorter duration of episodes, even though it has led to marked alleviation and suppression of depressive manifestations. Graphs of the number of patients in remission over time are typically logarithmic in shape,

i.e. remission becomes increasingly unlikely the longer depressive manifestations have been present (Coryell and Winokur 1992).

M.B. Keller et al. (1992) found that 54% of the patients they studied recovered within 6 months, 70% within 1 year, and 88% within 5 years. The available data are as yet insufficient to determine whether the duration of phases is intraindividually stable across episodes in patients suffering from multiple phases without chronification. While Angst (1987), Hautzinger (1997), and Coryell et al. (1994), among other authors, postulated such stability, the studies conducted by Maj et al. (1992) and Solomon et al. (1997) revealed no consistent relationship between the duration of phases of individual patients. An epidemiological study revealed a median duration of 3 months for the first episode and of only 2 months for subsequent episodes (Eaton et al. 1997).

Cycles are 4.5–5 years long on average (Angst 1986). It cannot be definitively stated whether cycles become shorter after several phases have occurred. It has been postulated that a logarithmic relationship obtains, whereby the second interval is much shorter than the first, but the shortening of subsequent intervals is much slower. Several studies, however, have failed to show any regular pattern of this kind (Cutler and Post 1982; Solomon et al. 1997).

It was primarily the findings of recent, large-scale epidemiologic studies that led to the revision of previous ideas about several descriptive features of the course of depression. These studies dealt with representative population samples and not, as in previous studies, with samples of patients who had presented for psychiatric treatment. The inclusion of milder and hitherto untreated cases of depression led to the correction and refinement of several previous conclusions. The mean duration of episodes is thus shorter, and monophasic courses are more common, than previously assumed. On the other hand, chronic depression is also more common than once thought.

A question that has received much attention recently is that of how frequently a chronic course of depression may be expected. Chronic depression appears in DSM-IV(APA 1994) under the category of "chronic major depression." A chronic course is said to obtain when the criteria for major depression are fulfilled over a period of 2 years. Older reviews stated that the frequency of chronic courses lies between 12% and 15% (E. Robins and Guze 1972; Weissman and Klerman 1977). Some of the more recent studies, however, have revealed higher rates, on the order of 25% (Angst 1990; M.B. Keller et al. 1986). The frequency of chronic courses in hospital-treated depression is 21% for patients with endogenous depression, according to the studies conducted by Wittchen and von Zerssen (1987), and 30% for patients with neurotic depression (Bronisch et al. 1985). Marneros et al. (1991) found that one third of the patients with affective psychoses in the Cologne Study underwent a lasting change to either a chronified subdepressive syndrome or a mild asthenic insufficiency syndrome. Although the treatment of depressive disorders has significantly improved, the percentage of chronic courses appears to have remained relatively stable over several decades (Paykel 1994).

The long-held assumption that chronic depression largely remains on the intermediately severe symptomatic level of a dysthymia must now be regarded as disproved. A total of 22% of patients with an episode of major depression without preceding dysthymia were found to have persistent manifestations at the level of severity of major depression over an interval of 2 years (M.B. Keller et al. 1983). The chronification of depressive manifestations does not represent a residual state in which no further improvement can be expected; on the contrary, several recent studies have shown that recovery is particularly likely to occur after a long symptomatic period (Keller and Shapiro 1982). The reported figures for complete remission in severe chronic cases vary between 27% and 43% (Gonzales et al. 1985; Keller and Lavori 1984). The cause of chronification seems to be multifactorial; empirical studies on this question are lacking. There are indications, however, that a comorbid mental disorder is associated with a higher risk of chronification (Zimmer 1991).

Alongside chronification, suicide is a common and feared outcome of depressive disorders. Individuals with depression are considered to be the leading risk group for suicidal behavior. It is estimated that 56% of depressed patients attempt suicide at least once (Goodwin and Jamison 1990). The Cologne Study found suicidal manifestations in 30% of the episodes of patients with melancholic manifestations; suicidal manifestations were less common in married and employed patients (Marneros et al. 1991). Over the entire course of illness, almost 60% of the patients had suicidal manifestations, 35% had suicidal thoughts and intentions, and 20% attempted suicide at least once. The probability of occurrence of suicidal manifestations was higher in women and in patients with a higher number and frequency of episodes (Marneros et al. 1991).

Even today, 10%–15% of severely depressed patients die by suicide (Angst 1980; Miles 1977; Winokur and Tsuang 1975). Guze and Robins (1970), in their review of the literature, found that 12%–19% of deceased patients who had suffered from an affective disorder had committed suicide. Patients with affective disorders thus have a risk of suicide 30 times higher than that of the general population (Guze and Robins 1970). The period of greatest risk for suicide is the first year after discharge from inpatient psychiatric treatment

(Fawcett et al. 1987, 1990). Avery and Winokur (1978) found that two thirds of suicides in previously hospitalized depressive patients occurred within 8 months of discharge.

Risk factors for suicide have been identified in several studies, including hopelessness (Beck 1986), feelings of worthlessness and guilt (Barraclough and Pallis 1975; Hole 1973), delusional manifestations, and sleep disturbances (Barraclough and Pallis 1975). A particularly high risk of suicide was found in depressed patients who simultaneously suffer from anxiety disorders (Rudd et al. 1993). A small number of studies are available regarding possible differences between depressive patients who do and do not attempt suicide (Bronisch and Hecht 1987; Sonneck et al. 1976; Wolfersdorf et al. 1996). These studies show differences in the severity of depressive manifestations and in the presence of relationship problems, threats of separation, and subjectively experienced stress in the workplace (Wolfersdorf et al. 1996). Angst (1987) still held that long-term treatment had no effect on the suicide rate of depressed patients, but more recent longitudinal studies have shown that long-term treatment with lithium can significantly reduce suicide-related mortality in patients with affective disorders (Coppen et al. 1991; Ahrens et al. 1995).

6.3 Incipient Depression

While research on the long-term course and ultimate outcome of depression has a long history, the study of incipient depression and prodromal states is still only in its initial stages. This area of research will become more important, because the early recognition of depressive disorders is a precondition for preventive interventions. The distinction between prodromal manifestations and indicators of vulnerability must be borne in mind (Mundt 1998). Prodromal manifestations occur immediately before the beginning of the depressive episode, while indicators of vulnerability, according to prevailing models (Nuechterlein 1987; Zubin and Steinhauer 1981), are present long before and persist in the interval between depressive episodes. It remains unclear at present whether prodromal manifestations are an intensification of these vulnerability traits or are related to them in some other way. Earlier authors worked out a characteristic sequence of precursor manifestations (Mayer-Gross et al. 1969); more recent empirical studies have added further to our understanding. In what follows, we will discuss findings concerning the prodromal manifestations of depressive disorders. More information on indicators of vulnerability can be found in Chaps. 18–21 (Vol. 3, Part 1).

The prodromal syndrome is understood to be that state in which individual manifestations of a mental illness are present, but the criteria for the presence of a depressive disorder are not fulfilled (Frank et al. 1991). In their review of the literature, Klosterkötter and Steinmeyer (1996) estimated that one third of individuals with depression have prodromal manifestations. The most common of these are pain syndromes, adynamia, central-vegetative disorders, generalized anxieties or phobias, body dysesthesias, easy fatigability, and disorders of sleep and appetite. The wide divergences among studies in the estimated frequency of prodromal manifestations among depressed patients are the result of the absence of a uniform, operationalized definition.

Divergent results have also been obtained for the duration of prodromal manifestations, probably also because of the use of variable criteria in determining their presence. The study by Carlson and Goodwin (1973) found that they lasted from less than 2 to more than 4 weeks; another study gave the mean duration of prodromal manifestations of depression as just over 2 years (Hopkins 1965; Young and Grabler 1985). The mean duration of prodromal manifestations appears to be significantly higher in mania and in schizophrenia than in depression (Carlson and Goodwin 1973; Molnar et al. 1988). The prodromal manifestations of schizophrenic, schizoaffective, and affective psychoses have many common features. These findings have been cited in support of the continuity hypothesis of psychotic illnesses (Angst and Scharfetter 1990; Crow 1986).

A number of studies indicate that prodromal manifestations have both a high interindividual variability and a high intraindividual consistency (Fava et al. 1985; Fava and Kellner 1991; Molnar et al. 1988; Paykel et al. 1976). It remains unclear whether such a relationship also obtains between prodromal and residual manifestations (Fava and Kellner 1991).

Almost 2000 subjects were studied in the long term in the Baltimore Epidemiologic Catchment Area Program Study (Eaton et al. 1997). Almost all of the 71 patients with newly incident major depression had had prodromal manifestations. The individual manifestations had been present for variable periods before the onset of illness; overall, they had been present for a median time of approximately 1 year, but the median period was significantly higher – up to 5 years – for loss of appetite, suicidal thoughts, and dysphoria. The longer duration in comparison to earlier studies is presumably the result of the operational definition of the onset of the prodromal syndrome as the time of occurrence of the first psychiatric manifestation, as assessed by the Diagnostic Interview Schedule (DIS; L.N. Robins et al. 1981). A close relationship between precursor manifestations and the onset of a depressive

episode was found for diminished sexual interest, feelings of worthlessness and guilt, a brooding tendency, and sleep disturbances (Dryman and Eaton 1991; Eaton et al. 1995).

Horwarth et al. (1992) showed, in an epidemiological study, that 50% of newly incident cases of major depression had had early depressive manifestations. Depressive manifestations were found in one quarter of all participants in this study over the course of their lives; most of them had had these precursor manifestations without ever developing major depression. The responsible factors have not yet been identified. The hypothesis has been advanced that the experience of such mild depressive phases is a kind of training in self-directed coping and may thus have a preventive effect (Hautzinger 1997).

The conflicting nature of these findings makes it clear that the concept of a prodromal syndrome is in need of clarification and uniform definition. In the affective disorders, in particular, a problem arises in the differentiation of prodromes from the disorder itself. Future prospective studies must clarify whether this categorization is truly helpful, or whether it would be more reasonable to assume a continuity model for the depressive disorders. A particular methodological problem arises from the fact that retrospective studies of patients or prospective studies of populations at risk may yield valid high-risk profiles for depression but, because of their lack of specificity, remain of little use for primary prevention in practice. The grouping of coarser bundles of factors as risk constellations thus seems to be more promising than the use of prodromes alone (see Mundt 1998). Research in this area proceeds in the hope that the early recognition of depression will lead to the more effective application of primary and secondary preventive treatment strategies.

6.4 Predictors of Disease Course

The large degree of heterogeneity in the course of the depressive disorders has led to an intensive search for favorable and unfavorable predictive factors – unfavorable meaning those that predispose to high rates of relapse, chronification of the disorder, and suicide. Many studies have dealt primarily with clinical and sociodemographic variables and, later, with psychosocial variables in relation to the onset and course of depressive disorders. The preceding clinical course has been considered to be the most important factor for the prognostication of the further course of the disorder (Angst 1988a).

Many studies have been performed with the purpose of making the course of illness easier to predict on the basis of multiple factors. Angst (1987) concluded his discussion of predictors of the course of depression by stating that the identification of risk factors, and the improvement of prognostication this would bring about, were still unfulfilled goals at that time, and that the question had not yet been adequately studied. In recent years, however, many relevant studies have been performed. Review articles on predictors of the course of depression not uncommonly contain tables listing more than 100 such factors, each one characterized by the strength of the supporting empirical evidence (Kaelber et al. 1995).

The numerous review articles on prognostic factors in depression are also confusing, because the analysis has not been carried out in a systematic fashion. There must be systematization, at the very least, with respect to the course of illness. A distinction must be made, for instance, between predictors, in the sense of risk factors, for the first appearance of depression and predictors of the time at which recovery or relapse will occur; the latter must, in turn, be distinguished from predictors of chronification. A second aspect of systematic analysis is the classification of prognostic factors according to the period of time over which they are operative. A distinction is usually drawn between short-term predictors, which apply over a time interval of 1–2 years, and long-term predictors. The type of treatment must also be taken into account: it must be made clear whether predictors are discussed in connection with the natural course of illness or with a particular type of treatment.

A further reason for caution in the interpretation of the findings of these studies is that we know from several other studies that depressive disorders often go unrecognized and, when they are recognized, are often not treated. Most of the research findings in this area were made with respect to samples of hospitalized depressed patients. Yet another difficulty arises from the fact that it is not made clear in most publications whether data are provided for all of the variables that were studied as potential predictors or only for those where positive findings were made. As the published findings are almost exclusively positive, publication bias must be suspected. Furthermore, explicit hypotheses for investigation are not stated in any of these studies. The purely exploratory research strategy that this suggests is not compatible with present standards for psychiatric research.

A number of sociodemographic, clinical, personality-psychological, and social factors have been studied as possible risk factors for the unfavorable course of depressive disorders. The importance of illness-specific factors is universally accepted. Angst (1988a,b) concluded that the previous course of depression is the best predictor of its further course. In addition to these illness-specific factors, the importance of psychosocial

factors such as life events, personality factors, and relationship variables was also demonstrated; but the results obtained by Paykel et al. (1996) provoked a new debate about the significance of psychosocial predictor variables for the course of depressive disorders. In their large-scale study, Paykel et al. (1996) showed that illness-specific prognostic factors have significant predictive value, but they did not find any further improvement of predictive value from the consideration of psychosocial traits. Similar findings had already been made long before by Angst and Weis (1967) by means of multivariate analysis of individual predictors. All of these findings are compatible with either of two hypotheses with regard to predictors of depressive disorder. It remains to be clarified, first, whether different predictive factors are operative for milder and more severe cases of depression; and, second, whether there are different predictive factors for the first, second, and later appearances of the illness (R.M. Post 1992).

Maj (1994) attempted to give structure to the findings regarding predictors of disease course in depression by methodological assessment of the design of the studies in which these findings had been made. An ideal predictor study, according to Maj (1994), is one which deals with a specific depressive disorder, ascertained in operationalized fashion by means of a structured interview. The study design should be prospective, and possible predictors should be assessed with standard instruments. Dropouts from the study should be reported, and no systematic selection effects should be present. Furthermore, the study must define clearly what is meant by remission, both with regard to the severity and number of depressive manifestations, and with regard to the duration of the required asymptomatic interval. Relevant information must be provided about treatments carried out during the period of investigation. Significant predictors must be shown to be free of possible confounding effects. For the analysis of predictors of recurrence, Maj (1994) further requires a clear definition of recurrence, a distinction between recurrence and relapse, and the exclusion of patients who developed bipolar illness in the course of the study.

Such a meta-analysis of the findings regarding the prediction of relapses can never be better than the original publications on which it is based and is, in addition, even more subject to the problem of publication bias. Significant predictors are often reported in published articles without mention of which putative predictors were found not to be significant. This inevitably leads to an overestimation of significant predictors. Future studies on the course of illness must, therefore, conform to present standards of research by testing explicitly stated hypotheses and by providing data on all of the variables studied.

In his review, Maj (1994) stated that the following are documented predictors of a longer time to recovery: the duration of the disease before the beginning of the study and before the beginning of treatment (M.B. Keller et al. 1986), the severity of depressive manifestations, and comorbid psychiatric disorders, particularly dysthymia (M.B. Keller et al. 1992; Wells et al. 1992). Sociodemographic predictors of the time to recovery include old age, female sex, and low family income. Neuroticism is the personality trait most clearly demonstrated to be a predictor. Recent studies have shown that an accompanying personality disorder is predictive of delayed recovery. Multiple studies have revealed an influence of life events on disease course. Partnership and family dysfunctionality is a further predictor. In the study by Maj et al. (1992), patients with three or more episodes had an elevated risk of relapse; in the Heidelberg study, patients with a single previous episode had a risk of relapse comparable to that of patients with multiple previous episodes (Mundt et al. 1998a).

The predictors for relapse of major depression overlap with those for recovery; the most important predictive feature of the disease itself seems to be the number of previous depressive episodes. A large body of evidence suggests that different predictors are operative for initial disease phases and for later recurrences. Thus, F. Keller et al. (1984) reported that there was a relationship between old age and recurrence only in patients in the initial phase of the illness. The findings concerning the influence of life events, social support, and interpersonal relationships are heterogeneous. Kronmüller and Mundt (1999), in a review article, pointed out that half of all studies supported the hypothesis of a predictive value of the concept of "expressed emotion" in depressed patients, while the other half did not. This finding contrasts with findings in bipolar patients, among whom the expressed emotion index has been demonstrated to be predictive for the course of the disorder by all of the studies performed on the question.

An almost totally neglected issue in predictor studies is the consideration of the treatment of depression as a moderating variable. Studies of disease course are almost always conceived exclusively in relation to features of the patients, and of their illnesses, as potential predictive factors. By analogy with the suggested model of Kiesler (1969) for research in psychotherapy, the question should rather be: What factors predict the further course of illness, in which patients, with what disease characteristics, and under what kinds of treatment? This is all the more important at present, now that we have a large number of treatment strategies for depression that have been demonstrated to be effective. The empirical findings in this area are largely of a global nature: continuous

treatment with antidepressants at adequate doses is associated with more rapid recovery and less frequent relapse. The duration before treatment with antidepressants accounts for almost half of the variance in the duration of episodes (Scott et al. 1992). There are hardly any available research findings to date on the influence of psychotherapeutic treatment on predictors of long-term course (Elkin 1994).

More recent findings, in addition to those dealt with by Maj (1994), include primarily those of the Cambridge Study (Paykel et al. 1996), the Heidelberg Study on the prediction of relapses in depression (Mundt et al. 1998a), and the NIMH multicenter study (Solomon et al. 1997). The Heidelberg Study identified numerous predictors for the 1- and 2-year course of depression. Half of the patients studied experienced a relapse within 2 years. Patients in the initial phase of the illness had a significantly more favorable course than those with one or more previous episodes. In addition, the severity of depression at the time of discharge was a further important predictor. Among personality features, neuroticism was shown to be an unfavorable predictor of disease course. A "typus melancholicus" personality structure (Tellenbach 1961) was associated with a better 2-year course. In contrast to the findings of Paykel et al. (1996), the course-predictive significance of life events was confirmed (Reck et al. 1999).

It was also possible to identify interpersonal traits of relationships with partners as predictors of relapse (Fiedler et al. 1998a,b). With the aid of multivariate weighting of individual predictors, it was shown that disease-specific and sociodemographic features indeed accounted for the largest fraction of the variance in the course of illness, but personality factors and features of interpersonal behavior accounted for a further portion (Backenstrass 1998; Mundt et al. 1998a). Marneros et al. (1991), in the Cologne Study, found that the previous course of illness and personality traits were the most important predictors of the further course of illness and the social adaptation of individuals with depression. These findings contrast with those of the studies by Paykel et al. (1996), Andrew et al. (1993), and Solomon et al. (1997), which, with very similar study designs, confirmed the influence of disease-specific features on the course of illness, but not that of psychosocial features.

There is thus a relatively broad consensus on the prognostic significance of disease-specific features. While it is agreed that psychosocial features also play a role, their importance for individual subgroups of patients with depressive disorders is subject to debate. Many differential findings of this nature have been obtained in recent years, but no unified picture has yet been constructed (Mundt et al. 1996). The further progress of research on the course of depression seems to depend both on the performance of hypothesis-based studies (an essential precondition) and on the investigation of differential effects on disease course. Further perspectives for research in this area may be derived from the analysis of the relationship of individual predictors to each other and their possible grouping into provocative factors, vulnerability factors, and symptom-determining factors, as suggested by Brown and Harris (1978), and from the question of whether latent biological susceptibility factors also play a role in the course of illness (R.M. Post 1992). The successful integration and application of these perspectives in future research studies will advance our understanding of the depressive disorders and their course.

7 References

*Abramson LY, Seligman MEP, Teasdale J (1978) Learned helplessness in humans: critique and reformulation. J Abnorm Psychol 87: 49–74

Ahrens B, Müller-Oerlinghausen B, Schou M et al (1995) Excess cardiovascular and suicide mortality of affective disorders may be reduced by lithium prophylaxis. J Affect Disord 33: 67–75

Aitken RCB (1969) Measurement of feelings using visual analogue scales. Proc R Soc Med 62: 989–993

Akiskal HS, King D, Rosenthal TL, Robinson D, Scott-Strauss A (1981) Chronic depressives. I. Clinical and familial characteristics in 137 probands. J Affect Disord 3: 297–315

AMDP (Arbeitsgemeinschaft für Methodik und Dokumentation in der Psychiatrie) (1995) Das AMDP-System: Manual zur Dokumentation psychiatrischer Befunde. Hogrefe, Göttingen

Andreasen NC, Black DW (1993) Lehrbuch Psychiatrie. Beltz, Weinheim

Andreasen NC, Grove WM (1982) The classification of depression: traditional views versus mathematical approaches. Am J Psychiatry 139: 45–52

*Andreasen NC, Grove WM, Maurer R (1980) Cluster analysis and the classification of depression. Br J Psychiatry 137: 256–265

*Andrew B, Horton K, Fagg F, Westbrook D (1993) Do psychosocial factors influence outcome in severely depressed female psychiatric in patients? Br J Psychiatry 163: 747–754

*Angst J (1966) Zur Ätiologie und Nosologie endogener depressiver Psychosen. Eine genetische, soziologische und klinische Studie. Springer, Berlin Heidelberg New York

*Angst J (1978) The course of affective disorders. II. Typology of bipolar manic-depressive illness. Arch Psychiatr Nervenkrankheiten 22: 65–73

Angst J (1980) Verlauf unipolar depressiver, bipolar manisch-depressiver und schizoaffektiver Erkrankungen und Psychosen. Ergebnisse einer prospektiven Studie. Fortschr Neurol Psychiatr 48: 3–30

Angst J (1986) The course of affective disorders. Psychopathology 19[SSuppl 2]: 47–52

Angst J (1987) Verlauf der affektiven Psychosen. In: Kisker KP, Lauter H, Meyer JE, Müller C, Strömgren E (eds) Psychiatrie

der Gegenwart, vol 5. Affektive Psychosen. Springer, Berlin Heidelberg New York, pp 115–133

Angst J (1988a) Clinical course of affective disorders. In: Helgason T, Daly RJ (eds) Depressive illness: prediction of course and outcome. Springer, Berlin Heidelberg New York, pp 1–44

*Angst J (1988b) Risikofaktoren für den Verlauf affektiver Störungen. In: von Zerssen D, Möller HJ (eds) Affektive Störungen: diagnostische, epidemiologische, biologische und therapeutische Aspekte. Springer, Berlin Heidelberg New York, pp 99–110

Angst J (1990) Natural history and epidemiology of depression. Results of community studies. In: Cobb J, Goeting N (eds) Current approaches. Prediction and treatment of recurrent depression. Duphar Medical Relations, Southampton, pp 1–11

Angst J (1993) Die depressive Verstimmung als Schaltstelle psychiatrischer Störungen. In: Hell D (ed) Ethologie der Depression. Fischer, Stuttgart, pp 3–15

Angst J, Frey R (1977) Die Prognose endogener Depressionen jenseits des 40. Lebensjahres. Nervenarzt 48: 571–574

*Angst J, Merikangas K (1997) The depressive spectrum: diagnostic classification and course. J Affect Disord 45: 31–40

Angst J, Scharfetter C (1990) Schizoaffektive Psychosen – ein nosologisches Ärgernis. In: Lungershausen E, Kaschka WP, Witkowski RJ (eds) Affektive Psychosen. Schattauer, Stuttgart, pp 23–31

Angst J, Weis P (1967) Periodicity of depressive psychoses. In: Brill H, Cole JO, Deniker P, Hippius H, Bradley PB (eds) Neuro-psychopharmacology. Excerpta Medica, Amsterdam (Proceedings of the Fifth International Congress of the Collegium Internationale Neuro-Psychopharmacologicum, Washington DC, 1966, pp 703–710)

Angst J, Vollrath M, Merikangas KR, Ernst C (1990) Comorbidity of anxiety and depression in the Zurich cohort study of young adults. In: Maser JD, Cloninger CR (eds) Comorbidity of mood and anxiety disorders. American Psychiatric Press, Washington DC, pp 123–137

APA (1994) Diagnostic and statistical manual of mental disorders, 4th edn. American Psychiatric Association, Washington DC

*Arolt V (1997) Psychische Störungen bei Krankenhauspatienten. Springer, Berlin Heidelberg New York

Arolt V, Driessen M, Bangert-Verleger A, Neubauer H, Schürmann A, Seibert W (1995) Psychische Störungen bei internistischen und chirurgischen Krankenhauspatienten. Nervenarzt 66: 670–677

Astrup C, Fossum A, Holmboe R (1959) A follow-up study of 270 patients with acute affective psychoses. Acta Psychiatr Neurol Scand 135: 7–65

Avery D, Winokur G (1978) Suicide, attempted suicide, and relapse rates in depression. Occurrence after ECT and antidepressant therapy. Arch Gen Psychiatry 35: 749–753

Backenstrass M (1998) Depression und partnerschaftliche Interaktion. Waxmann, Münster

Barraclough B, Pallis D (1975) Depression followed by suicide: a comparison of depressed suicides with living depressives. Psychol Med 5: 55–61

Baumann U (1976) Methodische Untersuchungen zur Hamilton Depressions Skala. Arch Psychiatr Nervenkrankheiten 222: 359–375

Bech P (1992) Symptoms and assessment of depression In: Paykel ES (ed) Handbook of affective disorders. Churchill Livingstone, Edinburgh, pp 3–13

*Bech P (1993) Rating scales for psychopathology, health status and quality of life. Springer, Berlin Heidelberg New York

Bech P, Rafaelsen OJ (1986) The Melancholia Scale: development, consistency, validity, and utility. In: Sartorius N, Ban TA (eds) Assessment of depression. Springer, Berlin Heidelberg New York, pp 259–269

Beck AT (1978) The depression inventory. Center for Cognitive Therapy, Philadelphia

Beck AT (1986) Hopelessness as a predictor of suicide. Ann NY Acad Sci 487: 90–96

Beck AT, Ward CH, Mendelson M, Mock J, Erbaugh J (1961) An inventory for measuring depression. Arch Gen Psychiatry 4: 561–571

*Beck AT, Rush BF, Shaw AJ, Emery G (1979) Cognitive therapy of depression. Wiley, Chichester

*Beckham EE, Leber WR (1995) Handbook of depression. Guilford, New York

*Bellini L, Gatti F, Gasperini M, Smeraldi E (1992) A comparison between delusional and non-delusional depressives. J Affect Disord 25: 129–138

Berrios GE, Samuel C (1987) Affective disorder in the neurological patient. J Nerv Ment Dis 173: 173–176

Blashfield RK, Morey LC (1979) The classification of depression through cluster analysis. Compr Psychiatry 20: 516–527

Bleuler E (1916) Lehrbuch der Psychiatrie. Springer, Berlin

Bortz J (1985) Lehrbuch der Statistik für Sozialwissenschaftler. Springer, Berlin Heidelberg New York

Boulenger JP, Lavallée YJ (1993) Mixed anxiety and depression: Diagnostic issues. J Clin Psychiatry 54: 3–8

Bouman TK (1993) Einschätzung von Stimmungsstörungen. In: Albersnagel FA, Emmelkamp PMG, Van den Hoofdakker RH (eds) Depression. Verlag für Angewandte Psychologie, Göttingen, pp 45–62

Bouman TK, Kok AR (1987) Homogeneity of Beck's Depression Inventory (BDI): applying Rasch Analysis in conceptual exploration. Acta Psychiatr Scand 76: 568–573

Boyd JH, Bruke JD, Gruenberg E et al (1984) Exclusion criteria of DSM-III: a study of co-occurence of hierarchy-free syndromes. Arch Gen Psychiatry 41: 983

Bridges KW, Goldberg DP (1985) Somatic presentations of DSM-III psychiatric disorders in primary care. J Psychosom Res 29: 563–569

Bronisch T, Hecht H (1987) Comparison of depressed patients with and without suicide attempts in their past history. Acta Psychiatr Scand 76: 438–449

Bronisch T, Wittchen HU, Krieg C, Rupp HU, von Zerssen D (1985) Depressive neurosis. A long-term prospective and retrospective follow-up study of former inpatients. Acta Psychiatr Scand 71: 237–248

*Brown GW, Harris T (1978) Social origins of depression. A study of psychiatric disorders in women. Tavistock, London

*Burch EA Jr, Anton RF, Carson WH (1994) Mood congruent and incongruent psychotic depressions: are they the same? J Affect Disord 31: 275–280

Carlson GA, Goodwin FK (1973) The stages of mania: the longitudinal analysis of the manic episode. Arch Gen Psychiatry 28: 221–228

Cassidy F, Murry E, Forest K, Carroll BJ (1997) The performance of DSM-III-R major depression criteria in the diagnosis of bipolar mixed states. J Affect Disord 46: 79–81

CIPS (1986) Internationale Skalen für Psychiatrie. Beltz, Göttingen

Clarke DC, Von Ammon Cavanaugh S, Gibbons RD (1983) The core symptoms of depression in medical and psychiatric patients. J Nerv Ment Dis 171: 705–713

Clayton PJ (1981) The epidemiology of bipolar affective disorder. Compr Psychiatry 22: 31–43

Clayton PJ, Guze SB, Cloninger CR, Martin RL (1992) Unipolar depression: diagnostic inconsistency and its implications. J Affect Disord 26: 111–116

Coppen A, Stadish-Barry H, Bailey J, Houston G, Silcocks P, Hermon C (1991) Does lithium reduce the mortality of recurrent mood disorders? J Affect Disord 23: 1–7

*Coryell W (1997) Do psychotic, minor and intermittent depressive disorders exist on a continuum? J Affect Disord 45: 75–83

*Coryell W, Winokur G (1992) Course and outcome. In: Paykel ES (ed) Handbook of affective disorders. Guilford, New York, pp 89–110

*Coryell W, Winokur G, Shea T, Maser JD, Endicott J, Akiskal HS (1994) The long-term stability of depressive subtypes. Am J Psychiatry 151: 199–204

*Coyne JC, Schwenk TL, Fechner-Bates S (1995) Nondetection of depression by primary care physicians reconsidered. Gen Hosp Psychiatry 17: 267–276

Crow TJ (1986) The continuum of psychosis and its implication of the structure of the gene. Br J Psychiatry 149: 419–429

Cullen W (1800) Nosology: or, a systematic arrangement of diseases, by classes, orders, genera, and species. Creech, Edinburgh

Cutler NR, Post RM (1982) Life course of illness in untreated manic-depressive patients. Compr Psychiatry 23: 101–115

Degwitz R, Helmchen H, Kockott G, Mombour W (1980) Diagnoseschlüssel und Glossar psychiatrischer Krankheiten. Deutsche Ausgabe der internationalen Klassifikation der Krankheiten der WHO, ICD 9. Revision, Kapitel V. Springer, Berlin Heidelberg New York

Devanand DP, Sano M, Tang MX et al (1996) Depressed mood and the incidence of Alzheimer's disease in the elderly living in the community. Arch Gen Psychiatry 53: 175–182

Dilling H, Dittmann V, Freyberger HJ (1990) ICD-10-field trial in German-speaking countries. Pharmocopsychiatry [Suppl] 23: 135–216

Dilling H, Mombour W, Schmidt MH (eds) (1991) Internationale Klassifikation psychischer Störungen. ICD-10, Kapitel V (F), Klinisch-diagnostische Leitlinien. Huber, Bern

Dilling H, Mombour W, Schmidt MH, Schulte-Markwort E (eds) (1994) Internationale Klassifikation psychischer Störungen. ICD-10, Kapitel V (F) Forschungskriterien. Huber, Bern

Dryman A, Eaton WW (1991) Affective symptoms associated with the onset of major depression in the community: findings from the US National Institute of Mental Health Epidemiologic Catchment Area Program. Acta Psychiatr Scand 84: 1–5

*Eaton WW, Badawi M, Melton B (1995) Prodromes and precursors: epidemiologic data for primary prevention of disorders with slow onset. Am J Psychiatry 152: 967–972

Eaton WW, Anthony JC, Gallo J et al (1997) Natural history of diagnostic interview schedule/DSM-IV major depression. Arch Gen Psychiatry 54: 993–999

*Elkin I (1994) The NIMH treatment of depression collaborative research program: where we began and where we are. In: Bergin AE, Garfield SL (eds) Handbook of psychotherapy and behavior change. Wiley, New York, pp 114–139

Erdberg P (1990) The projective assessment of affective disorders. In: Wolman BB, Stricker G (eds) Depressive disorders: facts, theories, and treatment methods. Wiley, New York, pp 248–254

Ernst C, Angst J (1995) Depression in old age. Is there a real decrease in prevalence? A review. European Arch Psychiatry Clin Neurosci 245: 272–287

Fähndrich E, Helmchen H, Linden M (1986) Standardized instruments used in the assessment of depression in German-speaking countries. In: Sartorius N, Ban TA (eds) Assessment of depression. Springer, Berlin Heidelberg New York, pp 1–8

Farmer R, Nelson-Gray R (1990) Personality disorders in depression: hypothetical relations, empirical findings and methodological considerations. Clin Psychol Rev 10: 453–476

*Fava GA, Kellner R (1991) Prodromal symptoms in affective disorders. Am J Psychiatry 148: 823–830

Fava GA, Grandi S, Canestrari R, Molnar G (1985) Prodromal symptoms in primary major depressive disorders. J Affect Disord 19: 149–152

*Fawcett J, Scheftner WA, Clark D, Hedeker D, Gibbons R, Coryell W (1987) Clinical predictors of suicide in patients with major affective disorders: a controlled prospective study. Am J Psychiatry 144: 35–40

*Fawcett J, Scheftner WA, Fogg L, Clark DC, Young MA, Hecker D, Gibbons R (1990) Time-related predictors of suicide in major affective disorders. Am J Psychiatry 147: 1189–1194

Feighner JP, Robins E, Guze SB, Woodruff RA, Winokur G, Munoz R (1972) Diagnostic criteria for use in psychiatric research. Arch Gen Psychiatry 26: 57–63

Feinstein AR (1970) The pre-therapeutic classification of co-morbidity in chronic disease. J Chron Dis 23: 455–468

Fiedler P (1997) Therapieplanung in der modernen Verhaltenstherapie: Von der allgemeinen zur phänomen- und störungsspezifischen Behandlung. In: Reinecker H, Fiedler P (eds) Therapieplanung in der modernen Verhaltenstherapie. Pabst, Lengerich, pp 1–27

Fiedler P, Backenstraß M, Kronmüller KT, Mundt C (1998a) "Expressed Emotion" (EE): Ehequalität und das Rückfallrisiko depressiver Patienten. Nervenarzt 69: 600–608

Fiedler P, Backenstraß M, Kronmüller KT, Mundt C (1998b) Eheliche Interaktion und das Rückfallrisiko depressiver Patienten: Eine Strukturanalyse ehelicher Beziehungsmuster mittels SASB. Verhaltenstherapie 8: 4–13

*Frank E, Prien RF, Jarrett RB, Keller MB (1991) Conceptualization and rationale for consensus definitions of terms in major depressive disorder: remission, recovery, relapse, and recurrence. Arch Gen Psychiatry 48: 851–855

Freyberger HJ, Dilling H (1996) Das Konzept affektiver Störungen im Kapitel V der ICD-10. In: Peters UH, Schifferdecker M, Krahl A (eds) 150 Jahre Psychiatrie, vol 2. Martini, Cologne, pp 580–584

Freyberger HJ, Dittmann V, Stieglitz RD, Dilling H (1990) ICD-10 in der Erprobung: Ergebnisse einer multizentrischen Feldstudie in den deutschprachigen Ländern. Nervenarzt 61: 271–275

*Fuchs T (1994) Uprooting and late-life psychosis. Eur Arch Psychiatry Clin Neurosci 244: 126–130

*Fydrich T, Laireiter AR, Saile H, Engberding M (1996a) Diagnostik und Evaluation in der Psychotherapie: Empfehlungen zur Standardisierung. Z Klin Psychol 25: 161–168

Fydrich T, Schmitz B, Dietrich G, Heinicke S, König J (1996b) Prävalenz und Komorbidität von Persönlichkeitsstörungen. In: Schmitz B, Fydrich T, Limbacher K (eds) Persönlichkeitsstörungen: Diagnostik und Psychotherapie. Psychologie Verlags Union, Weinheim, pp 56–90

Gonzales LR, Lewinsohn PM, Clarke GN (1985) Longitudinal follow-up of unipolar depressives: an investigation of predictors of relapse. J Consult Clin Psychol 53: 461–469

Goodwin F, Jamison K (1990) Manic-depressive illness. Oxford University Press, London

Guze SB, Robins E (1970) Suicide and primary affective disorders. Br J Psychiatry 117: 437–438

Hamilton M (1960) A rating scale for depression. J Neurol Neurosurg Psychiatry 12: 56–62

Hamilton M (1989) Frequency of symptoms in melancholia (depressive illness). Br J Psychiatry 154: 201–206

Hautzinger M (1994) Kognitive Verhaltenstherapie bei Depressionen. In: Hautzinger M (ed) Kognitive Verhaltenstherapie bei psychischen Erkrankungen. Quintessenz, Berlin, pp 39–62

*Hautzinger M (1997) Affektive Störungen. In: Ehlers A, Hahlweg K (eds) Enzyklopädie der Psychologie, vol 2. Grundlagen der Klinischen Psychologie. Hogrefe, Göttingen, pp 155–239

Hautzinger M, Bailer M (1993) Allgemeine Depressions-Skala. Beltz, Weinheim

Hautzinger M, Bailer M (1994) Das Inventar Depressiver Symptome. Beltz, Weinheim

Hautzinger M, Bailer M, Keller F, Worall H (1993) Das Beck Depressions-Inventar. Huber, Bern

Heinroth JCA (1818) Lehrbuch der Störungen des Seelenlebens oder der Seelenstörungen und ihrer Behandlung. Vogel, Leipzig

Helmchen H, Linden M, Wernicke T (1996) Psychiatrische Morbidität bei Hochbetagten. Ergebnisse aus der Berliner Altersstudie. Nervenarzt 67: 739–750

Herpertz S, Saß H, Steinmeyer EM (1996) Subaffektive Persönlichkeitsstörung und affektive Psychosen. In: Gross G, Huber G, Morgner J (eds) Persönlichkeit, Persönlichkeitsstörung, Psychose. Schattauer, Stuttgart, pp 133–140

Hiller W, Dichtl G, Hecht H, Hundt W, von Zerssen D (1993) An empirical comparison of diagnoses and reliabilities in ICD-10 and DSM-III-R. Eur Arch Psychiatry Clin Neurosci 242: 209–217

Hinrichsen GA (1992) Recovery and relapse from major depressive disorder in the elderly. Am J Psychiatry 149: 1575–1579

Hirschfeld RMA (1994) Major depression, dysthymia and depressive personality disorder. Br J Psychiatry 165[Suppl 26]: 23–30

Hole G (1973) Suizidalität und Selbstwertverlust im Erleben depressiver Patienten. Grenzen der phänomenologischen Erfaßbarkeit. Z Psychother Psychol 23: 233–238

Hopkinson G (1965) The prodromal phase of the depressive psychosis. Psychiatr Neurol 149: 1–6

Horwath E, Johnson J, Klerman GL, Weissman MM (1992) Depressive symptoms as relative and attributable risk factors for first-onset major depression. Arch Gen Psychiatry 49: 817–823

Jenkins JH, Kleinman A, Good BJ (1990) Cross-cultural studies of depression. In: Becker J, Kleinman A (eds) Advances in mood disorders. Erlbaum, Hillsdale

Judd LL (1997) Pleomorphic expressions of unipolar depressive disease: summary of the 1996 CINP President's Workshop. J Affect Disord 45: 109–116

Kaelber CT, Moul DE, Farmer ME (1995) Epidemiology of Depression. In: Beckham EE, Leber WR (eds) Handbook of depression. Guilford, New York, pp 3–35

*Katz R, Shaw BF, Vallis TM, Kaiser AS (1995) The assessment of severity and symptom patterns in depression. In: Beckham EE, Leber WR (eds) Handbook of depression. Guilford, New York, pp 61–85

Keller F, Kempf W, Straub R (1984) Zur Differenzierung agitierter und nicht agitierter depressiver Syndrome anhand von Selbstbeurteilungen. In: Wolfersdorf M, Straub R, Hole G (eds) Depressiv Kranke in der Psychiatrischen Klinik. Roderer, Regensburg, pp 336–349

*Keller MB (1994) Depression: a long-term illness. Br J Psychiatry 165[Suppl 26]: 9–15

*Keller MB, Lavori PW (1984) Double depression, major depression, and dysthymia: distinct entities or different phases of a single disorder? Psychopharmacol Bull 20: 399–402

Keller MB, Shapiro RW (1982) "Double depression": superimposition of acute depressive episodes on chronic depressive disorders. Am J Psychiatry 139: 438–442

Keller MB, Lavori PW, Endicott J, Coryell W, Klerman GL (1983) "Double depression": two-year follow-up. Am J Psychiatry 140: 689–694

Keller MB, Lavori PW, Rice J, Coryell W, Hirschfeld RMA (1986) The persistent risk of chronicity in recurrent episodes of nonbipolar major depressive disorder: a prospective follow-up. Am J Psychiatry 143: 24–28

Keller MB, Lavori PW, Mueller TI, Endicott J, Coryell W, Hirschfeld RMA, Shea T (1992) Time to recovery, chronicity and levels of psychopathology in major depression. A 5-year prospective follow-up of 431 subjects. Arch Gen Psychiatry 49: 809–816

Keller MB, Klein DN, Hirschfeld RMA, Kocsis JH (1995) Results of the DSM-IV mood disorders field trial. Am J Psychiatry 152: 843–849

Kendell RE (1976) The classification of depressions: a review of contemporary confusion. Br J Psychiatry 129: 15–28

Kendell RE, Gourlay J (1970) The clinical distinction between psychotic and neurotic depressions. Br J Psychiatry 117: 257–266

Kendler KS (1991) Mood-congruent psychotic affective illness. Arch Gen Psychiatry 48: 362–369

Kessler RC, McGonagle KA, Nelson CB, Hughes M, Swartz M, Blazer DG (1994) Sex and depression in the National Comorbidity Survey. J Affect Disord 30: 15–26

Kielholz P (1971) Diagnose und Therapie der Depression für den Praktiker. Lehmann, Munich

Kiesler DJ (1969) A grid model for theory and research in the psychotherapies. In: Eron LD, Callahan R (eds) The relation of theory to practice in psychotherapy. Aldine, Chicago, pp 115–145

Kinkelin M (1954) Verlauf und Prognose des manisch-depressiven Irreseins. Schweiz Arch Neurol Neurochir Psychiatr 73: 100–146

*Klein MH, Wunderlich S, Shea MT (1993) Models of relationships between personality and depression: toward a framework for theory and research. In: Klein MH, Kupfer DJ, Shea MT (eds) Personality and depression. Guilford, New York, pp 1–54

*Kleinman A, Good B (1985) Culture and depression. University of California Press, Los Angeles

*Klerman GL, Weissman MM (1992) The course, morbidity, and costs of depression. Arch Gen Psychiatry 49: 831–834

Klerman GL, Weissman MM, Rounsaville BJ, Chevron ES (1984) Interpersonal psychotherapy for depression. Basic Books, New York

Klosterkötter J, Steinmeyer EM (1996) Die neuen Ansätze zur Früherkennung von Psychosen. In: Peters UH, Schifferdecker M, Krahl A (eds) 150 Jahre Psychiatrie, vol 2. Martini, Cologne, pp 558–562

Knäuper B, Wittchen HU (1995) Epidemiologie der Major Depression: Nehmen depressive Erkrankungen zu? Z Klin Psychol 23: 8–24

Kraepelin E (1913) Psychiatrie. Ein Lehrbuch für Studierende und Ärzte, vol 3. Klinische Psychiatrie, part 2. Barth, Leipzig

Kraus A (1980) Psychopathologie und Klinik der manisch-depressiven Psychosen. In: Peters UH (ed) Die Psychologie des 20. Jahrhunderts, vol X: Ergebnisse für die Medizin, part 2: Psychiatrie. Kindler, Zurich, pp 437–464

*Kronmüller KT, Mundt C (1999) Interaktionsmuster bei unipolaren und bipolaren Patienten. In: Marneros A (ed) Handbuch der unipolaren und bipolaren Erkrankungen. Thieme, Stuttgart, pp 390–431

*Kuhs H (1990) Depression und Angst. Springer, Berlin Heidelberg New York

*Kuhs H, Tölle R (1987) Symptomatik der affektiven Psychosen (Melancholien und Manien). In: Kisker KP, Lauter H, Meyer JE, Müller C, Strömgren E (eds) Psychiatrie der Gegenwart, vol 5. Affektive Psychosen. Springer, Berlin Heidelberg New York, pp 69–113

Kurz A (1997) Depression im Alter: Klassifikation, Differentialdiagnose und Psychopathologie. In: Radebold H, Hirsch RD, Kipp J, Kortus R, Stoppe G, Struwe B, Wächtler C (eds) Depressionen im Alter. Steinkopff, Darmstadt, pp 33–40

Lang H (1991) Depression und organische Erkrankung. In: Mundt C, Fiedler P, Lang H, Kraus A (eds) Depressionskonzepte heute: Psychopathologie oder Pathopsychologie? Springer, Berlin Heidelberg New York, pp 188–197

Leonhard K, Korff I, Schulz H (1962) Die Temperamente in den Familien der monopolaren und bipolaren phasischen Psychosen. Psych Neurol 143: 416–434

*Lewinsohn PM, Duncan EM, Stanton AK, Hautzinger M (1986) Age at first onset for nonbipolar depression. J Abnorm Psychol 95: 378–383

Lewinsohn PM, Rohde P, Seeley JR, Hops H (1991) Comorbidity of unipolar depression. I. Major depression with dysthymia. J Abnorm Psychol 100: 205–213

Liebowitz MR (1993) Mixed anxiety and depression: should it be included in DSM-IV? J Clin Psychiatry 54: 4–7

Linden M (1979) Psychiatrische und psychologische Klassifikation depressiver Störungen. In: Hautzinger M, Hoffmann N (eds) Depression und Umwelt. Müller, Salzburg, pp 95–124

Lobo A, Campos, R (1997) Managing the psychiatry/primary care interface. In: Robertson MM, Katona CLE (eds) Depression and physical illness. Wiley, Chichester, pp 39–66

Lobo A, García-Campayo JJ, Campos R et al (1996) Somatization in primary care in Spain. I. Estimates of prevalence and clinical characteristics. Br J Psychiatry 168: 344–348

Luderer HJ (1994) Himmelhoch jauchzend, zum Tode betrübt. Thieme, Stuttgart

Lundquist G (1945) Prognosis and course in manic-depressive psychoses. A follow-up study of 319 first admissions. Acta Psychiatr Scand 35: 1–96

*Maier W, Philipp M (1993) Reliabilität und Validität der Subtypisierung und Schweregradmessung depressiver Syndrome. Springer, Berlin Heidelberg New York

*Maj M (1994) Predictors of course of depression. Curr Opin Psychiatry 7: 22–25

Maj M, Veltro F, Pirozzi R, Lobrace S, Magliano L (1992) Pattern of recurrence of illness after recovery from an episode of major depression: a prospective study. Am J Psychiatry 149: 795–800

Manson SM (1996) Culture and DSM-IV: implications for the diagnosis of mood and anxiety disorders. In: Mezzich JE, Kleinman A, Fabrega H, Perron DL (eds) Culture and psychiatric diagnosis. American Psychiatric Press, Washington DC, pp 99–113

*Marneros A, Deister A, Rohde A (1991) Affektive, schizoaffektive und schizophrene Psychosen. Springer, Berlin Heidelberg New York

Maser JDE, Cloninger CRE (1990a) Comorbidity of mood and anxiety disorders. American Psychiatric Press, Washington DC

Maser JDE, Cloninger CRE (1990b) Comorbidity of mood and anxiety disorders: introduction and overview. In: Maser JDE, Cloninger CRE (eds) Comorbidity of mood and anxiety disorders. American Psychiatric Press, Washington DC, pp 3–12

Maser JD, Weise R, Gwirtsman H (1995) Depression and its boundaries with selected axis I disorders. In: Beckham EE, Leber WR (eds) Handbook of depression. Guilford, New York, pp 86–106

Matussek P (1983) Clusteranalyse als Methode psychopathologischer Forschung. Nervenarzt 54: 363–371

Mayer-Gross W, Slater E, Roth M (1969) Clinical Psychiatry, 3rd revised edn. Baillière and Tindall, London

Mayou RA (1997) Depression and types of physical disorders and treatment. In: Robertson MM, Katona CLE (eds) Depression and physical illness. Wiley, New York, pp 21–38

*Merikangas KR, Wicki W, Angst J (1994) Heterogeneity of depression: classification of depressive subtypes by longitudinal course. Br J Psychiatry 164: 342–348

Mezzich JE, Ahn C, Fabrega HJ, Pilkonis PA (1990) Evidence for comorbidity: treated samples and longitudinal studies. In: Maser JD, Cloninger CR (eds) Comorbidity of mood and anxiety disorders. American Psychiatric Press, Washington DC, pp 189–294

Mezzich JE, Kleinman A, Fabrega H, Parron DL (1996) Culture and Psychiatric Diagnosis. American Psychiatric Press, Washington DC

Miles C (1977) Conditions predisposing to suicide: a review. J Nerv Ment Dis 164: 231–246

*Millon T, Kotik-Harper D (1995) The relationship of depression to disorders of personality. In: Beckham EE, Leber WR (eds) Handbook of depression. Guilford, New York, pp 107–146

Moffic HS, Paykel ES (1975) Depression in medical in-patients. Br J Psychiatry 126: 346–353

*Moldin S, Scheftner WA, Rice JP, Nelson E, Knesevich MA, Akiskal HS (1993) Association between major depressive disorder and physical illness. Psychol Med 2: 755–761

Möller HJ, von Zerssen D (1983) Psychopathometrische Verfahren. II. Standardisierte Beurteilungsverfahren. Nervenarzt 54: 1–16

Molnar G, Feeney G, Fava G (1988) Duration and symptoms of bipolar prodromes. Am J Psychiatry 145: 1576–1578

Montgomery SA, Asberg M (1979) A new depression scale designed to be sensitive to change. Br J Psychiatry 134: 382–389

Mundt C (1991) Endogenität von Psychosen – Anachronismus oder aktueller Wegweiser für die Pathogeneseforschung? Nervenarzt 62: 3–15

*Mundt C (1996) Die Psychotherapie depressiver Erkrankungen: Zum theoretischen Hintergrund und seiner Praxisrelevanz. Nervenarzt 67: 183–197

*Mundt C (1998) Psychopathologische und psychosoziale Frühindikatoren depressiver Erkrankungen. In: Klosterkötter J (ed) Frühdiagnostik und Frühbehandlung psychischer Störungen. Springer, Berlin Heidelberg New York, pp 171–183

Mundt C, Saß H (1992) Für und wider die Einheitspsychose. Thieme Stuttgart

*Mundt C, Goldstein MJ, Hahlweg K, Fiedler P (eds) (1996) Interpersonal factors in the origin and course of affektive disorders. Gaskell, London

Mundt C, Backenstraß M, Kronmüller KT, Fiedler P, Kraus A, Stanghellini G (1997) Personality and endogenous/major depression: an empirical approach to typus melancholicus. 2. Validation of typus melancholicus core-properties by personality inventory scales. Psychopathology 30: 130–139

Mundt C, Kronmüller KT, Backenstraß M, Reck C, Fiedler P (1998a) The influence of psychopathology, personality, and marital interaction on the short-term course of major depression. Psychopathology 31: 29–36

Mundt C, Richter P, Hess H, von Stumpf T (1998b) Zeiterleben und Zeitschätzung depressiver Patienten. Nervenarzt 69: 38–45

Nelson JC, Charney DS (1981) The symptoms of major depressive illness. Am J Psychiatry 138: 1–13

NIMH-NIH (National Institute of Mental Health – National Institutes of Health) (1985) Consensus Development Conference Statement: mood disorders. Pharmacologic prevention of recurrences. Consensus Development Panel. Am J Psychiatry 142: 469–476

Nuechterlein KH (1987) Vulnerability models for schizophrenia: State of the art. In: Häfner H, Gattaz WF, Janzarik W (eds) Search for the causes of schizophrenia. Springer, Berlin Heidelberg New York, pp 297–316

*Ormel J, Koeter MW, Brink W, van den Willige G (1991) Recognition, management, and course of anxiety and depression in general practice. Arch Gen Psychiatry 48: 700–706

*Parker G, Hadzi-Pavlovic D, Hickie I, Boyce P, Mitchell P, Wilhelm K, Brodaty H (1991) Distinguishing psychotic and non-psychotic melancholia. J Affect Disord 22: 135–148

Paykel ES (1971) Classification of depressed patients: a cluster analysis derived grouping. Br J Psychiatry 118: 275–288

*Paykel ES (1992) Handbook of affective disorders, 2nd edn. Churchill Livingstone, Edinburgh

*Paykel ES (1994) Historical overview of outcome of depression. Br J Psychiatry 164: 6–8

Paykel ES, Henderson AJ (1977) Application of cluster analysis in the classification of depression: a replication study. Neuropsychobiology 3: 111–119

Paykel ES, Norton KRW (1986) Self-report and clinical interview in the assessment of depression. In: Sartorius N, Ban TA (eds) Assessment of depression. Springer, Berlin Heidelberg New York, pp 356–366

Paykel ES, Prusoff BA, Tanner J (1976) Temporal stability of symptom patterns in depression. Br J Psychiatry 128: 369–374

*Paykel ES, Cooper Z, Ramana R, Hayhurst H (1996) Life events, social support and marital relationships in the outcome of severe depression. Psychol Med 26: 121–133

*Perris C (1966) A survey of bipolar and unipolar recurrent depressive psychoses. Acta Psychiatr Scand 194[Suppl]: 1–189

Pfeiffer WM (1971) Transkulturelle Psychiatrie. Ergebnisse und Probleme. Thieme, Stuttgart

Pfeiffer WM (1990) Depression in kulturvergleichender Sicht. In: Lungershausen E, Kaschka WP, Witkowski RJ (eds) Affektive Psychosen. Schattauer, Stuttgart, pp 35–39

*Pfeiffer WM (1994) Transkulturelle Psychiatrie. Ergebnisse und Probleme, 2nd edn. Thieme, Stuttgart

Pfohl B, Black DW, Noyes R, Coryell WH, Barrash J (1991) Axis I and axis II comorbidity findings: implications for validity. In: Oldham JM (ed) Personality disorders: new perspectives on diagnostic validity. American Psychiatric Press, Washington DC, pp 145–162

Philipp M, Maier W (1987) Diagnosensysteme endogener Depression. Springer, Berlin Heidelberg New York

Philipp M, Maier W, Delmo CD, Buller R, Winter P, Schwarze H (1991) Das depressive Kernsyndrom im Vergleich der operationalisierten Klassifikationssysteme. In: Mundt C, Fiedler P, Lang H, Kraus A (eds) Depressionskonzepte heute: Psychopathologie oder Pathopsychologie? Springer, Berlin Heidelberg New York, pp 145–156

*Phillips KA, Gunderson JG, Hirschfeld RM, Smith LE (1990) A review of the depressive personality. Am J Psychiatry 147: 830–837

*Piccinelli M, Wilkinson G (1994) Outcome of depression in psychiatric settings. Special issue: depression. Br J Psychiatry 164: 297–304

Pilcz A (1901) Die periodischen Geistesstörungen. Fischer, Jena

Post F (1962) The significance of affective symptoms in old age. London University Press, London

*Post RM (1992) Transduction of psychosocial stress into neurobiology of recurrent affective disorder. Am J Psychiatry 149: 999–1010

*Prien RF, Carpenter LL, Kupfer DJ (1991) The definition and operational criteria for treatment outcome of major depressive disorder: a review of the current research literature. Arch Gen Psychiatry 48: 796–800

Rabins PV, Merchant A, Nestadt G (1984) Criteria for diagnosing reversible dementia caused by depression: validation by 2-year follow-up. Br J Psychiatry 144: 488–492

Radloff LS (1977) The CES-D Scale: a self-report depression scale for research in the general population. Appl Psychol Measure 1: 385–401

Reck C, Backenstraß M, Kronmüller KT, Sommer G, Fiedler P, Mundt C (1999) Zur Bedeutung kritischer Lebensereignisse im 2-Jahresverlauf der "Major Depression". Nervenarzt 70: 637–644

Richter P, Werner J, Heerlein A, Kraus A, Sauer H (1998) On the validity of the Beck Depression Inventory. A review. Psychopathology 31: 160–168

*Robertson MM, Katona CLE (1997) Depression and physical illness. Wiley, Chichester

Robins E, Guze SB (1972) Classification of affective disorders: the primary-secondary, the endogenous-reactive, and the neurotic-psychotic concepts. In: Williams TA, Katz MM, Shield JA (eds) Recent advances in the psychobiology of the depressive illnesses. Government Printing Office, Washington DC, pp 283–293

Robins LN, Helzer JE, Croughan J, Ratcliff KS (1981) National Institute of Mental Health Diagnostic Interview Schedule: its history, characteristics, and validity. Arch Gen Psychiatry 38: 381–389

Rodin G, Craven J, Littefield C (1991) Depression in the medically ill – an integrative approach. Brunner, New York

Rohde P, Lewinsohn PM, Seeley JR (1991) Comorbidity of unipolar depression. II. Comorbidity with other mental disorders in adolescents and adults. J Abnorm Psychol 100: 214–222

Röhrle B (1988) Fragebogen zur verhaltenstherapeutischen Diagnostik depressiver Störungen: Ein Kompendium. Deutsche Gesellschaft für Verhaltenstherapie, Tübingen

Rudd MD, Dahm PF, Rajab MH (1993) Diagnostic comorbidity in persons with suicidal ideation and behavior. Am J Psychiatry 150: 928–934

Sanderson WC, Beck AT, Beck J (1990) Syndrome comorbidity in patients with major depression or dysthymia: prevalence and temporal relationships. Am J Psychiatry 147: 1025–1028

Sartorius N, Ban TA (1986) Assessment of depression. Springer, Berlin Heidelberg New York

*Sartorius N, Üstün TB (1995) Mixed anxiety and depressive disorder. Nosology and research methods in psychiatry. Psychopathology 28: 21–25

Sartorius N, Jablensky A, Gulbinat W, Ernberg G (1980) WHO collaborative study: assessment of depressive disorders. Psychol Med 10: 743–749

Sartorius N, Kaelber CT, Cooper JE et al (1993) Progress toward achieving a common language in psychiatry: results from the field trial of the clinical guidelines accompanying the WHO classification of mental and behavioral disorders in ICD-10. Arch Gen Psychiatry 50: 115–124

*Saß H (1987) Die Krise der psychiatrischen Diagnostik. Fortschr Neurol Psychiatr 55: 355–360

Saß H, Wittchen HU, Zaudig M (1996) Diagnostisches und statistisches Manual psychischer Störungen DSM-IV. Hogrefe, Göttingen

*Schatzberg AF, Rothschild AJ (1992) Psychotic (delusional) major depression: should it be included as a distinct syndrome in DSM-IV? Am J Psychiatry 149: 733–745

Schmidt-Degenhard M (1983) Melancholie und Depression. Kohlhammer, Stuttgart

*Schmidt-Degenhard M (1991) Phänomenologische Begriffsbestimmung der Melancholie. In: Mundt C, Fiedler P, Lang H, Kraus A (eds) Depressionskonzepte heute: Psychopathologie oder Pathopsychologie? Springer, Berlin Heidelberg New York, pp 17–32

Schneider K (1932) Über Depressionszustände. Z Ges Neurol Psychiatr 138: 584–589

Schulte D (1994) Vom zunehmenden Einfluß klassifikatorischer Diagnostik auf psychotherapeutische und psychodiagnostische Forschung und Praxis. Diagnostica 40: 262–269

Schulte D, Wittchen HU (1988) Wert und Nutzen klassifikatorischer Diagnostik für Psychotherapie. Diagnostica 34: 85–98

Schulte W (1961) Nichttraurigseinkönnen im Kern melancholischen Erlebens. Nervenarzt 32: 314–320

Scott J, Eccleston D Boys R (1992) Can we predict the persistence of depression? Br J Psychiatry 161: 633–637

Seidenstücker G, Baumann U (1978) Multimethodale Diagnostik. In: Baumann U, Berbalk H, Seidenstücker G (eds) Klinische Psychologie: Trends in Forschung und Praxis. Huber, Bern, pp 134–182

*Shea MT, Hirschfeld RM (1996) Chronic mood disorder and depressive personality. Psychiatr Clin North Am 19: 103–120

Sherman Y (1995) Depressive personality disorder. J Clin Psychiatry 56: 266

Siebel U, Michels R, Hoff P, Schaub RT, Droste R, Freyberger HJ, Dilling H (1997) Multiaxiales System des Kapitels V (F) der ICD-10. Nervenarzt 68: 231–238

Solomon DA, Keller MB, Leon AC et al (1997) Recovery from major depression: a 10-year prospective follow-up across multiple episodes. Arch Gen Psychiatry 54: 1001–1006

Sonneck J, Grüneberger J, Ringel E (1976) Experimental contribution to the evaluation of the suicidal risk of depressive patients. Psychiatr Clin 9: 84–96

Speier PL, Sherak DL, Hirsch S, Cantwell DP (1995) Depression in children and adolescents. In: Beckham EE, Leber WR (eds) Handbook of depression. Guilford, New York, pp 467–493

Spitzer RL, Fleiss JL (1974) A re-analysis of the reliability of psychiatric diagnosis. Br J Psychiatry 125: 341–347

Spitzer RL, Endicott J, Robins E (1978) Research diagnostic criteria: rationale and reliability. Arch Gen Psychiatry 35: 773–782

Stavrakaki C, Vargo B (1986) The relationship of anxiety and depression: a review of the literature. Br J Psychiatry 149: 7–16

Steck P (1988) Sind endogene und neurotische Depressionen psychopathologisch unterscheidbar? Ergebnisse statistischer Analysen. Z Klin Psychol Psychopathol Psychother 36: 337–356

Steinmeyer EM (1980) Depression: Ätiologie, Diagnostik und Therapie. Kohlhammer, Stuttgart

Stenstedt A (1952) A study in manic-depressive psychosis: clinical, social, and genetic investigations. Acta Psychiatr Scand 79: 1–112

Stevens DE, Merikangas KR, Merikangas JR (1995) Comorbidity of depression and other medical conditions. In: Beckham EE, Leber WR (eds) Handbook of depression. Guilford, New York, pp 147–199

*Stieglitz RD (1997) Depressionsdiagnostik heute – Aktuelle Ansätze. In: Wolfersdorf M (ed) Depressionsstationen/stationäre Depressionsbehandlung. Springer, Berlin Heidelberg New York

*Tellenbach H (1961) Melancholie. Springer, Berlin Göttingen Heidelberg

Thase E (1992) Long-term treatments of recurrent depressive disorders. J Clin Psychiatry 53: 32–44

Thiel A, Hoff P, Scherbaum N (1996) Mood (affective) disorders (F3). Psychopathology 5: 285–291

von Gülick-Bailer M, Hautzinger M (1990) Veränderungsverläufe bei depressiven Patienten unter Antidepressivatherapie und Verhaltenstherapie. In: Baumann U, Fähndrich E, Stieglitz RD, Woggon B (eds) Veränderungsmessung in Psychiatrie und klinischer Psychologie. Profil, Munich, pp 85–98

von Riso LP, Thase ME, Howland RH, Friedman ES, Simons AD, Tu XM (1997) A prospective test of criteria for response, remission, relapse, recovery, and recurrence in depressed patients treated with cognitive behavior therapy. J Affect Disord 43: 131–142

von Zerssen D (1976a) Paranoid-Depressivitäts-Skala. Beltz, Weinheim

von Zerssen D (1976b) Die Beschwerdenliste. Beltz, Weinheim

Weissman MM, Klerman GL (1977) The chronic depressive in the community: unrecognized and poorly treated. Compr Psychiatry 18: 523–532

Weitbrecht HJ (1972) Depressive und manische endogene Psychosen. In: Kisker KP, Meyer JE, Müller M, Strömgren E (eds) Psychiatrie der Gegenwart, vol 1. Klinische Psychiatrie. Springer, Berlin Heidelberg New York, pp 83–140

Wells KB, Burnam MA, Rogers W, Hays R, Camp P (1992) Course of depression in adult outpatients. Results from the Medical Outcome Study. Arch Gen Psychiatry 49: 788–794

Westhoff G (1993) Handbuch psychosozialer Meßinstrumente. Hogrefe, Göttingen

WHO (1978) Mental disorders: glossary and guide to their classification in accordance with the ninth revision of the International Classification of Diseases. World Health Organization, Geneva

WHO (1992a) International classification of mental disorders ICD-10, Chap. V (F). Clinical descriptions and diagnostic guidelines. World Health Organization, Geneva

WHO (1992b) SCAN Schedules for Clinical Assessment in Neuropsychiatry. World Health Organization, Geneva

WHO (1993) International classification of mental disorders ICD-10, Chapt. V (F). Diagnostic criteria for research (DCR). World Health Organization, Geneva

Wing JK, Babor T, Brugha T, Burke J (1990) SCAN: Schedules for Clinical Assessment in Neuropsychiatry. Arch Gen Psychiatry 47: 589–593

*Winokur G (1979) Unipolar depression. Is it divisible into autonomous subtypes? Arch Gen Psychiatry 36: 47–52

Winokur G, Tsuang M (1975) The Iowa 500: suicide in mania, depression and schizophrenia. Am J Psychiatry 132: 650–651

Winokur G, Clayton P, Reich T (1969) Manic depressive illness. Mosby, St Louis

Wittchen HU, Essau CA (1993) Comorbidity and mixed anxiety-depressive disorders: is there epidemiologic evidence? J Clin Psychiatry 54: 9–15

Wittchen HU, Semler G (1991) CIDI. Composite International Diagnostic Interview. Beltz, Weinheim

Wittchen HU, Unland H (1991) Neue Ansätze zur Symptomerfassung und Diagnosestellung nach ICD-10 und DSM-III-R: Strukturierte und standardisierte Interviews. Z Klin Psychol 20: 321–342

*Wittchen HU, von Zerssen D (1987) Verläufe behandelter und unbehandelter Depressionen und Angststörungen. Eine klinisch-psychiatrische und epidemiologische Verlaufsuntersuchung. Springer, Berlin Heidelberg New York

Wittchen HU, Zaudig M, Spengler P et al (1991) Wie zuverlässig ist operationalisierte Diagnostik? Die Test-Retest-Reliabilität des Strukturierten Klinischen Interviews für DSM-III-R. Z Klin Psychol 20: 136–153

Wittchen HU, Möller HJ, Vossen A, Hautzinger M, Kasper S, Heuser I (1995) Depression. Wege aus der Krankheit. Karger, Freiburg

Wittchen HU, Wunderlich U, Gruschwitz S, Zaudig M (1997) Strukturiertes Klinisches Interview für DSM-IV, Achse-I (SKID). Hogrefe, Göttingen

Woggon B (1986) AMDP-III in the assessment of depression. In: Sartorius N, Ban TA (eds) Assessment of depression. Springer, Berlin Heidelberg New York, pp 82–89

Wolfersdorf M (1994) Depressionen. Verstehen und bewältigen. Springer, Berlin Heidelberg New York

*Wolfersdorf M (1995) Depressive Störungen. Psychotherapeut 40: 330–347

Wolfersdorf M, Niehus EM, Keller F (1996) Depression und Suizidalität – Ein Kontrollgruppenvergleich zur Psychopathologie bei suizidalen und nicht-suizidalen Depressiven. In: Peters UH, Schifferdecker M, Krahl A (eds) 150 Jahre Psychiatrie, vol 2. Martini, Cologne, pp 585–588

Young MA, Grabler P (1985) Rapidity of symptom onset in depression. Psychiatry Res 16: 309–315

Zeiss AM, Lewinsohn PM, Rohde P (1996) Functional impairment, physical disease and depression in older adults. In: Kato JN, Mann T (eds) Handbook of diversity issues in health psychology: issues of age, gender and orientation and ethnicity. Plenum, New York

Zerbin-Rüdin E (1969) Zur Genetik depressiver Erkrankungen. In: Hippius H, Selbach H (eds) Das depressive Syndrom. Urban and Schwarzenberg, Munich, pp 37–56

Ziehen T (1896) Die Erkennung und Behandlung der Melancholie in der Praxis. Marhold, Halle

Zimmer FT (1991) Konzepte und Aspekte der Chronifizierung von Depressionen. In: Mundt C, Fiedler P, Lang H, Kraus A (eds) Depressionskonzepte heute: Psychopathologie oder Pathopsychologie? Springer, Berlin Heidelberg New York, pp 249–267

*Zimmer R, Lauter H (1984) Zum Problem der depressiven Pseudodemenz. Z Gerontol 17: 109–112

Zubin J, Steinhauer SR (1981) How to break the logjam in schizophrenia: a look beyond genetics. J Nerv Ment Dis 169: 477–492

Zung WWK (1965) A self-rating depression scale. Arch Gen Psychiatry 12: 63–70

CHAPTER

15

P.J. Clayton

Clinical Picture and Course of Bipolar Affective Disorder

1 Definition

Bipolar I affective disorder is a disorder of mood characterized by periods of mania and depression. During the manic phase, the patient feels elated, euphoric, ecstatic, and/or irritable, along with associated symptoms of pressured speech, racing thoughts, distractibility, agitation or overactivity, decreased need for sleep, inflated self-esteem and grandiosity, and excessive involvement in pleasurable activities. In order to meet criteria for mania, the patient must have this mood and symptoms for at least 1 week. During the depressive phase, the patient feels depressed, sad, low, down in the dumps, fearful, anxious, or irritable, with the associated symptoms of diminished interest and pleasure, anorexia or weight loss or increased appetite and weight gain, insomnia or hypersomnia, psychomotor retardation or agitation, fatigue, feelings of worthlessness and guilt, hopelessness, inability to concentrate and recurrent thoughts of death, suicidal ideation, or suicide attempts. These symptoms must occur for a period of at least 2 weeks. Many patients actually experience mixed episodes (sometimes called dysphoric mania). The most conservative definition requires them to meet criteria for both mania and depression. With these criteria, about 5%–10% of episodes are mixed (Bauer et al. 1994). In addition, many patients have brief episodes during the manic phase of tearfulness and sadness and other depressive symptoms. If the patient has a manic episode for a shorter time, e.g. at least 4 days and with less impairment in functioning, the episode is called hypomania and the illness, if embedded in depressive phases, is called bipolar II. For the most part, this illness is distinct from bipolar I (Coryell 1996). The third disorder, cyclothymic disorder, characterizes a patient who has, for at least 2 years, episodes of hypomanic and depressive symptoms which do not meet criteria for either mania or major depression. Whether there should also be a diagnostic category for hyperthymia has not been evaluated. Some recurrent depressives are hyperthymic in recovery.

Although mania and depression as mental disorders were recognized by the Greeks and the bipolarity of the illness was characterized in the 1600s (Clayton 1994), the separation of bipolar illness from unipolar illness did not occur until the late 1960s (Perris 1966; Angst 1966; Winokur and Clayton 1967), and it did not become recognized in our nomenclature until the DSM-III (American Psychiatric Association 1980). Because of that, most studies are modern and still unfolding.

2 Epidemiology (see also Vol. 3, Part 1, Chap. 17)

The lifetime prevalence for bipolar I is 0.8% and bipolar II is 0.5% (Robins and Regier 1991). Kessler et al. (1997) estimated the lifetime prevalence of mania to be 0.9%. Hypomania or bipolar II were not ascertained, and there are no data on cyclothymia.

Bipolar I and bipolar II are nonoverlapping disorders. As compared to bipolar I disorder, data from five family studies of bipolar II disorder indicate that the most prevalent illness in relatives of bipolar II probands is bipolar II disorder, and rates of all major affective disorders (including major depression) are higher in bipolar II probands (Coryell 1996). Those bipolar II probands with bipolar I relatives may be misclassified. In addition, in a 10-year follow-up of bipolar II patients (Coryell et al. 1995), a maximum of 15% developed mania, again indicating the small percentage who belong in the bipolar I category. On the other hand, age of onset, numbers of episodes, propensity for seasonal affective disorder (SAD), and rapid cycling are similar in the two disorders. How these likenesses and differences relate to pharmacotherapy for the two disorders is unknown, and how bipolar II relates to other disorders, including somatoform disorders and borderline personality disorders, is not clear.

Bipolar I has a slight preponderance of women over men, although women, as expected, have more episodes of depression, and men of mania. Bipolar II, because it is mainly depressive, has a preponderance of women over men. There is little ethnic or urban/rural variation in the bipolar disorders.

3 Age of Onset and First Episode

The age of onset of the first episode is between 10 and 60, although new cases have been reported beyond the usual range (Geller and Luby 1997; Young and Klerman 1992). Approximately a third of patients have their first episode in their teenage years, and 1% have their first episode after 60 (Clayton 1981). More likely, after 60, a patient has had a depressive episode earlier and then manifests mania for the first time at an older age. The first episode in either bipolar I or II disorder is usually a depressive episode, particularly for women (Angst 1978). The average time from the first depression to the switch is 6 years (Akiskal et al. 1983), and known precipitants of such a switch are pregnancy and antidepressant medication (Altshuler et al. 1995). As many as 15% of patients with three previous episodes

of depression switch to mania years later. Other symptoms or syndromes known to be associated with an outcome of bipolarity are adolescent psychotic or severe depression, the occurrence of a postpartum psychosis, a history of attention deficit disorder in childhood, and perhaps the symptoms of hypersomnia and retardation.

4 Clinical Picture

Although the typical bipolar patient starts with an episode of depression, the defining feature of the illness is an episode of mania. Mania can begin suddenly with the development of symptoms over hours, or more likely over days, but it seldom takes weeks. Early in the course of illness (Ambelas 1987), there is a strong correlation between stressful life events and manic admissions, which lessen as the illness progresses. We have shown (Clayton 1998) a relationship between stress and disrupted sleep patterns, and Wehr et al. (1987) suggested that sleep reduction is the final common pathway to the precipitation of mania. In fact, in watching videotapes of depressed patients treated with sleep deprivation, it is apparent that some patients develop an elevated mood. What percentage do and whether they should be questioned for symptoms of hypomania has yet to be studied. Thus manic patients frequently have insomnia as a major symptom, which definitely worsens their episode, if indeed it is not precipitated by it. The triad of hyperactivity, flight of ideas, and push of speech is the key to the manic syndrome, although any one of these symptoms could be altered if the patient is taking antipsychotics or mood stabilizers. In addition to the mood of euphoria or irritability or both, patients show distractibility, circumstantiality, intrusiveness, attention to detail, and restlessness or extreme motor activity. They may be hypersexual, make excessive phone calls, and be extravagant by overspending or giving away items. Their speech is laced with sentences that clearly show elevated self-esteem and grandiosity. Fifty-three percent of bipolar I patients reported delusion and/or hallucination (Guze et al. 1975). These delusions are usually congruent with their high mood, but sometimes they are paranoid as an outgrowth of their overly active interpretation of their surroundings. The themes of their grandiosity and delusions are often religious, political, or business schemes or related to sex. Catatonic symptoms are also possible. A recent study (Bräunig et al. 1998) reported that 31% of manics had such symptoms and that they were markers for a more severe course and outcome. Patients may frequently lose interest in eating just because they are distracted by so many other thoughts and activities. They certainly sleep less and need less sleep. Because they do not have insight into the nature of their illness, they are frequently impatient with those around them, including family and spouses, who frequently bear the brunt of their irritability. Their impatience is often shown during the psychiatric interview. They have no regard for boundaries and therefore get themselves into trouble, e.g. by putting people down, by inappropriately touching, by arguing with someone in a bar, or trying to enter the White House to deliver a message to the president. Manic patients frequently drink more during an episode and are at risk of developing comorbid alcoholism or substance abuse. It is unclear whether they drink to treat their symptoms or because of their increased sociability. Most manic patients need to be hospitalized during the manic episode. One symptom that is often overlooked is that between 35% and 55% of manics, during their most acute phase, show disorientation or confusion. This was recognized by Kraepelin (1921), who termed it "delirious mania." It is a common symptom in postpartum mania.

A recent factor analysis of symptoms (Cassidy et al. 1998) reported that, in 237 bipolar patients (204 manic; 33 mixed, 14%), five factors emerged. The first and strongest was a depressive factor – depressed mood, anxiety, guilt, mood lability, suicide, and negative euphoria. The second was racing thoughts, pressured speech, increased motor activity, and increased contact. The third was grandiosity, psychosis (any delusion or hallucination), lack of insight, and paranoia (probably also delusional). Factor four was euphoric mood, increased sexuality, humor, and grandiosity, and the fifth was paranoia (more likely hypervigilance and suspiciousness), aggression, and irritability. The depressive factor remained even when the mixed manics were removed, again illustrating that most manic patients have depressive symptoms during an episode. One assumes that only the 14% who met the DSM-III-R criteria for both depression and mania were severely depressed. The duration and severity of these symptoms were not measured. We (Winokur et al. 1969) described depressive symptoms and found them not to be enduring in many patients, a finding consistent with the low rate of completed suicide during manic episodes. Strakowski et al. (1996a) reported that the severity of depressive symptoms was associated with "suicidality," which is confirmed by the factor analysis above.

Although the bulk of the literature supports the similarities rather than the differences between bipolar and unipolar depression, the most controversial issue is whether bipolar depressives have more psychomotor retardation than unipolar depressives. Early studies (Dunner et al. 1976) demonstrated differences in

depressive symptomatology between the two, particularly in psychomotor retardation. Casper et al. (1985) found minimal differences when symptoms were correlated for age. Mitchell et al. (1992) did not confirm these differences either. They only demonstrated that bipolar depressives had shorter episodes of depression and were more likely to be agitated and less likely to demonstrate slowed movements. There was a trend for retardation to be less common in the bipolar depressives. In the collaborative study of depression (P.J. Clayton, unpublished data), we compared the symptoms of 31 bipolar depressives with 327 unipolar depressives. All of the interview items from the Schedule for Affective Disorder and Schizophrenia (SADS) were compared. There were almost no differences between the unipolar and bipolar depressives, including hypersomnia, weight gain, or psychomotor retardation, although the bipolar depressives were more likely to be retarded rather than agitated – a finding not true for the unipolar depressives. The only significant symptom differences were that the unipolar depressives reported more severe insomnia and the bipolar depressives reported more use of alcohol in the current episode. This question needs to be further studied and the comparison needs to include bipolar I and bipolar II depressives as well as unipolar depressives.

5 Course and Outcome

Bipolar illness is a recurring illness. Virtually every patient who has an episode of mania has a recurrence, although the number of episodes in a lifetime varies from two to more than 30 (Grof et al. 1974). Between 15% and 20% of bipolar patients exhibit rapid cycling, meaning four or more episodes of mania, hypomania, or depression in a single year (Coryell et al. 1992).

The usual manic phase lasts approximately 3 months, and the depressive phase 4 months. A short depressive phase frequently precedes the mania, and depression often follows a mania. The episodes may be seasonal, with the depression occurring in the autumn and winter, and the mania in spring and summer, or they may have their own distinct patterns, e.g. mania in February, May, August, and November. At least 10% of patients have manias only and, as previously indicated, there are more men in this category. At the beginning of the illness, the interval between the episodes tends to decrease, e.g. it may be 4 years between the first and second episode, then 2 years until the next, and then 1 year until the next. Once the patient has gone through a number of episodes, however, the duration of the cycle, defined as the time from the beginning of one episode to the beginning of the next, becomes stable and bottoms out with a cycle of about 6–9 months. The best way to predict a patient's future course is by examining his or her course in the past 2 years (Grof et al. 1974).

Chronic mania is rare. Welner et al. (1977) were the first to highlight the debilitating nature of the illness by showing that chronicity, if defined as the presence of symptoms, social decline, or both, occurred in at least one third of bipolar patients. The symptoms are almost always depressive.

In addition to chronicity, other complications of the illness include heavy drinking, substance abuse, pathologic gambling, failure to marry, divorce, and some cognitive decline (Van Gorp et al. 1998).

Bipolar patients have higher suicide attempt rates than unipolar patients (Kessler et al. 1997) and similar completed suicide rates. Despite the availability of adequate treatments, there still are indications that between 10% and 15% of patients diagnosed as bipolar kill themselves during their lifetime. It may be that the trend for suicide in bipolar disease, as in unipolar illness, occurs early in the illness and less frequently as the disease continues (Tsuang and Woolson 1977; Tsuang 1978). There is also an increased mortality from other causes in bipolar illness, and some (Yates and Wallace 1987; Weeke et al. 1987; Sharma and Markar 1994) suggested that it is due to an excess in cardiovascular mortality.

6 Premorbid Personality

Studies on premorbid personality are confounded because they frequently report on interepisode assessment of personality or on those stabilized on lithium or bipolars who rate their "usual selves" after recovery. The other difficulty in most studies is that, although they compare them to similarly assessed unipolar patients, they do not have control groups. In two separate analyses (Angst and Clayton 1986; Clayton et al. 1994), we reported on the premorbid personality traits of Swiss men who later developed unipolar or bipolar disorders and compared them to controls. In both these instances, the bipolar men rated themselves as normal, i.e. their answers to the personality inventory were not different from those of men who did not develop psychiatric disorders. These findings are similar to those of Akiskal et al. (1995, 1998), who used 17 self-rated personality scales for patients and found that, in most instances, the postmorbid usual self of the bipolar I patient was sanguine. The bipolar II patient, in contrast, was cyclothymic and labile, and the depressed patient was always subanxious and subdepressive. This last finding is also similar to what

we reported as the premorbid personality of depressives in the Swiss studies. Sauer et al. (1997) reported that treated bipolar patients, when rating their "usual selves," compared to treated unipolar patients had higher scores on extroversion, even in remission. This is compatible with earlier work previously summarized (Clayton 1994) that also indicated that manics evaluated themselves in a positive way, even in remission.

7 Differential Diagnosis

Although differential diagnosis depends on the age of onset, two points need to be emphasized. Although obviously a biased group, in a recent survey sampling some of the members of National Depressive and Manic-Depressive Association (Lish et al. 1994), 59% of respondents indicated that they had their first symptoms in childhood or adolescence. Fifty percent (less so with early symptoms) consulted a professional within a year or so after the onset of symptoms. There was a substantial delay, however, in receiving the correct diagnosis. About one third were ill for more than 10 years before this was done. Ten percent had consulted seven or more professionals before receiving the correct diagnosis. So while this seems like a rather specific and obvious disorder, it is frequently misdiagnosed. The second important point is that there is considerable evidence now that psychotic mania, which occurs in 50% of the bipolar I patients, is still misdiagnosed, particularly in Afro-Americans (Strakowski et al. 1996b). Despite the fact that there were very few differences in affective symptoms, more Afro-Americans were misdiagnosed as schizophrenic. These two things need to be kept in mind when considering differential diagnosis.

7.1 Childhood, Adolescence, and Early Adulthood

In the occasional bipolar disorder that begins in childhood, the most common misdiagnosis is attention deficit/hyperactivity disorder (ADHD). It may precede mania, be comorbid with mania, or be mania instead of ADHD. These misdiagnoses occur during the manic presentation, which more frequently occurs in boys, as does ADHD. For a diagnosis of ADHD, at least one symptom must be present before age 7. In adolescence and early adulthood, it may still be possible to misdiagnose a patient as having ADHD, but other more important misdiagnoses are schizophrenia, especially when patients present with psychotic mania or psychotic depression, substance abuse of any kind, and conduct disorder.

7.2 Adulthood

The key differentiation in adulthood has to be between schizophrenia and bipolar affective disorder. The schizophrenic subcategory may be catatonic, paranoid, or disorganized. Schizophrenics certainly have grandiose delusions, so grandiosity per se does not help in the differential diagnosis. Moreover, manics have paranoid delusions, although usually the manic delusions are multiple and fleeting as opposed to the methodical, intricately developed scheme observed in schizophrenics. If the patient presents as a schizoaffective manic, he or she is more likely to be manic in the follow-up, whereas schizoaffective depressed patients are as frequently schizophrenic as bipolar. In this age-group, bipolar illness may be masked by a substance abuse disorder or misdiagnosed as a personality disorder such as borderline or antisocial disorder. There is also a smattering of patients with anxiety symptoms (including obsessive compulsive disorder usually embedded in a presenting depression) who have bipolar illness.

7.3 Elderly

Because most bipolar patients present at a younger age, the appearance of mania for the first time after 60, without a history of previous depression, must be considered secondary or induced mania (organic affective syndrome) until proven otherwise. Excellent reviews of the subject can be found in Winokur (1991) and Shulman (1997). The most common induced manias are those involved with some neurologic abnormality such as epilepsy or neoplasm, but clearly metabolic disturbances, infections, and various drugs such as steroids and levodopa can precipitate mania. This should be considered at any time when mania appears for the first time, but is an absolutely essential consideration in the elderly.

7.4 Summary

Although the differential diagnosis for each age-group is somewhat unique, the best underlying way to confirm a diagnosis of bipolar I disorder is by family history of mania or recurrent depression. The triad of push of speech, flight of ideas, and hyperactivity also

could differentiate mania from the other syndromes; however, if the patient is medicated, particularly with mood stabilizers or antipsychotic drugs, these symptoms are muted. Catatonia is more frequently a part of mania than schizophrenia. Two other variables are helpful: (1) age of onset – one third of consecutive bipolar patients are identified before age 20, whereas only 10% of schizophrenics are identified by that age (Clayton 1981); and (2) acuteness versus chronicity – there should be an "onset" for depression and mania and usually an "offset." ADHD, schizophrenia, and personality disorders are usually insidious in onset and continuous and/or unremitting in course.

8
Biologic Markers (see also Vol. 3, Part 1, Chap. 19)

The two most consistent findings in bipolar patients, mostly in the depressed state but also in the manic state, are also found in unipolar patients. These are hypercortisolemia (mostly as measured by dexamethasone suppression) and sleep abnormalities, including shortened rapid eye movement (REM) latency, reduced amount of slow-wave sleep, and increased frequency of eye movements during REM sleep. There are, however, no unequivocal findings reported for bipolar I patients. Recently, Bellivier et al. (1998) used the candidate gene strategy to look at the tryptophan hydroxylase (TPH) gene, which codes for the rate-limiting enzyme in serotonin biosynthesis. They suggested that there was increased presence of at least one copy of the TPH allele in patients with bipolar disorder. Unfortunately, they included both bipolar I and bipolar II patients. Perhaps, however, this method will be more successful than the previous assessment of various enzymes, neurotransmitters, hormones, or other biochemical markers in the illness. (For the current state of knowledge on genetics of affective disorders, see Vol. 3, Part 1, Chap. 18; for current treatment strategies for the bipolar illness, see Vol. 3, Part 1, Chaps. 23–25.)

9
Areas for Further Study

Although bipolar illness was separated from unipolar illness on the basis of differences in age of onset, course, family history, and response to treatment, this separation may not, in the end, prove valid. There are data that suggest, as Kraepelin (1921) did, that the two illnesses may be different forms of the same disorder, with bipolar illness being the more severe, earlier-onset form and the recurrent unipolar the later-onset form. The place of bipolar II in this continuum is between the two. In bipolar II disease, if 15% of patients become bipolar I and another percentage have family members who have bipolar I disease, still others may be similar to recurrent depressive disorder. It is clear that mania marks the illness distinctly, hypomania less so, and depression not at all. Embedded in the depression category are a solid group of patients who probably belong to the bipolar continuum. Lithium augmentation (Álvarez et al. 1997) for depressives who have characteristics similar to bipolar patients and lithium maintenance (Greil et al. 1996) for recurrent depressives add support to this hypothesis. How patients with SAD and psychotic depression fit into this continuum is unclear. In addition, as mentioned earlier, a large number of manic patients have depressive symptoms during their mania, and the most frequent affective disorder in the relatives of bipolar I and II patients (the latter more than the former) is depression. We must continue to search for those unipolar depressed patients who belong to the bipolar category. Thus we still need prospective studies of mania, hypomania, bipolar I depression, bipolar II depression, and major depression. The early-onset group, adolescents and young adults, would be ideal to study. In all five groups, we should study sleep (looking for decreased REM latencies and slow-wave sleep deficits), biologic markers (such as nonsuppression in the dexamethasone suppression test, DST), and treatment response. This would answer questions such as whether there are symptoms in bipolar I and II depression that are distinct from recurrent unipolar depression, what the intermorbid personalities of such patients are, and what the best treatments for these disorders are. We still also need to continue to explore the outcomes of patients with cyclothymia and hyperthymia, if there are enough of both or either.

The ideal treatment for the maintenance of this illness still needs to be found. Lithium is most certainly the drug of choice, but contradictory reports, owing perhaps to no or improper comparison groups, of the effect of lithium on the kidney need to be resolved. Perhaps a proper control in 1999 would be a group of similarly matched bipolar patients treated with different mood stabilizers rather than untreated bipolars, other psychiatric patients, or no comparison. Bipolar patients must be the controls, and they need to be followed for years. There are still unanswered questions about the efficacy of valproate in maintenance and the relationship between valproate and polycystic ovaries in young women who receive it for long-term treatment. Efficacy of optimal treatments during pregnancy and postpartum, the treatment of bipolar I and II depression, and the treatment of rapid cyclers all need systematic studies.

None of these issues deal with the comorbidity of this disorder with other disorders, particularly substance abuse. In a naturalistic study, the percentage of patients who abuse alcohol, smoke, or use cocaine or other drugs and the nature of the comorbid childhood conditions in these groups identified as teenagers or young adults need to be recorded. Certainly there, the retrospective diagnoses could be validated by sibs, parents, and teachers, something that cannot be done with older adults.

Finally, as well as looking for candidate genes for this disorder, its neurocircuitry needs to be mapped.

10 References

Akiskal H, Wlaker P, Puzantian V, King D, Rosenthal T, Dranon M (1983) Bipolar outcome in the course of depressive illness: phenomenologic, familial, and pharmacologic preditors. J Affective Disord 5: 1115–128

Akiskal HS, Maser JD, Zeller PJ, Endicott J, Coryell W, Keller M, Warshaw M, Clayton P, Goodwin F (1995) Switching from "unipolar" to bipolar. II. An 11 year study of clinical and temperamental predictors in 559 patients. Arch Gen Psychiatry 52: 114–123

Akiskal H, Kilzieh N, Zeller P, Maser J, Clayton P, Shea MT, Endicott J, Scheftner W, Hirschfeld R, Keller M (1998) The distinct temperamental profiles of bipolar I, bipolar II and unipolar patients. Arch Gen Psychiatry

Altshuler L, Robert M, Leverich G, Mikalauskas K, Rosoff A, Ackerman L (1995) Antidepressant-induced mania and cycle acceleration: a controversy revisited. Am J Psychiatry 152: 8

Álvarez E, Pérez-Solá V, Pérez-Blanco J, Queraltó J, Torrubia R, Noguera R (1997) Predicting outcome of lithium added to antidepressants in resistant depression. J Affective Disord 42: 179–186

Ambelas A (1987) Life events and mania: a special relationship? Br J Psychiatry 150: 235–240

American Psychiatric Association (1980) Diagnostic and statistical manual of mental disorders, 3rd edn (DSM-III). American Psychiatric Association, Washington DC

**Angst J (1966) Zur Ätiologie und Nosologie endogenerativer depressiver Psychosen. Springer, Berlin Heidelberg New York

*Angst J (1978) The course of affective disorders. II. Typology of bipolar manic-depressive illness. Arch Psychiatr Nervenkr 226: 65–73

Angst J, Clayton PJ (1986) Premorbid personality of depressive, bipolar, and schizophrenic patients. With special reference to suicidal issues. Compr Psychiatry 27: 511–532

Bauer M, Whybrow P, Gyulai L, Gonnel J, Yeh H (1994) Testing definitions of dysphoric mania and hypomania: prevalence, clinical characteristics and inter-episode stability. J Affect Disord 32: 201–211

Bellivier F, Leboyer M, Courtet P, Buresi C, Beaufil B, Samolyk D, Allilair H, Feingold J, Mallet J, Malafosse A (1998) Association between the tryptophan hydroxylase gene and manic-depressive illness. Arch Gen Psychiatry 55(1): 33–37

Bräunig P, Krüger S, Shugar G (1998) Prevalence and clinical significance of catatonic symptoms in mania. Compr Psychiatry 39(1): 35–46

Casper R, Redmond E, Katz M, Schaffer C, Davis J, Koslow S (1985) Somatic symptoms in primary affective disorder. Arch Gen Psychiatry 42: 1098–1104

Cassidy F, Forest K, Murry E, Carroll B (1998) A factor analysis of the signs and symptoms of mania. Arch Gen Psychiatry 55(1): 27–32

**Clayton P (1981) The epidemiology of bipolar affective disorder. Compr Psychiatry 22: 31–43

*Clayton P (1994) Bipolar illness. In: Winokur G, Clayton P (eds) The medical basis of psychiatry, 2nd edn. Saunders, Philadelphia, pp 47–67

Clayton P (1998) The model of stress: the bereavement reaction. In: Dohrenwend B (ed) Adversity, stress and psychopathology. Oxford University Press, Oxford, pp 96–110

*Clayton PJ, Ernst C, Angst J (1994) Premorbid personality traits of men who develop unipolar or bipolar disorders. Eur Arch Psychiatry Clin Neurosci 243: 340–346

Coryell W (1996) Bipolar II disorder: a progress report. J Affect Disord 41(3): 159–161

Coryell W, Endicott J, Keller M (1992) Rapidly cycling affective disorder: demographics, diagnosis, family history and course. Arch Gen Psychiatry 49: 126–131

Coryell W, Endicott J, Maser J, Keller M, Leon A, Akiskal H (1995) Long-term stability of polarity distinctions in the affective disorders. Am J Psychiatry 152: 385–390

Dunner D, Dwyer T, Fieve R (1976) Depressive symptoms in patients with unipolar and bipolar affective disorder. Compr Psychiatry 17(3): 447–451

Geller B, Luby J (1997) Child and adolescent bipolar disorder: review of the past 10 years. J Am Acad Child Adolesc Psychiatry 36: 9

Greil W, Ludwig-Mayerhofer W, Erazo N, Engel R, Czernik A, Giedke H, Müller-Oerlinghausen, Osterheider M, Rudolf G, Sauer H, Tegeler J, Wetterling T (1996) Comparative efficacy of lithium and amitriptyline in the maintenance treatment of recurrent unipolar depression: a randomized study. J Affect Disord 40: 179–180

**Grof P, Angst J, Haines T (1974) The clinical course of depression: practical issues. In: Angst J (ed) Classification and prediction of outcome of depression. Schattauer, New York

Guze S, Woodruff R Jr, Clayton P (1975) The significance of psychotic affective disorders. Arch Gen Psychiatry 32: 1147–1150

Kessler R, Rubinow D, Holmes C, Abelson J, Zhao S (1997) The epidemiology of DSM-III-R bipolar I disorder in a general population survey. Psychol Med 27: 1079–1089

**Kraepelin E (1921) Manic-depressive insanity and paranoia. E & S, Edinburgh

Lish J, Dime-Meenan S, Whybrow P, Price R, Hirschfeld R (1994) The National Depressive and Manic-Depressive Association (DMDA) survey of bipolar members. J Affect Disord 31: 281–294

Mitchell P, Parker G, Jamieson K, Wilhelm K, Hickie I, Brodaty H, Boyce P, Hadzi-Pavlovic D, Roy K (1992) Are there any differences between bipolar and unipolar melancholia? J Affect Disord 25: 97–106

**Perris C (1966) A study of bipolar (manic depressive) and unipolar recurrent depressive psychoses. Acta Psychiatr Scand 42: 1–18

Robins L, Regier D (eds) (1991) Psychiatric disorders in America: the Epidemiologic Catchment Area Study. Free Press, New York

Sauer H, Richter P, Czernik A, Ludwig-Mayerhofer W, Schöchlin, Greil W, von Zerssen D (1997) Personality differences between patients with major depression and bipolar disorder – the impact of minor symptoms on self-ratings of personality. J Affect Disord 42: 166–177

Sharma R, Markar H (1994) Mortality in affective disorder. J Affect Disord 31: 91–96

Shulman K (1997) Disinhibition syndromes, secondary mania and bipolar disorder in old age. J Affect Disorder 46(3): 175–182

Strakowski S, McElroy S, Keck P, West S (1996a) Suicidality among patients with mixed and manic bipolar disorder. Am J Psychiatry 153(5): 674–676

Strakowski S, McElroy S, Keck P, West S (1996b) Racial influence on diagnosis in psychotic mania. J Affect Disord 39: 157–162

Tsuang M (1978) Suicide in schizophrenics, manics, depressives, and surgical controls. Arch Gen Psychiatry 35: 153–155

Tsuang M, Woolson R (1977) Mortality in patients with schizophrenia, mania, depression and surgical conditions. Br J Psychiatry 130: 162–166

Van Gorp W, Altshuler L, Theberge D, Wilkins J, Dixon W (1998) Cognitive impairment in euthymic bipolar patients with and without prior alcohol dependence. Arch Gen Psychiatry 55: 41–46

Weeke A, Juel K, Vaeth M (1987) Cardiovascular death and manic-depressive psychosis. J Affect Disord 13: 287–292

Wehr TA, Sack DA, Rosenthal NE (1987) Sleep reduction as a final common pathway in the genesis of mania. Am J Psychiatry 144: 201–204

Welner A, Welner Z, Leonar MA (1977) Bipolar manic-depressive disorder: a reassessment of course and outcome. Compr Psychiatry 18: 327–332

Winokur G (1991) Mania and depression: a classification of syndrome and disease. Johns Hopkins University Press, Baltimore

**Winokur G, Clayton P (1967) Family history studies. I. Two types of affective disorders separated according to genetic and clinical factors. In: Wortis J (ed) Recent advances in biological psychiatry. Plenum, New York, pp 25–30

**Winokur G, Clayton P, Reich T (1969) Manic depressive illness. Mosby, St Louis

Yates W, Wallace R (1987) Cardiovascular risk factors in affective disorder. J Affect Disord 12: 129–134

Young R, Klerman G (1992) Mania in late life: focus on age at onset. Am J Psychiatry 149: 7

CHAPTER

16

R.J. Boland, M.B. Keller

Other Affective Disorders

1 Introduction

This chapter includes mood disorders that are not included in the category "major affective disorder." Some, such as dysthymia, have become an accepted part of the official nomenclature. Others, such as recurrent brief depression (RBD) and minor depression, remain controversial.

Many of these disorders are new to the nomenclature. In the past, disorders such as minor depression were not necessarily treated as distinct from other depressive disorders, but were instead considered on a continuum with the major mood disorders. DSM-III (American Psychiatric Association 1980), however, used criterion-based categorical diagnoses, separating major depression and bipolar disorder into distinct disorders. These criterion-based diagnoses required the presence of a minimum number of symptoms and a minimum duration of symptoms. This fundamental change in the method of diagnosing psychiatric disorders has been very useful in improving the reliability and validity of mood disorders. This new system has, however, left unrecognized several disorders which are no longer included under the umbrella of the major mood disorders.

These "other affective disorders" can differ from the major mood disorders in either cross-sectional criteria or longitudinal criteria. The disorders that differ in cross-sectional criteria usually have milder symptoms than major depression and are occasionally called "subsyndromal" disorders. These include minor depressive disorder and mixed anxiety–depressive disorder. Those that differ in a longitudinal sense generally have a shorter duration of symptoms than that required for major depression; they may also represent a subtype of major depression with a specialized course (as in seasonal depression). Some disorders, such as dysthymia and double depression, differ in both cross-sectional and longitudinal aspects.

Although organizing these disorders by their cross-sectional and longitudinal criteria helps to structure our discussion of these disorders, the lack of established criteria for many of these disorders leads to continued confusion and mixed terminology. For example, a variety of terms, such as "subsyndromal" and "subthreshold" may be used to describe disorders that do not meet the criteria of established diagnoses. More general categories, such as the use of "not otherwise specified" categories of the official nomenclature cast a wider net and can be used for both cross-sectional and longitudinal diagnoses that do not have an official status.

The fact that these disorders may be less severe than the major mood disorders does not mean that they are benign. Many of these disorders have a chronic course and may, as with double depression, imply a worse outcome than other forms of depression. As with other chronic disorders, such as diabetes or hypertension, considerable functional impairment, morbidity, and mortality is associated with chronic depressive disorders, even when they are relatively mild in severity (Keller and Hanks 1995).

2 Dysthymia

2.1 Definition

Dysthymia, or dysthymic disorder, is characterized by cross-sectional symptoms that are less severe than those of major depression. The longitudinal course, however, is usually chronic and unremitting.

2.2 Epidemiology and Etiology

Dysthymia is common in community samples, and 4% of the general population (Keller and Hanks 1995) are estimated to have this disorder. It may have a prevalence as high as 15% in primary care settings (Sansone and Sansone 1996).

The etiology of this disorder is not clear. Classically, this disorder was considered one of the "neurotic" disorders and was considered synonymous with depressive neurosis. This neurotic–psychotic distinction, though influential in theoretical and clinical approaches to dysthymia, suffered from questionable validity (Klerman 1984). As a result, the association with depression neurosis was abandoned, and the term "neurosis" was removed altogether from the official U.S. nomenclature.

Subsequent research has focused on the neurobiological aspects of this disorder. The emergence of dysthymia has been reported after subcortical lesions, particularly when accompanied by other focal neurological signs (such as dystonia or extrapyramidal symptoms) (Lauterbach et al. 1997). Other neurological disorders, such as multiple sclerosis, may be associated with the emergence of dysthymia (Moller et al. 1994). As with depression, dysthmic patients may have an abnormal cortisol response to cortisol stimulation tests (Leake et al. 1989). Treatment studies have also changed the way we look at dysthymia.

Classically, in keeping with the psychotic–neurotic distinction, dysthymia was thought less responsive to somatic treatment. Although there is inadequate research on this subject, available research suggests that this long-held assumption is not true. Compared with major depression, this disorder may respond preferentially to serotonin reuptake inhibitors, arguing for a serotonin-specific mechanism of the disorder. The result has been a gradual change in the view of dysthymia from being a type of personality disorder to a chronic form of mood disorder.

2.3 Diagnosis

In DSM-IV, dysthymia is included in the official nomenclature as a mood disorder and termed dysthymic disorder (American Psychiatric Association 1994). It is defined as a "depressed mood, for most of the day, for more days than not" which lasts for at least 2 years. Along with a depressed mood, the individual must have at least two out of six possible depressive symptoms (see Appendix A). In children, the mood may also be irritable instead of depressed and may last for only 1 year.

In DSM-IV, dysthymic disorder cannot directly follow an episode of major depression. This is required to distinguish dysthymia from an episode of major depression that is in partial remission. Dysthymia can be diagnosed after an episode of major depression if there has been a full remission from the major depressive episode.

A distinction is made between dysthymia that occurs before or after age 21. This may have some prognostic significance.

DSM-IV also includes alternative diagnostic criteria in the appendix, in which the depressive symptoms (criterion "B") are replaced with a different group. These alternative criteria focus more on cognitive symptoms than the current set (Appendix B) and may better describe the quality of mood symptoms described by dysthmic patients.

In ICD-10, the disorder is termed dysthymia and is described similarly to in DSM-IV (Appendix C).

2.4 Course and Prognosis

By definition, dysthymia is a chronic disorder. It usually waxes and wanes over time. In children, the average length of a dysthmic episode is about 3 years, which is much longer than that for a major depressive episode (Keller and Hanks 1995). In adults, the duration may range from 2 to 20 years, with a mean of 5 years' duration. An early onset of the disorder may imply a more chronic course: one study of dysthymia beginning in childhood or adolescence reported a mean duration of 30 years (Shelton et al. 1997).

The outcome is also problematic. Most patients with dysthymia go on to develop major depression (Keller and Hanks 1995). The disorder may show a high comorbidity with other psychiatric disorders, such as personality disorders. Similarly, comorbid medical disorders, such as chronic gastrointestinal complaints and neurological disorders, are common. One large-scale study in the United States followed depressed individuals for 2 years and found that functional outcome was worse for those individuals with dysthymia (with or without major depression) than for those with major depression alone (Wells et al. 1992).

3 Double Depression

3.1 Definition

Double depression refers to the concurrent presence of both dysthymia and major depression. In this disorder, the episodes of major depression are superimposed on a more chronic depressive disorder.

3.2 Epidemiology and Etiology

Double depression is common: between 25% and 35% of patients with a major depressive episode also have a more chronic depression (Keller et al. 1982).

It is not clear from an etiological perspective what would predispose an individual to double depression and another to dysthymia or unipolar major depression alone. One study examining inheritance patterns in depressed patients and their first-degree relatives was unable to (genetically) distinguish major depression, recurrent depression, minor depression, and double depression from each other (Remick et al. 1996). Another study used single photon emission computed tomography (SPECT) to compare individuals with major depression alone and double depression. The SPECT scans showed significant differences in the frontal and posterior regional cerebral blood flow ratio (Thomas et al. 1993). Such a finding suggests that the cerebral dysfunctions may be different for these two disorders.

3.3 Diagnosis

Currently, double depression is not considered an independent diagnosis by either DSM-IV or ICD-10. In both systems, the disorder would be recorded as two comorbid diagnoses.

This disorder should be distinguished from a minor depression that continues after partial remission from a major depressive episode. To make such a distinction, DSM-IV requires that the dysthymia must clearly precede the episode of major depression or, if it occurs after an episode of major depression, there must be a clear remission of major depressive symptoms for at least 2 months. In practice, it can be difficult to distinguish a chronic residual state of major depression from an return to dysthymia. A history of minor depression preceding the major depressive episode – often by years – is a key feature in differentiating the two.

3.4 Course and Prognosis

By definition, double depression is chronic. Compared with major depressive disorder alone, double depression has an earlier age of onset. Furthermore, double depression affects the outcome of depressive episodes. Keller and colleagues (1983) found that patients with double depression recovered more rapidly from episodes of major depression than those with major depression alone. However, the recovery tends not to be to a state of "normalcy," but to one of dysthymia. Relapse is more frequent in patients with double depression than those with major depression alone – almost twice as likely in one study of 32 double depressed subjects followed for 2 years (Keller et al. 1983).

Individuals with double depression may be more likely to have a personality disorder than those with major depression or dysthymia alone. Double depression also causes greater social morbidity than major depression or dysthymia alone. It may also be a risk factor for suicidality.

4 Recurrent Brief Depression

4.1 Definition

RBD is characterized by brief depressive episodes that do not meet the 2-week criteria for major depressive episodes. The longitudinal course is characterized by multiple and frequent recurrences.

4.2 Prevalence

The prevalence of this disorder in the community is not clear. Angst (1990), in an epidemiological survey of young adults in Switzerland, estimated the prevalence of RBD to be approximately equal to that seen for major depression (estimating a 1-year prevalence of about 4.5%; males, 3.9%; females, 4.9%). Amore and colleagues (1995) suggested a 1-year prevalence in the general population of about 5% and a lifetime prevalence of 16%. The Mood Disorders Field Trial for DSM-IV (Keller et al. 1995), using a clinical sample of depressed patients in the United States, found that 17% of the subjects met the criteria for RBD.

RBD may be more common in samples drawn from primary care or other medical settings. One study using a primary care sample found that almost 8% of primary care patients met criteria for RBD, with about 4% meeting strict criteria (Weiller et al. 1994).

4.3 Diagnosis

Definitions for this disorder vary: the most liberal describes the disorder as meeting all of the DSM criteria for major depressive episodes except the time criterion. Stricter definitions, used in some research studies, require an episode length of less than 1 week consisting of at least four depressive symptoms. Such episodes are required to occur at least monthly for 1 year, with evidence of social or occupational dysfunction (Angst 1990).

Currently, RBD is included in the appendix of DSM-IV under "Criteria Sets and Axes Provided for Further Study" as research criteria (Appendix D). Individuals having these criteria under DSM-IV would be diagnosed as having a depressive disorder, not otherwise specified. ICD-10 includes a category of recurrent depressive disorder, but patients must meet full criteria of major depression to be included in this category. RBD would be diagnosed in ICD-10 under the general category of "Other persistent mood [affective] disorders."

It remains open whether this diagnosis is necessary to capture individuals with debilitating disorders that do not meet the criteria for major depression or dysthymia. Studies comparing such clinical variables as family history, age at onset, and psychiatric comorbidity do not distinguish RBD from major depression, nor do such biological parameters as the

dexamethasone suppression test or thyroid-releasing hormone response. In the DSM-IV Mood Disorders Field Trial (Keller et al. 1995), the majority (91%) of the depressed patients met the criteria for either major depression or dysthymia. This suggests that additional diagnoses are not necessary.

There also remains some debate as to the validity of this diagnosis and its relationship to other diagnoses, such as personality disorders. In patients in which recurrence is frequent with brief interepisode recovery, distinguishing this disorder from dysthymia may be difficult. One study in a primary care sample, however, found that individuals with RBD had not had another affective disorder diagnosis within the last year (Maier et al. 1994).

4.4
Course and Prognosis

As implied by the name, the course of RBD is chronic and recurrent, and it may be particularly unresponsive to usual treatments for depression. The mean episode length is about 3–5 days, with a mean of 20 episodes in a year (Amore et al. 1995). Onsets are sudden, with rapid progression to peak symptoms. Episodes vary in their time to recurrence, averaging about 18 days. About a third of such episodes are sufficiently severe to meet all but the time criteria for severe major depression (Montgomery and Montgomery 1992). Angst (1996) reported that, in his Swiss cohort, RBD remitted in 41%, recurred in 35%, progressed to major depression in 22%, and progressed to bipolar disorder in only 7% of the group.

RBD may have a seasonal pattern, similar to seasonal depression, and subjects with RBD may be more likely to have certain medical complaints, such as functional gastrointestinal complaints. Studies of suicidality in individuals suggest that RBD is a risk factor for suicide. For example, Weiller and colleagues (1994) found that, among primary care patients with RBD, almost a quarter had a history of suicide attempts.

5
Seasonal Depression

5.1
Definition

Seasonal Depression, also called seasonal affective disorder, describes a mood disorder that occurs predictably at a certain time of year. Most commonly, the onset of a depressive episode occurs during the winter, but it can occur at other times of year (such as the spring) as well. With winter depression, the disorder is often associated with episodes of hypomania occurring in the spring.

Symptoms of depression in this disorder are often "atypical" or "reverse vegetative symptoms": hypersomnia, hyperphagia, and weight gain being common (Allen et al. 1993). Fatigue is often the most frequent complaint.

5.2
Epidemiology and Etiology

The prevalence of this disorder varies with age, gender, and location. It is most common (particularly the winter variety) in the young, and 60%–90% of those with this disorder are women. It is more prevalent at higher latitudes: one community survey reported a rate of almost 10% in Fairbanks, Alaska (Booker and Hellekson 1992).

Although the cause of seasonal depression is not known, an association with the diminished periods of daytime is presumed for winter depression. This is supported both by the fact that the disorder is more common at higher latitudes and by the successful use of light therapy to treat the disorder. A serotonergic role, involving rapid tryptophan depletion, has been suggested, but investigations of this hypothesis have had mixed results. Circadian rhythm disturbances have also been suggested as a cause, but hypotheses have disagreed on the exact nature of the disturbance. No consistent evidence exists to support a family predisposition for this disorder.

5.3
Diagnosis

Past diagnostic criteria differ in the degree to which depressive symptoms should be limited to a seasonal pattern (to the exclusion of other episodes during the year). In DSM-IV, seasonal depression is included as a longitudinal course specifier for major depression and requires a pattern of seasonal episodes for at least 2 years, without any other episodes of major depression during this period (Appendix E). Though included as a subtype of major depression, evidence exists for a subsyndromal variety of this disorder.

5.4
Course and Prognosis

The disorder usually has a chronic and remitting course. As many as a third of patients suffering from

seasonal depression also have symptoms typical of RBD.

It appears to vary in whether or not individuals continue to have exclusive seasonal depression. In one retrospective study (Schwartz et al. 1996), less than half the patients (42%) studied with a seasonal pattern had a recurrent disorder limited to the seasonal variety. In that study, a similar percentage of patients had episodes of nonseasonal depression over the same 8-year period, and only 14% had a full remission. Another study, carried out in Japan, reported that 22% of patients with seasonal depression continued to show a stable pattern over a mean of 10 years (Sakamoto et al. 1995).

6
Minor Depressive Disorder

6.1
Definition

Minor depressive disorder is, as the name implies, a more minor form of major depression - a subthreshold depression. It differs from major depression in having fewer symptoms and less functional impairment than the major disorder. Both disorders, however, share similar durations and are otherwise similar.

6.2
Epidemiology

The prevalence of this disorder is unclear, as little consensus exists on criteria for this disorder. It is thought to be at least as common as major depression. In the DSM-IV Mood Disorder Field Study (Keller et al. 1995), 4% of the sample met the criteria for minor depressive disorder. Estimates in primary care populations range from 6% to 10%.

6.3
Diagnosis

Currently, minor depressive disorder is included in the appendix of DSM-IV under "Criteria Sets and Axes Provided for Further Study" as research criteria (Appendix F). Individuals meeting these criteria would currently be diagnosed in DSM-IV as having adjustment disorder with depressed mood. There remains a reluctance to introduce a diagnosis that may be no more than a subthreshold form of an existing diagnosis. In ICD-10, the diagnosis would be included under "other mood [affective] disorders."

6.4
Course and Prognosis

The course of this disorder is not clear, but it is assumed to be similar to that for major depression. Patients with this disorder have higher medical comorbidity, medical costs, and multiple unexplained physical symptoms.

7
Mixed Anxiety-Depressive Disorder

7.1
Definition

Mixed anxiety-depressive disorder represents a disorder in which both anxiety and depressive symptoms are present, but do not meet the threshold for any clinical diagnosis. Although this disorder does not meet strict criteria for either anxiety or depressive diagnoses, it still causes substantial impairment (Zinbarg et al. 1994).

7.2
Epidemiology

In the DSM-IV field trials (Zinbarg et al. 1994), mixed anxiety-depressive disorder was common - at least as prevalent as standard mood and anxiety disorders - particularly in primary care settings.

7.3
Diagnosis

Currently, mixed anxiety-depression is included in the appendix of DSM-IV under "Criteria Sets and Axes Provided for Further Study" as research criteria. In the current nomenclature, individuals meeting these criteria would be defined in "DSM-IV as having anxiety disorder, not otherwise specified."

The presence of this apparently common syndrome may predict a poorer course of illness. It is not clear, however, whether such poor outcomes are simply an additive effect of the comorbid syndromes found in this group. Joffe and colleagues (1993) found a poor outcome to be related to the increased functional and symptomatic severity existing in patients with both anxious and depressive symptoms. Clayton and colleagues (1991), however, found the delayed treatment response to be independent of illness severity. Greater

use of this tentative diagnosis may help clarify these apparent discrepancies.

8
Treatment Issues

Perhaps the most interesting and controversial area of research on the "other affective disorders" is that of treatment. In the past, treatment of most of these disorders (particularly dysthymia and the subthreshold depressive disorders) was based more on theoretical understandings of the disease than on empirical studies. The result has been an undertreatment of these patients. In one study of depressed patients with a mean duration of 30 years of dysthymia, about 40% ever received pharmacotherapy, and only 56% received psychotherapy (Shelton et al. 1997). Excluding seasonal depression (which will be discussed separately), the bulk of data concerns the treatment of dysthymia.

8.1
Psychotherapy

Psychotherapy has been the traditional treatment for dysthymia and subthreshold depressive disorders. This is due more to a traditional understanding of dysthymia as the result of developmental and other personality factors rather than being "physiological" in nature.[a] There is little empirical research, however, to support this practice. The studies that do exist tend to have inadequate methodologies (small samples, lack of a control group; Markowitz 1994; Conte and Karasu 1992). In addition, there is some difficulty in the studies in defining what the goal of therapy is in this chronic population: should it be symptom reduction, improved psychosocial functioning, or relapse prevention? Most existing studies examine the first goal (symptom reduction). Not surprisingly, psychotherapeutic techniques that lend themselves to standardization, such as cognitive therapy, are the most frequent studied. In addition to cognitive therapy, a variety of other therapies have been studied, including interpersonal therapy, marital therapy, group therapy, and family therapy (Markowitz 1994; Conte and Karasu 1992; Paykel 1994). The majority report as least modest response; however, the methodological difficulties make any conclusions impossible.

[a]If only certain types of depression are physiological, what are the others? The philosophical concept of such a mind–body dichotomy is clearly flawed. However, in the opinion of the authors, this dichotomy still forms the basis of much of our clinical practice.

8.2
Pharmacotherapy

As noted above, for many years pharmacotherapy was not considered the treatment of choice for dysthymia, minor depression, and other affective disorders that did not meet the criteria for major depression. This was largely due to the conceptualization of dysthymia and other disorders as more akin to personality disorders than to major depression. This bias continues; even recent studies show that dysthymia is more likely to be treated with antidepressants if the patient has a history of major depression than if they have dysthymia alone (Shelton et al. 1997). There may be at least partial basis for this bias; Akiskal and others have suggested a subtyping of primary dysthymia into disorders that appear more "characterological" (i.e. a type of personality disorder) and those that appear to be a "subaffective disorder" (i.e. a less severe, but otherwise similar version of major depression) (Ravindran et al. 1994).

Similar to psychotherapy, there is a paucity of rigorous studies on the pharmacotherapy of dysthymia and other affective disorders not meeting criteria for major depression. Most available studies investigate the treatment of dysthymia. Most of the studies investigating the treatment of dysthymia show at least a modest response to medication (Howland 1991). Overall, randomized controlled studies exist to show efficacy for most available agents, including tricyclic antidepressants (Kocsis et al. 1985; J.W. Stewart et al. 1993), serotonin reuptake inhibitors (Hellerstein et al. 1993; Thase et al. 1996; Ravindran et al. 1994; Vanelle 1997), and such atypical agents as ritanserin (Bakish et al. 1994), moclobemide (Botte et al. 1992), and amisulpride (Boyer and Lecrubier 1996). Some attempts have been made to investigate whether some agents are preferentially efficacious, e.g. some earlier data suggested that monoamine oxidase inhibitors were more useful for dysthymia (Howland 1991). However, there is no definitive data to suggest that any one agent is more efficacious that the other – the bulk of data suggests, instead, that any available agent used for major depression is likely to be effective for these other disorders. As discussed above, it would appear that the major problem continues to be the undertreatment of these disorders rather than a lack of efficacy for any specific treatment.

8.3
Light Therapy for Seasonal Depression

A full review of therapy for seasonal depression will not be undertaken here, as the use of light therapy for

seasonal depression is widely known and reported. As with the discussion of possibility etiologies of seasonal depression, no proposed mechanism of treatment (such as circadian phase adjustments) adequately explains all empirical observations regarding this treatment. Several questions regarding light therapy are of interest:

1. Is light therapy specific to the treatment of seasonal depression?
2. What aspects of the procedure are necessary for effect?
3. Is it the treatment of choice for seasonal depression, or merely an alternative treatment?

Light therapy is difficult to study, given the challenges of designing a convincing placebo condition. A great number of studies having been done, however, with large sample sizes and reasonable methodology. The majority of these studies strongly support the efficacy of light therapy as a first-line treatment for seasonal depression (Lam et al. 1989). Whether light therapy is specific to seasonal depression is not as clear. The majority of studies seem to support a specificity for seasonal depression (Thalen et al. 1997), with less or no antidepressant response seen in patients with nonseasonal depression. In studies that do show a positive response to light therapy for nonseasonal depression (e.g. Mackert et al. 1990), it is notable that these studies do not show the dose (light intensity) - response effect that has been observed in seasonal depression studies, which raises the question of a placebo effect. Still, more studies need to be done, and light therapy may prove to have a role in nonseasonal depression, either alone or as an adjunct to pharmacotherapy (Lam et al. 1989).

Several aspects of the treatment appear to be directly related to a positive effect. The patient's eyes must be directly exposed to the light, as the therapeutic effect appears to be optically mediated rather then through some other organ (such as the skin) (Attar Levy 1997). Likewise, the duration of treatment and the intensity for the light appear to be directly related to the effectiveness of treatment. The effect does not seem to require the ultraviolet A wavelengths of light (Lam et al. 1992), a fortunate finding, given the potential side effects of chronic ultraviolet exposure. Broadband white light may be preferable to other narrower spectra (K.T. Stewart 1991); however, the intensity of light is of more importance than the spectra.

The proper timing of light therapy is controversial, and studies have conflicted greatly on whether morning or evening light is preferable (or whether it makes no difference). These discrepancies are usually blamed on the varying study designs or on the general difficulty of devising rigorous studies. Many use a crossover design, which may confound the results, as the order in which the treatment is introduced may affect the outcome. In a randomized controlled study using a parallel design on 40 patients with seasonal depression, Wirz-Justice and colleagues (1993) found no difference in the timing of treatment. No subsequent study has convincingly refuted this finding.

Little is known about the final question of how light therapy compares to other treatments for seasonal depression. As light therapy is usually more difficult to obtain and implement than medication treatment, it is an important question. Virtually no comparative studies exist. Research on the pharmacotherapy of seasonal depression is in a relatively nascent stage. A number of medications have been shown to be useful for seasonal depression (at least in open studies), including buproprion (Dilsaver et al. 1992), citalopram (Wirz-Justince et al. 1992), fluoxetine (Ruhrmann et al. 1993), d-fenfluramine (O'Rourke et al. 1989), and the natural remedy hypericum (Kasper 1997).

Appendix A. DSM-IV Diagnostic Criteria for 300.4 Dysthymic Disorder

A. Depressed mood for most of the day, for more days than not, as indicated either by subjective account or observation by others, for at least 2 years. Note: In children and adolescents, mood can be irritable and duration must be at least 1 year.
B. Presence, while depressed, of two (or more) of the following:
   1. Poor appetite or overeating
   2. Insomnia or hypersomnia
   3. Low energy or fatigue
   4. Low self-esteem
   5. Poor concentration or difficulty making decisions
   6. Feelings of hopelessness
C. During the 2-year period (1 year for children or adolescents) of the disturbance, the person has never been without the symptoms in criteria A and B for more than 2 months at a time.
D. No major depressive episode has been present during the first 2 years of the disturbance (1 year for children and adolescents), i.e. the disturbance is not better accounted for by chronic major depressive disorder or by major depressive disorder, in partial remission.

Note: There may have been a previous major depressive episode provided there was a full remission (no significant signs or symptoms for 2 months) before development of the dysthymic disorder. In addition, after the initial 2 years (1 year in children or adolescents) of dysthymic disorder, there may be superimposed episodes of major depressive disorder, in which

case both diagnoses may be given when the criteria are met for a major depressive episode.

E. There has never been a manic episode, a mixed episode, or a hypomanic episode, and criteria have never been met for cyclothymic disorder.
F. The disturbance does not occur exclusively during the course of a chronic psychotic disorder, such as schizophrenia or delusional disorder.
G. The symptoms are not due to the direct physiological effects of a substance (e.g. a drug of abuse, a medication) or a general medical condition (e.g. hypothyroidism).
H. The symptoms cause clinically significant distress or impairment in social, occupational, or other important areas of functioning.

Specify if:
- Early onset: if onset is before age 21 years
- Late onset: if onset is age 21 years or older

Specify (for most recent 2 years of dysthymic disorder):
- With Atypical Features

Appendix B. Alternative DSM-IV Research Criterion B for Dysthymic Disorder

B. Presence, while depressed, of three (or more) of the following:
1. Low self-esteem or self-confidence, or feelings of inadequacy
2. Feelings of pessimism, despair, or hopelessness
3. Generalized loss of interest or pleasure
4. Social withdrawal
5. Chronic fatigue or tiredness
6. Feelings of guilt, brooding about the past
7. Subjective feelings of irritability or excessive anger
8. Decreased activity, effectiveness, or productivity
9. Difficulty in thinking, reflected by poor concentration, poor memory, or indecisiveness

Appendix C. ICD-10 Diagnostic Guidelines for Dysthymia

The essential feature is a very long-standing depression of mood which is never, or only very rarely, severe enough to fulfil the criteria for recurrent depressive disorder, mild or moderate severity. It usually begins early in adult life and lasts for at least several years, sometimes indefinitely. When the onset is later in life, the disorder is often the aftermath of a discrete depressive episode and associated with bereavement or other obvious stress.

It includes the following:

- Depressive neurosis
- Depressive personality disorder
- Neurotic depression (with more than 2 years' duration)
- Persistent anxiety depression

It excludes the following:

- Anxiety depression (mild or not persistent)
- Bereavement reaction, lasting less than 2 years (prolonged depressive reaction)

Appendix D. DSM-IV Research Criteria for Recurrent Brief Depressive Disorder

A. Criteria, except for duration, are met for a major depressive episode.
B. The depressive periods in criterion A last at least 2 days but less than 2 weeks.
C. The depressive periods occur at least once a month for 12 consecutive months and are not associated with the menstrual cycle.
D. The periods of depressed mood cause clinically significant distress or impairment in social, occupational, or other important areas of functioning.
E. The symptoms are not due to the direct physiological effects of a substance (e.g. a drug of abuse, a medication) or a general medical condition (e.g. hypothyroidism).
F. There has never been a major depressive episode, and criteria are not met for dysthymic disorder.
G. There has never been a manic episode, a mixed episode, or a hypomanic episode, and criteria are not met for cyclothymic disorder.

Note: This exclusion does not apply if all of the manic-, mixed-, or hypomanic-like episodes are substance or treatment induced.

H. The mood disturbance does not occur exclusively during schizophrenia, schizophreniform disorder, schizoaffective disorder, delusional disorder, or psychotic disorder not otherwise specified.

Appendix E. DSM-IV Criteria for Seasonal Pattern Specifier

Specify if with seasonal pattern (can be applied to the pattern of major depressive episodes in bipolar I disorder, bipolar II disorder, or major depressive disorder, recurrent):

A. There has been a regular temporal relationship between the onset of major depressive episodes in bipolar I or bipolar II disorder or major depressive disorder, recurrent, and a particular time of the year (e.g. regular appearance of the major depressive episode in the fall or winter).

Note: Do not include cases in which there is an obvious effect of seasonal-related psychosocial stressors (e.g. regularly being unemployed every winter).

B. Full remissions (or a change from depression to mania or hypomania) also occur at a characteristic time of the year (e.g. depression disappears in the spring).

C. In the last 2 years, two major depressive episodes have occurred that demonstrate the temporal seasonal relationships defined in criteria A and B, and no nonseasonal major depressive episodes have occurred during that same period.

D. Seasonal major depressive episodes (as described above) substantially outnumber the nonseasonal major depressive episodes that may have occurred over the individual's lifetime.

Appendix F. DSM-IV Research Criteria for Minor Depressive Disorder

A. A mood disturbance, defined as follows:

1. At least two (but less than five) of the following symptoms have been present during the same 2-week period and represent a change from previous functioning; at least one of the symptoms is either (a) or (b):
   a) Depressed mood most of the day, nearly every day, as indicated by either subjective report (e.g. feels sad or empty) or observation made by others (e.g. appears tearful)

Note: In children and adolescents, can be irritable mood.

   b) Markedly diminished interest or pleasure in all, or almost all, activities most of the day, nearly every day (as indicated by either subjective account or observation made by others)
   c) Significant weight loss when not dieting or weight gain (e.g. a change of more than 5% of body weight in a month), or decrease or increase in appetite nearly every day.

Note: In children, consider failure to make expected weight gains.

   d) Insomnia or hypersomnia nearly every day
   e) Psychomotor agitation or retardation nearly every day (observable by others, not merely subjective feelings of restlessness or being slowed down)
   f) Fatigue or loss of energy nearly every day
   g) Feelings of worthlessness or excessive or inappropriate guilt (which may be delusional) nearly every day (not merely self-reproach or guilt about being sick)
   h) Diminished ability to think or concentrate, or indecisiveness, nearly every day (either by subjective account or as observed by others)
   i) Recurrent thoughts of death (not just fear of dying), recurrent suicidal ideation without a specific plan, or a suicide attempt or a specific plan for committing suicide
2. The symptoms cause clinically significant distress or impairment in social, occupational, or other important areas of functioning.
3. The symptoms are not due to the direct physiological effects of a substance (e.g. a drug of abuse, a medication) or a general medical condition (e.g. hypothyroidism).
4. The symptoms are not better accounted for by bereavement (i.e. a normal reaction to the death of a loved one).

B. There has never been a major depressive episode, and criteria are not met for dysthymic disorder.

C. There has never been a manic episode, a mixed episode, or a hypomanic episode, and criteria are not met for cyclothymic disorder.

Note: This exclusion does not apply if all of the manic-, mixed-, or hypomanic-like episodes are substance or treatment induced.

D. The mood disturbance does not occur exclusively during schizophrenia, schizophreniform disorder, schizoaffective disorder, delusional disorder, or psychotic disorder not otherwise specified.

9 References

Allen JM, Lam RW, Remick RA, Sadovnick AD (1993) Depressive symptoms and family history in seasonal and nonseasonal mood disorders. Am J Psychiatry 150: 443–448

American Psychiatric Association (1980) Diagnostic and statistical manual of mental disorders, 3rd edn. American Psychiatric Association, Washington, DC

American Psychiatric Association (1994) Diagnostic and statistical manual of mental disorders, 4th edn. American Psychiatric Association, Washington, DC

Amore M, Ricci M, Giorgetti G (1995) Recurrent brief depression. Minerva Psichiatr 36: 83–89

**Angst J (1990) Recurrent brief depression. A new concept of depression. Pharmacopsychiatry 23: 63–66

Angst J (1996) Comorbidity of mood disorders: a longitudinal prospective study. Br J Psychiatry 30[Suppl]: 31–7

Attar Levy D (1997) Seasonal depression. Rev Prat 47: 1899–903

Bakish D, Ravindran A, Hooper C, Lapierre Y (1994) Psychopharmacological treatment response of patients with a DSM-III diagnosis of dysthymia disorder. Psychopharm Bull 30: 53–59

Booker JM, Hellekson CJ (1992) Prevalence of seasonal affective disorder in Alaska. Am J Psychiatry 149: 1176–1182

Botte J, Evrard JL, Gilles C, Stenier P, Wolfrum C (1992) Controlled comparison of RO-11–1163 (moclobemide) and placebo in the treatment of depression. Acta Psychiatry Belg 92: 355–69

Boyer P, Lecrubier Y (1996) Atypical antipsychotic drugs in dysthymia: placebo controlled studies of amisulpride versus imipramine, versus amineptine. Eur Psychiatry 11[Suppl 3]: 135S–140S

Clayton PJ, Grove WM, Coryell W, Keller M, Hirschfeld R, Fawcett J (1991) Follow-up and family study of anxious depression. Am J Psychiatry 148: 1512–1517

**Conte HR, Karasu TB (1992) A review of treatment studies of minor depression: 1980–1991. Am J Psychother 46: 58–74

Dilsaver SC, Qamar AB, De Medico VJ (1992) The efficacy of buproprion in winter depression: results of an open trial. J Clin Psychiatry 53: 252–255

*Hellerstein DJ, Yanowitch P, Rosenthal J et al (1993) A randomized double-blind study of fluoxetine versus placebo in the treatment of dysthymia. Am J Psychiatry 150: 1169–1175

*Howland RH (1991) Pharmacotherapy of dysthymia: a review. J Clin Psychopharmacol 11: 83–92

Joffe RT, Bagby RM, Levitt A (1993) Anxious and nonanxious depression. Am J Psychiatry 150: 1257–1258

Kasper S (1997) Treatment of seasonal affective disorder (SAD) with hypericum extract. Pharmacopsychiatry 30[Suppl 2]: 89–93

*Keller MB, Hanks DL (1995) Course and natural history of chronic depression. In: Kocsis JH, Klein DN (eds) Diagnosis and treatment of chronic depression. Guilford, New York

Keller MB, Shapiro RW, Lavori PW, Wolfe N (1982) Recovery in major depressive disorder: analysis with the life table. Arch Gen Psychiatry 39: 905–910

**Keller MB, Lavori PW, Endicott J, Coryell W, Klerman GL (1983) "Double depression": two-year follow-up. Am J Psychiatry 140: 689–694

Keller MB, Klein DN, Hirschfeld RMA et al (1995) Results of the DSM-IV Mood Disorders Field Trial. Am J Psychiatry 152: 843–849

**Klerman GL (1984) History and development of modern concepts of affective illness. In: Post RM, Ballenger JC (eds) Neurobiology of mood disorders. Williams and Wilkins, Baltimore, pp 1–19

Kocsis JH, Frances AJ, Voss C, Mann JJ, Mason BJ, Sweeney J (1985) Imipramine treatment for chronic depression. Arch Gen Psychiatry 45: 253–257

Lam RW, Kripke DF, Gillin JC (1989) Phototherapy for depressive disorders: a review. Can J Psychiatry 34: 140–147

Lam RW, Buchanan A, Mador JA, Corral MR, Remick RA (1992) The effects of ultraviolet-A wavelengths in light therapy for seasonal depression. J Affective Disord 24: 237–243

Lauterbach EC, Jackson JG, Price ST, Wilson AN, Kirsh AD, Dever GE (1997) Clinical, motor, and biological correlates of depressive disorders after focal subcortical lesions. J Neuropsychiatry Clin Neurosci 9: 259–266

Leake A, Griffiths HW, Ferrier IN (1989) Plasma N-POMC, ACTH and cortisol following hCRH administration in major depression and dysthymia. J Affective Disord 17: 57–64

Maier W, Herr R, Lichtermann D, Gansicke M, Benkert O, Faust G (1994) Brief depression among patients in general practice. Prevalence and variation by recurrence and severity. Eur Arch Psychiatry Clin Neurosci 244: 190–195

*Markowitz JC (1994) Psychotherapy of dysthymia. Am J Psychiatry 151: 1114–1121

Moller A, Wiedemann G, Rohde U, Backmund H, Sonntag A (1994) Correlates of cognitive impairment and depressive mood disorder in multiple sclerosis. Acta Psychiatr Scand 89: 117–121

Montgomery SA, Montgomery D (1992) Features of recurrent brief depression. Encephale 18/4: 521–523

O'Rourke D, Wurtman JJ, Wurtman RJ, Chebli R, Gleason R (1989) Treatment of seasonal depression with d-fenfluramine. J Clin Psychiatry 50: 343–347

Paykel ES (1994) Dysthymia in clinical practice: psychological therapies. Acta Psychiatr Scand [Suppl] 383: 35–41

Ravindran AV, Bialik RJ, Lapierre YD (1994) Therapeutic efficacy of specific serotonin reuptake inhibitors (SSRIs) in dysthymia. Can J Psychiatry 39: 21–26

Remick RA, Sadovnick AD, Lam RW, Zis AP, Yee IM (1996) Major depression, minor depression, and double depression: are they distinct clinical entities? Am J Med Genet 67: 347–353

Ruhrmann S, Kasper SD, Hawellek B et al (1993) Fluoxetine versus light therapy in the treatment of SAD [Abstract]. Biol Psychiatry 33: 83A

Sakamoto K, Nakadaira S, Kamo K, Kamo T, Takahashi K (1995) A longitudinal follow-up study of seasonal affective disorder. Am J Psychiatry 152: 862–868

Sansone RA, Sansone LA (1996) Dysthymic disorder: the chronic depression. Am Fam Physician 53: 2588–2596

*Schwartz PJ, Brown C, Wehr TA, Rosenthal NE (1996) Winter seasonal affective disorder: a follow-up study of the first 59 patients of the National Institute of Mental Health Seasonal Studies Program. Am J Psychiatry 153: 1028–1036

*Shelton RC, Davidson J, Yonkers KA, Koran L, Thase ME, Pearlstein T, Halbreich U (1997) The undertreatment of dysthymia. J Clin Psychiatry 58: 59–65

Stewart JW, McGrath PJ, Quitkin FM et al (1993) Chronic depression: response to placebo, imipramine and phenelzine. J Clin Psychopharmacol 13: 391–396

Stewart KT, Gaddy JR, Byrne B, Miller S, Brainard GC (1991) Effects of green or white light for treatment of seasonal depression. Psychiatry Res 38: 261–270

Thalen BE, Morkrid L, Kjellman BF, Wetterberg L (1997) Cortisol in light treatment of seasonal and non-seasonal depression: relationship between melatonin and cortisol. Acta Psychiatr Scand 96: 385–394

Thase ME, Fava M, Halbreich U et al (1996) A placebo-controlled, randomized clinical trial comparing sertraline and imipramine for the treatment of dysthymia. Arch Gen Psychiatry 53: 777–784

Thomas P, Vaiva G, Samaille E (1993) Cerebral blood flow in major depression and dysthymia. J Affective Disord 29: 235–242

Vanelle JM (1997) Controlled efficacy study of fluoxetine in dysthymia. Br J Psychiatry 170: 345–350

*Weiller E, Lecrubier Y, Maier W, Ustun TB (1994) The relevance of recurrent brief depression in primary care. A report from the WHO project on Psychological Problems in General

Health Care conducted in 14 countries. Eur Arch Psychiatry Clin Neurosci 244: 182–189

Wells KB, Burnam MA, Rogers W, Hays R, Camp P (1992) The course of depression in adult outpatients. Results from the Medical Outcomes Study. Arch Gen Psychiatry 49: 788–794

Wirz-Justice A, ven der Velde P, Bucher A, Nil R (1992) Comparison of light treatment with citalopram in winter depression: a longitudinal single case study. Int Clin Psychopharmacol 7: 109–116

Wirz-Justice A, Graw P, Krauchi K (1993) Light therapy in seasonal affective disorder is independent of time of day or circadian phase. Arch Gen Psychiatry 50: 929–937

*Zinbarg RE, Barlow DH, Liebowitz M et al (1994) The DSM-IV field trial for mixed anxiety–depression. Am J Psychiatry 151: 1153–1162

CHAPTER

17

H.-U. Wittchen

Epidemiology of Affective Disorders

Translator: E. Taub

1 Introduction

The major forms of affective disorder with which this chapter is concerned, namely, major depression, bipolar disorders, and dysthymia, have been well studied epidemiologically on an international scale. Nonetheless, this is so only with respect to descriptive epidemiology, i.e. the quantitative measurement of disease prevalence, putative risk factors, and possible complications (e.g. comorbidity). Less is known about which risk factors are in fact relevant, as well as the epidemiological aspects of mental health care delivery, the natural course of disease, and the incidence of affective disorders in different age-groups. There is also a lack of knowledge regarding other illnesses that can be counted among the affective disorders, including adaptive disorder, mixed disorders, brief, recurrent depressive disorders, and "subthreshold" or "minor" depression – all of which, like the major affective illnesses, may be associated with major psychosocial impairment.

Only a few, mostly regional, studies from the 1980s and 1990s are available for the German-speaking countries; these do not allow any reliable conclusions on a national level, nor do they allow possible differences between the former East and West Germany to be reliably determined. The present review is limited to population-based studies of the prevalence of affective illnesses, as narrowly defined by means of structured and standardized diagnostic instruments according to the criteria of DSM-III, DSM-III-R, ICD-10, and DSM-IV. We have deliberately chosen not to discuss the older studies that were already dealt with comprehensively by Shepherd (1975) in the earlier (German) edition of this text.

ICD-10 and DSM-IV, the two diagnostic classifications currently in use, define affective disorders in very similar ways, not only at the level of disease manifestations, but also at the level of diagnostic classification. This convergence has the great advantage for epidemiologic studies that artificial differences related to the particular system of classification used are kept to a minimum, while core epidemiologic data can be obtained with reference to both systems through the use of the same diagnostic instruments.

2 Studies on the Prevalence of Depressive and Bipolar Disorders in the General Population

As shown in Table 1, five population-based studies have been carried out in the German-speaking countries in recent years. These are the Munich Follow-Up Study (MFS) of Wittchen and von Zerssen (1987), carried out between 1974 and 1981 in the federal states (*Länder*) of the former West Germany; the regional follow-up study performed by Fichter (1990) as a component of the Traunstein study (Upper Bavaria); the Swiss regional age-cohort study conduced by Angst et al. (1984); the Early Developmental Stages of Psychopathology (EDSP) study, performed in 1995, in which 3021 subjects drawn from the general population of Munich were investigated (Wittchen et al. 1998); and the Basel study carried out by Wacker et al. (1992).

Because all of these studies, with the exception of the MFS, were restricted to single regions, and most of them dealt with narrowly defined age ranges, they cannot be considered representative of the population of the German-speaking countries as a whole, or even of individual German states. A number of North American, South American, African, and Asian studies of disease prevalence are available for comparative purposes, of which the American National Comorbidity Survey (NCS; Kessler et al. 1994) is particularly comprehensive and carefully performed.

In the great majority of these studies, the affective illnesses were defined according to the diagnostic criteria of DSM-III, DSM-III-R, or DSM-IV, all of which are similar, and most of the studies employed the Composite International Diagnostic Interview (CIDI; WHO 1990; Wittchen and Semler 1990; Wittchen and Pfister 1997) or its precursor, the Diagnostic Interview Schedule (DIS; Robins et al. 1982; Wittchen et al. 1985), a standardized diagnostic instrument for case ascertainment whose reliability and validity have been well studied (Wittchen 1994a). There is thus a relatively homogeneous database available for the last two decades, although the interpretation of these data requires careful attention to a number of methodological differences between the studies.

Nearly all of these studies dealt not with the entire German population, but only with that of smaller regions (e.g. the Traunstein study; Fichter 1990). In addition, a considerable, artificial variation in findings may result from differences between the studies with respect to the population sample studied (e.g. age-groups), the diagnostic system used (DSM-III, DSM-III-R, DSM-IV, ICD-10), and the instruments and methods of evaluation applied (e.g. weighting procedures).

In Table 1, data are given regarding four prevalence figures: lifetime prevalence, i.e. the percentage of individuals who have had an affective illness at any time in their lives until the present, and three different types of cross-sectional prevalence (12-month, 6-month, and point prevalence, the last of which is defined in practice as prevalence over a period of 2–4 weeks, depending on the study).

Table 1. Prevalence (%) of affective disorders according to the DSM-III, DSM-III-R, and DSM-IV criteria

Study	Period	Affective disorders	Subtypes		
			Bipolar disorder	Major depression	Dysthymia
ECA, USA (Regier et al. 1990b)	Point	5.2	0.6	2.3	–
	Interval	5.8	1.0	3.0	–
	Lifetime	8.3	1.3	5.9	3.3
NCS, USA[a] (Kessler et al. 1994)	Point	–	–	–	–
	Interval	11.3	1.3	10.3	2.5
	Lifetime	19.3	1.6	17.1	6.4
Munich, Germany (Wittchen and von Zerssen 1987)	Point	5.6	–	1.7	–
	Interval	6.9	0.2	3.0	–
	Lifetime	12.9	0.2	9.0	4.0
Upper Bavaria, Germany (Fichter 1990)	Point	8.7	0.2	1.7	5.9
	Interval	–	–	–	–
	Lifetime	–	–	–	–
Puerto Rico (Canino et al. 1987)	Point	–	–	–	–
	Interval	2.9	0.3	3.0	–
	Lifetime	7.9	0.5	4.6	4.7
Edmonton, Canada (Bland et al. 1988a,b)	Point	5.1	0.1	2.3	–
	Interval	6.8	0.2	4.6	–
	Lifetime	10.2	0.6	8.6	3.7
Seoul, Korea (Lee et al. 1990a,b)	Point	–	–	–	–
	Interval	5.5	–	–	–
	Lifetime	–	0.4	3.3	2.4
Christchurch, New Zealand (Wells et al. 1989)	Point	8.5	0.1	3.7	–
	Interval	10.4	0.2	6.7	–
	Lifetime	14.7	0.7	12.6	6.4
Florence, Italy (Faravelli et al. 1990)	Point	–	0.4	2.8	1.0
	Interval	–	1.3	6.3	3.0
	Lifetime	–	–	–	–
Zurich, Switzerland (Angst et al. 1984)	Point	–	–	1.8	–
	Interval	–	0.8	7.0	–
	Lifetime	–	3.3	14.4	0.9
Basel, Switzerland[a] (Wacker et al. 1992)	Point	–	–	3.2	1.7
	Interval	–	–	7.2	2.1
	Lifetime	–	0.4	15.7	7.2
Munich, Germany[b] (Wittchen et al. 1998)	Point	–	–	–	–
	Interval	10.1	1.7	5.3	2.9
	Lifetime	16.8	1.8	11.8	3.0
Median (in %, scatter)	Point	–	0.4 (0.1–0.6)	3.1 (1.5–4.9)	2.1 (1.2–3.9)
	Interval	–	1.1 (1.0–1.7)	6.5 (2.6–9.8)	3.3 (2.3–4.6)
	Lifetime	–	1.3 (0.6–3.3)	16.1 (4.4–18)	3.6 (3.1–3.9)

ECA, Epidemiologic Catchment Area Study; NCS, National Comorbidity Study.
[a]DSM-III-R.
[b]DSM-IV.

2.1 Point Prevalence

As shown in summary form in Table 1, there is at first sight a remarkable variation between studies with respect to lifetime prevalence, but most of the studies, despite major differences in the age-groups studied, yielded a figure near 3.1% for the point prevalence of episodes of major depression (range, 1.7%–3.7%). This implies more generally that approximately 3.1% (conservatively estimated) of the population between the ages of 15 and 65 suffered from depression at the time the study was performed. The median value

across studies for the point prevalence of dysthymia was 2.1% (range, 1.0%–5.9%), and that of bipolar disorders, 0.4% (0.1%–0.6%). The latter figure concerns only bipolar I disorder, i.e. fully developed manic episodes.

2.2 Prevalence over Defined Intervals and Lifetime Prevalence

As expected, the lifetime prevalence of bipolar and dysthymic disorders (1.3% and 3.6%, respectively) is somewhat higher than their point prevalence, while the lifetime prevalence of major depression, estimated at 16.1%, is much higher than its point prevalence. Data obtained in the MFS and EDSP studies, which were broken down according to severity of illness and type of course, imply that nearly one third of all individuals suffering from major depression fulfill the criteria for recurrent disease. Furthermore, moderate and severe episodes are clearly more common than mild episodes.

Despite the apparent high degree of scatter in the findings on major depression across studies, a closer analysis reveals that nearly all of the variation can be accounted for by methodological differences, particularly the choice of age-group and the manner of ascertaining and evaluating disease manifestations; cultural factors account for comparatively little of the variation (Wittchen et al. 1994). Thus studies employing the DIS, whose criteria for disease manifestations are stricter, and those with a very wide age range extending into old age yielded the lowest prevalence figures. The studies in which the 6-month prevalence of affective illnesses was determined yielded median figures of 6.5% for major depression, 3.3% for dysthymia, and 1.1% for bipolar disorders.

It is also worthy of note that the more recent studies yielded higher prevalence figures than the older studies. All studies carried out in the 1990s yielded considerably higher rates of depression than older studies, for both men and women. Perhaps a small fraction of the increase in prevalence can be attributed to the more sensitive diagnostic criteria of DSM-III-R, ICD-10, and DSM-IV (inclusion of less severe depressive episodes), but there may also be a "real" increase in the prevalence of depressive illnesses, particularly in younger age-groups (see below).

If we consider the figures for prevalence in different age-groups, we find that, according to most studies, e.g. Epidemiologic Catchment Area (ECA) Study, NCS, MFS (Weissman et al. 1991; Kessler et al. 1994; Wittchen and von Zerssen 1987), the highest prevalence of both bipolar and depressive disorders was between the ages of 30 and 44. In a small number of studies, however, the highest prevalence of major depression was found in the somewhat younger 25- to 30-year-old group (Wells et al. 1989; Bland et al. 1988c; Wittchen 1988; Lee et al. 1990a; Wittchen et al. 1992). The EDSP, in which only 14- to 25-year-olds were studied, further revealed that major depression is very common in adolescents and young adults, with a lifetime prevalence of 11.8% and a 12-month prevalence of 5.3%.

In contrast, the prevalence of dysthymia differs little across age-groups up to age 64 according to the aggregate ECA data and those of most other studies. Past this age, the prevalence of dysthymia declines considerably (Weissman et al. 1991). There is evidence, however, that the ages of 45 to 64 may be a period of especially high risk for this disease; the elevated risk seems to be related to comorbid, chronic physical illnesses (Canino et al. 1987; Wittchen 1988; Wells et al. 1989; Lee et al. 1990a; Wittchen et al. 1992).

Subthreshold (minor) depression, which may be associated with significant psychosocial impairment even though not all of the strict criteria for major depression are fulfilled, and a number of other types of depressive disorder have only rarely been studied epidemiologically. Oldehinkel et al. (1999) estimate the prevalence of subthreshold depression at 5.3% on the basis of a new, carefully performed analysis. Adaptive disorders have been studied to date only in a few, mostly older studies employing the clinical criteria of ICD-9. Wittchen and von Zerssen (1987), in the MFS, arrived at a cross-sectional prevalence of 1.2% and a lifetime prevalence of 2.3%; a somewhat higher prevalence of 4.9% was found by Fichter in the Upper Bavarian study (Fichter 1990). Mixed states of anxiety and depression not fulfilling the criteria for any specific disorder are estimated to have a lifetime prevalence of approximately 1% (Wittchen and Essau 1993a) when only individuals who have never had major depression, dysthymia, or a bipolar disorder are counted. Brief, recurrent depressive disorders have a 12-month prevalence of 4.2%–7.2% according to the studies conducted by Angst et al. (1990).

2.3 Grades of Severity and Subtypes of Depressive Illnesses

Only a few reliable estimates of the prevalence of different grades of severity (mild, moderate, severe) and subtypes (single episode, recurrent) of the depressive disorders are available to date. The EDSP study of adolescents and young adults (Oldehinkel et al. 1999) showed that most individuals with major depression had either moderate or severe depression; mild depressive episodes were present in only 18%. Moreover, the EDSP study and the NCS revealed that 25% (EDSP, in 14- to 24-year-olds) to 46% (NCS, in

15- to 55-year-olds) of all depressive illnesses are of the recurrent type.

2.4 Prevalence of Depressive Disorders in General Medical Practice

General medical practice is an important venue for epidemiologic study, both because of the regular medical help-seeking behavior of the public and because of the opportunity to assess health economic stress indicators of a psychiatric nature. According to the findings of the WHO multicenter study (Üstün and Sartorius 1995), the point prevalence of affective disorders is actually considerably higher in general medical practice than in the population at large. This study was carried out in 15 regions in 14 different countries; two general medical practices in Germany (one in Mainz and one in Berlin) were included in the study. On the basis of clinical interviews employing the CIDI, it was determined that 10.4% of all patients visiting a general medical practitioner currently suffer from a depressive illness according to the ICD-10 criteria (Üstün and Sartorius 1995). As seen in Table 2, the specific figures for Mainz and Berlin are 11.1% and 6.1%, respectively.

An additional 5.4% of these patients fulfilled the ICD-10 criteria for neurasthenia, while 6.5% fulfilled the criteria for subthreshold depression with major psychosocial impairment. The prevalence of bipolar disorders was not assessed. Other noteworthy findings of this study included the high comorbidity rates with anxiety disorders and also with somatoform and addictive illnesses. This study serves to emphasize not only the high prevalence of affective disorders in general medical practice, and the burden of care they represent for the treating physicians (Sartorius et al. 1989), but also the fact that there is a major deficiency in the delivery of mental health care to such patients. Barely 50% of patients found to have current, verified major depression by the research psychiatrist had been diagnosed as such by the general medical practitioner, and only one in ten had received specific treatment for depression.

3 Risk Factors

3.1 Sex and Age at Onset of Illness

All relevant epidemiologic studies in all age-groups have consistently revealed that women are affected by depression much more commonly than men. In the relevant studies (Robins et al. 1984; Canino et al. 1987; Wittchen 1988; Lee et al. 1990a; Bland et al. 1988a; Wacker et al. 1992; Wittchen et al. 1992; Kessler et al. 1994), the lifetime prevalence of major depression has been found to be two to three times higher in women; the prevalence figures ranged from 4.1% to 21.3% for women and from 2.3% to 12.7% in men. The differences between men and women in the point and interval prevalence of depression are just as marked.

Dysthymia, also, has a higher lifetime prevalence in women (2.3%–10.3% in women, 1.2%–4.8% in men). This sex difference has been found to be statistically significant as early as puberty (Oldehinkel et al. 1999) and has been reconfirmed by the recent American representative study, the NCS (Kessler et al. 1994), which employed the more sensitive DSM-III-R criteria. The NCS revealed a lifetime prevalence of depression of 21.3% for women and 12.7% for men, while the corresponding figures for dysthymia were 4.8% for women and 0.8% for men. An even sex distribution has been consistently reported for the bipolar disorders.

The findings regarding the age of onset of illness are less consistent. The earlier studies all showed an earlier age of onset for bipolar disorders than for major depression. According to the findings of the ECA studies (Weissman et al. 1991), the average age of onset of bipolar disorders is 18 years, while hypomanic forms appear somewhat later on average (21.7 years) but still considerably earlier than the average age of

Table 2. Point prevalence of depressive disorders in general medical practice (weighted prevalence in percent) (WHO Cooperative Study after Üstün and Sartorius 1995)

Disorder	Findings in general medical practices	
	Berlin (n=400)	Mainz (n=400)
Depression (ICD-10: F32/33)	6.1	11.1
Dysthymia (ICD-10: F34)	0.5	0.9
Recognized cases	56.7	55.6
Treated cases		
With antidepressants	11.4	10.5
With sedatives	5.7	23.7

onset of major depression, which is 26.5 years. A bimodal distribution of bipolar disorders, with peaks in the third and fifth decades of life, has frequently been postulated (Angst 1987).

Studies have usually revealed a wide variation in the age of onset of depressive disorders. Periods of significantly increased risk for the onset of disease include late adolescence and early adulthood, with progressively increasing incidence up to the age of 45. The age of 30 is usually given as a typical age of onset, although more recent investigations imply that the average age of onset is considerably earlier. Kessler et al. (1994), using multivariate risk analysis, found that the relative risk of developing depression for the first time (compared to the over-45 age-group) is 1.7 between the ages of 15 and 24, 1.3 between the ages of 25 and 34, and 1.4 between the ages of 35 and 44.

3.2 Differences Between Age-Groups

Contrary to the traditional clinical conception that the highest rates of depression are found in older persons (over age 45), more recent studies have found the highest rates of depression in younger age cohorts.

The cumulative rates of a "major depressive episode" in different age cohorts, as found in various studies and listed in order of the age of onset, are shown in Fig. 1. Although the overall prevalence varies among studies, it is nonetheless clear that the youngest age cohorts have the highest prevalence, and the oldest age cohorts the lowest. The difference is considerable: the prevalence of depressive illnesses is barely over 2% in older cohorts, but approximately 10% in younger cohorts. Furthermore, the age of onset of the disorder is clearly earlier in the younger cohorts, i.e. an increasing number of young patients are developing depression at earlier and earlier ages.

Moreover, it is interesting to note that the older studies from the years 1980–1985 consistently revealed a lower prevalence than the more recent studies. This has been taken as an indication that the higher prevalence in younger age cohorts is continuing to rise. Taking the NCS (Kessler et al. 1996) as an example, we can see from Fig. 1 that a difference in prevalence between age cohorts is demonstrable, though not marked, while the age of onset of depression has clearly shifted more strongly toward the younger age cohorts. As the separate curves for men and women reveal, this difference between age cohorts holds regardless of sex.

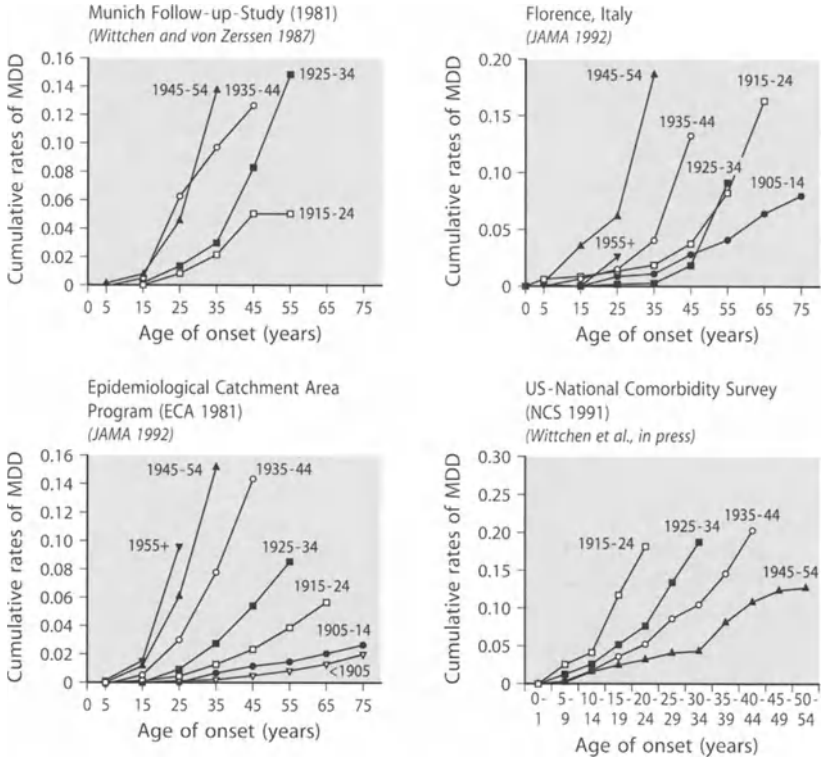


Fig. 1. Cumulative rates of major depressive disorder (*MDD*) in different age cohorts as a function of the age of onset of disease (data obtained from different studies)

These findings have been the subject of thorough review in a number of studies (Cross-National Collaborative Group 1992; Knäuper 1994; Wittchen et al. 1994; Kessler et al. 1996). Possible sources of statistical artifact were examined, including differential morbidity, institutionalization, selective migration, different diagnostic habits and criteria, varying age-specific attitudinal patterns, and methodological aspects of the study instruments. The ultimate conclusion of all of these studies, taken together, is that a large number of methodological artifacts might contribute to the observed differences in prevalence, but no single factor or combination of factors appears able to account for the size of the differences found.

It is of interest that the criticism of these findings from the clinical perspective has been primarily with respect to the lower prevalence among older individuals, rather than the high prevalence among adolescents and young adults. The latter finding appears to be adequately supported by repeated clinical observation and by special studies in childhood and adolescence; the studies carried out Lewinsohn et al. (1993), Garrison et al. (1992), and Klerman (1988) clearly reveal a high prevalence of more than 10%–18% in younger age-groups. As far as the older age-groups are concerned, the Berlin Old Age Study (BAS; Mayer and Baltes 1996) confirmed the absence of any notable increase of prevalence in old age. Studies in our own institution, however, revealed statistically significant evidence of a systematic generation of artifact (Knäuper 1994); thus it might be the case that our instruments for the assessment of depression lead to a systematic underestimation of its prevalence in older individuals.

3.3
Family Situation

Affective disorders, particularly depression, occur more commonly in subjects who are separated from their partners, divorced, or widowed (Wittchen et al. 1987; Bland et al. 1988a; Weissman et al. 1991). Major depression is 2.5 times more common in separated people living alone (Kessler et al. 1994), although living alone per se cannot be considered a risk factor (Weissman et al. 1991). The elevation of risk due to a separation of any kind is significantly higher in women than in men (relative risk, 4.2). The risk analyses performed in a number of studies already mentioned (the NCS, ECA, and MFS) also confirmed that married women are twice as likely as men to develop major depression and that this risk becomes even higher (relative risk, 4.8) if they are not regularly employed (Wittchen 1994b). Even after these factors are statistically controlled for, the main finding that women have a higher prevalence of depression than men remains valid.

3.4
Other Psychosocial Factors

People who were unemployed during at least 6 months out of the 5 years preceding assessment have a threefold elevation of the risk of developing an episode of major depression. Moreover, those with low income who are dependent on government financial aid have a threefold elevation of the prevalence of bipolar disorders or major depression. There are no significant differences between urban and rural areas in this respect (Weissman et al. 1991).

Critical life events with respect to interpersonal relationships, particularly the death of a person with whom the individual has a relationship, and physical illnesses in association with a paucity of social resources (e.g. inadequate social support) and personal resources (e.g. dysfunctional coping behavior) are found primarily in depressive cases (Wittchen 1987; Angst and Dobler-Mikola 1985) and have also been found in analytical epidemiologic studies with small sample groups and in clinical studies. Most of these findings were obtained retrospectively, and the causal relationships are still relatively unclear (see Katschnig 1980; Lin et al. 1986).

3.5
Psychopathological Indicators

Epidemiologic studies have also shown that psychopathological and treatment variables are closely associated with the risk of depressive episodes. For example, with respect to chronicity, Wittchen (1994b) stated, on the basis of the findings of the longitudinal MFS, that an earlier onset of illness, a slower rather than acute onset, the presence of dysthymia, the presence of chronic physical illnesses, and above all the presence of a chronic anxiety disorder (see Sect. 4) all seriously increase the risk of chronic depression.

4
Comorbidity

Although no comprehensive evaluation of comorbidity is yet available from the ongoing longitudinal studies (e.g. the EDSP), a preceding anxiety disorder seems to be one of the more important risk factors for major depression (Thompson et al. 1989; Wittchen and Essau

1989, Maser and Cloninger 1990; Wittchen 1996; Merikangas et al. 1996). Panic disorders are frequently associated with depression, but supposedly mild phobias have also been repeatedly found to be highly significant risk factors for depression (Stein et al. 1990; Vollrath et al. 1990; Merikangas et al. 1996; Kessler et al. 1996). It is worth noting that the association between dysthymia and anxiety disorders is also statistically significant, but of a much smaller magnitude.

There is controversy over the interpretation of these findings. They may be explained by the hypothesis that depression is a frequent secondary complication of chronic anxiety disorders or, alternatively, that other mechanisms are responsible for the association, such as familial genetic linkage, different stages of a single disease process, and so forth (for a review, see Maser and Cloninger 1990; Wittchen and Essau 1993b; Wittchen and Vossen 1995). A further important observation from the comorbidity perspective is that depression takes a different and less favorable course when it is clearly coupled with anxiety (see below).

There are also significant associations between the affective disorders and disorders involving psychotropic substances (abuse, dependency). In the ECA studies, 32% of those with affective disorders also had a disorder involving a psychotropic substance (Regier et al. 1990b). Although the diagnoses of substance abuse and dependency were common in individuals with all subtypes of affective disorder, they were especially common in those suffering from bipolar disorder (60.7%).

An association with somatoform disorders has also been found: 55% of patients with somatoform disorders meet the criteria for major depression at some point in their lives, and 19% meet the criteria for dysthymia (Swartz et al. 1991).

Comorbidity within the group of the affective disorders, particularly that of dysthymia with major depression, is difficult to evaluate because of methodological problems. The differential diagnosis of the affective disorders is based less on the clinical manifestations of disease than on the nature of disease course. It is therefore controversial at present whether dysthymic disorders represent an independent disorder, with a specific type of course, and can thus be considered a valid subtype (Bronisch 1990; Angst and Wicki 1991). Increased clarity on this point can perhaps be expected as a result of the DSM-IV criteria, with which finer distinctions can now be made.

5 Course

The course of the affective disorders is highly variable across the different disorders. Affective disorders may take a phasic course, but they may also be singular occurrences. The duration of episodes varies, but several prospective studies of disease course (the MFS, the ECA, and the Zurich study) have shown that the affective disorders remit completely much more often than the anxiety disorders (range, 32%–46%). As a rule, bipolar forms are thought to be of shorter duration than unipolar forms (Wittchen et al. 1991). A further difficulty in the study of disease course is that disorders initially diagnosed as dysthymia may turn out to be bipolar disorders on further follow-up, i.e. these diagnoses cannot be considered very stable (Angst 1987). Moreover, only a few longitudinal studies of disease course are now available.

Bronisch et al. (1985) found that patients with endogenous depression and with neurotic depression had different courses of disease. A total of 71% of the patients with endogenous depression, but only 37% of the patients with neurotic depression had a favorable course. The rate of remission (defined as 5 years without relapse), according to Angst (1986), is 29% for bipolar disorders and 42% for unipolar depressive illnesses. Robins et al. (1991) found that 42% of those who had ever been diagnosed has having major depression, but only 28% of those who had manic episodes in the preceding year, were asymptomatic at the time of investigation. Bland et al. (1988b), on the other hand, determined that 56% of those with manic episodes and 46% of those with at least one "major depressive episode" were asymptomatic for at least 1 year.

The risk of suicide has been estimated at 15% (Hautzinger and de Jong-Meyer 1994) and is thus considerably higher than in the normal population. More recent findings from the MFS (Wittchen 1993) and the EDSP (Wunderlich et al. 1998) further indicate that comorbid depressive episodes last significantly longer than non-comorbid depression and that they also have a considerably higher rate of relapse.

6 Use of Mental Health Care Services

Although most depressive individuals and most people suffering from a bipolar disorder in the general population report having discussed the symptoms of their disorder with a physician at least once in the course of the illness (Robins et al. 1991), more recent findings from Germany imply that only a third of such persons receive the medical diagnosis of depression or receive any kind of psychological or pharmacological intervention (Wittchen et al. 1999a,b). Nonetheless, among all mental illnesses, affective disorders are second only to the schizophrenic disorders in the rate

of utilization of mental health care services. According to one study, approximately three quarters of affected individuals had visited some type of public health care facility in the 6 months preceding questioning; among these, 31% had visited psychological, psychiatric, and psychosocial facilities, 14% general medical care facilities, and 18% psychiatric specialists (Shapiro et al. 1984). At all levels of mental health care delivery, the rate of treatment of bipolar disorders was somewhat higher than that of major depression (Weissman et al. 1991).

7 Concluding Remarks

The foregoing overview of the currently available epidemiologic studies has shown that depressive disorders are a common and geographically widespread phenomenon. Multivariate analyses have confirmed the existence of multiple risk factors of a sociodemographic nature, particularly age, sex, family situation, and employment; these factors may also interact with one another. Other types of risk factors have been studied only rarely to date, but one consistent and significant finding is that psychopathological precursor states allow considerably better prediction than sociodemographic predictors. Our own working group (Wittchen 1996; Wittchen and Vossen 1995; Kessler et al. 1996) has been able to demonstrate that primary anxiety disorders, in particular, can be considered an important and hitherto underappreciated risk factor for the appearance of depression. Odds ratios in the primary anxiety disorders are typically in the range of 7–12, i.e. subjects with primary anxiety disorders have a 7- to 12-fold risk of developing depression in comparison to the general population.

Another important and stable finding of epidemiologic research is that the highest prevalence of depressive disorders is not, as expected, in old age, but rather in the younger age-groups. Because studies performed in the 1980s yielded considerably lower figures for prevalence than more recent studies, especially in younger age cohorts, it may be concluded that the prevalence of depressive disorders has risen, particularly in the young, at least since the early 1980s. This surprisingly large increase has not yet been definitively explained. Proposed explanations primarily concern psychosocial factors and precursor states, as suggested by the increased prevalence of substance abuse and dependency over the same period.

This overview of the epidemiology of the affective disorders must inevitably remain incomplete. For example, because of space limitations, it has not been possible to discuss the perspectives of health psychology, public health, care evaluation and planning, and pharmacoepidemiology in detail or to pay greater attention to the recent findings of genetic and family genetic research. Nonetheless, we have been able to present some of the major contributions made by descriptive epidemiologic studies to the construction of theories and to the field of psychopathology, e.g. the identification of specific anxiety syndromes as significant risk factors for the onset and relapse of depressive illnesses.

As we have also discussed, there is a major lack of clear, current epidemiological reference data for the German-speaking countries; our knowledge of the epidemiology of affective disorders in this area is largely derived from studies performed in the early 1980s. We thus cannot show any clear secular trend toward an increasing prevalence of mental disorders, as has been found in the United States, although a number of studies make such a trend appear likely. Partly because of the lack of relevant data, it is also not yet possible to establish a rational plan for the delivery of mental health care, e.g. with respect to the demand for psychologists.

8 References

Angst J (1986) The course of affective disorders. Psychopathology 19[Suppl 2]: 47–52

Angst J (1987) Verlauf der affektiven Psychosen. In: Hautzinger M (ed) Psychiatrie der Gegenwart, vol 5. Affektive Psychosen. Springer, Berlin Heidelberg New York, pp 115–133

Angst J, Dobler-Mikola A (1985) The Zurich study. V. Anxiety and phobia in young adults. Eur Arch Psychiatry Neurol Sci 234: 408–418

Angst J, Wicki W (1991) The Zurich Study. XI. Is dysthymia a separate form of depression? Results of the Zurich cohort study. Eur Arch Psychiatry Clin Neurosci 240: 349–354

Angst J, Dobler-Mikola A, Binder J (1984) The Zurich Study – a prospective epidemiological study of depressive, neurotic and psychosomatic syndromes. I. Problem, methodology. Eur Arch Psychiatry Neurol Sci 234: 13–20

Angst J, Vollrath M, Merikangas KR, Ernst C (1990) Comorbidity of anxiety and depression in the Zurich Cohort Study of young adults. In: Maser JD, Cloninger CR (eds) Comorbidity of mood and anxiety disorders. American Psychiatric Press, Washington DC, pp 123–153

Bland RC, Newman SC, Orn H (1988a) Lifetime prevalence of psychiatric disorders in Edmonton. Acta Psychiatr Scand 77 [Suppl 338]: 24–32

Bland RC, Newman SC, Orn H (1988b) Period prevalence of psychiatric disorders in Edmonton. Acta Psychiatr Scand 77 [Suppl 338]: 33–42

Bland RC, Newman SC, Orn H (1988c) Age of onset of psychiatric disorders. Acta Psychiatr Scand 77[Suppl 338]: 43–49

Bronisch T (1990) Dysthyme Störungen. Nervenarzt 61: 133–139

Bronisch T, Wittchen HU, Krieg JC, Rupp HU, von Zerssen D (1985) Depressive neurosis – a long-term prospective and retrospective follow-up study. Acta Psychiatr Scand 71: 237–248

Canino GS, Bird HR, Shrout PE et al (1987) The prevalence of specific psychiatric disorders in Puerto Rico. Arch Gen Psychiatry 44: 27–735

Cross-National Collaborative Group (1992) The changing rate of major depression. Cross-national comparisons. JAMA 268 (21): 3098–3105

Faravelli C, Degl'Innocenti BG, Aiazzi L, Incerpi G, Pallanti S (1990) Epidemiology of mood disorders: a community survey in Florence. J Affect Disord 20: 135–141

Fichter M M (1990) Verlauf psychischer Erkrankungen in der Bevölkerung. Springer, Berlin Heidelberg New York

Garrison CZ, Addy CL, Jackson KL, McKeowon RE, Waller JL (1992) Major depressive disorder and dysthymia in young adolescents. Am J Epidemiol 135(7): 792–802

Hautzinger M, de Jong-Meyer R (1994) Depressionen. In: Reinecker H (ed) Lehrbuch der Klinischen Psychologie. Hogrefe, Göttingen, pp 177–218

Katschnig H (1980) Methodische Probleme der Life Event Forschung. Nervenarzt 51: 332–343

*Kessler RC, McGonagle KA, Zhao S et al (1994) Lifetime and 12-month prevalence of DSM-III-R psychiatric disorders in the United States: results from the National Comorbidity Survey. Arch Gen Psychiatry 51: 8–19

Kessler RC, Nelson CB, McGonagle KA, Liu I, Swartz M, Blazer DG (1996) Comorbidity of DSM-III-R major depressive disorder in the general population: results from the National Comorbidity Survey. Br J Psychiatry 168[Suppl 30]: 17–30

Klerman GL (1988) The current age of youthful melancholia. Br J Psychiatry 152: 4–14

Knäuper B (1994) Depressionsdiagnostik im Alter. Verständnis und Verständlichkeit standardisierter diagnostischer Interviewfragen. Roderer, Regensburg

Lee CK, Kwak YS, Yamamoto J et al (1990a) Psychiatric epidemiology in Korea. I. Gender and age differences in Seoul. J Nerv Ment Dis 178(4): 242–246

Lee CK, Kwak YS, Yamamoto J et al (1990b) Psychiatric epidemiology in Korea. II. Urban and rural differences. J Nerv Ment Dis 178(4): 247–252

Lewinsohn PM, Hops H, Roberts RE, Seeley JR, Andrews JA (1993) Adolescent psychopathology. I. Prevalence and incidence of depression and other DSM-III-R disorders in high school students. J Abnorm Psychol 102(1): 133–144

Lin N, Dean A, Ensel WM (1986) Social support, life events and depression. Academic, Orlando

*Maser JD, Cloninger CR (eds) (1990) Comorbidity of mood and anxiety disorders. American Psychiatric Press, Washington DC

Mayer KU, Baltes PB (eds) (1996) Die Berliner Altersstudie. Akademie, Berlin

Merikangas K, Angst J, Eaton W et al (1996) Comorbidity and boundaries of affective disorders with anxiety disorders and substance abuse: results of an international task force. Br J Psychiatry 168[Suppl 30]: 49–58

Oldehinkel AJ, Wittchen HU, Schuster P (1999) Prevalence, 20-month incidence, and outcome of unipolar depressive disorders in a community sample of adolescents. Psychol Med 29: 655–668

Regier DA, Burke JD, Burke KC (1990a) Comorbidity of affective and anxiety disorders in the NIMH Epidemiologic Catchment Area Program. American Psychiatric Press, Washington DC

Regier DA, Farmer ME, Rae DS, Locke BZ, Keith SJ, Judd LL, Goodwin FK (1990b) Comorbidity of mental disorders with alcohol and other drug abuse. Results from the Epidemiologic Catchment Area (ECA) Study. JAMA 264(19): 2511–2518

Robins LN, Helzer JE, Ratcliff KS, Seyfried W (1982) Validity of the Diagnostic Interview Schedule, version II: DSM-III diagnoses. Psychol Med 12: 55–870

Robins LN, Helzer JE, Weissman MM, Orvaschel H, Gruenberg E, Burke JD Jr, Regier DA (1984) Lifetime prevalence of psychiatric disorders at three sites. Arch General Psychiatry 41: 949–959

Robins LN, Locke BZ, Regier DA (1991) An overview of psychiatric disorders in America. In: Robins LN, Regier DA (eds) Psychiatric disorders in America. The Epidemiologic Catchment Area Study. Free Press, New York, pp 328–366

Sartorius N, Nielsen JA, Strömgren E (eds) (1989) Changes in frequency of mental disorder over time: results of repeated surveys of mental disorders in the general population. Acta Psychiatr Scand 79[Suppl 348]

Shapiro S, Skinner EA, Kessler LG et al (1984) Utilization of health and mental health services. Arch Gen Psychiatry 41: 971–978

Shepherd M (1975) Epidemiologische Psychiatrie. In: Kisker KP, Meyer JP, Müller C, Strömgren E (eds) Psychiatrie der Gegenwart, vol 3: Soziale und angewandte Psychiatrie. Springer, Berlin Heidelberg New York

Stein MB, Tancer ME, Uhde TW (1990) Major depression in patients with panic disorder: factors with course and recurrence. J Affect Disord 19: 287–296

Swartz M, Landerman R, George LK, Blazer DG, Escobar J (1991) Somatization disorder. In: Robins LN, Regier DA (eds) Psychiatric disorders in America. The Epidemiologic Catchment Area Study. Free Press, New York, pp 220–255

Thompson AH, Bland RC, Orn HT (1989) Relationship and chronology of depression, agoraphobia, and panic disorder in the general population. J Nerv Ment Dis 177(8): 456–463

*Üstün TB, Sartorius N (1995) Mental illness in general health care: an international study. Wiley, Chichester

Vollrath M, Koch R, Angst J (1990) The Zurich Study. IX. Panic disorder and sporadic panic: symptoms, diagnosis, prevalence, and overlap with depression. Eur Arch Psychiatry Neurol Sci 239(4): 221–230

Wacker HR, Müllejans R, Klein KH, Battegay R (1992) Identification of cases of anxiety disorders and affective disorders in the community according to ICD-10 and DSM-III-R using the Composite International Diagnostic Interview (CIDI). Int J Method Psychiatr Res 2: 91–100

Weissman MM, Bruce ML, Leaf PJ, Florio LP, Holzer C (1991) Affective disorders. In: Regier DA, Robins LN (eds) Psychiatric disorders in America: the Epidemiologic Catchment Area Study. Free Press, New York, pp 53–80

Wells JE, Bushnell JA, Hornblow AR, Joyce PR, Oakley-Browne MA (1989) Christchurch Psychiatric Epidemiology Study. I. Methodology and lifetime prevalence for specific psychiatric disorders. Aust NZ J Psychiatry 23: 315–326

Wittchen HU (1987) Chronic difficulties and life events in the long-term course of affective and anxiety disorders: results from the Munich Follow-up Study. In: Angermeyer MC (ed) From social class to social stress – new developments in psychiatric epidemiology. Springer, Berlin Heidelberg New York, pp 176–196

Wittchen HU (1988) Natural course and spontaneous remissions of untreated anxiety disorders – results of the Munich Follow-up Study (MFS). In: Hand I, Wittchen HU (eds) Panic and phobias. 2. Treatments and variables affecting course and outcome. Springer, Berlin Heidelberg New York, pp 3–17

Wittchen HU (1993) Komorbidität bei Angststörungen – Häufigkeit, ätiologische und klinische Implikationen. In: Kasper S, Möller HJ (eds) Angst und Panikerkrankungen. Diagnose – Therapie. Socio Medico, Gräfelfing, pp 60–69

Wittchen HU (1994a) Reliability and validity studies of the WHO Composite International Diagnostic Interview (CIDI): a critical review. J Psychiatr Res 28(1): 57–84

Wittchen HU (1994b) Who becomes chronically depressed? WPA Teach Bull Depress 2(3): 1–2

*Wittchen HU (ed) (1996) Comorbidity of mood disorders. Br J Psychiatry 168[Suppl 30]

Wittchen HU, Essau CA (1989) Comorbidity of anxiety disorders and depression: does it affect course and outcome? J Psychiatry Psychobiol 4: 315–323

Wittchen HU, Essau CA (1993a) Comorbidity and mixed anxiety-depressive disorders: is there epidemiological evidence? J Clin Psychiatry 54(1): 9–15

Wittchen HU, Essau CA (1993b) Epidemiology of anxiety disorders. In: Michels R (ed) Psychiatry. Lippincott, Philadelphia, pp 1–25

Wittchen HU, Pfister H (1997) DIA-X-Interviews: Manual für Screening-Verfahren und Interview; PC-Programm zur Durchführung des Interviews; Auswertungsprogramm. Swets and Zeitlinger, Frankfurt am Main

Wittchen HU, Semler G (1990) Composite International Diagnostic Interview (CIDI), version 1.0. Beltz, Weinheim

Wittchen HU, von Zerssen D (eds) (1987) Verläufe behandelter und unbehandelter Depressionen und Angststörungen. – Eine klinisch-psychiatrische und epidemiologische Verlaufsuntersuchung. Springer, Berlin Heidelberg New York

Wittchen HU, Vossen A (1995) Implikationen von Komorbidität bei Angststörungen. Ein kritischer Überblick. Verhaltenstther Prax Forsch Perspekt 5(3): 120–133

Wittchen HU, Semler G, von Zerssen D (1985) A comparison of two diagnostic methods – clinical ICD-10 diagnoses vs DSM-III and research diagnostic criteria using the Diagnostic Interview Schedule, version 2. Arch Gen Psychiatry 42: 677–684

Wittchen HU, Hecht H, Zaudig M, Vogl G, Semler G, Pfister H (1987) Häufigkeit und Schwere psychischer Störungen in der Bevölkerung – Eine epidemiologische Feldstudie. In: Wittchen HU, von Zerssen D (eds) Verläufe behandelter und unbehandelter Depressionen und Angsstörungen. Springer, Berlin Heidelberg New York, pp 232–251

Wittchen HU, Essau CA, Krieg JC (1991) Anxiety disorders: similarities and differences of comorbidity in treated and untreated groups. Br J Psychiatry 159[Suppl 12]: 23–33

Wittchen HU, Essau CA, von Zerssen D, Krieg CJ, Hecht H (1992) Lifetime and six-month prevalence of mental disorders in the Munich Follow-Up Study. Eur Arch Psychiatry Clin Neurosci 241: 247–258

Wittchen HU, Knäuper B, Kessler RC (1994) Lifetime risk of depression. Br J Psychiatry 165[Suppl 26]: 16–22

Wittchen HU, Perkonigg A, Lachner G, Nelson CB (1998) The early developmental stages of psychopathology study (EDSP) – objectives and design. Eur Addict Res 4: 18–27

Wittchen HU, Schuster P, Pfister H, Müller N, Storz S, Isensee B (1999a) Nichtbehandelte Depressionen in der Allgemeinbevölkerung – Schlecht erkannt und selten behandelt. Nervenheilkunde 18: 202–209

Wittchen HU, Schuster P, Pfister H, Gander F, Müller N (1999b) Warum werden Depressionen häufig nicht erkannt und selten behandelt? Patientenverhalten und Erklärungswert des "Sisi-Syndroms". Nervenheilkunde 18: 210–217

WHO (1990) Composite International Diagnostic Interview (CIDI), version 1.0. World Health Organization, Geneva

Wunderlich U, Bronisch T, Wittchen HU (1998) Comorbidity patterns in adolescents and young adults with suicide attempts. Eur Arch Psychiatry Clin Neurosci 248: 87–95

CHAPTER 18

W. Maier, S. Schwab, M. Rietschel

Genetics of the Affective Disorders

Translator: E. Taub

1 Introduction

The causation of the affective disorders is the subject of much speculation, but little secure knowledge. The most reliable and replicable findings on this topic to date are the following:

- Affective disorders occur with increased frequency in patients' relatives as compared with the population at large, which implies that familial factors play a role in their causation.
- Concordance rates for affective disorders are higher in monozygotic than in dizygotic twins, which implies that these familial factors are at least partly genetic.

Molecular genetic methods that have been available for the past decade now enable the study of single genetic mutations as putative causative factors for genetically determined disorders. More recently, it has also become possible to identify genomic candidate regions that may contain genes influencing the occurrence of affective disorders. The extremely rapid development of molecular genetic methods has given genetic research on the causation of mental illness a decisive push forward.

In this chapter, the classical methods of family genetics that do not rely upon genetic markers, i.e. family, twin, and adoption studies, will be discussed first. These methods serve to establish the fact of familial occurrence, identify familially determined subtypes, and test the relevance of putatively influential genetic factors. The elements of molecular genetic technique will be presented next, followed by a discussion of the present state of molecular genetic research in the affective disorders.

2 Familial Genetic Determination

All studies performed on the subject have revealed that affective disorders occur more frequently in patients' relatives than in the population at large. Table 1 shows the more recent family studies. These studies meet current methodological standards, i.e.:

- The majority of living first-degree relatives were personally studied with standardized interviews and compared with control subjects not affected by mental illness.
- All diagnoses were made in accordance with one of the current diagnostic manuals, which enable operational diagnosis based on well-defined criteria.

The familial lifetime risk of developing an affective disorder, i.e. the risk that a patient's first-degree relative will develop such a disorder at any time of life, varies considerably among studies, with figures ranging from 7% to more than 31%; a similarly wide variation of the lifetime risk of unipolar depression has been found in samples of the general population and in groups of control subjects (0%–23%). Patients with bipolar disorders have a stronger familial loading of affective disorders (of all types, considered as a single group) than patients with unipolar depression. The variation of the lifetime risk of unipolar depression is partly attributable to variable methods of identifying cases (for an overview, see Maier and Lichtermann 1993).

As shown in Table 2, the concordance rates for monozygotic twins are higher than those for dizygotic twins, which implies an at least partly genetic determination of the disease. The degree of heritability (i.e. the percentage of variance attributable to genetic factors) is 60%–80% for bipolar affective disorders and 28%–52% for unipolar depression. The latter figure varies depending on the definition used; maximal heritability is found when the definition requires an episode of unipolar depression to last a minimum of 2 weeks (Kendler et al. 1992b). A few studies support the hypothesis that heritability is lower for less severe degrees of unipolar depression (Torgersen 1986).

The concordance rate among monozygotic twins is 50% for unipolar depression and 80% for bipolar affective disorder. The fact that both of these figures lie under 100% indicates that nongenetic etiological factors are also at work. Putative nongenetic, or at least partly nongenetic, factors under discussion include perinatal complications (particularly for bipolar disorders; Kinney et al. 1993) and critical life events, as well as familial factors such as modes of upbringing, parental neglect, and parental conflict (mainly for unipolar depression; McGuffin et al. 1988; Angst and Wicki 1990; Murray and Sines 1996). The unusually large Virginia Twin Study revealed an interaction between genetic etiological factors and specific environmental factors. The environmental risks were not equally distributed; instead, it was found that the patients at elevated genetic risk for unipolar depression were more exposed to environmental risks and either sought out these risk-enhancing environmental conditions (critical life events) or contributed to them (Kendler and Karkowski-Shuman 1997).

All subtypes of affective disorder appear more commonly in relatives of patients. There are also subtype-specific patterns of familial occurrence, as will be discussed below.

Table 1. Age-adjusted lifetime prevalence of affective disorders in first-degree relatives, compared to control subjects from the general population

Diagnosis	Study	Lifetime prevalence of the illness of the index patient in first-degree relatives (%)	Lifetime prevalence in the general population (%)
Bipolar disorder	Gershon et al. (1988) RDC/DSM-III	7.2	0.3
	Maier et al. (1993) RDC/DSM-III-R	7.0	1.8
Unipolar depression	Gershon et al. (1988) RDC/DSM-III	16.7	6.7
	Kendler et al. (1993c) DSM-III-R	31.1	22.8
	Maier et al. (1993) RDC/DSM-III-R	21.6	10.6
	Winokur et al. (1995a) RDC	10.4	4.9

RDC, Research Diagnostic Criteria.

Table 2. Results of recent twin studies

Diagnosis of the index patient	Study	Pairs of twins studied (*n*)		Concordance rates for the illness of the proband (%)	
		MZ	DZ	MZ	DZ
Bipolar disorder	Bertelsen et al. (1977) ICD-7	32	37	79	19
Unipolar depression	Torgersen (1986) DSM-III	28	46	25%	11%
	McGuffin et al. (1991) DSM-III	62	79	53%	28%
	Kendler et al. (1992b) DSM-III	FGP	FGP	49%	42%
	Kendler et al. (1993a) DSM-III-R	*FGP	FGP	48%	42%

MZ, monozygotic; DZ, dizygotic; FGP, twin study in the female general population, not selected by diagnosis (MZ, 510; DZ, 440).

3 Diagnostic Subclassification of the Affective Disorders

3.1 Unipolar Depression Versus Bipolar Illnesses

The subclassification of affective illnesses into unipolar depressive and bipolar manic-depressive illnesses is highly relevant to patterns of familial occurrence and genetic determination. The great majority of published family studies document a major elevation of the lifetime risk of bipolar disorders in first-degree relatives of patients with a bipolar affective illness (3%–10%), while the lifetime risk of bipolar affective illnesses in relatives of unipolar depressive patients is merely 0.5%–2%, which is not significantly different from that of normal controls (0.5%–1%) (Angst 1966; Gershon et al. 1988; Maier et al. 1993; Winokur et al. 1995b); the relative risk (quotient of lifetime risks) is therefore approximately 7.

This relative diagnostic specificity of increased familial occurrence is found particularly in bipolar subtype I, in which manic episodes are fully developed (Winokur et al. 1995a). Among relatives of such patients, the risk of developing the disease is the same for both sexes (as it is in the general population). It was recently found in a sample of families with a strong history of affective disorders that bipolar affective disorder is preferentially transmitted by the mother of the index patient (McMahon et al. 1995), although this finding has been disputed (Kato et al. 1996).

In families of patients with fully developed bipolar affective disorder, all other variants of bipolar disorder are found more commonly than in the general

population: cyclothymia and hypomania, which are characterized by frequently recurring, but less than fully developed manic and/or depressive episodes (Winokur et al. 1969), and schizoaffective disorders with manic episodes (Maier et al. 1992a). The frequently replicated increased familial incidence of bipolar disorders indicates the validity of the diagnostic differentiation of affective disorders into two types, as found in the currently accepted diagnostic manuals (ICD-10, DSM-IV).

Unipolar depression also occurs more frequently in patients' relatives: while the lifetime prevalence of unipolar depressive episodes (major depression) in the general population is approximately 10%, according to recent epidemiological studies, the comparable figure for first-degree relatives of patients with unipolar depressive episodes is approximately 20% (relative familial risk, approximately 2). Families of patients with unipolar depressive episodes are also subject to a higher-than-normal frequency of less severe or shorter-lasting forms of unipolar depression (including recurrent, brief depression) (Maier et al. 1992b). In such families, as in the general population, all variants of unipolar depression occur approximately twice as frequently in women as in men. As male patients with depressive episodes do not have a stronger family history than female patients, these findings cannot be the result of the so-called Carter effect (sex-specific threshold values in diseases of polygenic etiology, e.g. pyloric stenosis; Merikangas et al. 1985a).

The familial transmission of the unipolar depressive subtype, unlike that of the bipolar affective subtype, is not highly specific. The lifetime risk of unipolar depression is elevated even in relatives of patients with bipolar disorders and is approximately 20%, just as in relatives of unipolar depressives. As far as is known at present, patients with unipolar depression in the familial genetic context of bipolar affective disorders have no distinguishing clinical features (Blacker et al. 1993). Particularly high rates of unipolar depression and suicidal behavior are found in adolescents with a parent suffering from a bipolar disorder (Merikangas 1993). Although bipolar disorders are not often seen before age 18 in the children of parents with bipolar affective disorders, a large fraction of cases of unipolar depression in these children are later transformed into bipolar affective disorder when manic episodes appear. Bipolar affective disorders do not occur more frequently in the families of patients with unipolar depression than in the general population. An analogous specificity for the relatives of patients with unipolar depressive illnesses cannot be demonstrated; the relatives of patients with bipolar affective disorders, as well as those of unipolar depressives, have an increased frequency of unipolar depression, and the lifetime risk in the two groups is practically identical (Gershon et al. 1988; Maier et al. 1993). As shown in Table 2, both subtypes, but particularly the bipolar affective disorders, are subject to genetic influences. As with bipolar depressive syndromes among the families of bipolar depressives, the families of unipolar depressives have an increased frequency of unipolar depressive syndromes of both greater and lesser degrees of severity – the more severe types including depressive episodes with melancholic traits, or somatic syndromes according to ICD-10 and DSM-III-R/IV, the less severe including "minor depression" according to the Research Diagnostic Criteria (RDC), dysthymia according to ICD-10 or DSM-III-R/IV, and repeated brief depressive episodes.

3.2 Subtypes of Bipolar Affective Disorders

The differentiation of bipolar affective disorders with fully developed manic episodes (bipolar I) from those with exclusively less pronounced, so-called hypomanic episodes (bipolar II, if these episodes alternate with depressive episodes) shows long-term interindividual stability (Coryell et al. 1995). Both subtypes occur more commonly in particular families (Endicott et al. 1985). Several studies have shown a surprisingly high subtype specificity of the pattern of familial occurrence of the bipolar II subtype: relatives of patients bearing this diagnosis had a higher risk than relatives of normal controls for the development of bipolar II disorders and unipolar depression, while their risk of developing bipolar I disorders was less markedly increased (see, e.g. DePaulo et al. 1990; Heun and Maier 1993a,b).

Frequent recurrence of the affective disturbance has been repeatedly discussed as a possible indication of familial, genetically transmitted etiology (see, e.g. Goodwin and Ghaemi 1998). This hypothesis is supported by prospective studies of the course of bipolar affective disorders (Winokur et al. 1994). Nonetheless, studies of the families of patients with frequently recurring bipolar affective disorder ("rapid-cycling" type) have repeatedly failed to confirm this hypothesis (Nurnberger et al. 1988; Coryell et al. 1992b; Lish et al. 1993): the relatives of such patients have neither bipolar affective disorders nor unipolar depression more frequently than the relatives of patients with bipolar affective disorders of other types. It is possible, however, that the relatives of rapidly cycling patients are at elevated risk for substance abuse (Lish et al. 1993).

Whether or not an association with rapid cycling can be confirmed, there does appear to be a secure associ-

ation with a positive response to the phasic prophylactic effect of lithium: several studies have shown that a positive response is associated with a stronger family history of bipolar affective disorders (see, e.g. Mendlewicz et al. 1973; Grof et al. 1994), while lack of response may be associated with a stronger family history of schizophrenia (Grof et al. 1994).

3.3 Subclassification of Unipolar Depression

According to a classical hypothesis, reactive or neurotic depression is more likely to occur sporadically, while endogenous depression and particularly bipolar affective disorders are associated with a strongly positive family history (Maier and Philipp 1993). More recent research shows that this hypothesis is probably not correct. None of the more recent family studies employing the current, operationalized diagnostic classifications (i.e. those of the RDC, DSM-III-R, and ICD-10) have shown a higher incidence of depressive illnesses in the relatives of patients with endogenous unipolar depression than in those of patients with nonendogenous unipolar depression.

A further hypothesis, that of subtype-specific homogeneity of the pattern of familial increased occurrence, has also failed to be confirmed. The relatives of patients with endogenous depression, for example, were found to be at increased risk not only for endogenous, but also for nonendogenous depression (Maier and Philipp 1993; Fanous et al. 1996). The same holds true for all of the possible definitions of endogenous or melancholic subtypes of unipolar depression that have been tested to date (RDC, DSM-III-R, ICD-10, Newcastle scales). Known risk factors for unipolar depression, such as elevated neuroticism values and critical life events, are themselves, in turn, subject to genetic influences.

Alongside major depression, the dysthymias are a second common subtype of unipolar depression. Major depression and the dysthymias are both associated with an elevated risk of depressive episodes, recurrent depression, and chronic depression in relatives; this fact indicates that they may share a common etiology, at least in part. Nonetheless, dysthymias, and particularly early-onset dysthymias, are found at significantly increased frequency in the families of patients with early-onset dysthymias, which implies that this subtype may be etiologically distinct (Klein et al. 1995; Donaldson et al. 1997). The combination of both diagnoses ("double depression") is also more commonly found in relatives of patients with this combination (Donaldson et al. 1997).

3.4 Affective Disorders with Psychotic Features and Schizoaffective Disorders

The familial association of affective and psychotic disorders is controversial. No controlled study has shown a significantly elevated risk of developing schizophrenia among relatives of patients suffering from affective disorders without psychotic features (Kendler et al. 1996). Nonetheless, the majority of controlled studies have shown that the relatives of patients suffering from affective disorders *with* psychotic features have an elevated prevalence not only of affective disorders, but also of all types of psychotic disorder (including affective disorders with psychotic features, and schizophrenia). Male relatives of female patients with schizoaffective disorders are at particularly increased risk (Goldstein et al. 1993).

Intrafamilial similarity is usually observed: the more severe the psychotic manifestations are in the index patient, the more likely mental illness will occur in relatives, and the more severe it is likely to be; but the more severe the affective disturbance is in the index patient, the less marked the psychotic manifestations are in the relatives. In general, bipolar disorders with psychotic episodes are more commonly found in relatives of index patients who also suffer from bipolar disorders with psychotic episodes, and less commonly in relatives of index patients suffering from unipolar depression with psychotic features (Maier et al. 1992a).

3.5 Age at Onset of Disease

The age at onset of bipolar affective disorders is usually between 18 and 30 years. Because this range is rather narrow, it is hard to make any firm and replicable statements concerning a possible association between the age at onset of the disease and family history of this or related diseases; furthermore, a well-documented secular trend in the age at onset of the disease (Gershon et al. 1987) makes it still more difficult to demonstrate any such association, if one exists. Unipolar depression, in contrast, may begin at any age, from childhood to advanced age. Several family studies have shown that the increase in the frequency of affective disorders among relatives of index patients with unipolar depressive episodes is inversely proportional to the age at onset of disease in the index patient. In particular, patients who had become ill before the age of 40 had a strongly positive family history, while patients with senile depression (age at onset greater than 60 years) had a negative or only weakly positive family history of affective

disorders (Maier et al. 1991). A number of studies have also shown a weakly positive intrafamilial correlation of the age of onset of unipolar depression (Weissman et al. 1984, 1986; Bland et al. 1986).

As mentioned, the age at onset of affective disorders is subject to a strong secular trend: all over the world, the age at onset of both unipolar depression and bipolar affective disorders is steadily decreasing (Gershon et al. 1987; Cross-National Collaborative Group 1992). Hypothetically, this might be explained by an increasing effect of nongenetic etiological factors. Within families with strongly positive family histories, a further secular trend may be observed: the reported age at onset of the disease decreases with each generation. If early onset of the disease can be regarded as a marker of severity, then this constellation of findings is consistent with the well-known genetic mechanism of anticipation (see below; Penrose 1948). It is generally interpreted in this way.

It is conceivable, however, that this intrafamilial downward shift of the age at onset of disease is merely a methodological artifact (Hodge and Wickramaratne 1995) resulting from retrospective ascertainment, faulty memory, the reduced availability of elderly relatives for questioning, and generation-specific stratification effects. The fact that a small number of families show a secular trend in the opposite direction seems to make this type of artifact less likely to be the cause of the phenomenon (Grigoroiu-Serbanescu et al. 1997).

4
Cosegregating Disorders and Traits

4.1
Nonaffective Disorders Cosegregating with Unipolar Depression

Families of patients with affective disorders have a higher frequency not only of affective disorders, but also of nonaffective disorders. Several family studies have shown that anxiety disorders (Kendler et al. 1992a; Weissman et al. 1997) and alcohol dependence (Winokur et al. 1971; Winokur and Coryell 1991) occur at increased frequency in families of unipolar depressive patients, without themselves being sequelae of unipolar depression. Above all, unipolar depression and anxiety disorders tend to occur in the same families more often than would be predicted by chance (Merikangas et al. 1985b; Weissman et al. 1997). Moreover, different subtypes of anxiety disorder have different patterns of cosegregation. Panic disorders are transmitted independently, or nearly independently, of unipolar depression (Weissman et al. 1993; Maier et al. 1995a). Unipolar depression and generalized anxiety disorder, however, seem to be alternative expressions of a single genetic predisposition: it has been shown in two groups of twin pairs that the genetic causative factors of depressive disorders and generalized anxiety disorders overlap to a considerable extent (Kendler et al. 1992a). The phenotypic differentiation of the two disorders depends primarily on the specific environmental conditions affecting the individual rather than on genetic factors.

The putative familial/genetic connection between unipolar depression and alcohol dependence and abuse is controversial (Merikangas et al. 1985b; Coryell et al. 1992a; Maier et al. 1994). A familial connection between a subtype of alcoholism in men and a subtype of unipolar depression (so-called spectrum depression) in women was postulated by Winokur, but could not be confirmed in the majority of subsequent studies. Family studies do, however, show that patients with chronic depression of early-onset type (dysthymia, particularly subaffective dysthymia) have a more strongly positive family history of alcoholism than patients with other subtypes of affective disorder (Anderson et al. 1996). A twin study employing younger female subjects revealed a moderately elevated cosegregation of unipolar depression of early onset and alcoholism, which implies a moderately strong correlation of the genetic risk factors for these entities (Kendler et al. 1992c, 1993d).

The familial relationship between unipolar depression and nicotine dependence has also been investigated. A twin study (Kendler et al. 1993b) revealed that these two disorders occur in the same families more frequently than would be predicted by chance. This finding implies that there are genetic risk factors common to the two disorders, which manifest themselves alternatively (or simultaneously) as unipolar depression and nicotine dependence.

The putative familial relationship between unipolar depression and eating disorders is controversial. The most recent family study (Lilenfeld et al. 1998), in contrast to earlier studies (Strober et al. 1990), yielded no evidence to support such a relationship.

The first-degree relatives of patients with unipolar depression of any clinical subtype have an elevated frequency of clinical personality disorders (Coryell and Zimmermann 1989; Maier et al. 1992c; Klein et al. 1995). This association is due in part to comorbidity of depression and personality disorders in the index patient. In some subtypes of unipolar depression, however, the association is particularly strong, implying the existence of common familial risk factors. This is true especially of the association between dysthymia and all of the cluster B personality disorders (borderline, antisocial, histrionic, and narcissistic personality disorders) (Klein et al. 1995; Riso et al. 1996). The hypothesis that borderline personality disorders (or

the subtype of borderline personality disorder characterized by lability of affect) are variants of affective disorders, for which most of the earlier evidence was derived from sleep-polygraphic studies (Akiskal et al. 1980), thus obtains limited further support from family genetic studies.

4.2 Nonaffective Disorders Cosegregating with Bipolar Disorders

Family studies have been performed to investigate the putative familial association of bipolar affective disorders with alcoholism, anxiety disorders, and childhood hyperkinetic syndrome (Winokur et al. 1993; Maier et al. 1995b; Wozniak et al. 1995; MacKinnon et al. 1997). All of these studies revealed that the latter disorders are transmitted largely independently of bipolar affective disorders, despite the occurrence of comorbidity more frequently than would be predicted by chance. It was also observed that the occurrence of two disorders in the index patient is associated with comorbidity of both disorders in relatives more frequently than would be predicted by chance. Thus comorbidity of bipolar disorders with the other illnesses listed seems to define an etiologically distinct subtype of bipolar disorder. Moreover, there is discussion over a possibly increased frequency of bipolar disorders in the biological relatives of children with fragile-X syndrome and Gilles de la Tourette syndrome (Jeffries et al. 1993; Kerbeshian et al. 1995).

The pattern of familial association of unipolar depression with other diseases is less diagnosis specific than that of bipolar affective disorders (with particular reference to nonaffective disorders). The morbidity of all types in families of unipolar depressive patients is, however, no greater than that in families of patients with bipolar affective disorders.

4.3 Cosegregating Personality Traits

A number of personality traits occurring more commonly in depressed patients also occur more commonly in biological relatives of depressed patients who have never themselves suffered from any mental disorder. Such individuals at risk have been found to have an elevated mean value for neuroticism (Maier et al. 1992c; Kendler et al. 1993a; Lauer et al. 1997; less marked: Ouimette et al. 1996), which was also found to be a risk factor for the later appearance of depressive disorders. Prospective twin studies in the general population have revealed that the genetic determination of unipolar depression is partly mediated by the genetic determination of the personality dimension of neuroticism or critical life events (Kendler et al. 1993a).

In addition to neuroticism, rigidity has also been found more frequently in the healthy relatives of unipolar depressives (Maier et al. 1992c; Lauer et al. 1997). The personality traits of healthy relatives of patients with bipolar affective disorders have not been studied as extensively. An elevated average value for rigidity, but not neuroticism, was found in a recent study (Maier et al. 1995c) as well as in the earlier study by Klein and Depue (1985).

Several aspects of personality subsumed under the concept of "temperament" may have a genetic relationship to the affective disorders. This term denotes, among other things, the basic energetic conditions for emotional behavior. Temperamental factors are strongly genetically determined, in contrast to character factors, which are subject to the influences of family environment and upbringing. Certain temperamental factors are under consideration as possible genetically determined risk factors (also termed "endophenotypes" or "intermediate phenotypes") that may confer a large part of the genetic risk for affective disorders. Temperamental factors such as depressivity and "harm avoidance" (the latter strongly correlated with neuroticism) have been discussed as possible intermediate phenotypes for unipolar depression, and hyperthymic temperament as a possible intermediate phenotype for bipolar disorders (Cassano et al. 1992); these relationships are implied by elevated mean values of these factors in the (still) healthy relatives of patients with these disorders. Analogous relationships were postulated long ago by Kraepelin (1921) and Kretschmer (1936). A further discussion of personality factors is found in Chap. 20 (Vol. 3, Part 1).

4.4 Cosegregating Normal Biological Variants (Vulnerability Markers)

Pathophysiological correlates of unipolar and bipolar depression also occur more frequently in the relatives of depressed patients who have never suffered from any mental illness themselves:

- Sleep patterns characteristic of depression (Lauer et al. 1995) – elevated "rapid eye movement" (REM) density, lesser percentage of "slow-wave sleep" – are more common in healthy individuals with a family history of depression than in the general population.
- Cholinergic supersensitivity: A greater than normal shortening of REM latency in response to the administration of cholinomimetic agents is seen in healthy individuals with a family history of affective disorders more commonly than in the general

population (Schreiber et al. 1992). This finding supports the hypothesis that a familial, genetically transmitted adrenergic or cholinergic imbalance is a marker of vulnerability to affective disorders (Janowsky et al. 1994).
- The healthy children of patients with affective disorders (unipolar-depressive and bipolar), like depressed patients themselves, react to certain stresses with overactivity of the autonomic nervous system; the children's reactions are, however, less intense (Zahn et al. 1989).
- The intensification of depression and dysphoria that may be induced in depressed patients by tryptophan depletion is also seen in the healthy relatives of depressed patients; this finding implies that the serotonergic system is dysfunctional not only in depressed patients, but also in individuals with an elevated risk of developing the disease (Benkelfat et al. 1994).
- Nonspecific disturbances of slow ocular pursuit appear to be present at higher than normal frequency in the healthy children of individuals suffering from affective disorders (Rosenberg et al. 1997). In addition, prolongation of the P300 latency, which is more frequently observed in patients with affective disorders, is considered a possible vulnerability marker cosegregating with bipolar affective disorders (Blackwood et al. 1996a).

These normal biological variants perhaps represent prodromal states of affective disorders that will later become clinically overt or, alternatively, attenuated forms of disease that are not severe enough to be clinically manifest because of favorable environmental factors, despite the presence of genetic risk factors. Follow-up studies are needed to determine which, if either, of these explanations is correct.

5 Affective Disorders as Genetically Complex Disorders

In monogenic diseases, a single gene or several variants at a single genetic locus are responsible for the occurrence of a disease. If this is the case, the gene is referred to as the causative gene. Such diseases are transmitted, according to the mendelian rules, in either a dominant or a recessive pattern. The observed patterns of familial occurrence of the affective disorders are, however, not compatible with mendelian inheritance.

In the polygenic transmission model, genetic variants at multiple loci may contribute to the severity of the disease phenotype. These "risk-modulating" genes may have cumulative (additive) or synergistic (mutually potentiating) effects, or their effects may be physiologically interchangeable. In this model, an abnormality of any single gene is neither necessary nor sufficient to produce the disease, and there is no single causative gene; rather, each of the contributing genes incrementally affects the probability of the individual's developing the disease, and they are therefore referred to as "susceptibility genes" (Greenberg 1993). It is thus entirely conceivable that such genes may be found in individuals not suffering from the given disease who lack the additional factors that would be necessary to produce the disease (i.e. further susceptibility genes or environmental influences). If both environmental and (poly)genetic factors are important for the generation of the disease, its transmission is said to be multifactorial. There may also be interactions among genetic factors or between genetic factors and nongenetic environmental factors.

Family studies have shown that no simple mode of inheritance can account for the pattern of familial occurrence of affective disorders:

- Biometric analyses of the pattern of familial transmission of bipolar affective disorders imply that monogenic transmission is unlikely. Several segregation studies suggest that some genes may account for a significant portion, though not all, of the genetic variance ("major genes") (Rice et al. 1987; Spence et al. 1995), but other segregation analyses make this seem implausible (Craddock et al. 1995c, 1997). Without using genetic markers, however, segregation analyses cannot adequately differentiate between competing complex models of transmission. The linkage analyses performed to date for the bipolar affective disorders indicate that genes influencing the risk of disease are located in several regions (see below). This finding suggests that the transmission is polygenic and that each of several genes contributes only a small amount to the total (thus a variant of complex genetic inheritance). The increased occurrence of unipolar depression in the families of patients with bipolar affective disorder may thus indicate a partial commonality of the genes contributing to the genesis of both types of disorder. There is, however, as yet no verified hypothesis concerning the mode of complex genetic transmission.
- Bipolar disorders and unipolar depression have imperfect concordance (well under 100%) in pairs of monozygotic twins, which implies that nongenetic factors have an influence on the occurrence of the disease.
- Biometric segregation analyses have determined that the patterns of inheritance of bipolar disorders and of unipolar depression are incompatible with

monogenic transmission (see above). Nonetheless, genetic heterogeneity of these disorders cannot be ruled out; it may be that they are indeed transmitted, in some of the multiply affected families, by a single, causative gene, in accordance with the rules of mendelian inheritance. It is also possible that different combinations of susceptibility genes are responsible for the development of illness in different families.

- A progression of disease severity across the generations may be observed in families with multiple cases of bipolar affective disorders (see below). The age at onset of the disease is taken as the most important indicator of severity (McInnis et al. 1993). This progressive pattern in affected families is hypothetically attributed to an anticipation mechanism (see below).
- The probability of familial occurrence may depend on the sex of the transmitting parent, particularly in affective disorders (see, e.g. McMahon et al. 1995), where the disorder is more likely to be transmitted maternally. This has been shown for families strongly affected by bipolar disorders (McMahon et al. 1995) and for unipolar depression in the families of patients treated for depression (Keller et al. 1986). In view of the fact that mothers influence familial environmental factors more strongly than fathers, it may be that this sex-specific mechanism of transmission is due to nongenetic factors.

A number of plausible human genetic models have been developed to explain complex genetic transmission (Lander and Schork 1994):

- Incomplete penetrance: some individuals bearing the mutation that leads to a given phenotype may not, in fact, possess that phenotype. A mutation of the BRCA1 gene, for example, leads to the development of breast cancer by age 80 in no more than 85% of women.
- Phenocopy: some individuals bear the phenotype, but not the mutation; thus the phenotype may be produced by other factors, e.g. nongenetic environmental factors.
- Genetic heterogeneity: different mutations at different genetic loci might produce the same phenotype. Mutations at different loci might thus lead to different patterns of familial occurrence: the different mutations giving rise to early-onset familial Alzheimer's disease (Van Broeckhoven 1995) are an example of this phenomenon. Alternatively, there may be allelic heterogeneity, i.e. different mutations of the same gene may produce the phenotype, e.g. β-amyloid precursor protein (APP) mutations (Fidani et al. 1992) and presenilin mutations in early-onset familial Alzheimer's disease (Van Broeckhoven 1995). In monogenic disorders, which are very rare, allelic heterogeneity generally does not induce a complex pattern of familial occurrence.
- Polygenic transmission: the simultaneous presence of several mutations at different genetic loci may be required for the production of a given phenotype. Each one of the genes involved makes only a small contribution to the production of the phenotype, and a mutation in any individual gene is not necessarily required. Genes that play a more important role than others are called major genes. Genetically determined quantitative phenotypic variations are generally the result of polygenic transmission (e.g. hypertension).
- Anticipation: most of the known mutations are stable and are transmitted according to the classical mendelian rules. There are, however, so-called dynamic mutations whose base-pair sequence changes during mitosis, with the result that parents and children may have different mutations. Multiply repeated base-pair sequences (usually triplets, such as CAG) with variable numbers of repeats are an example. Such mutations are stably transmitted when the number of repeats lies below a certain threshold; however, when there are more than this number of repeats, the number may change – usually with an increase ("expansion") – during mitosis. Unstable mutations have been found to date in several disease-conferring genes, e.g. Huntington's disease (Huntington's Disease Collaborative Research Group 1993) and fragile-X syndrome (Verkerk et al. 1991). The number of repeats beyond the critical threshold for the disease is correlated with disease severity. Thus so-called full mutations of the FMR-1 gene locus (more than 150 CGG repeats) lead to the full development of fragile-X syndrome in boys, with mental debility and severe behavioral disturbances; the presence of 60–150 repeats ("premutation") leads to reduced intelligence and mild behavioral abnormalities in adult men; and alleles with fewer than 60 repeats are stably transmitted. The progression of disease severity (i.e. progressive reduction of age at onset of disease; see above) in families affected with bipolar disorders over several generations has given rise to the hypothesis that these disorders, too, are produced by dynamic mutations (Petronis and Kennedy 1995).

6 Strategies for the Identification of Genetic Variants of Mental Disorders

For monogenic inherited diseases, there is a well-charted path to the discovery of the causative gene,

Table 3. Comparison of linkage and association analysis

	Linkage analysis	Association analysis
Purpose of investigation	Cosegregation, i.e. common transmission of "gene" and disorder in multiply affected families	More frequent occurrence of an allele in the presence of a disorder
Sample	Multiply affected families over several generations	Independent patients and independent control subjects
Strategy	Genome scan, candidate genes	Candidate genes
Genetic markers	Highly polymorphous; distributed throughout the genome with the shortest possible intervening distances	If possible, only biallelic markers within or immediately adjacent to candidate genes; polymorphism should be expressed, if possible, as functional variation
Product of investigation	Candidate region	Candidate gene or linkage imbalance
Dependence on sample ("ascertainment bias")	Low	High
Sensitivity for:		
Genes responsible for a high percentage of the variance	High	Less high
Genes responsible for a low percentage of the variance	Very low	Low

beginning with gene localization by means of linkage studies in different families and ending with the demonstration of mutations in the gene causing the illness. The causative gene is localized to a given chromosomal segment of the genome by means of linkage analyses. Once such a segment is identified, it is subdivided into smaller DNA fragments. These fragments are then searched for genes, which, in turn, are searched for mutations in patients with the disease (e.g. by sequence analysis). If corresponding changes are found only in patients with the disease and not in healthy subjects, then the gene in which they are found may be the causative gene for the disease.

In principle, the same route may be taken to determine the causative genes of more complex diseases, such as the psychiatric diseases under discussion here. It is much more difficult, however, to find genetic abnormalities associated with these diseases (see, e.g. Baron 1996), for the following reasons:

- The mode of inheritance is non-mendelian and not yet determined.
- The phenotype transmitted in families is not clearly defined, but rather has unsharp boundaries and may be heterogeneous.
- Complex disorders are usually common illnesses, so phenocopies may be common (i.e. an identical clinical picture may be the result of nongenetic causes; see above).
- The disease may be genetically heterogeneous.
- Nongenetic environmental factors may play a significant role.
- Common genetic variants may increase an individual's risk of developing the disease, while the commonness of these variants has the effect of reducing the information available to the investigator (Lander and Schork 1994).

There are two ways to search for disease-associated genetic changes at the level of the genome: association studies and linkage studies. Both approaches are used to search for susceptibility genes for complex illnesses. Both start from the assumption of a categorical designation of patients as having the disease. A side-by-side comparison of the most important aspects of the two approaches is given in Table 3. A more extensive discussion of these methods is found in Chap. 4 (Vol. 1, Part 1).

7 Results of Genetic Marker Studies

7.1 Linkage Studies

The results of initial linkage studies using DNA markers were published as early as 1987 (Table 4); these studies were of a large Old Order Amish family (Egeland et al. 1987; Kelsoe et al. 1989). Bipolar affective disorders in this family were found to be

Table 4. Genetic linkage studies in bipolar affective disorders

Localization	Study	Findings
Dopamine D_3 receptor	Mitchell et al. (1993)	–
Dopamine D_5 receptor	Byerley et al. (1994)	–
Chromosome 4p	Blackwood et al. (1996b)	+
	M.M. Nöthen et al. unpublished	+
Chromosome 5q	Jensen et al. (1992)	–
	Curtis et al. (1993a)	–
	Mirow et al. (1994)	–
	Shah et al. (1995)	+
Chromosome 7 (DOPA decarboxylase gene, DDL) (7p11-p13)	Ewald et al. (1995a)	–
Chromosome 9 (dopamine β-hydroxylase)	Ewald et al. (1994a)	–
Chromosome 11		
11p (INS, HRAS)	Egeland et al. (1987)	+
11p (INS, HRAS)	Kelsoe et al. (1989)	–
11q (DRD_2)	Holmes et al. (1991)	–
11p (TH, HRAS, INS)	Mendlewicz et al. (1991)	–
11p (TH, INS, HRAS)	Mitchell et al. (1991)	–
11p (INS, HRAS)	Nanko et al. (1991)	–
11p	Pauls et al. (1991)	–
11p (TH)	Byerley et al. (1992)	–
11p (INS, HRAS)	Law et al. (1992)	–
11q (DRD_2)	Mitchell et al. (1992)	–
11q	Kelsoe et al. (1993)	–
11p (TH, DRD_4), 11q (DRD_2)	De Bruyn et al. (1994)	–
11p (TH), 11q (DRD_2)	Ewald et al. (1994d)	–
11p (TH, DRD_4)	Sidenberg et al. (1994)	+
11q	Ewald et al. (1995c)	–
11p (TH)	Kawada et al. (1995a)	–
Chromosome 12 (Darier disease region) (12q23–q24.1)	Craddock et al. (1994a)	+
	Ewald et al. (1994b)	–
	Dawson et al. (1995b)	+
	LaBuda et al. (1996)	+
Chromosome 15 (Prader-Willi region) (15q11-q13)	Ewald et al. (1994c)	–
Chromosome 16p (phosphoglycolate phosphatase)	Eiberg et al. (1993)	+
	Ewald et al. (1995d)	+
	LaBuda et al. (1996)	–
	Adams et al. (1997)	–
	Edenberg et al. (1997)	+
Chromosome 17 (serotonin transporter, 5-HTT)	Kelsoe et al. (1996)	–
Chromosome 18	Berrettini et al. (1994)	+
	Maier et al. (1995d)	–
	Pauls et al. (1995)	–
	Stine et al. (1995)	+
	Coon et al. (1996)	+
	De Bruyn et al. (1996)	+
	Freimer et al. (1996)	+
	Gershon et al. (1996)	+
	Ginns et al. (1996)	–
	LaBuda et al. (1996)	–
	McInnes et al. (1996)	+
	Nöthen et al. (1996)	+
	Claes et al. (1997)	–
	Detera-Wadleigh et al. (1997)	–
	Ewald et al. (1997)	+

Table 4 (Continued)

Localization	Study	Findings
	Mynett-Johnson et al. (1997)	–
	Knowles et al. (1998)	–
Chromosome 20	Le et al. (1994)	–
	Ewald et al. (1995b)	–
Chromosome 21q	Straub et al. (1994)	+
	Byerley et al. (1995)	–
	Gurling et al. (1995)	+
	Detera-Wadleigh et al. (1996)	+
	Ewald et al. (1996)	–
	Vallada et al. (1996)	+
Chromosome X (Xq27-q28)	Baron et al. (1987)	+
	Baron et al. (1993)	–
	Mendelbaum et al. (1995)	–
	Pekkarinen et al. (1995)	+
Genome scan	Coon et al. (1993)	
	Curtis et al. (1993b)	
	Gejman et al. (1993)	
	Detera-Wadleigh et al. (1994)	
	Ewald et al. (1994e)	
	Ginns et al. (1996)	
	NIMH Genetics Initiative Group:	
	Detera-Wadleigh et al. (1997)	
	Edenberg et al. (1997)	
	Rice et al. (1997)	
	Stine et al. (1997)	

INS, polymorphism of the insulin gene; HRAS, polymorphism of the Harvey ras gene; DRD2, polymorphism of the dopamine D2 receptor gene; DRD4, polymorphism of the dopamine D4 receptor gene; TH, polymorphism of the tyrosine hydroxylase gene; +, positive finding; –, negative finding.

linked to a region on the short arm of chromosome 11, in which not only the dopamine D4 receptor, but also the enzyme tyrosine hydroxylase (TH) are encoded. Nonetheless, further studies of this region in other families and in a more extensive sample of the initial family, including collateral branches, failed to replicate the finding of linkage.

Replicable findings concerning the genetic linkage of bipolar affective disorders have become available only in the last 2–3 years. The most important of these are listed below.

Chromosome 18

Berrettini et al. (1994) were the first to report a linkage of pericentromeric markers on chromosome 18 with bipolar affective illnesses in 22 families. This region and also the long arm of the chromosome were then intensively studied by several research groups. Both positive and negative findings were published; an overview of these studies is given in Table 4. It is particularly noteworthy that some of the research groups found clearer evidence for a predisposing genetic locus when the families were chosen in consideration of whether the illness was maternally or paternally transmitted (Stine et al. 1995). Evidence for a genetic locus on chromosome 18q was found mainly in families in which the transmission was largely paternal. This constellation of findings was replicated by a German research group (Nöthen et al. 1996). The genetic transmission of the risk-conferring gene may depend on the sex of the transmitting parent. If this is true, then "imprinting," a previously described mode of molecular genetic transmission, might be the mechanism by which it occurs, although the participation of an imprinting factor has not yet been demonstrated (Stine et al. 1995; Nöthen et al. 1996). A further susceptibility gene may lie on the long arm of chromosome 18 (18q23) (Freimer et al. 1996).

Chromosome 21

Straub et al. (1994) found a possibly predisposing genetic locus on chromosome 21q22.3 by means of a genome scan of North American and Israeli families. This finding was replicated in a separate family by Gurling et al. (1995) and then by Detera-Wadleigh et al. (1996). A region on chromosome 21q overlapping with this original region was found independently in a sample of North American twin pairs affected by the illness; this finding was made in the course of a genome-wide search for predisposing genetic loci (as

described further below), which was a component of the extensive NIMH Collaborative Study (Detera-Wadleigh et al. 1997). Negative linkage findings were obtained, however, in one North American (Byerley et al. 1995) and one Danish family sample (Ewald et al. 1996).

Chromosome 16

In 1993, Eiberg et al. reported the linkage of a phenotypic marker on chromosome 16p (phosphoglycolate phosphatase) with bipolar affective illnesses. This finding was replicated by two research groups using polymorphic genotypic markers (Ewald et al. 1995d; Edenberg et al. 1997), while two further groups (LaBuda et al. 1996; Adams et al. 1997) found no evidence for a predisposing genetic locus in this region. Further studies are needed before the significance of this region can be determined conclusively.

Chromosome 12

The 12q23–q24.1 region of chromosome 12 has been intensively studied. Craddock et al. (1994a) reported the cosegregation of bipolar illness and Darier's disease with markers in this region. Studies by two independent groups of investigators provided further evidence for a linkage of this region with bipolar illness. Weak evidence for a linkage was also provided by the genome-wide search for predisposing genetic loci for affective illness, which is described further below (Rice et al. 1997).

Chromosome 4

Blackwood et al. (1996b), studying a large Scottish family with affective disorders, found clear evidence for a predisposing genetic locus on chromosome 4p. This study was replicated by a German group (M.M. Nöthen et al. unpublished). No evidence for a linkage was found in the NIMH Collaborative Study of ill twin pairs, which employed identical markers (Detera-Wadleigh et al. 1997).

X Chromosome

X-chromosomal transmission is characterized in the formal genetic sense by the lack of father–son transmission of the disorder. This pattern is only rarely observed; in a subgroup of pedigrees with familial affective disorders, the pattern of inheritance is reportedly consistent with sex-linked transmission (Pekkarinen et al. 1995). A possible linkage of affective disorders with the genetic locus for red–green color blindness and glucose-6-phosphate dehydrogenase on chromosome Xq28 has been under study for three decades, but the findings to date have not been consistently positive. These families were first studied with clinically defined markers and later with molecular genetic markers (for a review, see Berrettini et al. 1990). Table 4 provides an overview of recently published reports.

The findings of Pekkarinen et al. (1995) are particularly noteworthy. Affective disorders in a large Finnish family were found to be linked to a candidate region overlapping with Xq28. The extensive NIMH Collaborative Study revealed further evidence for such a linkage in a group of ill North American twin pairs (Stine et al. 1997). Linkage to a gene in the Xq28 region thus remains possible for a subgroup of families with familial affective disorders, despite the numerous negative findings to date.

Furthermore, the region of genetic linkage appears to depend on whether the affected twin pairs consist of brothers or sisters. Stine et al. (1997), for example, reported a possible linkage with Xp22 for pairs of brothers and Xq26–q28 for pairs of sisters. Here, too, further analyses are needed before the question of linkage to these regions can be definitively answered.

The initial findings of a multicenter study involving a genome-wide search for genetic loci predisposing to bipolar affective illnesses were published in 1997 (NIMH Genetics Initiative Bipolar Group 1997; see Table 4). Four centers in the United States collected a total of 97 families (540 individuals) with affective illnesses for investigation, and 301 highly informative microsatellite markers, spaced an average of 10 cM apart, were used in the study. Both parametric and nonparametric statistical analyses were performed. A total of 34 regions of the genome were identified as possibly containing a genetic locus predisposing to affective illness. None of these regions, however, contained a locus accounting for more than 50% of the variance in bipolar affective disorder. The regions identified included segments of chromosomes 1p31, 7q21–31, 10p12, and 16p12.

This study revealed no evidence for a genetic locus for affective disorders on either chromosome 4 or chromosome 18; the regions on chromosomes 21, 16, and 12 mentioned in the above discussion were, however, among those identified as possibly containing a locus predisposing to affective illness.

A summary of currently available data on genetic linkage in the affective disorders is found in Table 5.

7.2 Association Studies

Recent genetic association studies have involved genes participating in both the dopaminergic and the serotonergic systems. Polymorphisms of the serotonin 2A receptor and of the serotonin transporter have been investigated for a possible association with bipolar illness (Table 6).

Tyrosine Hydroxylase

TH, the key enzyme in cathecholamine synthesis, may play a role in the pathophysiology of affective

Table 5. Summary of the most important linkage findings for affective illnesses to date

Chromosome	Study	Family sample(s)	Linkage method		
			Parametric lod score	Nonparametric *p* value	lod score
18p/q	Berrettini et al. (1994)	22	2.38	<0.01	–
	Stine et al. (1995)	28	1.45	0.0006	–
	Coon et al. (1996)	6	2.22	–	2.60
	De Bruyn et al. (1996)	1	1.34	–	–
	Freimer et al. (1996)	2	4.06	–	–
	Gershon et al. (1996)	22	–	<0.00001	–
	McInnes et al. (1996)	2	1.6	–	–
	Nöthen et al. (1996)	61	2.48	–	–
	Ewald et al. (1997)	2	1.83	–	–
21q	Straub et al. (1994)	1	3.41	–	–
	Gurling et al. (1995)	17	1.33	–	–
	Detera-Wadleigh et al. (1996)	22	–	0.0008	–
	Vallada et al. (1996)	60	1.2	–	–
12q23–q24.1	Dawson et al. (1995b)	45	0.69	<0.007	–
4p	Blackwood et al. (1996b)	12	4.1	–	–
Genome scan	NIMH Genetics Initiative Group				
1	Rice et al. (1997)	97	1.94[a]	<0.05	–
6	Rice et al. (1997)	97	2.37[a]	<0.05	–
7	Detera-Wadleigh et al. (1997)	97	–	0.002	
10	Rice et al. (1997)	97	3.47[a]	<0.001	–
12	Rice et al. (1997)	97	1.89[a]	<0.05	–
16	Edenberg et al. (1997)	97	–	0.006	–
21	Detera-Wadleigh et al. (1997)	97	–	0.008	–
22	Edenberg et al. (1997)	97	–	<0.001	2.46
X	Stine et al. (1997)	97	–	–	1.34

[a]mod score.

disorders. From 1990 onward, a group of investigators in France found evidence supporting a possible association of TH alleles with bipolar affective disorders (Leboyer et al. 1990; Meloni et al. 1995). These studies all involved a comparison of patients with independent control subjects. A large number of other groups were unable to replicate this finding, which may actually have been an artifact produced by stratification effects (see, e.g. Körner et al. 1994; Rietschel et al. 1995; Turecki et al. 1997).

Serotonin Transporter

The serotonin transporter gene (5-HTT) is located on chromosome 17 and is known to have two sites of polymorphism:

- A site in the gene promoter region, where the shorter of two variant forms is associated with reduced expression of the transporter
- A highly variable site ("variable number of tandem repeats"; VNTR) in intron 2 of the gene

Both polymorphisms have been investigated by multiple groups for a possible association (see Table 6). No final conclusion can yet be drawn as to whether or to what extent the serotonin transporter participates in the generation of bipolar illnesses.

Serotonin 2A Receptor

The fact that the serotonin 2A receptor (5-HT-2A) participates in the mechanism of action of the newer antidepressants has led to intensive investigation of this receptor for a possible association with bipolar illnesses. No such association has yet been found, and it is thus not yet known whether 5-HT-2A plays a role in the generation of bipolar illnesses. The published findings to date are summarized in Table 6.

Tryptophan Hydroxylase

Tryptophan hydroxylase (TPH) is the rate-determining enzyme in serotonin synthesis and is thus regarded as a candidate gene with respect to affective disorders. A group of investigators in France (Bellivier et al.

Table 6. Genetic association studies in bipolar affective disorders

Localization	Study	Findings
Dopamine D_1 and D_2 receptor	Nöthen et al. (1992)	–
Dopamine D_2 receptor (Cys311 variant)	Craddock et al. (1995a)	–
	Arinami et al. (1996)	+
	Sasaki et al. (1996)	–
	Souery et al. (1996)	–
Dopamine D_3 receptor	Rietschel et al. (1993)	–
	Shaikh et al. (1993)	–
	Parsian et al. (1995)	–/+
	Souery et al. (1996)	–
	Piccardi et al. (1997)	–
Dopamine D_4 receptor	Lim et al. (1994)	–
	Perez de Castro et al. (1994)	–
γ-Aminobutyric acid, A3 receptor gene (GABRA3)	Puertollano et al. (1995)	–
Serotonin 2A receptor	Ozaki et al. (1996)	–
	Arranz et al. (1997)	–
	Gutierrez et al. (1997a)	–
	Mahieu et al. (1997)	–
Serotonin transporter, VNTR variant	Battersby et al. (1996)	+
	Collier et al. (1996a)	+
	Rees et al. (1997)	+
	Esterling et al. (1998)	–
	Gutierrez et al. (1998)	–
Serotonin transporter, promoter variant	Collier et al. (1996b)	+
	Rees et al. (1997)	–
	Esterling et al. (1998)	–
	Gutierrez et al. (1998)	–
	Mendes de Oliveira et al. (1998)	–
Tyrosine hydroxylase	Leboyer et al. (1990)	+
	Inayama et al. (1993)	–
	Körner et al. (1994)	–
	Kawada et al. (1995a)	–
	Meloni et al. (1995)	+
	Perez de Castro et al. (1995)	+
	Rietschel et al. (1995)	–
	Souery et al. (1996)	–
	Turecki et al. (1997)	–
MAO-A/-B	Craddock et al. (1995b)	–
	Kawada et al. (1995b)	+
	Lim et al. (1995)	+
	Nöthen et al. (1995)	–
	Rubinsztein et al. (1996)	+
	Parsian and Todd (1997)	–
Phospholipase A_2	Dawson et al. (1995a)	+
	Jacobsen et al. (1996)	–
Pseudoautosomal region	Yoneda et al. (1992)	+
	Parsian and Todd (1994)	–
Catechol *O*-methyltransferase (COMT)	Gutierrez et al. (1997b)	–
	Lachman et al. (1997)	–
FRA-X locus	Craddock et al. (1994b)	–

+, positive finding; –, negative finding.

1998) recently reported an association between a noncoding variant and bipolar disorder. Replication studies are needed to determine whether this finding is merely an artifact produced by stratification effects. A way to counteract this possibility might be the use of internal controls. In this technique, not only the index subjects, but also their parents are studied. The alleles transmitted to the ill index subjects are used as the study sample, while the parental alleles that are not transmitted are used as a control sample.

Monoamine Oxidase A and B

The enzymes monoamine oxidase (MAO) A and B serve to degrade neurotransmitter substances; their pharmacologic inhibition has an antidepressant effect. The genes for both enzymes lie on the long arm of the X chromosome and are polymorphic. A nonsense mutation preventing the synthesis of MAO-A has been described. This mutation has been shown to be responsible for the occurrence of a form of mental retardation, along with a tendency toward aggression and manic behavior, in a Dutch family with many affected members (Brunner et al. 1993). A further restriction fragment length polymorphism (RFLP) of the MAO-A gene has been found to be associated with elevated MAO-A activity in vitro (Lim et al. 1995). This MAO-A gene polymorphism and a further one were found to be weakly associated with bipolar disorder in several British study cohorts and in one Japanese cohort (Kawada et al. 1995b; Lim et al. 1995; Rubinsztein et al. 1996), but these findings were not definitively replicated in a further British study (Craddock et al. 1995b).

All association studies of MAO genes yielding positive findings to date have employed groups of independent control subjects that were not shown to be sufficiently comparable to the patient group to rule out the possibility of a false-positive result (see above). The few studies employing internal controls (Nöthen et al. 1995; Parsian and Todd 1997), which are not subject to this criticism, revealed no positive association for alleles of either the MAO-A or MAO-B gene. It therefore seems unlikely that the MAO genes play a role in the generation of the illness.

Other Possible Associations

Several studies have concerned the catechol *O*-methyl transferase (COMT) gene on chromosome 22 and the phospholipase A2 (PLA-2A) gene on chromosome 12 (which lies in the region of the genetic locus for Darier's disease). Their findings have been inconsistent for the PLA-2A locus and uniformly negative for the COMT locus. No definitive conclusion can yet be drawn regarding the relevance of the PLA-2A gene to the generation of the disease.

7.3 Repeat Expansion Detection

A British group working mainly with the repeat expansion detection (RED) technique demonstrated, in two separate studies, that patients with bipolar illness have expanded trinucleotide repeat sequences more frequently than control subjects (O'Donovan et al. 1995, 1996). Oruc et al. (1997) arrived at the same result for familial cases, while Lindblad et al. (1995), too, found expanded trinucleotide repeats in patients. Vincent et al. (1996) attempted, but failed, to confirm these findings. O'Donovan and colleagues (see Guy et al. 1997) studied 50 specific loci containing trinucleotide repeats, but failed to find expanded repeats of any of them in patients. An American group employed a hybridization technique to search specific loci for polymorphic trinucleotide units and then look for corresponding expansions in patients and control subjects. By means of this method, an association of expanded trinucleotide repeats in 13 different loci with bipolar illness was ruled out (Jain et al. 1996), but a partly unstable CTG repeat on chromosome 18q21.1 was found, which is sometimes enormously expanded (Breschel et al. 1997). Unfortunately, no association of this site with bipolar illness has yet been demonstrated.

In the coming years, the extent to which expanded trinucleotide repeats participate in the generation of the disease will become clearer. An assessment of the findings obtained by RED will be possible only after the expanded trinucleotide repeats have been localized in the genome.

8 Concluding Remarks

Clinical epidemiological studies have shown that affective disorders occur more frequently in relatives of affective patients and that they are partly genetically determined, but the mechanism of genetic determination remains unclear.

To summarize the subclassification of affective disorders, it may be stated that the patterns of familial occurrence of the various subtypes of unipolar depression and of bipolar disorder have certain similarities: all subtypes of affective disorder studied to date are associated with a higher frequency of depressive episodes (major depression) in patients' relatives than in the general population. On the other hand, subtype-specific patterns of increased occurrence have also been observed: bipolar affective disorders occur more frequently only in families of patients with bipolar

affective and bipolar schizoaffective disorders. The latter fact distinguishes this group of diagnoses etiologically from all variants of unipolar depression and indicates that genetic factors play a major causative role.

While unipolar depressive episodes occur with increased frequency in relatives of patients suffering from any type of affective disorder, early-onset dysthymias are more common in relatives of patients with dysthymia, but not in relatives of patients without dysthymia who suffer from depressive episodes.

Bipolar I disorders occur more commonly in relatives of patients with bipolar I, and similarly for bipolar II disorders, although there is some overlap in both directions.

The earlier model ("multiple threshold model") of a familial and genetic continuum, in which schizomanic disorder is the most severe and most highly familial element, followed in sequence by bipolar I disorder, bipolar II disorder, and then unipolar depression (Gershon et al. 1982), is no longer plausible in view of the partial specificity of the subtypes of the affective disorders. Rather, it must be assumed that the familial and genetic risk factors of the various subtypes are partly subtype specific and partly overlapping. There is relatively little genetic overlap between unipolar depressive disorders and bipolar affective disorders.

The pattern of familial transmission of affective disorders, and of all of their known subtypes, is determined by both genetic and nongenetic factors. Bipolar I disorders are the type most strongly influenced by genetic factors.

The patterns of familial transmission of the affective disorders do not correspond to the known patterns of transmission of monogenic diseases, and they also make the existence of a single, major gene appear unlikely for the majority of multiply affected families. The affective disorders are, therefore, genetically complex disorders. Specification of the causative genetic factors of such disorders is obtained by the identification of susceptibility genes through linkage and association studies; the latter technique is more likely than the former to yield false-positive results. Present knowledge with regard to susceptibility genes is largely limited to bipolar affective disorders; only a few genetic association studies have been performed on the familially transmitted phenotype of unipolar depression, both because it is imprecisely defined and because it is more weakly genetically determined than bipolar disorder, as discussed above. No study to date concerning genetic markers in unipolar depression has yielded replicable positive findings.

Linkage studies have revealed several candidate regions for bipolar affective disorders, and findings have been replicated. The majority of the findings are consistent with polygenic transmission and thus accord with the findings of segregation analysis performed without genetic markers.

The linkage signal is sharply delimited and distinct in the case of monogenic diseases; therefore, as a rule, the candidate regions revealed by linkage studies can be narrowed down further by the study of critical recombinations located between the disease and marker genes (Kruglyak and Lander 1995). The replicable candidate regions emerging from linkage studies of complex disorders are rather broad (approximately 20 cM); further narrowing of candidate regions and the identification of susceptibility genes will therefore require either good fortune or a long-term, systematic search.

The candidate regions may contain candidate genes whose products play a role in the pathophysiology of the disorder. If these candidate genes are in fact susceptibility genes, decisive evidence for the fact can be obtained from mutation analysis of the candidate genes, suitable further linkage and association studies, and studies of the functional correlates of mutations. The study of the genetics of schizophrenia and the affective disorders has not yet reached this desirable state, but new candidate genes are continually being discovered as the pathophysiology of these disorders becomes better understood and the associated gene products are identified.

Under current conditions, candidate regions can be narrowed down by the following means:

- Enlargement of the family sample and thus of the number of recombinations; this enhances the power of the statistical analysis and enables a more precise definition of the region of the genome that might contain a susceptibility gene.
- Analyses of linkage to narrowly spaced, highly polymorphic markers (N.B.: the identification of susceptibility genes in this manner may require a very large sample).
- Determination of linkage disequilibrium between the marker and disease loci in nuclear family samples (ill index patient and both parents). Linkage disequilibrium represents the number of recombinations that occurred between the marker and disease loci in earlier generations. Linkage disequilibrium is more informative than linkage, insofar as it is more pronounced at the disease locus when the relative risk is low and falls off rapidly to insignificance as a function of the separation between marker and disease loci. This strategy is particularly promising in isolated or inbred populations, in which linkage disequilibrium is strong (Jorde 1995).
- Linkage disequilibrium of candidate gene loci in regions linked to the locus of the susceptibility gene.

This strategy may attain its objective rapidly if mutations of the candidate gene locus directly influence the risk of developing the disease (e.g. the fibrillin gene in Marfan's syndrome).

The advantages just mentioned of determining linkage disequilibrium, as opposed to determining linkage, will become particularly evident once the Human Genome Project has succeeded in its goal of mapping all genes on the genome and once alternatives have been developed to the relatively cumbersome electrophoretic sequencing techniques currently available (e.g. chip technology). All alleles of polymorphous genes (of which there are about 100,000) can then be studied for linkage disequilibrium. If certain assumptions are made as part of the model, the number of nuclear families (ill index patients and their parents) that will have to be enrolled or recruited for such studies will remain within practical limits (Risch and Merikangas 1996).

Genes shown by linkage analysis to possess mutations associated with an elevated risk of developing an affective disorder will certainly be accessible to identification, the more so in view of the steady progress being made in gene sequencing techniques and the speed at which the base pair sequence of the human genome is already being determined at present.

9 References

Adams LJ, Salmon JA, Kwok JB, Vivero C, Donald JA, Mitchell PB, Schofield PR (1997) Exclusion of linkage between bipolar affective disorder and chromosome 16 in 12 Australian pedigrees. Am J Med Genet 74: 304–310

Akiskal HS, Rosenthal TL, Haykal RF, Lemmi H, Rosenthal RH, Scott-Strauss A (1980) Characterological depressions: clinical and sleep EEG findings separating "subaffective dysthymias" from "character spectrum disorders". Arch Gen Psychiatry 37: 777–783

Anderson RL, Klein DN, Riso LP, Ouimette PC, Lizardi H, Schwartz JE (1996) The subaffective-character spectrum subtyping distinction in primary early-onset dysthymia: a clinical and family study. J Affect Disord 38: 13–22

Angst J (1966) Zur Ätiologie und Nosologie endogener depressiver Psychosen. Springer, Berlin Heidelberg New York

Angst J, Wicki W (1990) The Zurich study. XI. Is dysthymia a separate form of depression? Results of the Zurich cohort study. Eur Arch Psychiatry Clin Neurosci 240: 349–354

Arinami T, Itokawa M, Aoki J et al (1996) Further association study on dopamine D2 receptor variant S311C in schizophrenia and affective disorders. Am J Med Genet 67: 133–138

Arranz MJ, Erdmann J, Kirov G et al (1997) 5-HT2A receptor and bipolar affective disorder: association studies in affected patients. Neurosci Lett 224: 95–98

Baron M (1996) Further reflections on linkage results in schizophrenia. Am J Med Genet 67: 430–432

Baron M, Risch N, Hamburger R et al (1987) Genetic linkage between X-chromosome markers and bipolar affective illness. Nature 326: 289–292

Baron M, Freimer NF, Risch N et al (1993) Diminished support for linkage between manic depressive illness and X-chromosome markers in three Israeli pedigrees. Nat Genet 3: 49–55

Battersby S, Ogilvie AD, Smith CA et al (1996) Structure of a variable number tandem repeat of the serotonin transporter gene and association with affective disorder. Psychiatr Genet 6: 177–181

Bellivier F, Leboyer M, Courtet P et al (1998) Association between the tryptophan hydroxylase gene and manic-depressive illness. Arch Gen Psychiatry 55: 33–37

Benkelfat C, Ellenbogen MA, Dean P, Palmour RM, Young SN (1994) Mood-lowering effect of tryptophan depletion. Arch Gen Psychiatry 51: 687–697

Berrettini WH, Goldin LR, Gelernter J, Gejman PV, Gershon ES, Detera-Wadleigh S (1990) X-chromosome markers and manic-depressive illness. Arch Gen Psychiatry 47: 366–373

Berrettini WH, Ferraro TN, Goldin LR, Weeks DE, Detera-Wadleigh S, Nurnberger J, Gershon ES (1994) Chromosome 18 DNA markers and manic-depressive illness: evidence for a susceptibility gene. Proc Natl Acad Sci USA 91: 5918–5921

**Bertelsen A, Harvald B, Hauge M (1977) A Danish twin study of manic-depressive disorders. Br J Psychiatry 130: 330–351

Blacker D, Lavori PW, Faraone SV, Tsuang MT (1993) Unipolar relatives in bipolar pedigrees: a search for indicators of underlying bipolarity. Am J Med Genet 48: 192–199

Blackwood DH, Sharp CW, Walker MT, Doody GA, Glabus MF, Muir WJ (1996a) Implications of comorbidity for genetic studies of bipolar disorder: P300 and eye tracking as biological markers for illness. Br J Psychiatry 168: 85–92

Blackwood DH, He L, Morris SW et al (1996b) A locus for bipolar affective disorder on chromosome 4p. Nat Genet 12: 427–430

Bland RC, Newman SC, Orn H (1986) Recurrent and nonrecurrent depression: a family study. Arch Gen Psychiatry 43: 1085–1089

Breschel TS, McInnis MG, Margolis RL et al (1997) A novel, heritable expanding CTG repeat in an intron of the SEF2–1 gene on chromosome 18q21.1. Hum Mol Genet 6: 1855–1863

Brunner HG, Nelen M, Breakefield XO, Ropers HH, Oost BA van (1993) Abnormal behavior associated with a point mutation in the structural gene for monoamine oxidase A. Science 262: 578–580

Byerley W, Plaetke R, Hoff M et al (1992) Tyrosine hydroxylase gene not linked to manic-depression in seven of eight pedigrees. Hum Hered 42: 259–263

Byerley W, Hoff M, Holik J, Coon H (1994) A linkage study with D5 dopamine and alpha 2C-adrenergic receptor genes in six multiplex bipolar pedigrees. Psychiatr Genet 4: 121–124

Byerley W, Holik J, Hoff M, Coon H (1995) Search for a gene predisposing to manic-depression on chromosome 21. Am J Med Genet 60: 231–233

Cassano GB, Akiskal HS, Perugi G, Musetti L, Savino M (1992) The importance of measures of affective temperaments in genetic studies of mood disorders. J Psychiatr Res 26: 257–268

Claes S, Raeymaekers P, van den Broeck MV et al (1997) A chromosome 18 genetic linkage study in three large Belgian pedigrees with bipolar disorder. J Affect Dis 43: 195–205

Collier DA, Arranz MJ, Sham P et al (1996a) The serotonin transporter is a potential susceptibility factor for bipolar affective disorder. Neuroreport 7: 1675–1679

Collier DA, Stober G, Li T et al (1996b) A novel functional polymorphism within the promoter of the serotonin transporter gene: possible role in susceptibility to affective disorders. Mol Psychiatry 1: 453–460

Coon H, Jensen S, Hoff M et al (1993) A genome-wide search for genes predisposing to manic-depression, assuming autosomal dominant inheritance. Am J Hum Genet 52: 1234–1249

Coon H, Hoff M, Holik J et al (1996) Analysis of chromosome 18 DNA markers in multiplex pedigrees with manic depression. Biol Psychiatry 39: 689–696

Coryell W, Zimmerman M (1989) Personality disorders in the families of depressed, schizophrenic and never-ill probands. Am J Psychiatry 146: 496–502

Coryell W, Winokur G, Keller M, Scheftner W, Endicott J (1992a) Alcoholism and primary major depression: a family study approach to co-existing disorders. J Affect Disord 24: 93–99

*Coryell W, Endicott J, Keller M (1992b) Rapidly cycling affective disorder: demographics, diagnosis, family history, and course. Arch Gen Psychiatry 49: 126–131

Coryell W, Endicott J, Maser JD, Keller MB, Leon AC, Akiskal HS (1995) Long-term stability of polarity distinctions in the affective disorders. Am J Psychiatry 152: 385–390

Craddock N, Owen M, Burge S, Kurian B, Thomas P, McGuffin P (1994a) Familial cosegregation of major affective disorder and Darier's disease (keratosis follicularis). Br J Psychiatry 164: 355–358

Craddock N, Daniels J, McGuffin P, Owen M (1994b) Variation at the fragile X locus does not influence susceptibility to bipolar disorder. Am J Med Genet 54: 141–143

Craddock N, Roberts Q, Williams N, McGuffin P, Owen MJ (1995a) Association study of bipolar disorder using a functional polymorphism (Ser311->Cys) in the dopamine D2 receptor gene. Psychiatr Genet 5: 63–65

Craddock N, Daniels J, Roberts E, Rees M, McGuffin P, Owen MJ (1995b) No evidence for allelic association between bipolar disorder and monoamine oxidase A gene polymorphisms. Am J Med Genet 60: 322–324

Craddock N, Khodel V, Van Eerdewegh P, Reich T (1995c) Mathematical limits of multilocus models. The genetic transmission of bipolar disorder. Am J Hum Genet 57: 690–702

*Craddock N, Eerdewegh P van, Reich T (1997) Single major locus models for bipolar disorder are implausible. Am J Med Genet 74: 18

Cross-National Collaborative Group (1992) The changing rate of depression. JAMA 268: 3098–3105

Curtis D, Brynjolfsson J, Petursson H et al (1993a) Segregation and linkage analysis in five manic depression pedigrees excludes the 5-HAT-1a receptor gene (HTR1A). Ann Hum Genet 57: 27–39

Curtis D, Sherrington R, Brett P et al (1993b) Genetic linkage analysis of manic depression in Iceland. J R Soc Med 86: 506–510

Dawson E, Gill M, Curtis D, Castle D, Hunt N, Murray R, Powell J (1995a) Genetic association between alleles of pancreatic phospholipase A2 gene and bipolar affective disorder. Psychiatr Genet 5: 177–180

Dawson E, Parfitt E, Roberts Q et al (1995b) Linkage studies of bipolar disorder in the region of the Darier's disease gene on chromosome 12q23–24.1. Am J Med Genet 60: 94–102

De Bruyn A, Mendelbaum K, Sandkuijl LA et al (1994) Non-linkage of bipolar illness to tyrosine hydroxylase, tyrosinase, and D2 and D4 dopamine receptor genes on chromosome 11. Am J Psychiatry 151: 102–106

De Bruyn A, Souery D, Mendelbaum K, Mendlewicz J, Van Broeckhoven C (1996) Linkage analysis of families with bipolar illness and chromosome 18 markers. Biol Psychiatry 39: 679–688

DePaulo JR, Simpson SG, Gayle JO, Folstein SE (1990) Bipolar II disorder in six sisters. J Affect Disord 19: 259–264

Detera-Wadleigh SD, Hsieh WT, Berrettini WH et al (1994) Genetic linkage mapping for a susceptibility locus to bipolar illness: chromosome 2, 3, 4, 7, 9, 10p, 11p, 22, and Xpter. Am J Med Genet 54: 206–218

Detera-Wadleigh SD, Badner JA, Goldin LR et al (1996) Affected-sib-pair analyses reveal support of prior evidence for a susceptibility locus for bipolar disorder on 21q. Am J Hum Genet 58: 1279–1285

*Detera-Wadleigh SD, Badner JA, Yoshikawa T et al (1997) Initial genome scan of the NIMH genetics initiative bipolar pedigrees: chromosomes 4, 7, 9, 18, 19, 20, and 21q. Am J Med Genet 74: 254–262

Donaldson SK, Klein DN, Riso LP, Schwartz JE (1997) Comorbidity between dysthymic and major depressive disorders: a family study analysis. J Affect Disord 42: 103–111

*Edenberg HJ, Foroud T, Conneally PM et al (1997) Initial genomic scan of the NIMH genetics initiative bipolar pedigrees: chromosomes 3, 5, 15, 16, 17, and 22. Am J Med Genet 74: 238–246

*Egeland JA, Gerhard D, Pauls DL et al (1987) Bipolar affective disorders linked to DNA markers on chromosome 11. Nature 325: 783–787

Eiberg H, Ewald H, Mors O (1993) Suggestion of linkage between manic-depressive illness and the enzyme phosphoglycolate phosphatase (PGP) on chromosome 16p. Clin Genet 44: 254–257

Endicott J, Nee J, Andreasen N, Clayton P, Keller M, Coryell W (1985) Bipolar II. Combine or keep separate? J Affect Disord 8: 17–28

Esterling LE, Yoshikawa T, Turner G et al (1998) Serotonin transporter (5-HTT) gene and bipolar affective disorder. Am J Med Genet 81: 37–40

Ewald H, Mors O, Flint T, Eiberg H, Kruse TA (1994a) Linkage analysis between manic depressive illness and the dopamine beta-hydroxylase gene. Psychiatr Genet 4: 177–183

Ewald H, Mors O, Flint T, Kruse TA (1994b) Linkage analysis between manic depressive illness and the region on chromosome 12q involved in Darier's disease. Psychiatr Genet 4: 195–200

Ewald H, Mors O, Flint T, Kruse TA (1994c) Linkage analysis between manic-depressive illness and the region on chromosome 15q involved in Prader-Willi syndrome, including two GABA-A receptor subtype genes. Hum Hered 44: 287–294

Ewald H, Mors O, Friedrich U, Flint T, Kruse TA (1994d) Exclusion of linkage between manic depressive illness and tyrosine hydroxylase and dopamine D2 receptor genes. Psychiatr Genet 4: 13–22

Ewald H, Mors O, Eiberg H (1994e) Linkage analysis between manic-depressive illness and 35 classical markers. Am J Med Genet 54: 144–148

Ewald H, Mors O, Eiberg H, Flint T, Kruse TA (1995a) No evidence of linkage between manic depressive illness and the dopa decarboxylase gene or nearby region on chromosome 7p. Psychiatr Genet 5: 161–169

Ewald H, Eiberg H, Mors O (1995b) A search for genes predisposing to manic depressive illness on chromosome 20. Psychiatr Genet 5: 105–111

Ewald H, Mors O, Flint T, Friedrich U, Eiberg H, Kruse TA (1995c) Linkage analysis between manic-depressive illness and markers on the long arm of chromosome 11. Am J Med Genet 60: 386–392

Ewald H, Mors O, Flint T, Koed K, Eiberg H, Kruse TA (1995d) A possible locus for manic depressive illness on chromosome 16p13. Psychiatr Genet 5: 71–81

Ewald H, Eiberg H, Mors O, Flint T, Kruse TA (1996) Linkage study between manic-depressive illness and chromosome 21. Am J Med Genet 67: 218–224

Ewald H, Mors O, Koed K, Eiberg H, Kruse TA (1997) Susceptibility loci for bipolar affective disorder on chromosome 18? A review and a study of Danish families. Psychiatr Genet 7: 1–12

Fanous AH, Walsh D, Kendler KS (1996) Do endogenous features in depression predict the risk of psychiatric illness in relatives? Acta Psychiatr Scand 94: 56–59

Fidani L, Rooke K, Chartier-Harlin M-C, Hughes D, Tanzi R, Mullan M, Roques P, Rossor M, Hardy J, Goatz A (1992) Screening for mutations in the open reading frame and promotor of the β-amyloid-precursor protein gene in familial Alzheimer's disease: identification of a further family with APP717 Val → Ilc. Hum Mol Genet 1: 165–168

Freimer NB, Reus VI, Escamilla MA et al (1996) Genetic mapping using haplotype, association and linkage methods suggests a locus for severe bipolar disorder (BPI) at 18q22-q23. Nat Genet 12: 436–441

Gejman PV, Martinez M, Cao Q et al (1993) Linkage analysis of fifty-seven microsatellite loci to bipolar disorder. Neuropsychopharmacology 9: 31–40

*Gershon ES, Hamovit J, Guroff JJ et al (1982) A family study of schizoaffective disorder, bipolar I, bipolar II, unipolar and normal control probands. Arch Gen Psychiatry 39: 1157–1167

*Gershon ES, Hamovit JH, Guroff JJ, Nurnberger JI (1987) Birth-cohort changes in manic and depressive disorders in relatives of bipolar and schizoaffective patients. Arch Gen Psychiatry 44: 314–319

Gershon ES, DeLisi LE, Hamovit J et al (1988) A controlled family study of chronic psychoses. Schizophrenia and schizoaffective psychoses. Arch Gen Psychiatry 45: 328–336

Gershon ES, Badner JA, Detera-Wadleigh SD, Ferraro TN, Berrettini WH (1996) Maternal inheritance and chromosome 18 allele sharing in unilineal bipolar illness pedigrees. Am J Med Genet 67: 202–207

Ginns EI, Ott J, Egeland JA et al (1996) A genome-wide search for chromosomal loci linked to bipolar affective disorder in the Old Order Amish. Nat Genet 12: 431–435

Goldstein JM, Faraone SV, Chen WJ, Tsuang MT (1993) The role of gender in understanding the familial transmission of schizoaffective disorder. Br J Psychiatry 163: 763–768

Goodwin FK, Ghaemi SN (1998) Understanding manic-depressive illness. Arch Gen Psychiatry 55: 23–25

*Greenberg DA (1993) Linkage analysis of "necessary" disease loci versus "susceptibility" loci. Am J Hum Genet 52: 135–143

Grigoroiu-Serbanescu G, Wickramaratne PJ, Hodge A, Milea S, Mihailescu (1997) Genetic anticipation and imprinting in bipolar I illness. Br J Psychiatry 170: 162–166

Grof P, Alda M, Grof E, Zvolsky P, Walsh M (1994) Lithium response and genetics of affective disorders. J Affect Disord 32: 85–95

Gurling H, Smyth C, Kalsi G et al (1995) Linkage findings in bipolar disorder (letter). Nat Genet 10: 8–9

Gutierrez B, Bertranpetit J, Collier D et al (1997a) Genetic variation of the 5-HT2A receptor gene and bipolar affective disorder. Hum Genet 100: 582–584

Gutierrez B, Bertranpetit J, Guillamat R, Valles V, Arranz MJ, Kerwin R, Fananas L (1997b) Association analysis of the catechol O-methyltransferase gene and bipolar affective disorder. Am J Psychiatry 154: 113–115

Gutierrez B, Arranz MJ, Collier DA et al (1998) Serotonin transporter gene and risk for bipolar affective disorder: an association study in Spanish population. Biol Psychiatry 43: 843–847

Guy C, Bowen T, Daniels JK et al (1997) Exclusion of expansion of 50 CAG/CTG trinucleotide repeats in bipolar disorder. Am J Psychiatry 154: 1146–1147

*Heun R, Maier W (1993a) Bipolar II disorders in six first degree relatives. Biol Psychiatry 34: 274–276

Heun R, Maier W (1993b) The distinction of bipolar II disorder from bipolar I and recurrent unipolar depression: results of a controlled family study. Acta Psychiatr Scand 87: 279–284

Hodge SE, Wickramaratne PJ (1995) Statistical pitfalls in detecting age-of-onset anticipation and detecting ascertainment bias. Psychiatr Genet 5: 43–47

Holmes D, Brynjolfsson J, Brett P, Curtis D, Petursson H, Sherrington R, Gurling H (1991) No evidence for a susceptibility locus predisposing to manic depression in the region of the dopamine (D2) receptor gene. Br J Psychiatry 158: 635–641

**Huntington's Disease Collaborative Research Group (1993) A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. Cell 72: 971–983

Inayama Y, Yoneda H, Sakai T et al (1993) Lack of association between bipolar affective disorder and tyrosine hydroxylase DNA marker. Am J Med Genet 48: 87–89

Jacobsen N, Daniels J, Moorhead S et al (1996) Association study of bipolar disorder at the phospholipase A2 gene (PLA2A) in the Darier's disease (DAR) region of chromosome 12q23-q24.1. Psychiatr Genet 6: 195–199

Jain S, Leggo J, DeLisi LE et al (1996) Analysis of thirteen trinucleotide repeat loci as candidate genes for schizophrenia and bipolar affective disorder. Am J Med Genet 67: 139–146

Janowsky DS, Overstreet DH, Nurnberger JI (1994) Is cholinergic sensitivity a genetic marker for the affective disorders? Am J Med Genet 54: 335–344

Jeffries FM, Reiss AL, Brown T, Meyers DA, Glicksman A, Bandyopadhyay S (1993) Bipolar spectrum disorder and fragile X syndrome: a family study. Biol Psychiatry 33: 213–216

Jensen S, Plaetke R, Holik J et al (1992) Linkage analysis of the D1 dopamine receptor gene and manic depression in six families. Hum Hered 42: 269–275

*Jorde LB (1995) Linkage disequilibrium as a gene-mapping tool. Am J Hum Genet 56: 11–14

Kato T, Winokur G, Coryell W, Keller MB, Endicott J, Rice J (1996) Parent-of-origin effect in transmission of bipolar disorder. Am J Med Genet 67: 546–550

Kawada Y, Hattori M, Fukuda R, Arai H, Inoue R, Nanko S (1995a) No evidence of linkage or association between tyrosine hydroxylase gene and affective disorder. J Affect Disord 34: 89–94

Kawada Y, Hattori M, Dai XY, Nanko S (1995b) Possible association between monoamine oxidase gene and bipolar affective disorder. Am J Hum Genet 56: 335–336

Keller MB, Beardslee WR, Dorer DJ, Lavori PW, Samuelson H, Klerman GR (1986) Impact of severity and chronicity of parental affective illness on adaptive functioning and psychopathology in children. Arch Gen Psychiatry 43: 930–937

*Kelsoe JR, Ginns EI, Egeland JA et al (1989) Reevaluation of the linkage relationship between chromosome 11p loci and the gene for bipolar affective disorder in the Old Order Amish. Nature 342: 238–243

Kelsoe JR, Kristbjanarson H, Bergesch P et al (1993) A genetic linkage study of bipolar disorder and 13 markers on chromosome 11 including the D2 dopamine receptor. Neuropsychopharmacology 9: 293–301

Kelsoe JR, Remick RA, Sadovnick AD et al (1996) Genetic linkage study of bipolar disorder and the serotonin transporter. Am J Med Genet 67: 215–217

Kendler KS, Karkowski-Shuman L (1997) Stressful life events and genetic liability to major depression: genetic control of exposure to the environment? Psychol Med 27: 539–547

Kendler KS, Neale MC, Kessler RC, Heath AC, Eaves LJ (1992a) Major depression and generalized anxiety disorder. Same genes, (partly) different environments? Arch Gen Psychiatry 49: 716–722

Kendler KS, Neale MC, Kessler RC, Heath AC, Eaves LJ (1992b) A population-based twin study of major depression in women. The impact of varying definitions of illness. Arch Gen Psychiatry 49: 257–266

Kendler KS, Heath AC, Neale MC, Kessler RC, Eaves LJ (1992c) A population-based twin study of alcoholism in women. JAMA 268: 1877–1882

Kendler KS, Kessler RC, Neale MC, Heath AC, Eaves LJ (1993a) The prediction of major depression in women: toward an integrated etiologic model. Am J Psychiatry 150: 1139–1148

Kendler K, Nenale MC, MacLean CL, Heath AC, Eaves LJ, Kessler RC (1993b) Smoking and major depression: a causal analysis. Arch Gen Psychiatry 50: 36–43

**Kendler KS, McGuire M, Gruenberg AM, O'Hare A, Spellman M, Walsh D (1993c) The Roscommon Family Study. IV. Affective illness, anxiety disorders, and alcoholism in relatives. Arch Gen Psychiatry 50: 952–960

Kendler KS, Heath AC, Neale MC, Kessler RC, Eaves LJ (1993d) Alcoholism and major depression in women. A twin study of the causes of comorbidity. Arch Gen Psychiatry 50: 690–698

Kendler KS, Karkowski-Shuman L, Walsh D (1996) The risk for psychiatric illness in siblings of schizophrenics: the impact of psychotic and non-psychotic affective illness and alcoholism in parents. Acta Psychiatr Scand 94: 49–55

Kerbeshian J, Burd L, Klug MG (1995) Comorbid Tourette's disorder and bipolar disorder: an etiologic perspective. Am J Psychiatry 152: 1646–1651

Kinney DK, Yurgelun-Todd DA, Levy DL, Medoff D, Lajonchere CM, Radford-Paregol M (1993) Obstetrical complications in patients with bipolar disorder and their siblings. Psychiatry Res 48: 47–56

Klein DN, Depue RA (1985) Obsessional personality traits and risks for bipolar affective disorder: an offspring study. J Abnorm Psychol 94: 291–297

Klein DN, Riso LP, Donaldson SK et al (1995) Family study of early-onset dysthymia. Mood and personality disorders in relatives of outpatients with dysthymia and episodic major depression and normal controls. Arch Gen Psychiatry 52: 487–496

Knowles JA, Rao PA, Cox-Matise T et al (1998) No evidence for significant linkage between bipolar affective disorder and chromosome 18 pericentromeric markers in a large series of multiplex extended pedigrees. Am J Hum Genet 62: 916–924

Körner J, Rietschel M, Hunt N et al (1994) Association and haplotype analysis at the tyrosine hydroxylase locus in a combined German-British sample of manic-depressive patients and controls. Psychiatr Genet 4: 167–175

Kraepelin E (1921) Einführung in die Psychiatrische Klinik, 4th edn. Barth, Leipzig

Kretschmer E (1936) Körperbau und Charakter, 11th edn. Springer, Berlin

*Kruglyak L, Lander ES (1995) High-resolution genetic mapping of complex traits. Am J Hum Genet 56: 1212–1223

*LaBuda MC, Maldonado M, Marshall D, Otten K, Gerhard DS (1996) A follow-up report of a genome search for affective disorder predisposition in the Old Order Amish. Am J Hum Genet 59: 1343–1362

Lachman HM, Kelsoe J, Moreno L, Katz S, Papolos DF (1997) Lack of association of catechol-O-methyltransferase (COMT) functional polymorphism in bipolar affective disorder. Psychiatr Genet 7: 13–17

**Lander ES, Schork N (1994) Genetic dissection of complex traits. Science 265: 2037–2048

Lauer CJ, Schreiber W, Holsboer F, Krieg JC (1995) In quest of identifying vulnerability markers for psychiatric disorders by all-night polysomnography. Arch Gen Psychiatry 52: 145–153

Lauer CJ, Bronisch T, Kainz M, Schreiber W, Holsboer F, Krieg JC (1997) Pre-morbid psychometric profile of subjects at high familial risk for affective disorder. Psychol Med 27: 355–362

Law A, Richard CW, Cottingham RW, Lathrop GM, Cox DR, Myers RM (1992) Genetic linkage analysis of bipolar affective disorder in an Old Order Amish pedigree. Hum Genet 88: 562–568

Le F, Mitchell P, Vivero C et al (1994) Exclusion of close linkage of bipolar disorder to the Gs-alpha subunit gene in nine Australian pedigrees. J Affect Disord 32: 187–195

Leboyer M, Malafosse A, Boularan S et al (1990) A tyrosine hydroxylase polymorphism reveals an association with manic-depressive illness. Lancet 335: 1219

Lilenfeld LR, Kaye WH, Greeno CG et al (1998) A controlled family study of anorexia nervosa and bulimia nervosa. Arch Gen Psychiatry 55: 603–610

Lim LC, Nöthen MM, Körner J et al (1994) No evidence of association between dopamine D4 receptor variants and bipolar affective disorder. Am J Med Genet 54: 259–263

Lim LC, Powell J, Sham P, Castle D, Hunt N, Murray R, Gill M (1995) Evidence for a genetic association between alleles of monoamine oxidase A gene and bipolar affective disorder. Am J Med Genet 60: 325–331

Lindblad K, Nylander PO, De Bruyn A et al (1995) Detection of expanded CAG repeats in bipolar affective disorder using the repeat expansion detection (RED) method. Neurobiol Dis 2: 55–62

Lish JD, Gyulai L, Resnick SM, Kirtland A, Amsterdam JD, Whybrow PC, Arlen-Price R (1993) A family history study of rapid-cycling bipolar disorder. Psychiatry Res 48: 37–45

MacKinnon DF, McMahon FJ, Simpson SG, McInnis MG, DePaulo JR (1997) Panic disorder with familial bipolar disorder. Biol Psychiatry 42: 90–95

Mahieu B, Souery D, Lipp O et al (1997) No association between bipolar affective disorder and a serotonin receptor (5-HT2A) polymorphism. Psychiatry Res 70: 65–69

Maier W, Lichtermann D (1993) Die familiäre Häufung affektiver Erkrankungen. Eine Übersicht über neuere familiengenetische Arbeiten. Nervenheilkunde 12: 34–40

Maier W, Philipp M (1993) Reliabilität und Validität der Subtypisierung und Schweregradmessung depressiver Syndrome. Springer, Berlin Heidelberg New York

Maier W, Lichtermann D, Minges J, Heun R, Hallmayer J, Klingler T (1991) Unipolar depression in the aged: determinants of familial aggregation. J Affect Disord 23: 53–61

Maier W, Lichtermann D, Minges J, Heun R, Hallmayer J, Benkert O (1992a) Schizoaffective disorder and affective disorders with mood-incongruent psychotic features: keep separate or combine? Evidence from a family study. Am J Psychiatry 149: 1666–1673

Maier W, Lichtermann D, Minges J, Heun R, Hallmayer J (1992b) The risk of minor depression in families of probands with major depression: sex differences and familiality. Eur Arch Psychiatry Clin Neurosci 242: 89–92

Maier W, Lichtermann D, Minges J, Heun R (1992c) The familial relation of personality disorders (DSM-III-R) to unipolar major depression. J Affect Disord 26: 151–156

*Maier W, Lichtermann D, Minges J, Hallmayer J, Heun R, Benkert O, Levinson DF (1993) Continuity and discontinuity of affective disorders and schizophrenia. Arch Gen Psychiatry 50: 871–883

Maier W, Lichtermann D, Minges J (1994) The relationship between alcoholism and unipolar depression – a controlled family study. J Psychiatr Res 28: 303–317

Maier W, Minges J, Lichtermann D (1995a) The familial relationship between panic disorder and unipolar depression. J Psychiatr Res 29: 375–388

Maier W, Lichtermann D, Minges J, Delmo C, Heun R (1995b) The relationship between bipolar disorder and alcoholism: a controlled family study. Psychol Med 5: 787–796

Maier W, Minges J, Lichtermann D, Heun R (1995c) Personality disorders and personality variations in relatives of patients with bipolar affective disorders. J Affect Disord 35: 173–181

Maier W, Hallmayer J, Zill P, Bondy B, Lichtermann D, Ackenheil M, Minges J, Wildenauer D (1995d) Linkage analysis between pericentromeric markers on chromosome 18 and bipolar disorder: a replication test. Psychiatry Res 59: 7–15

McGuffin P, Katz R, Aldrich J, Bebbington P (1988) The Camberwell collaborative depression study. II. Investigation of family members. Br J Psychiatry 152: 766–774

McGuffin P, Katz R, Rutherford J (1991) Nature, nurture and depression: a twin study. Psychol Med 21: 329–335

*McInnes LA, Escamilla MA, Service SK et al (1996) A complete genome screen for genes predisposing to severe bipolar disorder in two Costa Rican pedigrees. Proc Natl Acad Sci USA 93: 13060–13065

McInnis MG, McMahon FJ, Chase GA et al (1993) Anticipation in bipolar affective disorder. Am J Hum Genet 53: 385–390

McMahon FJ, Stine OC, Meyers DA, Simpson SG, DePaulo JR (1995) Patterns of maternal transmission in bipolar affective disorder. Am J Hum Genet 56: 1277–1286

Meloni R, Leboyer M, Bellivier F, Barbe B, Samolyk D, Allilaire JF, Mallet J (1995) Association of manic-depressive illness with tyrosine hydroxylase microsatellite marker. Lancet 345: 932

Mendelbaum K, Sevy S, Souery D et al (1995) Manic depressive illness and linkage reanalysis in the Xq27-Xq28 region of chromosome X. Neuropsychobiology 31: 58–63

Mendes de Oliveira JR, Otto PA, Vallada H et al (1998) Analysis of a novel functional polymorphism within the promoter region of the serotonin transporter gene (5-HTT) in Brazilian patients affected by bipolar disorder and schizophrenia. Am J Med Genet 81: 225–227

Mendlewicz J, Fieve RR, Stallone F (1973) Relationship between the effectiveness of lithium therapy and family history. Am J Psychiatry 130: 1011–1013

Mendlewicz J, Leboyer M, De Bruyn A et al (1991) Absence of linkage between chromosome 11p15 markers and manic-depressive illness in a Belgian pedigree. Am J Psychiatry 148: 1683–1687

Merikangas KR (1993) Genetic epidemiologic studies of affective disorders in childhood and adolescence. Eur Arch Psychiatry Clin Neurosci 243: 121–130

Merikangas KR, Weissman MM, Pauls D (1985a) Genetic factors in the sex ratio of major depression. Psychol Med 15: 63–69

Merikangas KR, Leckmann JF, Prusoff BA, Pauls DL, Weissman MM (1985b) Familial transmission of depression and alcoholism. Arch Gen Psychiatry 42: 367–372

Mirow AL, Kristbjanarson H, Egeland JA et al (1994) A linkage study of distal chromosome 5q and bipolar disorder. Biol Psychiatry 36: 223–229

Mitchell P, Waters B, Morrison N, Shine J, Donald J, Fisman J (1991) Close linkage of bipolar disorder to chromosome 11 markers is excluded in two large Australian pedigrees. J Affect Disord 21: 23–32

Mitchell P, Selbie L, Waters B, Donald J, Vivero C, Tully M, Shine J (1992) Exclusion of close linkage of bipolar disorder to dopamine D1 and D2 receptor gene markers. J Affect Disord 25: 1–11

Mitchell P, Waters B, Vivero C et al (1993) Exclusion of close linkage of bipolar disorder to the dopamine D3 receptor gene in nine Australian pedigrees. J Affect Disord 27: 213–224

Murray KT, Sines JO (1996) Parsing the genetic and non-genetic variance in children's depressive behavior. J Affect Disord 38: 23–34

Mynett-Johnson LA, Murphy VE, Manley P, Shields DC, McKeon P (1997) Lack of evidence for a major locus for bipolar disorder in the pericentromeric region of chromosome 18 in Irish pedigrees. Biol Psychiatry 42: 486–494

Nanko S, Kobayashi M, Gamou S et al (1991) Linkage analysis of affective disorder using DNA markers on chromosome 11 and X. Jpn J Psychiatry Neurol 45: 53–56

*NIMH Genetics Initiative Bipolar Group (1997) Genomic survey of bipolar illness in the NIMH genetics initiative pedigrees: a preliminary report. Am J Med Genet 74: 227–237

Nöthen MM, Erdmann J, Körner J et al (1992) Lack of association between dopamine D1 and D2 receptor genes and bipolar affective disorder. Am J Psychiatry 149: 199–201

Nöthen MM, Eggermann K, Albus M et al (1995) Association analysis of the monoamine oxidase A gene in bipolar affective disorder by using family-based internal controls. Am J Hum Genet 57: 975–978

Nöthen MM, Cichon S, Craddock N et al (1996) Linkage studies of bipolar disorder to chromosome 18 markers. Biol Psychiatry 39: 615

Nurnberger J, Guroff J, Hamovit J, Berrettini W, Gershon E (1988) A family study of rapid-cycling bipolar illness. J Affect Disord 15: 87–91

O'Donovan MC, Guy C, Craddock N et al (1995) Expanded CAG repeats in schizophrenia and bipolar disorder. Nat Genet 10: 380–381

O'Donovan MC, Guy C, Craddock N et al (1996) Confirmation of association between expanded CAG/CTG repeats and both

schizophrenia and bipolar disorder. Psychol Med 26: 1145–1153

Oruc L, Lindblad K, Verheyen GR et al (1997) CAG repeat expansions in bipolar and unipolar disorders (letter). Am J Hum Genet 60: 730–732

Ouimette PC, Klein DN, Pepper CM (1996) Personality traits in the first degree relatives of outpatients with depressive disorders. J Affect Disord 39: 43–53

Ozaki N, Rosenthal NE, Pesonen U et al (1996) Two naturally occurring amino acid substitutions of the 5-HT2A receptor: similar prevalence in patients with seasonal affective disorder and controls. Biol Psychiatry 40: 1267–1272

Parsian A, Todd RD (1994) Bipolar disorder and the pseudoautosomal region: an association study. Am J Med Genet 54: 5–7

Parsian A, Todd RD (1997) Genetic association between monoamine oxidase and manic-depressive illness: comparison of relative risk and haplotype relative risk data. Am J Med Genet 74: 475–479

Parsian A, Chakraverty S, Todd RD (1995) Possible association between the dopamine D3 receptor gene and bipolar affective disorder. Am J Med Genet 60: 234–237

Pauls DL, Gerhard DS, Lacy LG et al (1991) Linkage of bipolar affective disorders to markers on chromosome 11p is excluded in a second lateral extension of Amish pedigree 110. Genomics 11: 730–736

Pauls DL, Ott J, Paul SM et al (1995) Linkage analyses of chromosome 18 markers do not identify a major susceptibility locus for bipolar affective disorder in the Old Order Amish. Am J Hum Genet 57: 636–643

Pekkarinen P, Terwilliger J, Bredbacka PE, Lönnqvist J, Peltonen L (1995) Evidence for a predisposing locus to bipolar disorder on Xq24-q27.1 in an extended Finnish pedigree. Genome Res 5: 105–115

*Penrose LS (1948) The problem of anticipation in pedigrees of dystrophia myotonica. Ann Eugenics 14: 125–132

Perez de Castro I, Torres P, Fernandez-Piqueras J, Saiz-Ruiz J, Llinares C (1994) No association between dopamine D4 receptor polymorphism and manic-depressive illness. J Med Genet 31: 897–898

Perez de Castro I, Santos J, Torres P, Visedo G, Saiz-Ruiz J, Llinares C, Fernandez-Piqueras J (1995) A weak association between TH and DRD2 genes and bipolar affective disorder in a Spanish sample. J Med Genet 32: 131–134

Petronis A, Kennedy JL (1995) Unstable genes – unstable mind? Am J Psychiatry 152: 164–172

Piccardi MP, Severino G, Bocchetta A, Palmas MA, Ruiu S, Del Zompo M (1997) No evidence of association between dopamine D3 receptor gene and bipolar affective disorder. Am J Med Genet 74: 137–139

Puertollano R, Visedo G, Saiz-Ruiz J, Llinares C, Fernandez-Piqueras J (1995) Lack of association between manic-depressive illness and a highly polymorphic marker from GABRA3 gene. Am J Med Genet 60: 434–435

Rees M, Norton N, Jones I et al (1997) Association studies of bipolar disorder at the human serotonin transporter gene (hSERT; 5HTT). Mol Psychiatry 2: 398–402

*Rice J, Reich T, Andreasen NC et al (1987) The familial transmission of bipolar illness. Arch Gen Psychiatry 44: 441–447

*Rice JP, Goate A, Williams JT et al (1997) Initial genome scan of the NIMH genetics initiative bipolar pedigrees: chromosomes 1, 6, 8, 10, and 12. Am J Med Genet 74: 247–253

Rietschel M, Nöthen MM, Lannfelt L et al (1993) A serine to glycine substitution at position 9 in the extracellular N-terminal part of the dopamine D3 receptor protein: no role in the genetic predisposition to bipolar affective disorder. Psychiatry Res 46: 253–259

Rietschel M, Nöthen MM, Maier W, Albus M, Franzek E, Propping P (1995) Tyrosine hydroxylase gene and manic-depressive illness. Lancet 354: 1368

**Risch N, Merikangas K (1996) The future of genetic studies of complex human diseases. Science 273: 1516–1517

Riso LP, Klein DN, Ferro T, Kasch KL, Pepper CM, Schwartz JE, Aronson TA (1996) Understanding the comorbidity between early-onset dysthymia and cluster B personality disorders: a family study. Am J Psychiatry 153: 900–906

Rosenberg DR, Sweeney JA, Squires-Wheeler, Keshavan MS, Cornblatt BA, Erlenmeyer-Kimling L (1997) Eye-tracking dysfunction in offspring from the New York high-risk project: diagnostic specificity and the role of attention. Psychiatry Res 66: 121–130

Rubinsztein DC, Leggo J, Goodburn S, Walsh C, Jain S, Paykel ES (1996) Genetic association between monoamine oxidase A microsatellite and RFLP alleles and bipolar affective disorder: analysis and meta-analysis. Hum Mol Genet 5: 779–782

Sasaki T, Macciardi FM, Badri F et al (1996) No evidence for association of dopamine D2 receptor variant (Ser311/Cys311) with major psychosis. Am J Med Genet 67: 415–417

Schreiber W, Lauer CJ, Krumrey K, Holsboer F, Krieg JC (1992) Cholinergic REM sleep induction test in subjects at high risk for psychiatric disorders. Biol Psychiatry 32: 79–90

Shah M, Coon H, Holik J, Hoff M, Helmer V, Panos P, Byerley W (1995) Mutation scan of the D1 dopamine receptor gene in 22 cases of bipolar I disorder. Am J Med Genet 60: 150–153

Shaikh S, Ball D, Craddock N et al (1993) The dopamine D3 receptor gene: no association with bipolar affective disorder. J Med Genet 30: 308–309

Sidenberg DG, King N, Kennedy JL (1994) Analysis of new D4 dopamine receptor (DRD4) coding region variants and TH microsatellite in the Old Order Amish family (OOA110). Psychiatr Genet 4: 95–99

Souery D, Lipp O, Mahieu B et al (1996) Association study of bipolar disorder with candidate genes involved in catecholamine neurotransmission: DRD2, DRD3, DAT1, and TH genes. Am J Med Genet 67: 551–555

Spence MA, Flodman PL, Sadovnik AD, Bailey-Wilson JE, Ameli H, Remick RA (1995) Bipolar disorder: evidence for a major locus. Am J Med Genet 60: 370–376

*Stine OC, Xu J, Koskela R et al (1995) Evidence for linkage of bipolar disorder to chromosome 18 with a parent-of-origin effect. Am J Hum Genet 57: 1384–1394

Stine OC, McMahon FJ, Chen L et al (1997) Initial genome screen for bipolar disorder in the NIMH genetics initiative pedigrees: chromosomes 2, 11, 13, 14, and X. Am J Med Genet 74: 263–269

Straub RE, Lehner T, Luo Y et al (1994) A possible vulnerability locus for bipolar disorder on chromosome 21q22.3. Nat Genet 8: 291–296

Strober M, Lampert C, Morrell W, Burroughs J, Jacobs C (1990) A controlled family study of anorexia nervosa. Int J Eating Disord 9: 239–253

Torgersen S (1986) Genetic factors in moderately severe and mild affective disorders. Arch Gen Psychiatry 43: 222–226

Turecki G, Rouleau GA, Mari J, Joober R, Morgan K (1997) Lack of association between bipolar disorder and tyrosine hydroxylase: a meta-analysis. Am J Med Genet 74: 348–352

Vallada H, Craddock N, Vasques L et al (1996) Linkage studies in bipolar affective disorder with markers on chromosome 21. J Affect Disord 41: 217–221

Van Broeckhoven C (1995) Presenilins and Alzheimer disease. Nat Genet 11: 230–232

Verkerk AJMH, Pieretti M, Sutcliffe JS et al (1991) Identification of a gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. Cell 65: 905–914

Vincent JB, Klempan T, Parikh SS et al (1996) Frequency analysis of large CAG/CTG trinucleotide repeats in schizophrenia and bipolar affective disorder. Mol Psychiatry 1: 141–148

Weissman MM, Wickramaratne P, Merikangas KR et al (1984) Onset of major depression in early adulthood: increased familial loading and specificity. Arch Gen Psychiatry 41: 1136–1143

Weissman MM, Merikangas KR, Wickramaratne P, Prusoff BA, Leckman JF, Pauls DL (1986) Understanding the clinical heterogeneity of major depression using family data. Arch Gen Psychiatry 43: 430–434

Weissman MM, Wickramaratne P, Adams PB et al (1993) The relationship between panic disorder and major depression. A new family study. Arch Gen Psychiatry 50: 767–780

Weissman MM, Warner V, Wickramaratne P, Moreau D, Olfson M (1997) Offspring of depressed parents. 10 years later. Arch Gen Psychiatry 54: 932–940

*Winokur G, Coryell W (1991) Familial alcoholism in primary unipolar major depressive disorder. Am J Psychiatry 148: 184–188

Winokur G, Clayton PJ, Reich T (1969) Manic-depressive illness. Mosby, St Louis

Winokur G, Cadoret RJ, Dorzab J, Baker M (1971) Depressive disease: a genetic study. Arch Gen Psychiatry 24: 135–144

Winokur G, Cook B, Liskow B, Flowler R (1993) Alcoholism in manic depressive (bipolar) patients. J Stud Alcohol 54: 574–576

*Winokur G, Coryell W, Akiskal HS, Endicott J, Keller M, Mueller T (1994) Manic-depressive (bipolar) disorder: the course in light of a prospective ten-year follow-up of 131 patients. Acta Psychiatr Scand 89: 102–110

Winokur G, Coryell W, Keller M, Endicott J, Leon A (1995a) A family study of manic-depressive (bipolar I) disease: is it a distinct illness separable from primary unipolar depression? Arch Gen Psychiatry 52: 367–373

Winokur G, Coryell W, Endicott J, Akiskal H, Keller M, Maser JD, Warshaw M (1995b) Familial depression versus depression identified in a control group: are they the same? Psychol Med 25: 797–806

Wozniak J, Biederman J, Mundy E, Mennin D, Faraone SV (1995) A pilot family study of childhood-onset mania. J Am Acad Child Adolesc Psychiatry 34: 1577–1583

Yoneda H, Sakai T, Ishida T, Inayama Y, Nonomura Y, Kono Y, Asaba H (1992) An association between manic-depressive illness and a pseudoautosomal DNA marker. Am J Hum Genet 51: 1172–1173

Zahn TP, Nurnberger JI, Berrettini WH (1989) Electrodermal activity in young adults at genetic risk for affective disorder. Arch Gen Psychiatry 46: 1120–1124

CHAPTER 19

F. Henn

Neurobiology of Affective Disorders

1 Introduction

The initial evolution of a biological explanation of mood and mood disorders came from accidental observations in internal medicine. Two observations resulted in the initial framing of the catecholamine hypothesis of affective disorders. The first came about when iproniazid was introduced for the treatment of tuberculosis. Patients at that time were sent to large sanatoriums and the initial trials of iproniazid were therefore carried out in these large clinic settings. It was noticed that patients on the new medication had pronounced mood elevation with increased sexual interest and had a picture which was essentially hypomanic. It was soon established that iproniazid was a monoamine oxidase inhibitor (MAOI) and could be expected to raise amine levels. The second observation was made in the treatment of hypertension. Reserpine was used in the 1950s and 1960s as the major drug in treating hypertension. Internists noted that up to 20% of the patients treated with reserpine developed severe depressions, and several suicides were reported. The mechanism of action of reserpine appears to be a depletion of catecholamines from storage vesicles, leading to decreased release of catecholamines. Putting these two observations together led to the catecholamine hypothesis of affective disorders, which in simple form stated that catecholamines were elevated in mania and decreased in depression. The initial focus was on the role of norepinephrine (NE), but very quickly interest in the role of serotonin grew, and evidence for this transmitter playing a role in depression accumulated.

The overly simplistic nature of a one-transmitter view of the affective disorders has gradually given way to a more integrative approach in which many transmitter systems and endocrine systems are seen to interact to produce mood disturbances. At present, there is no single accepted biological theory of depression. In this chapter, we will briefly review the evidence implicating major transmitter systems in depression and the role of the hypothalamic–pituitary–adrenal (HPA) axis in mediating stress and depression and will then attempt to show how these results can be integrated and where additional data is still needed. The concept which will emerge suggests that a variety of vulnerability factors may combine to lead to affective disorders.

2 Norepinephrine

We will begin our overview with NE. This system involves only about 15,000 cells in the central nervous system (CNS), located in the locus ceruleus and scattered in the lateral ventricle tegmental fields. This system has widely distributed projection throughout the limbic system, midbrain, cerebellum, and cortex. The distribution suggests a modulatory role for norepinephrine rather than the primary transfer of information. The system appears to play a role in regulating mood and anxiety, controlling the sleep–wake cycle, and even in memory and learning.

In an effort to support a basic role in depressive pathophysiology for NE, considerable effort has gone into looking at metabolite levels of NE in blood and plasma. One serious problem with this is that both urine and blood metabolite levels reflect predominantly peripheral rather than central levels of NE release. These studies have in general yielded conflicting results; there is a tendency to find lower urinary 3-methoxy-4-hydroxyphenylglycol (MHPG) in bipolar probands than in depressive probands. This appears especially true for bipolar I patients, whereas bipolar II patients show no differences (Schatzberg et al. 1989). Similar uncertainty exists with the response of metabolites to antidepressant treatment and with cerebrospinal fluid (CSF) metabolites of NE. Taken as a whole, the metabolite studies thus have significant limitations and do not provide strong support for the role of NE in depression. An alternative strategy has recently been used by Delgado et al. (1993) to look at the role of both serotonin (5HT) and NE in depressive illness. Basically, these studies involve depleting the precursors to NE or 5HT synthesis, presumably resulting in lower CNS levels. What is found is that patients who respond to desipramine worsen when tyrosine is depleted using alpha methyl tyrosine, suggesting that those patients who respond to a medication which targets the NE system are sensitive to changes in the level of NE. Alternatively, patients who respond to selective serotonin reuptake inhibitors (SSRI) do not worsen when NE is depleted, but do worsen when tryptophan, the precursor of 5HT, is lowered, suggesting a particular susceptibility to levels of 5HT. These data suggest a heterogeneity in depressive illness which is repeatedly seen in pharmacological and biological studies.

At the receptor level, two NE receptors have been examined, the α_{a2} receptor and β_b receptor. The presynaptic α_{a2} receptor, which plays a role in controlling the release of NE, has been the focus of a variety of functional and binding studies. The results are marked by inconsistency. There were reports of increased platelet binding of clonidine (Garcia Sevilla et al. 1986; Pandey et al. 1989), which could not be reproduced (Georgotas et al. 1987). Studies of platelet aggregation and cyclic adenosine monophosphate (cAMP) response also have yielded inconsistent results. In the case of the β_b receptor, the initial interest came from the observation made by Sulser et al. (1978) that antidepressants down regulate β_b receptors. This led to

an investigation of postmortem levels of β_b receptors in suicide victims with documented depression compared to controls. Here, the balance of studies support increased β_b receptor density in cortex and hippocampus (Mann et al. 1986). Studies on peripheral cells such as lymphocytes show inconsistent results. The β_b receptor theory of antidepressant action was called into question when newer medications which did not appear to interact with the NE β_b system were introduced, including mianserin and the SSRI. In both cases, administration of the medication chronically to animals does not result in β_b receptor downregulation. This suggests that β receptor downregulation may not play a central role in the action of antidepressants. Work using the animal model of depression, learned helplessness, suggested that perhaps a central role for β receptors in the action of antidepressants should be considered again. In this model, β_b receptors are upregulated when the animals become helpless, and even medications such as SSRI or mianserin will downregulate the β_b receptors which are pathologically elevated, while having no effect on the basal level of receptors. Thus it appears that the postsynaptic change in β_b receptors may be part of a common final pathway in depressive illness, although it appears unlikely that this is a causal factor in the etiology of the illness (Henn et al. 1993).

There appear to be interactions between the NE system and several other CNS systems which could play a role in depression. In particular, the balance between cholinergic activity and noradrenergic activity was postulated by Janowsky et al. (1974) to play a role in depression. The role of the HPA axis in depression has been the focus of a great deal of research following the early observations that cortisol is elevated in depression (Sachar 1967) and that this is often not suppressible using dexamethasone (Carroll 1982). There also appears to be a relationship between adrenergic activity and HPA axis activity (Stokes et al. 1987). This could also be due to the effect of CRF on both systems, since CRF overactivity has also been suggested as having a role in depression (Nemeroff et al. 1984).

3 Serotonin

5HT, like NE, has the distribution of a modulatory system. It arises from a group of midbrain cells located in the raphe and has a widespread distribution through limbic system, cortex, and midbrain. The synapses of 5HT involve a wide number of specific receptors; at latest count 15 have been cloned, and the boutons often appear to release their contents into extracellular spaces without a neuronal connection. Thus it clearly appears to function as a modulatory system in the CNS and has documented roles in the control of sleep, appetite, libido, cognition, and impulsivity. This alone suggests a role in affective disorder in which the vegetative functions of sleep, appetite, libido, and concentration are disturbed. In order to summarize the enormous data available on the role of 5HT in depression, we will look at metabolic studies, receptor studies, and interactions with the HPA axis separately and then attempt to develop an overview of the role of 5HT in affective illnesses.

3.1 Metabolic Studies

The initial studies focusing on blood looked at 5HT levels; however, since the platelets have a powerful uptake system for removing free 5HT from plasma, a characterization of the uptake system in platelets proved more reasonable. Initially, imipramine binding was studied as a measure of uptake sites, but this proved to be somewhat nonspecific since the ligand measures two different receptors. Several studies, but not all, showed a decrease in platelet imipramine binding sites. More recent studies using the more specific ligand paroxetine have not shown consistent differences. Thus the best studies have looked at 5HT uptake by platelets; these studies are numerous and give variable results. A majority suggest that the rate of uptake for 5HT is significantly lower in depression than in controls. However, about 30% of these studies find no difference between patients and controls. Taken together, these data show sufficient variability both within studies and across studies to preclude any reliable interpretation (Meltzer and Arora 1991).

The metabolic studies began with an examination of 5HT turnover looking at the major metabolite 5-hydroxy indolacetic acid (5HIAA) in CSF. This has been done by many groups with and without probenicid treatment, which prevents the reuptake of the 5HT. The most interesting data on 5HIAA levels came from Asberg et al. (1976), who reported that a subgroup of depressive suicides had very low 5HIAA levels. A variety of studies on low 5HIAA showed that very low levels were associated with impulsivity (Faustman et al. 1991) and with such acts as fire-setting and impulsive violence (Virkkunen et al. 1987) rather than depression per se. A summary of the 5HIAA studies in the CSF of depressed patients does not show a lower mean concentration in depression; however, as with the platelet studies, there is a high degree of variability.

Another way of looking at the metabolic control of 5HT is to look at the precursor levels, in this case the levels of L-tryptophan. Here, suggestions that 5HT may well be an important factor in the vulnerability to

depression gain strength. There are suggestions that L-trypophan is lower in depressed patients than in controls (Maes et al. 1990). Repeated studies have shown that depleting L-tryptophan leads to a lowering of mood (Heninger et al. 1992), and when using dietary lowering of L-tryptophan levels in remitted depressed patients, a relapse was seen which was correlated to the lowered levels of CSF 5HIAA, suggesting that this had lowered the levels of central 5HT in these patients. Interestingly, only patients who had responded to SSRI showed this response; if patients only responded to desipramine, lowering L-tryptophan had no effect on mood. When looking at patients still in a depressed state, lowering L-tryptophan levels does not result in a worsening of symptoms. These studies suggest that central levels of 5HT must be one of several factors which influence mood in depressed patients. Another interesting feature of precursor studies is the possibility that the system which leads to the catabolism of L-tryptophan in the liver is induced in depression; one study is consistent with this idea (Maes et al. 1987).

3.2 Receptor Studies

The family of 5HT receptors has become increasingly well defined through molecular techniques; cloning has revealed at least 15 distinct receptor subtypes in three major classes. This has made the interpretation of older binding studies somewhat difficult, since the identification of a class of receptors using pharmacological binding techniques often lacked the specificity to cleanly separate receptor subtypes. Nevertheless, receptor pharmacology of the 5HT system has contributed to understanding the role of 5HT in depression and reinforces the hypothesis that 5HT has a role in this illness.

The $5HT_{1a}$ receptors have been examined in postmortem tissue from depressed patients, and although there is some variability in the data, most well-controlled studies fail to find significant differences in this receptor level. In addition, specific $5HT_{1a}$ agonists are not consistently effective antidepressants, though they have been reported to either speed up antidepressant actions or increase the effectiveness of other antidepressant treatments (Jenkins et al. 1990).

The $5HT_2$ receptor binding has been measured in platelets, and the majority of studies have found increased binding. This is consistent with increased phosphoinositol turnover, which was also shown in one study (Mikuni et al. 1992). The postmortem studies of this receptor in general have found increased levels of receptor binding in frontal cortex (Arango et al. 1992). One problem in assessing the role of this receptor in depression are changes due to active treatments. Most antidepressant drugs induce a down-regulation of this receptor, with the exception of ECT, which induces an upregulation (Lerer 1987). This suggests that the $5HT_2$ receptor changes are probably not primary in causing depression, but rather a secondary consequence of other changes.

3.2.1 *Endocrine Interactions*

In general, HPA axis activity is increased in depression with increased cortisol secretion and decreased feedback control. 5HT has been shown to interact with the HPA axis through partial control of corticotropin-releasing hormone (CRH) secretion. In general, 5HT can be shown to increase CRH release and stimulate HPA axis activity. On the surface, this appears contradictory, since depression appears to be associated with lowered 5HT activity and this should not result in increased HPA axis activity. One hypothesis is that 5HT may regulate feedback control in the HPA axis. Decreased 5HT leads to decreased glucocorticoid and mineralocorticoid receptor function, and this in turn would result in higher cortisol secretion.

3.2.2 *Pharmacological Evidence*

Many tricyclic antidepressants have potent inhibitory effects on 5HT reuptake. This suggests that increasing 5HT centrally may have beneficial effects on depression. This is consistent with the action of MAOI and ECT and was the basis for the design of SSRI. While some antidepressants such as desipramine or specific NE reuptake blockers have little or no effect on 5HT, the majority of effective agents increase the availability of 5HT centrally, consistent with a major role of this transmitter system in depression.

3.3 Role in the Etiology of Depression

The evidence briefly reviewed above suggests two things. First, 5HT plays a major role in the CNS changes which result in depressive illness, and second, 5HT is probably not a primary etiological factor in the onset of depression. As with norepinephrine, this is a major modulatory pathway which clearly plays a role in the regulation of mood. In the case of 5HT, it appears that it may have a central role in control of impulsivity and, as such, dysfunctions in this system may be directly related to suicide. However, the

balance of the evidence suggests that either depression is etiologically heterogeneous with only some cases involving 5HT directly, and/or serotonin is involved in regulating the vulnerability of the CNS to other factors which directly result in depression.

4 Dopamine

Although not classically considered as part of the catecholamine hypothesis, much of the clinical data suggesting that norepinephrine plays a role in depression could equally well apply to dopamine. In view of the central role of dopamine in the neurobiology of psychosis, it has been less discussed in affective disorders. Dopamine is distributed in three primary tracts in the CNS, the nigrostriatal tract, the mesolimbic pathway, and the mesocortical pathway, and innervation of the pituitary. In considering affective changes, we are concerned with the mesolimbic pathway, which has been postulated to be the center of reward behavior and closely involved in mediating the effects of stimulants. Since stimulants can induce hypomania, it appears clear that the dopamine system is related to the control of mood.

4.1 Metabolic Changes

Dopamine turnover has been measured looking at levels of homovanillic acid (HVA) in the CSF of patients using probenicid to block HVA transport. These studies consistently show decreased levels of HVA in the CSF of depressed patients, particularly those with psychomotor retardation (Willner 1983). In view of this, HVA has been looked at in other situations having motoric slowing, such as Parkinson's disease and Alzheimer's disease (Wolfe et al. 1990). In these cases it was also decreased, but in view of the cellular loss and dopamine dysfunction this is no surprise. Nevertheless, it appears that dopamine turnover can best be correlated with activity levels.

4.2 Pharmacology

Stimulants which elevate dopamine clearly have a transient beneficial effect on depression. Both methylphenidate and amphetamine also can serve to predict the response to tricyclic antidepressants (Fawcett and Siomopoulos 1971). Interestingly, this is not true for response to SSRI (Little 1988). In addition, the dopamine reuptake inhibitor nomifensine, which was withdrawn from the market due to side effect problems, was an effective antidepressant. Data on dopamine agonists is also positive; in one study, bromocreptine was found to be as effective as imipramine (Willner 1983). Thus there is some suggestion that stimulation of the dopamine system can relieve depression.

In summary, it is clear that the dopamine system also plays a role in control at least of some aspects of affect. It may be that dopamine is particularly involved in psychomotor symptoms, anhedonia, and emotional blunting. The chronic stress model presented by Willner et al. (1992), which features anhedonia, clearly involves an alteration in the dopamine system. It may be that there is heterogeneity throughout the affective disorders, and certain neurotransmitters or modulators contribute more to the development of one set of symptoms than another.

5 Acetylcholine

The role of acetylcholine in affective disorders really entered the discussion through a study of bipolar disease. It was found very early that cholinergic antagonists could cause a short-term reduction in manic symptoms (Janowsky et al. 1973), and these findings led Janowsky et al. (1974) to propose a cholinergic–adrenergic balance hypothesis of affective disorders. The data in support of this hypothesis are relatively weak and consist mostly of observations concerning the effects on mood of drugs which affect the cholinergic system. The best clinical data on patients showing a relationship between depression and the cholinergic system come from sleep studies. The problem is that these studies depend on the shortening of rapid eye movement (REM) latency and increase in REM density which are seen in major depression, but unfortunately are not specific to this disease. The use of the long-acting muscarinic agonist RS-86 had a greater effect in shortening REM in depressive patients than in controls (Berger et al. 1989), and this appears to be a state phenomena. This is consistent with the report that mitotic response to a cholinergic agonist in eyedrops is greater in depressive patients than in controls.

While the clinical data in this area consist of observations on the effect on mood of drugs with cholinergic activity, strong anticholinergic drugs, while occasionally reported to have anti depressive activity (see Janowsky and Overstreet 1995), are not considered as effective antidepressant treatments. This

suggests that the cholinergic axis may modulate mood, but does not play a primary role in depressive illness. However, animal studies suggest that manipulation of the cholinergic system can lead to a model depression. The Flinders sensitive line of animals bred for sensitivity to an anticholinesterase has been proposed by Overstreet (1993) as an animal model of depression. The animals show decreased weight, learning problems, reduced activity, increased REM sleep with reduced REM latency, and an exaggerated response to chronic mild stress. All this points in the direction of cholinergic modulation of factors related to mood.

6 Hypothalamic–Pituitary–Adrenal Axis and Depression

The human organism is programmed with a variety of adaptive responses to stress, including the sympathetic nervous system and the HPA axis (see Chap. 8, Vol. 1), which regulates the release of hormones from the adrenal medulla. Since depression often appears to be associated with increased life events (see Chap. 21, this volume) which can be considered stressors, and the agents of the sympathetic nervous system appear to play a role in the control of mood, it was logical to look at the HPA axis function in this disorder. Early on, Sachar et al. (1970) showed that cortisol levels in depression appear significantly higher than seen in control subjects. It was found that over 50% of all depressed patients failed to suppress cortisol secretion following a dexamthasone challenge (Stokes et al. 1975). These data suggested that perhaps the stress response was overactive and not appropriately terminated in depression. That there is a loss of feedback control of the system was illustrated by the development of the combined dexamethasone–CRH challenge test, in which a subject receives dexamethasone to suppress cortisol release and later a dose of CRH, which should stimulate release. In normal subjects, there is a dexamethasone dose-related decrease in adrenocorticotropic hormone (ACTH) and cortisol release, but not in depressed patients, in which there is increased ACTH and cortisol release, in spite of the fact that when depressives are directly challenged with exogenous CRH, they show decreased ACTH responses and normal cortisol secretion (Amsterdam et al. 1988).

An additional factor which could play a role is endogenous CRH release. Nemeroff et al. (1991) showed that CRH is elevated in depression and decreased after ECT treatment. This suggests that increased CRH may be present in depression and was supported by postmortem studies looking at CSF CRH(5) and tissue levels of the CRH receptor, which appears to be downregulated in the frontal cortex (Nemeroff et al. 1988). The CRH changes appear to be state-dependent changes which revert to normal following a depressive episode. It is currently not possible to determine whether they reflect primary pathology in depression or a secondary compensation. The fact that some patients have a normal HPA system during a depressive episode suggests either heterogeneous causes or that these abnormalities reflect secondary changes in depression.

7 Hypothalamic–Pituitary–Thyroid Axis and Depression

The hypothalamic–pituitary–thyroid axis (HPT), like the HPA axis, is a hierarchical system in which the hypothalamus produces thyroid-releasing hormone (TRH), which acts in the anterior pituitary to release thyrotropin (THS); this in turn goes to the thyroid gland, where it induces the synthesis of triiodothyronine (T_3) and thyroxin (T_4). It has long been known that thyroid status can have profound effects on mood. Hypothyroidism is part of the differential diagnosis for depression with patients often showing low mood and lethargy. Hyperthyroidism can also be mistaken for mania, with anxious overactive patients presenting in the clinic. This prompted an investigation of the thyroid axis in depression and the development of the concept of subclinical hypothyroidism. Here, elevated TSH at baseline is seen with normal T_3 and T_4. When a TRH stimulation test is carried out on depressed patients, between 25% and 50% will show a blunted TSH response. This reverts to normal in most patients as the depression resolves. Those patients in which the TRH stimulation test does not normalize have an increased risk of relapse.

8 Other Peptide Systems

The first peptide system which should be considered is the opiate peptide system. Historically, this is one of the first systems identified as being associated with depression through the use of morphine to treat depressed patients. This work goes back to the initial use of opiates and was referred to by Kraepelin. Interestingly, it appeared to be a moderately effective treatment with little addiction developing as a result of

exposure to the drug during the illness. With the awareness of the addictive potential of the opiates, this treatment was soon abandoned.

The distribution and action of opiates suggests they may play some role in hippocampal activity and learning as well as an extensive role in the activity of noradrenergic systems. Opiates acutely inhibit locus ceruleus cell firing, and these cells quickly develop tolerance to the inhibitory effects of opioids. The substances appear to act through activation of the second messenger system involving cAMP, and activation of cAMP response element binding protein (CREB) leads to gene activation. These established activities of opiate receptors certainly suggest that they may be involved in the circuits which control affect.

Postmortem studies of suicide victims without history of drug abuse revealed remarkably elevated levels of the mu opiate receptor in young adults (Gross-Iseroff et al. 1990). These receptors showed a manyfold increase over controls, results which should be reproducible since they represent far larger effects than usually seen in biological psychiatry. Small studies of dynorphin in treatment-resistant depression have shown effectiveness, but large-scale replication trials are not available. In the learned helplessness model of depression, inbred helpless strains have been established (Henn et al. 1993). In these strains, mu opiate receptors showed an up to fourfold regionally specific increase over non-helpless animals. These results suggest that the opiate system probably mediated through the mu opiate receptor may also play a role in the control of mood.

Recent data on substance P suggests that this peptide may also play a critical role in the control of mood. Again, the distribution of substance P suggests activity not only in the well-known pain pathways of the spinal cord, but also colocalization with both noradrenergic containing neurons and 5HT neurons. Antidepressant treatment has been reported to result in a downregulation of substance P (Barden et al. 1983), and substance P is raised in inescapable foot shock (Lisoprawski et al. 1981). The substance P antagonist MK-869 at high doses of 300 mg/day was tested in a single reported trial of 70 patients. It had the same time course and efficacy as paroxetine, but neither were statistically better than placebo in this limited study. This suggests that more work on this new avenue for the treatment of depression is necessary.

9 Intracellular Targets

Ultimately, the signals which carry information about stress and mood must act at the level of the individual neuronal cells. Here, changes in function and possibly structure must take place which lead to the behavioral changes associated with mood disorders. This must obviously take place in defined circuits with defined functions, but ultimately individual cells must change. In this section, we will look at what happens intracellularly when signals from amines or peptides alter receptor activity at the cell surface. There are two broad classes of intracellular signal transduction pathways. The first, which we will consider in slightly more detail, are the pathways which are activated by neurotransmitters and neuromodulators. These involve the major classes of second messenger systems, cAMP, cGMP, inositol phosphate, Ca^{2+}, and nitric oxide (NO). The second large class involves neurotrophin and cytokine receptors acting through a series of tyrosine kinases; these have previously not been central to a discussion of the neurobiology of depression, but newer evidence suggests we should consider these pathways as well. The effects of signal transduction pathways in neurons is to initiate changes which underlie neuronal plasticity and hence all behavioral changes.

The second messenger systems activated by the amines and peptides include many such as cAMP activation which are initiated through G protein activation. G proteins are a large class of proteins, with at least 15 different forms, which couple receptor activation to cellular effector function. Receptors which, when stimulated, lead to an increase in the effector molecule, e.g. cAMP, such as the adrenergic β_b receptor, work through a stimulatory G protein know as G_s, while receptors such as the cholinergic muscarinic receptor which inhibit this process work through inhibitory G proteins known as G_i or G_o. There are several other G proteins some of which interact with the phosphinositol system. All the G proteins appear to be made up of three subunits, α_a, β_b, and γ_g. The α_a subunit contains the binding site for guanine nucleotides and has a catalytic site as well which hydrolyzes GTP to GDP plus P_i. When GTP binds with the α_a subunit, the G protein complex dissociates into α_a and $\beta_b\gamma_g$ subunits, both of which bind adenyl cyclase (see Fig. 1). This activates adenyl cyclase and increases cAMP, which then acts on downstream kinases.

One such kinase is the cAMP-dependent phosphokinase A (PKA), which when activated is translocated into the cell nucleus and increases; this initiates gene activity resulting in new protein synthesis. CREB may play a central role in the control of mood, since it can be activated by several of the neuromodulators mentioned above, including NE and 5HT. In addition, Ca^{2+}-dependent kinases can also activate CREB. Two of the intracellular systems which can regulate CREB appear to be altered in a state-dependent way in depression, suggesting this pathway is part of a final

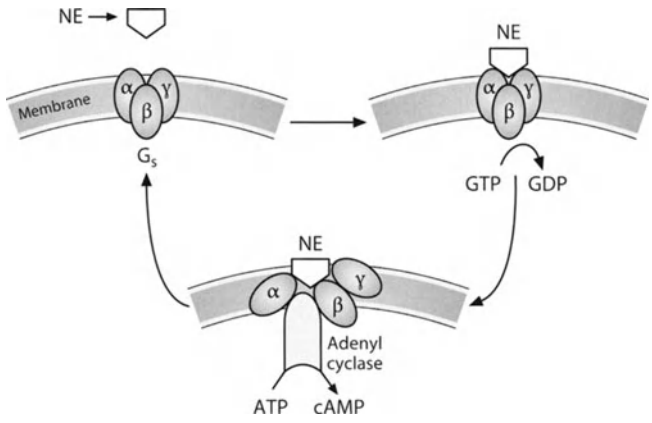


Fig. 1. G_s protein activation by norepinephrine (*NE*) and production of cyclic adenosine monophosphate (*cAMP*)

common pathway in the control of mood. Both changes have been measured in peripheral cells, mononuclear leukocytes, which could reflect similar changes in neurons. The first involves G proteins, which appear to be decreased in depression and elevated in mania, and the second involves internal Ca^{2+} flux in response to mitogen stimulation.

The G proteins have been investigated in mania (Schreiber et al. 1990) and depression (Avissar et al. 1997) and found to be increased in mania by about 25% using an immunoblot analysis and decreased in depression by about 20%. Looking at receptor-enhanced binding of a GTP analogue, depressed patients showed almost no overlap with controls when isoproterenol was used to stimulate binding. Under similar conditions, manic patients showed nearly a doubling of binding. A variety of antidepressant treatments including medication and ECT normalize the level of G_s in depression. In mania, the muscarinic G_i is increased, and this is normalized by lithium. Thus G protein function and level appear to be correlated with mood state.

Ca^{2+} flux also appears to be state dependent when looking at isolated lymphocytes. In general, when lymphocytes are stimulated with a mitogen, there is an increase in intracellular Ca^{2+}, which occurs with oscillating waves of Ca^{2+} appearing intracellularly. Recent studies have shown that, in depression, lymphocytes do not respond nearly as strongly to a standard mitogen challenge. When patients recover, even when treated with interpersonal psychotherapy, without medication, or when treated with a tricyclic medication, the Ca^{2+} response to mitogen normalizes (Aldenhoff et al. 1997).

The above data suggests that there are multiple intracellular state-dependent changes which occur in depression, all of which could work in the direction of less CREB activation. Consistent with this interpretation is the finding that chronic antidepressant treatment activates multiple components of the signal transduction pathway, including, in addition to G_s and cAMP, the cAMP protein kinases (42 Da) and phosphorylation of microtubules (43 Da). Long-term antidepressant treatment with fluoxetine directly has been shown to increase CREB mRNA levels in the hippocampus (D). Thus antidepressants appear to always work in the direction of increasing CREB and inducing gene activation. This suggests the possibility that the long-term plasticity which occurs with the development of depression is due to changes in gene expression leading to changes in neuronal function. The effected cellular pathways are shown in Fig. 2. What is clear is that a variety of inputs including NE, 5HT, and several peptides could react through signal transduction pathways to increase CREB, which may represent one step in a final common pathway necessary for the development of an effective antidepressant response. This suggests that, for depression to begin, some normally expressed pattern of gene activation must be inhibited. If this is a reasonable hypothesis, the next question is which gene expression is necessary to prevent the development of depression. To begin to look at this question, we need to look at those genes activated by CREB. Within the hippocampus, it has been shown that CREB plays a role in the control of some neurotrophin functions. This is of particular interest since the neurotrophins are natural candidates for regulating neuronal plasticity, this being merely an extension of their role in development. One neurotrophin induced by CREB is brain-derived neurotrophic factor (BDNF).

There is considerable information suggesting that changes in the level of BDNF may play a role in antidepressant action; thus, by implication, it may play a role in the initiation of depression. Studies of ECT and long-term administration of antidepressants, including SSRI, NE reuptake inhibitors, and MAOI, all show an upregulation of the expression of BDNF (Nibuya et al. 1995). In addition, the direct administration of BDNF in the learned helplessness model

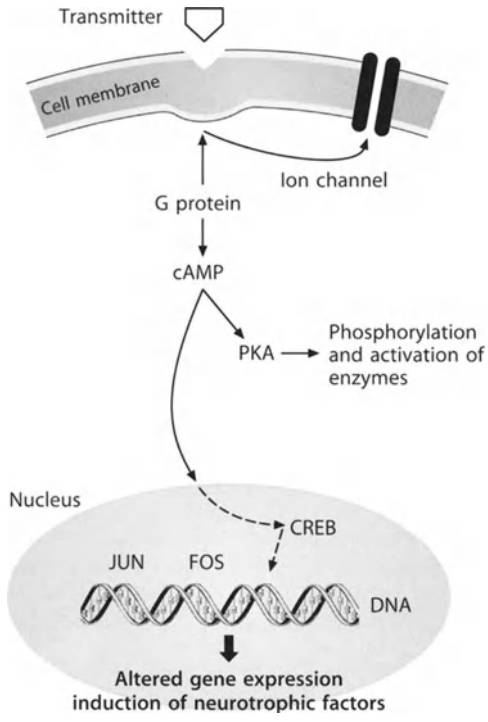


Fig. 2. Altered gene expression through receptor activation. *cAMP*, cyclic adenosine monophosphate; *PKA*, phosphokinase A; *CREB*, cAMP response element binding protein

of depression results in the reversal of depression (Siuciak et al. 1996). Stress also has a direct effect on BDNF levels, and restraint stress inhibits the expression of several neurotropic factors, including BDNF (Smith et al. 1995).

BDNF and other growth factors appear to play an ongoing role in neural plasticity in the fully developed brain, in addition to their more commonly studied role in development. That this plasticity has functional consequences which could play a role in the etiology of depression is suggested by the recent work by Korte et al. (1996), in which they were able to show that, in BDNF knockout mice, hippocampal long-term potentiation (LTP) was altered and could be normalized with the restoration of BDNF through virus-mediated gene transfer. Since LTP appears to be associated with learning in the hippocampus, this could be a relevant pathway for the induction of depression. Furthermore, it has been shown that BDNF can promote sprouting in 5HT (Mamounas et al. 1995) and NE neurons (Sklair-Tavron and Nestler 1995). Thus we can imagine a hypothetical pathway beginning with uncontrolled stress which leads to the activation and subsequent depletion of sympathetic pathways. Concurrently, we would expect that increased HPA axis activity leading to increased vulnerability of hippocampal neurons, the decreased production of cAMP, and changes in the activation of CREB with the subsequent change in neurotrophin levels could lead to subtle structural and functional changes which alter the perception of information, resulting in feelings of hopelessness and ultimately depression. This hypothesis suggests that a multifaceted pathway is possible involving multiple transmitter systems and is consistent with the variety of evidence presented above. It also suggests that, while there may be a final common pathway of changes leading to depression, the driving etiological factors may in fact be heterogeneous.

10 Conclusion

The above hypothesis suggests that alterations in how stress is perceived in the CNS or alterations in a variety of systems mediating input to the CNS, such as the aminergic modulatory systems, could lead to similar dysfunction in the maintenance of a balanced affective system. This system appears to control the processing of information in such a way as to allow the organism to formulate ways to control its environment; when this sense of control is lost, anxiety and helplessness take over. It appears likely that if this becomes chronic, a depressive state results. That this can occur through a variety of systems is not unexpected, and that a primary change in a given system, be it CRH production or control of the noradrenergic β receptor, will lead to compensatory changes in a wide variety of other CNS systems is also to be expected. It does appear that information is given its emotional coloring in the interplay between hippocampus, amygdala, hypothalamus, and frontal cortex and that the interactions between NE, 5HT, the HPA axis, and neuropeptides in this system are critical in establishing the set point for affective control. The data reviewed suggests that no single system appears to be etiologically responsible for all cases of depression and further suggests there may be new points for developing more effective therapies.

This chapter did not address one significant point which may have major neurobiological impact on the control of mood, and that is the issue of biological clocks. Clearly, depression is usually an episodic illness in which sleep patterns are disturbed and seasonal patterns of onset have been detected. Two therapies which have shown some utility, sleep deprivation and light therapy, appear to work in part through the mediation of biological rhythms. Much has been learned about the nature of biological clocks, but up till now no clear connection between the biology of rhythm control and mood disorder has been shown. This remains an area for future work.

In summary, depressive illnesses appear to result from a dysregulation of the systems which mediate the interpretation of stressful events, principally including the modulatory amine systems and the HPA axis. There may be a common final pathophysiological pathway involving these systems which results in changes in NE function, 5HT function, and the HPA axis such that a normal expression of gene products is interrupted. One possible final step which may allow integration of the variety of findings in depression is that patterns of neurotrophic factor production are altered. This hypothesis needs considerably more study, but offers a step beyond the previous amine theories of depression, in that it would reflect possible changes in a variety of modulatory systems. Though the vast amount of data collected on the various systems reviewed has not always been consistent, a general pattern emerges which is consistent with a group of illnesses with heterogeneous etiologies which for the most part effect a single final common pathway. Using the new tools of molecular biology and imaging, it is certain that considerable clarification in the details of such a pathway will emerge in the near future.

11 References

Aldenhoff JB, Dumais-Huber C, Fritzsche M, Sulger J, Vollmayr B (1997) Altered Ca^{2+} homeostasis in single T-lymphocytes of depressed patients. J Psychiatr Res 31(3): 315–322

Amsterdam JD, Maislin G, Winokur A, Berwish N, Kling M, Gold P (1988) The oCRH stimulation test before and after clinical recovery from depression. J Affect Disord 14(3): 213–222

Arango V, Underwood, MD, Mann JJ (1992) Alterations in monoamine receptors in the brain of suicide victims. J Clin Psychopharmacology 12: 8–12

*Asberg M et al (1976) 5-HIAA in the cerebrospinal fluid. A biochemical suicide predictor? Arch Gen Psychiatry 33(10): 1193–1197

Avissar S, Nechamkin Y, Roitman G, Schreiber G (1997) Reduced G protein functions and immunoreactive levels in mononuclear leukocytes of patients with depression. Am J Psychiatry 154(2): 211–217

Barden N, Daigle M, Picard V, Di Paolo T (1983) Perturbation of rat brain serotonergic systems results in an inverse relation between substance P and serotonin concentrations measured in discrete nuclei. J Neurochem 41(3): 834–840

Berger M, Riemann D, Hochli D, Spiegel R (1989) The cholinergic rapid eye movement sleep induction test with RS-86. Arch Gen Psychiatry 46: 421–428

*Carroll GJ (1982) The dexamethasone suppression test for melancholia. Br J Psychiatry 140: 292–304

Delgado PL, Miller HL, Salomon RM et al (1993) Monoamines and the mechanism of antidepressant action: effects of catecholamine depletion on mood of patients treated with antidepressants. Psychopharmacol Bull 29: 389–396

Faustman WO, King RJ, Faull KF, Moses JA Jr, Benson KL, Zarcone VP, Csernansky JG (1991) MMPI measures of impulsivity and depression correlate with CSF 5-HIAA and HVA in depression but not schizophrenia. J Affect Disord 22: 235–239

Fawcett J, Siomopoulos V (1971) Dextroamphetamine response as a possible predictor of improvement with tricyclic therapy in depression. Arch Gen Psychiatry 25: 247–255

Garcia-Sevilla JA, Guimon J, Garcia-Vallejo P et al (1986) Biochemical and functional evidence of supersensitive platelet α_{a2}-adrenoreceptors in major affective disorder: effect of long-term lithium carbonate treatment. Arch Gen Psychiatry 43: 51–57

Georgotas A, Schwertzer J, McCue RE et al (1987) Clinical and treatment effects on 3-H-clonodine and 3-H-imipramine binding in elderly depressed patients. Life Sci 40: 2137–2143

Gross-Iseroff R, Dillon KA, Israeli M, Biegon A (1990) Regionally selective increases in Mu opiate receptor density in the brains of suicide victims. Brain Res 530: 312–316

Heninger GR, Delgado PL, Charney DS, Price LH, Aghajanian GK (1992) Tryptophan-deficient diet and amino acid drink deplete plasma tryptophan and induce a relapse of depression in susceptible patients. J Chem Neuroanat 5: 347–348

*Henn FA, Edwards E, Muneyyirci J (1993) Animal models of depression. Clin Neurosci 1: 152–156

Janowsky DS, Overstreet DH (1995) The role of acetylcholine mechanisms in mood disorders. In: Bloom FE, Kupfer DJ (eds) Psychopharmacology: the fourth generation of progress. Raven, New York

Janowsky DS, El-Yousef MK, Davis JM, Sekerke HJ (1972) A cholinergic-adrenergic hypothesis of mania and depression. Lancet 2: 632–635

Janowsky DS, El-Yousef MK, Davis JM, Sekerke HJ (1973) Parasympathetic suppression of manic symptoms by physostigmine. Arch Gen Psychiatry 28: 542–547

Janowsky DS, El-Yousef MK, Davis JM (1974) Acetylcholine and depression. Psychosom Med 36: 248–257

Jenkins SW, Robinson DS, Fabre LF, Amdang JJ, Messina ME, Reich LA (1990) Gepirone in treatment of major depression. J Clinical Psychopharm 10(3)[Suppl]: 775–855

Korte M, Griesbeck O, Gravel C, Carroll P, Staiger V, Thoenen H, Bonhoeffer T (1996) Virus-mediated gene transfer into hippocampal CA1 region restores long-term potentiation in brain-derived neurotrophic factor mutant mice. Proc Natl Acad Sci USA 93(22): 12547–12552

Kramer MS, Gutler N, Feighner J, Shrivastava R, Carman J, Sramek JJ, Reines SA et al (1998) Distinct mechanism for antidepressant activity by blockade of central substance P receptors. Science 281: 1640–1645

Lerer B (1987) Neurochemical and other neurobiological consequences of ECT: implications for the pathogenesis and treatment of affective disorders. In: Meltzer HY (ed) Psychopharmacology: the third generation of progress. Raven, New York, pp 577–588

Lisoprawski G, Blanc J, Glowinski J (1981) Activation by stress of the habenulo-interpedunclar substance P neurons in rat. Neurosci Lett 25: 47–51

Little KY (1988) Amphetamine, but not methylphenidate, predicts antidepressant response. J Clin Psychopharmacol 8: 177–183

Maes M, De Ruyter M, Suy E (1987) The renal excretion of xanthurenic acid following L-tryptophan loading in depressed patients. Human Psychopharmacol 2: 231–235

Maes M, Jacobs M-P, Suy E, Minner B, Leclercq C, Christiaens F, Raus J (1990) Suppressant effects of dexamethasone on the availability of plasma L-tryptophan and tyrosine in healthy controls and in depressed patients. Acta Psychiatr Scand 81: 19–23

Mamounas LA, Blue ME, Siuciak JA, Altar CA (1995) BDNF promotes the survival and sprouting of serotonergic axons in the rat brain. J Neurosci 15: 7929–7939

Mann JJ, Stanley M, McBride PA et al (1986) Increased serotonin$_2$ and β_b-adrenergic receptor binding in the frontal cortices of suicide victims. Arch Gen Psychiatry 43: 954–969

Meltzer HY, Arora RC (1991) Platelet serotonin studies in affective disorders: evidence for a serotonergic abnormality. In: Sandler M, Coppen A, Harnett S (eds) 5-Hydroxytryptamine in psychiatry: a spectrum of ideas. Oxford University Press, New York, pp 50–89

Mikuni M, Kagaya A, Takahashi K, Meltzer HY (1992) Serotonin but not norepinephrine-induced calcium mobilization of platelets is enhanced in affective disorders. Psychopharmacology 106: 311–314

*Nemeroff CB, Widerlov E, Bissette G et al (1984) Elevated concentrations of CSF corticotropin-releasing factor-like immunoreactivity in depressed patients. Science 226: 1342–1344

Nemeroff CB, Owens MJ, Bissette G, Andorn AC, Stanley M (1988) Reduced corticotropin-releasing factor (CRF) binding sites in the frontal cortex of suicides. Arch Gen Psychiatry 45: 577–579

Nemeroff CB, Bissette G, Akil H, Fink M (1991) Neuropeptide concentrations in the cerebrospinal fluid of depressed patients treated with electroconvulsive therapy: corticotropin-releasing factor, βb-endorphin and somatostatin. Br J Psychiatry 158: 59–63

Nemeroff CB, Krishnan KRR, Reed D, Leder R, Beam C, Dunnick R (1992) Adrenal gland enlargement in major depression. Arch Gen Psychiatry 49: 384–387

Nibuya M, Morinobu S, Duman RS (1995) Regulation of BCMF and trkB mRNA in rat brain by chronic electroconvulsive seizure and antidepressant drug treatments. J Neurosci 15: 7539–7547

Overstreet DH (1993) The Flinders sensitive line rats: a genetic animal model of depression. Neurosci Biobehav Rev 17: 51–68

Pandey GN, Janicak PG, Javaid JI et al (1989) Increased 33-H-clonidine bindings in the platelets of patients with depressive and schizophrenic disorders. Psychiatry Res 28: 73–88

Sachar EJ (1967) Corticosteroids in depressive illness. II. A longitudinal psychoendocrine study. Arch Gen Psychiatry 17(5): 554–567

Sachar EJ, Hellman L, Fukushima DK, Gallagher TF (1970) Cortisol production in depressive illness. A clinical and biochemical clarification. Arch Gen Psychiatry 23(4): 289–298

Schatzberg AF, Samson JA, Bloomingdale KL, Schildkraut JJ (1989) Toward a biochemical classification of depressive disorders. X. Urinary catecholamines, their metabolites, and D-type scores in subgroups of depressive disorders. Arch Gen Psychiatry 46: 260–268

Schreiber G, Avissar S, Danon A, Belmaker RH (1990) Hyperfunctional G proteins in mononuclear leukocytes of patients with mania. Biol Psychiatry 29: 273–280

Siuciak JA, Lewis D, Wiegand SJ, Lindsay RM (1996) Antidepressant-like effect of brain-derived neurotrophic factor. Pharmacol Biochem Behav 56: 131–137

Sklair-Tavron L, Nestler EJ (1995) Opposing effects of morphine and the neurotrophins NT-3, NT-4, and BDNF, on locus coeruleus neurons in vitro. Brain Res 702: 117–125

Smith MA, Makino S, Kvetnansky R, Post RM (1995) Stress alters the expression of brain-derived neurotrophic factor and neurotrophin-3 mRNAs in the hippocampus. J Neurosci 15: 1768–1777

Stokes PE, Pick GR, Stoll PM, Nunn WD (1975) Pituitary-adrenal function in depressed patients: resistance to dexamethasone suppression. J Psychiatr Res 12: 271–281

Stokes PE, Maas JW, Davis JM et al (1987) Biogenic amine and metabolic levels in depressed patients with high versus normal hypothalamic-pituitary-adrenocortical activity. Am J Psychiatry 144(7): 868–872

*Sulser F, Vetulani J, Mobley PL (1978) Mode of action of antidepressant drugs. Biochem Pharmacol 27: 257–271

Virkkunen M, Nuutila A, Goodwin FK, Linnoila M (1987) Cerebrospinal fluid monoamine metabolite levels in male arsonists. Arch Gen Psychiatry 44(3): 241–247

Willner P (1983) Dopamine and depression: a review of recent evidence. Brain Res Rev 6: 211–246

*Willner P, Muscat R, Papp M (1992) Chronic mild stress-induced anhedonia: a realistic animal model of depression. Neurosci Biobehav Rev 16: 525–534

Wolfe N, Katz DI, Albert ML et al (1990) Neuropsychological profile linked to low dopamine: in Alzheimer's disease, major depression, and Parkinson's disease. J Neurol Neurosurg Psychiatry 53: 915–917

CHAPTER

20

D. von Zerssen

Personality and Affective Disorders

Translator: S. Davison

1 Introduction

This chapter examines the relationship between personality and affective disorders, the latter as described in ICD-10 (F30–39) (WHO 1992). Anxiety disorders will only be discussed in relation to their comorbidity with affective disorders. Schizoaffective disorder (Angst and Scharfetter 1990; Marneros et al. 1991; Sauer et al. 1989; Schützwohl et al. 1992; Thau et al. 1991) and suicidal behavior (Angst and Clayton 1998; Bronisch and Wolfersdorf 1996; Engström et al. 1997; Straub et al. 1992), including extended suicide (Marneros 1997; Okumura and Kraus 1996), will not be discussed at all. Finally, the role of personality disorders in affective disorders (e.g. Gunderson and Phillips 1991; see also Chap. 11, Vol. 3, Part 2) will only be dealt with peripherally in connection with the issue of comorbidity.

As in the previous edition of this textbook (see Möller and von Zerssen 1987), this chapter deals primarily with *pre*morbid patient characteristics, even though these may be difficult to distinguish retrospectively from morbid and postmorbid personality changes in patients who have already developed an affective disorder. We have chosen here to concentrate on studies which have contributed to progress in this field since the last edition. They represent significant changes in research methods, as well as in results and their interpretation. The reader is also referred to recent overviews by Akiskal (1996), Cloninger (1994), Hirschfeld and Shea (1992), Klerman and Hirschfeld (1988), Möller (1992), Nietzel and Harris (1990), Tölle (1987), and von Zerssen (1996a–c), specifically in relation to bipolar affective disorder to Goodwin and Jamison (1990, pp. 281–317) as well as Kröber (1988).

It is important to note some recent changes in the classification of affective disorders which are relevant to the subject under discussion. The classical division, in the earlier literature, of depression without an obvious organic basis into endogenous and neurotic forms has been largely abandoned. Instead, the term "major depression" is widely used (see APA 1994). This disorder is divided into depression with and without melancholic features. It is also often divided into primary depression (in the absence of a preexisting psychiatric disorder) and secondary depression (following a preexisting psychiatric disorder). In addition, the clinical picture of dysthymia has been elaborated upon. Both dysthymia and major depression without melancholic features are more commonly associated with personality disorders (Parker et al. 1998) and anxiety disorders (Parker et al. 1999) than is pure depression with melancholic features. The latter corresponds broadly to the clinical picture of endogenous depression. The other forms of depression such as secondary depression, depression without melancholic features, and dysthymic disorder all overlap to some extent with the older concept of neurotic depression.

These relationships explain why, even in more recent studies of patients with neurotic depression (e.g. Uluşahin and Uluğ 1997), questionnaire measures of neuroticism, which tend to be elevated in personality disorders (Assion et al. 1998; Davidson et al. 1985; Fiedler 1997; Pukrop et al. 1998; von Zerssen et al. 1988) and anxiety disorders (Alnæs and Torgersen 1990), are found to be raised in just the same way as they are in studies of the comorbidity of depression with anxiety disorders (Alnæs and Torgersen 1990; Bronisch and Hecht 1990). The reverse is also understandable: questionnaire scores relating to neuroticism, such as those measuring interpersonal sensitivity (Boyce et al. 1993), may not be elevated in major depression with melancholic features. However, they are raised in major depression without melancholic features, because this is the type of depression which is often associated with other mental disorders. Furthermore, a study of the validity of the operational definition of endogenous depression (Zimmerman et al. 1986), the form of depression most closely approximating to major depression with melancholic features, found a comparatively low rate of associated personality disorder (see also Parker et al. 1998). However, there have been some contradictory findings (Grove et al. 1987).

In recent years, there has been an effort to replace the classification of affective disorders based on Kraepelin's categorical concept by the concept of a spectrum, in which more emphasis is placed on the overlap between the recognizable clinical pictures than on their differences (e.g. Angst and Scharfetter 1990; von Zerssen 1996a). In particular, Akiskal (1988, 1989, 1992, 1996), starting from Kraepelin's concept of fundamental states of "manic depressive insanity" (Kraepelin 1913), has conceived a bipolar spectrum in which subclinical forms of an affective disorder are included as minor forms of a bipolar II disorder. According to Akiskal, the affective disorders overlap with four different temperaments (Kraepelin used the term "personal disposition" rather than temperament), namely, a depressive, a cyclothymic, an irritable, and a hyperthymic ("manic" according to Kraepelin) temperament.

2 Definition and Classification of Personality

Personality is defined implicitly in the majority of papers relevant to this chapter as a pattern of relatively persistent ways of experience and behavior, i.e. in the

Table 1. Classification of personality in psychiatric patients according to von Zerssen (in press)

The "pattern of characteristic thoughts, feelings, and behaviors that distinguishes one person from another and that persists over time and situations" (Phares 1988):	
over the whole lifespan or during the last part of it prior to the first symptoms of a mental state disorder	= premorbid personality or primary personality
during the prodromal or initial stages of a disorder	= initial change of personality or "preclinical" personality
during the course of a disorder	= intramorbid personality
after a mental disorder:	= postmorbid personality
with incomplete recovery	= pathological change of personality
with full recovery	= reconstituted personality
during largely symptom-free intervals of a relapsing illness	= intermorbid or interval personality

sense of a trait concept (Wiggins 1997), despite criticisms from certain psychologists (e.g. Mischel and Peak 1982). Phares (1988, p. 4) provides a good example of a definition of this kind: "Personality is that pattern of characteristic thoughts, feelings, and behaviors that distinguishes one person from another and that persists over time and situations." Developmental changes are not ruled out by this a definition.

The term "temperament" is currently used quite often in the U.S. literature. In contrast to the term "character," temperament is considered to be the "biological core" of personality and to be relatively independent of social learning. The term "temperament" is, however, used in different ways within different conceptual frameworks (von Zerssen and Akiskal 1998). It may refer to a personal predisposition to or a fundamental state of an affective disorder (Akiskal 1996). It is also used as a generic term for the predominantly genetically determined components of personality (Cloninger et al. 1993). Alternatively, it is applied to habitual patterns of experience and behavior that can already be observed in childhood (e.g. Merikangas et al. 1998).

It is particularly confusing that, in Cloninger's theory of temperament and character, the different types of temperament described in Akiskal's framework are derived from combinations of Cloninger's hypothetical character dimensions (Cloninger et al. 1998). In view of such discrepancies in the terminology, it is advisable to use the term "personality," which is neutral in relation to etiology, whether biological or social. This makes particular sense in view of research comparing monozygotic twins reared together and apart. The results suggest that it is hardly possible, at least with current research instruments, to distinguish between personality dimensions that are genetically determined and those which are acquired through social learning (Waller et al. 1990).

There are two conceptual classifications of personality which are specific for clinical psychiatry (von Zerssen, in press). One of them relates to the appropriateness of ways of thinking, feeling, and behaving in particular situations, with an emphasis on coping. This implies a ranking from normal, through accentuated, to abnormal personality. The latter includes the particularly maladaptive personality variants which, due to their dysfunctional lifestyles, are referred to as personality disorders. The other psychiatry-specific approach to the classification of personality is to consider the temporal relationship between personality and mental state disorders, such as an affective disorder (Table 1).

A third, more general way of classifying personality refers to typical constellations of personality traits which distinguish larger groups of individuals from one another (categorical classification), or to the common variation of such characteristics (dimensional classification, because the common tendencies in the variation derived by means of factor analysis relate to euclidean dimensions). Both categorical and dimensional classifications can be transformed into each other because categorical personality types usually refer to extremes in a dimensional system (von Zerssen 1994, in press). For the sake of simplicity, both forms of classification will be dealt with here as equivalent.

We have introduced a classification system which arose out of research into the premorbid personalities of patients with different types of an affective disorder, in particular: primary unipolar depression with melancholic features; predominantly manic form of bipolar affective disorder; so-called neurotic depression; anxiety disorders; and psychoses of the schizophrenic spectrum (Pössl and von Zerssen 1990a,b; von Zerssen 1996a). A series of studies on premorbid personality in affective disorders which were based on this classification have since been published (Hecht et al. 1997, 1998; von Zerssen and Pössl 1990; von Zerssen et al. 1994b, 1996, 1998b).

Personality types predominating in different forms of affective disorders were called "affective types." They consist of the "typus melancholicus" (melancholic type) as described by Tellenbach (Kraus 1991, 1996; Peters 1984); the "typus manicus" (manic type) (von Zerssen 1988, 1992), which corresponds to a large extent to

Akiskal's hyperthymic temperament; and a rarer variant of the manic type, called "happy-go-lucky" or "relaxed, easy-going type." In addition, there are the "neurotoid types," which are encountered frequently both in schizophrenic and neurotic disorders. They include an "anxious, insecure type," its rare variant, the "unrealistic, dreamy type," and, finally, a "nervous, tense type."

With the exception of the melancholic type, the names of the affective types give an indication of the most striking features of the personalities they describe (see also the descriptions of the affective types in table form in the articles cited above). They should not be confused with the "affective temperaments" described by Akiskal (1988; see above).

The affective types can be assessed operationally on the basis of biographical information about the premorbid development of psychiatric patients obtained from case notes or from an appropriate interview (von Zerssen et al. 1994a, 1996, 1998a).

The following dimensional personality models are currently given most weight within psychiatry:

- Eysenck's (1990) personality model consisting of three dimensions: extraversion, neuroticism, and psychoticism.
- The Five-Factor Model, consisting of the so-called Big Five: extraversion, neuroticism, conscientiousness, agreeableness, and openness to (new) experience (Wiggins and Trapnell 1997). (The negative of conscientiousness and agreeableness is combined in Eysenck's psychoticism dimension; see Eysenck 1992.)
- Cloninger's model of temperament and character. Temperament is conceptualized as being composed of four dimensions: harm avoidance, novelty seeking, reward dependence, and persistence. Character is conceptualized as consisting of three dimensions: self-directedness, cooperativeness, and self-transcendence (Cloninger et al. 1993).

It can only be briefly mentioned in this context that the dimensions from these three systems, even if named differently, may overlap to a considerable extent (Zuckerman and Cloninger 1996). The reader is referred to Chap. 11 (Vol. 3, Part 2) for a more extensive discussion of the recent personality classification systems.

3 Research Methods

3.1 Assessment Methods

The taxonomies described above are closely linked to the development of methods for the assessment of personality (von Zerssen, in press). Since the publication of the last edition of this textbook, a series of new approaches and techniques have been applied in the study of personality in individuals with affective disorders. In an improvement of previous methods (e.g. Marneros et al. 1991; Tölle 1988), the assessment of premorbid personality from case note data was systematized with the use of extracts from case notes relating exclusively to premorbid biographical patient information (von Zerssen and Pössl 1990). The raters were blind to clinical information about the patient, thus eliminating any diagnostic investigator bias. This procedure was later operationalized by using an extensive list of personality descriptors for the ratings, which could then be analyzed by means of a diagnostic algorithm (von Zerssen et al. 1994a). The findings thus obtained correlate well with the results of a global assignment to the types of premorbid personality (von Zerssen et al. 1994a). In addition to assigning patients to a category, a dimensional description of premorbid personality is also possible by using the scores for the different personality types (von Zerssen et al. 1994b).

This case note analysis served as a model for the development of an interview technique using the same principle of collecting biographical data. The data include demographic variables, family history (excluding psychiatric history of blood relatives), and patients' history prior to the onset of their disorder, subdivided into internal and external life history. The interview is conducted and documented in a verbal protocol by an investigator different from the one who then has to carry out the rating. After special training, the interviewer assesses patients in clinical remission, blind to their diagnosis and other clinical information, and/or normal controls.

This Biographical Personality Interview (BPI; von Zerssen et al. 1996, 1998a,b) has been used extensively in studies of patients with affective disorders in clinical remission, healthy controls (Hecht et al. 1997, 1998), and hitherto healthy close relatives of patients (Hecht et al. 1998). It was applied in conjunction with likewise newly developed personality questionnaires (von Zerssen 1994; von Zerssen et al. 1988) that had also been used in the validation of the BPI (von Zerssen et al. 1996, 1998a,b).

Just as the BPI assesses the affective types described above, as well as wider aspects of premorbid personality, so the semi-structured Affective Temperaments Interview (TEMPS-I; Akiskal et al. 1998; Placidi et al. 1998) has been used to assess the four Kraepelinian fundamental states of manic depressives (cyclothymic, depressive, hyperthymic, and irritable) according to criteria formulated by Akiskal and Mallya (1987). These are operationalized using a series of distinguishing characteristics that are inquired about in a

fairly direct manner. The instrument can be used in a modified form as a self-report questionnaire (Akiskal et al. in press). No findings from clinical studies applying either of these instruments had been published at the time of writing this chapter.

Cloninger's group has developed the Temperament and Character Inventory (TCI) (Cloninger et al. 1993) for the assessment of their postulated dimensions of temperament and character. It should be noted that the Kraepelinian fundamental states of manic depressive illness, called temperaments by Akiskal, are operationalized by particular combinations of values on the three character dimensions. It should also be mentioned that three of Cloninger's temperament dimensions, namely, harm avoidance, novelty seeking, and reward dependence, had formed the conceptual basis for designing the Tridimensional Personality Questionnaire (TPQ; Cloninger 1987).

Akiskal's instruments are based on clinical constructs, while Cloninger's questionnaires were derived from theoretical considerations. On the other hand, Costa and McCrae's (1992) NEO-Five Factor Inventory is based on the Five Factors derived from factor-analytic studies of healthy individuals. The academic origin of the NEO-FFI is clearly apparent from the content of some of the items and the complicated mode of answering.

For this reason, von Zerssen (1994) developed the Six-Factor Test (SFT) as a simple questionnaire to assess the "Big Five" as well as an additional factor of conventional piety. It has been shown to have higher construct validity than the NEO-FFI, particularly in psychiatric patients (Steinmeyer et al. 1996). In addition, a simple Visual Analogue Scale for Six Factors (VAS-SF) (the same factors as measured by the SFT) and a parallel version for use with an informant (VAS-SF) were designed, whose application and analysis are more economical, but whose reliability and validity are necessarily lower than those of the SFT.

The Munich Personality Test (MPT; von Zerssen et al. 1988), like Akiskal's questionnaire, was primarily clinically conceived, but was also developed on the basis of extensive item and factor analyses (as were the NEO-FFI and SFT) in accordance with the classical test theory. It exists in self-assessment and informant assessment forms and has already been applied in many hospital and outpatient studies (Bronisch and Hecht 1989, 1990; Hecht et al. 1997; Heerlein et al. 1996; Sakado et al. 1997; Sauer et al. 1997; Schäfer 1991, 1994; von Zerssen et al. 1996, 1997, 1998a,b) as well as in high-risk studies (Hecht et al. 1998; Lauer et al. 1997; Maier et al. 1992, 1995).

Interviews and questionnaires for the diagnosis of personality disorders are described in Chap. 11 (Vol. 3, Part 2).

In all retrospective studies, including those with patients in clinical remission, the possible relationship between symptoms of psychiatric disorders and self-reported personality traits must be taken into account. This particularly applies to scales for the assessment of neuroticism and similar constructs, as well as those measuring extraversion (Ouimette et al. 1996; Reich et al. 1987), but hardly at all for the rigidity scale of the MPT (Sauer et al. 1997). Older findings on this theme have been confirmed (von Zerssen et al. 1988). On the basis of a prospective study, Shea et al. (1996) concluded that values obtained postmorbidly by means of personality inventories in patients with depression could not be explained as "psychological scars," except possibly in patients with frequent or prolonged episodes.

One research approach, which has only been used once in the context of affective disorder and personality and which differs significantly from interview and questionnaire methods, is participant observation within the family settings of patients discharged from hospital (Peters 1991). The observation of marital interactions in the clinic (Mundt et al. 1994) presents an alternative to participant observation. In comparison, it balances the disadvantage of the unnatural research situation with greater economy and better opportunities for quantitative analysis. With both these methods, conclusions about *pre*morbid behavior can only be made with considerable reservations.

3.2 Research Design and Data Analysis

Clear progress has been achieved in the design of studies in this field. Recent retrospective investigations have nearly always been conducted with patients in remission. In some of these studies, the influence of residual symptoms on the personality measures has been statistically controlled for (e.g. Hecht et al. 1997, 1998; Sakado et al. 1997; Sauer et al. 1997). Moreover, an increasing number of prospective or prospectively designed high-risk studies of women prior to childbirth (Boyce et al. 1991; Marks et al. 1992) and in particular of first-degree relatives of patients with affective disorders have been performed (Hecht et al. 1998; Lauer et al. 1997, 1998; Maier et al. 1992, 1995; Nurnberger et al. 1988; see also Goodwin and Jamison 1990). Even whole families have been investigated (Merikangas et al. 1998). The participant observational study mentioned above falls into this category too (Peters 1991). High-risk studies of patients' close relatives have not yet been running long enough to reach firm and valid conclusions about the affective disorders which develop. However, there are longitudinal epidemiological field studies, some of which cover decades (Clayton

et al. 1994; Crow et al. 1995; Duncan-Jones et al. 1990). One particular type of prospective study was that conducted by a North American research group who investigated monozygotic twins (Kendler et al. 1993a,b). The authors were thus able to draw conclusions about the overall contribution of genetic factors to the variance of personality traits (neuroticism) and illness manifestation (depressive episodes).

Meta-analysis of published data has now also become an established method in the study of personality and affective disorders, in particular depression (Nietzel and Harris 1990). This has been made possible by the considerable increase in the number of original studies which have used quantitative measures of personality traits, in this case emotional dependence and achievement orientation. Increasingly complex statistical methods have also been applied in original studies. These have included so-called path analysis models, which are designed to elucidate causal relationships between variables, thus going beyond merely statistical interdependence. For example, this technique has been used to demonstrate causal relationships between neuroticism and major depression (Kendler et al. 1993b) or between scores on symptom and personality scales (Sauer et al. 1997).

Despite these methodological advances, the results produced do not form a completely consistent picture. The procedures for sampling patients, individuals at risk, and normal controls differ too greatly across studies, and some of the research instruments have too little validity to reach valid conclusions. The convergence of findings using different research approaches will always be the best evidence of their validity (von Zerssen, in press).

4 Results

4.1 Personality and Depression

Research into the relationship between personality and depression most often relates explicitly to unipolar forms of depression. It is usually not apparent whether primary or secondary depression or even so-called double depression (i.e. a major depression superimposed on dysthymic disorder) is being examined. This and other clinical differentiations will therefore only be made explicit below where it is essential in understanding the results, e.g. where particular subtypes of unipolar depression have been compared to one another. The controls in studies of unipolar depression have most frequently been patients with bipolar affective disorder (Hecht et al. 1997; Richter et al. 1993; Sauer et al. 1997; Tölle et al. 1987). Healthy individuals have sometimes been used in conjunction with bipolar patients (Hecht et al. 1997; Sakado et al. 1997) or alone as controls (Boyce et al. 1993; Sato et al. 1994). In contrast, other psychiatric patients have rarely been investigated as control samples (e.g. by Alnæs and Torgersen 1990; Heerlein et al. 1996; Sakado et al. 1997; von Zerssen et al. 1997). Thus results are often not directly comparable to one another. In addition, the clinical diagnostic terms, operationalization, and other methodological aspects differ between studies. The reader is strongly recommended to refer to the original papers where there is any doubt.

4.1.1 Retrospective Questionnaire Studies

Self-report questionnaires are by far the most commonly used tools for assessing personality traits. This is why results obtained by means of these methods will be dealt with first. The results will be discussed under the headings of the "Big Five" personality factors. Even though questionnaires which specifically and directly measure the Five Factors have only rarely been used (e.g. by Bagby et al. 1996a,b), these factors lend themselves most easily to an account which brings together the various studies, because of their similarity to the constructs underlying most of the personality inventory scales.

Neuroticism scores are raised in most retrospective studies of unipolar depression. Because of the well-recognized overlap between depressive symptoms and items on neuroticism scales (Duncan-Jones et al. 1990; von Zerssen et al. 1988), residual symptoms should be taken into consideration even following clinical remission and, whenever possible, controlled for statistically (Sauer et al. 1997). Scales measuring traits that are associated with neuroticism such as interpersonal sensitivity (Boyce et al. 1993), reduced tolerance to frustration (von Zerssen et al. 1997), self-criticism (Bagby et al. 1992), and emotional dependence (Bagby et al. 1992; Nietzel and Harris 1990) frequently show corresponding changes, in particular in comparison with healthy controls (Hecht et al. 1997, 1998) and to some extent also compared to patients with bipolar affective disorder (Richter et al. 1993).

However, no such deviation from the norm was found in a study of Japanese patients after residual depressive symptoms were controlled for (Sakado et al. 1997). In another similar study, also of Japanese patients, von Zerssen et al. (1997) observed a reduced tolerance to frustration but no raised neuroticism. It is difficult to determine whether these findings represent a transcultural difference compared with Western countries because, in at least one of these studies (von Zerssen et al. 1997), patients were specially

recruited who had experienced at least one episode of primary major depression with melancholic features. It seems generally less common to find elevated scores of neuroticism and similar characteristics in patients with than in those without such features (Boyce et al. 1993).

Measures of extraversion can be reduced even in remission from a depressive episode (Janowsky et al. 1998), probably due to residual symptoms (Sauer et al. 1997). However, there is no overall deviation from the norm. Nevertheless, lower scores on individual aspects of extraversion, such as sensation seeking, have been found in patients with unipolar depression (Carton et al. 1995).

There are no consistent findings with regard to aggressiveness. This may relate to the fact that patients in remission following a depressive episode with melancholic features show less marked outward-directed aggression than patients without melancholic features (Matussek et al. 1985). Here, too, the question arises as to the extent to which the results relate to comorbid disorders.

Openness to experience was found to be higher in seasonal depression than in other depressive disorders in one study (Bagby et al. 1996a). Because patients were not in remission and no healthy control group was included, these results may be state dependent. In addition, they might reflect a low degree of openness in other depressive disorders rather than a high degree in seasonal disorder.

Relatively consistent findings have been reported with respect to conscientiousness and other related constructs such as rigidity, orderliness, and obsessionality. Patients with unipolar depression, in particular those with melancholic features, obtain higher scores than healthy controls (Heerlein et al. 1996; Sakado et al. 1997; Schäfer 1991; von Zerssen et al. 1997), commonly also compared to other psychiatric patients (Sakado et al. 1997) and sometimes compared to patients with bipolar affective disorder (Sauer et al. 1997) or patients with depressive reactions (Bronisch and Hecht 1989). There are exceptions (e.g. Hecht et al. 1997, 1998), but the opposite finding, i.e. lower scores than comparison groups, has never been reported. These findings are also consistent with the reduced tolerance of ambiguity, which was postulated on theoretical grounds by Kraus (1988, 1991) and demonstrated empirically in studies of German (Heerlein and Richer 1991; Mundt et al. 1997) and Chilean patients (Heerlein et al. 1996). These findings are also concordant with earlier results relating to different aspects of orderliness referred to in the previous edition of this textbook (Möller and von Zerssen 1987).

The elevated or reduced scores on the traits mentioned so far build up a picture which corresponds by and large to the melancholic type of personality ("typus melancholicus"). Thus increased rigidity and conscientiousness in association with reduced aggressiveness (Furukawa et al. 1998), i.e. marked agreeableness, are correlates of this affective type. In this respect, the findings with regards to neuroticism and extraversion are contradictory and inconsistent (Furukawa et al. 1998; Hecht et al. 1997; Mundt et al. 1997). In contrast, low tolerance to frustration has repeatedly been found to be a correlate of this affective type (Hecht et al. 1997; Mundt et al. 1997; von Zerssen 1996a). Another correlate of the melancholic type is conventional piety as measured by the SFT (von Zerssen 1996a), a finding consistent with older studies of religiosity in patients with depression (Hole 1977).

A few older as well as more recent studies have found elevated scores on scales measuring social desirability (Mundt et al. 1997; von Zerssen et al. 1997). At present, it is not possible to determine whether such scales merely record a tendency to portray oneself in a socially desirable manner on questionnaires or whether they reflect an orientation toward social norms that is relevant to real behavior. The earlier term "lie scales" even suggested a conscious intention to deceive. This, however, is rather unlikely, especially as particularly low scores, thus suggesting "truthfulness," have been found in patients with personality disorders (Assion et al. 1998; von Zerssen et al. 1988). This is an argument that such scales are probably measures of a real orientation toward social norms. For this reason, the respective MPT scale is named accordingly (von Zerssen et al. 1988). Elevated scores in unipolar depressives seem to correspond to what Kraus (1991) has described as "hypernomia" of the melancholic type.

Orientation to social norms, orderliness, and rigidity are less marked, and neuroticism is more marked in patients with so-called neurotic depression (Uluşahin and Uluğ; 1997) or depression without melancholic features. As already mentioned, this subtype of unipolar depression tends to be associated with anxiety and/or personality disorders.

On the basis of clinical observation, Beck (1983) contrasts dependent depression with autonomous depression, which he equates more or less with endogenous depression. Accordingly, he describes the premorbid characteristics of patients with autonomous depression as strongly autonomous. This is, however, in clear contradiction to nearly all other studies in this area.

According to a meta-analysis by Nietzel and Harris (1990), the dependency of patients with depression is particularly well established, better even than achievement orientation, which was previously emphasized in the U.S. literature. The latter trait corresponds to the task orientation and high demands of one's own achievements seen in the melancholic type. A lack of assertiveness is supposedly characteristic of this form

of achievement orientation. This, in addition to the differences between the different types of depression in the prevalence of the melancholic type, might in part explain the contradictory findings in the literature. It is likely that the term "autonomy" does not describe a typical premorbid personality trait in any subgroup of depressive disorders and the postmorbid personality is more accurately described by an increased tendency to dependency. Akiskal (1988) even considers emotional dependency to be a direct consequence of the disorder, as prospective studies have not shown it to be a *pre*morbid personality characteristic (see Sect. 4.1.3).

The premorbid deviations from the norm of patients with so-called neurotic depression appear overall to be fairly nonspecific and at least partly explained by the comorbidity with anxiety and/or personality disorders. The greater specificity of the melancholic type for melancholic forms of primary unipolar depression is limited by the fact that the psychometric profile corresponding to this affective type has also been demonstrated in patients with migraine, albeit not in other pain patients, patients with neuroses, or patients with so-called psychosomatic disorders (Schäfer 1991, 1994).

4.1.2 Retrospective Studies Using Case History Analyses, Interviews, and Observational Methods

Analyses of premorbid developmental biographical data from case histories of psychiatric patients support Tellenbach's concept that the melancholic type is representative of the personality characteristics of a considerable proportion of patients with so-called endogenous depression (Marneros et al. 1991); so do studies blind to clinical diagnosis (Pössl and von Zerssen 1990a,b; von Zerssen and Pössl 1990) and others where the classification of personality types was even strictly operationalized (Ernst et al. 1996; von Zerssen et al. 1994b). There is no such close association with this personality type in neurotic depressive disorders (von Zerssen et al. 1994b). When Angst and Ernst (1996) examined childhood histories of depressives as recorded in Zurich case histories, the features which stood out were primarily those which represent the anxious, insecure and the nervous, tense (neurotoid) types. This is consistent with Söldner's findings (1994; see below).

On the basis of clinical interviews, Mundt et al. (1997) found that 50% of patients with endogenous depression could be classified unequivocally, and a further 25% with reservations (overall, approx. 75%), as being of the melancholic type, which agrees with earlier findings from case history analyses (von Zerssen 1991). This even applies to seemingly very different results (Tölle et al. 1987; see von Zerssen 1991). A method blind to diagnosis and strictly operationalized yielded similar figures (von Zerssen et al. 1998b). The results from a quantitative analysis of scores for the affective types concord with these findings.

One reason for discrepancies in the description of premorbid personality in patients with depression, other than differences in clinical classifications, could be that certain traits of the melancholic type, in particular the achievement orientation and obsessional orderliness, are not as marked at a young age as they are in later life. Thus Söldner (1994) found more similarities than differences in the childhood histories of patients with unipolar endogenous depression and those with neurotic depression on the basis of psychoanalytic interviews. Lack of independence, low self-confidence, high sensitivity, and low tolerance to frustration were predominant. These correspond largely to the anxious, insecure (neurotoid) type (Pössl and von Zerssen 1990b). However, excitability and a proneness to feelings of revenge and aggressive fantasies were also evident, i.e. features of the nervous, tense (neurotoid) type according to Pössl and von Zerssen (1990b). In general, the fantasy lives of the patients with neurotic depression were more strongly developed. Moreover, the patients who later developed unipolar endogenous depression spent more time at home working than playing. This could be an indication that they already had a stronger tendency to be achievement oriented. If this tendency becomes more marked in later life, they could eventually develop the full picture of the melancholic type of personality.

Marital interactions of patients with depression in clinical remission assessed by interview (Matussek et al. 1986) related to postmorbid personality. Such observations show that patients with endogenous depression often display a marked tendency toward autoaggressive processing of frustration. In the systematic observation of such interactions in patients of the melancholic type, it becomes evident that they try to avoid marital disharmony at any price (Mundt et al. 1994).

Participant observation of patients with endogenous depression discharged from hospital, in their natural home surroundings, demonstrates that their families are restricted by the patients' lifestyle intent on orderliness (Peters 1991). The author therefore speaks of a "familia melancholica" in analogy to the individual patients' melancholic type of personality.

4.1.3 Prospective Field and High-Risk Studies

The following types of study belong in the category of prospective field and high-risk studies: prospective epidemiological studies of individuals drawn from the general population who have not yet become ill

(Clayton et al. 1994); studies of female monozygotic twins in some of whom a major depressive episode has been diagnosed retrospectively (Kendler et al. 1993a,b); and prospective studies of individuals who have not yet become depressed but are at high risk because family members are affected or due to critical life events such as childbirth or death of partner. For economic reasons, questionnaire methods have predominantly been used to assess premorbid personality traits in such high-risk studies, whereas interviews have only rarely been applied (e.g. the BPI by Hecht et al. 1998).

The results agree broadly with those of retrospective studies. However, the conclusions that can be drawn about personality are less complex because the number and types of instruments used are limited, largely on economic grounds (exceptions to this include Hecht et al. 1998; Hirschfeld et al. 1989; Lauer et al. 1997). In addition, the breadth of research instruments has been further restricted by the fact that only a small number of studies have been conducted so far. In high-risk studies, global diagnostic categories such as major depression have to be applied, without considering features such as comorbidity and melancholic symptoms because of the relatively small number of patients who develop a clinical disorder. On the other hand, high-risk and prospective field studies have the advantage that the results do indeed relate relatively clearly to *pre*morbid personality. One limitation is that early subclinical disorders can hardly be ruled out with certainty, especially where the data is not collected by fully qualified psychiatrists or psychologists for economic reasons.

Most studies using neuroticism scales have shown elevated scores in individuals who later became ill (Boyce et al. 1991; Clayton et al. 1994; Hirschfeld et al. 1989; Kendler et al. 1993a,b; Marks et al. 1992) and in individuals with a genetically increased risk of developing a disorder (Lauer et al. 1997; Maier et al. 1992; exceptions: Hecht et al. 1998; Hirschfeld et al. 1989) compared with healthy controls or published test norms.

Obsessionality and rigidity of individuals at high familial risk have also been found in some studies to be increased (Lauer et al. 1997; Maier et al. 1992). Hecht et al. (1998), however, did not find this difference. Instead, they observed a reduced tendency to extraversion and a strong orientation to social norms compared with controls. Despite these differences, which are probably due to a selection bias (see above), the results fit in broadly with the picture of the melancholic type of personality. However, only a certain proportion of individuals at risk and individuals with a disorder fall within this category. Others exhibit predominantly "neurotoid" traits, as found in retrospective studies of patients with the diagnosis of major depression that has not been broken down into its subtypes (see above).

4.2 Personality and Bipolar Affective Disorder

4.2.1 Retrospective Studies

The results of recent studies on the relationship between personality and bipolar affective disorder are, as in the earlier literature (see Kröber 1988; Tölle 1987), in part very contradictory. Most retrospective studies of patients in remission using personality questionnaires reveal no, or very few, differences in scores between cases of bipolar affective disorder and unipolar depression (Angst and Ernst 1996; Hecht et al. 1997; Heerlein et al. 1996; Roy 1990; Tölle et al. 1987). Sometimes the scores of patients with bipolar affective disorder correspond more closely to those of healthy controls or to norms derived from the general population (Richter et al. 1993).

Differences from normal values are most likely to be found in terms of raised scores for neuroticism and related constructs. Sometimes they are found with respect to increased scores on measures of hysterical tendencies (Solomon et al. 1996), novelty seeking (Young et al. 1995), and extraversion in predominantly manic patients (von Zerssen 1988).

The results clearly depend on the sample studied. Thus Hecht et al. (1997) combined the scores from patients with unipolar depression and those with bipolar II disorder in their data analysis, as the two groups did not differ significantly on any of the questionnaire and BPI scales. On the other hand, Angst and Ernst (1996) separated out patients with pure mania from those with bipolar manic depressive disorder because, on one scale, their scores differed from the controls in opposition to the group with typical bipolar disorder and the group with unipolar depression – they showed an increased tendency to extraversion, as in the study by von Zerssen (1988). On other scales, they did not differ from healthy controls.

The size of samples studied, their comparability in terms of gender, age, stage of remission, the use or not of relevant statistics to control for confounding factors, and the type of scales used all have a considerable influence on the results obtained. For example, the study conducted by Sauer et al. (1997), which involved a large number of cases, found that even after careful statistical correction for sources of error, significant differences emerged between individuals with bipolar and unipolar disorders on three of the MPT scales. Individuals with bipolar disorder appeared more extraverted than those with unipolar depression or compared with normative values (see also Abou-Saleh

and Coppen 1984; Bagby et al. 1996b); von Zerssen (1988) and Angst and Ernst (1996) also found this with manic patients. Moreover, individuals with bipolar affective disorder also demonstrated less rigidity and orientation to social norms than subjects with unipolar depression.

Overall, the results described so far fit with the concept of a spectrum of affective disorders (von Zerssen 1996a). The individuals with unipolar depression who predominantly have features of the melancholic type of personality such as rigidity, orientation to social norms, a reduced tendency to extraversion, and certain neurotoid tendencies make up one pole of the spectrum. Patients whose illness has an entirely or predominantly manic course and who premorbidly display many features of the manic type of personality, such as a greater tendency to extraversion (von Zerssen 1988), make up the opposite pole. Between the two poles of the affective spectrum (i.e. from bipolar II cases through bipolar I cases with predominantly depressive phases, those with an approximately equal proportion of depressive and manic phases to cases with a slight excess of manic phases), features of the two extreme types of premorbid personality are mixed in a relation which is approximately proportional to the ratio of the manic and depressive components in the long-term course of the illness. The clearest differences from the norm could be expected at either end of the spectrum, namely in patients with unipolar depression or bipolar depression with almost exclusively manic episodes.

This concept is supported by studies using case history analysis blind to clinical diagnosis. As predicted, the proportion of representatives of the melancholic type and those of the manic type of personality does indeed range from unipolar depression of the endogenous type (with melancholic features) through bipolar II and classic bipolar I forms to the rare disorder with an almost purely manic course (von Zerssen and Pössl 1990). This is confirmed by an operationalized analysis of case history data (Ernst et al. 1996; von Zerssen et al. 1994b). Data published by other authors on this topic are by and large consistent with these findings (see von Zerssen 1991, 1992, 1996b).

Biographical data obtained from interviews with patients in clinical remission have led to similar findings (Hecht et al. 1997, 1998). These studies have also enabled a comparison with data from normal controls, which would not be possible with case history data. The scores of normal controls were most similar to those of patients displaying a balanced proportion of manic (M) and depressive (D) episodes in the course of their illness (i.e. $M \approx D$). However, they also did not differ significantly from patients with predominantly manic episodes ($M > D$). This could relate to the fact that the latter group accounted for the smallest number of cases and most of them did not actually fulfill the strict criteria ($M : D \geqslant 4 : 1$) set out by von Zerssen and Pössl (1990). In addition, the control group was not quite representative of the average population because of the exclusion of all cases with a lifetime diagnosis of any psychiatric disorder. This is probably one of the reasons why they tended toward the psychometric profile of the manic type with high scores on the MPT for extraversion and esoteric tendencies, often seen in manic patients (von Zerssen 1988).

In accordance with the results from case history analysis (Pössl and von Zerssen 1990a,b), among patients with affective disorders characterized by predominantly manic episodes, in addition to those with a tendency toward the manic type, there are occasional individuals with introverted, socially inhibited personalities. They can usually be distinguished from those with a melancholic or an anxious, insecure type by their rich fantasy life, which resembles that of the unrealistic, dreamy type. It can be inferred from observations of individual cases and a casuistic article from Japan by Sone and Ueki (1984) that in these manic patients, a constitutional expansiveness was suppressed at an early age by an excessively authoritarian upbringing and turned inward, i.e. introverted in the truest sense of the word. In any case, we are dealing here with a rare exception to the rule of a more extraverted manic type already described in similar terms in the older literature (e.g. Waters 1979; for overview, see von Zerssen 1988, 1992, 1996a).

A somewhat different picture of the premorbid personality of patients with bipolar affective disorder has emerged from studies by Akiskal (1988, 1996) and his Italian colleagues (Cassano et al. 1988, 1992), although these authors have also used a spectrum concept of affective disorders for their research. This concept, however, differs in a few relevant points from that described above (von Zerssen 1996a). Individuals with unipolar depression are not divided into those with and without melancholic features, but rather into those with so far only one episode and those with more than one. Patients with bipolar II disorder include unipolar depressive patients with premorbid hyperthymic personality traits (Cassano et al. 1988, 1992) and patients with cluster B personality disorders, i.e. cases of secondary depression (Akiskal 1988). In addition, patients with purely or predominantly manic episodes are not distinguished from other patients with bipolar I disorder. This grouping of disorders and the narrowing of the personality assessment to the Kraepelinian fundamental states of manic depressive disease makes it understandable why Cassano et al. (1988) only found a depressive temperament and not a temperament corresponding to the melancholic type in their patients with unipolar depression, and why the traits of the melancholic type were also not detected in

bipolar II disorder, a disorder described by Dunner (1983) without reference to premorbid personality characteristics. The broad definition of the bipolar II group, including clinically pure depression with hyperthymic temperament (which corresponds approximately to the manic type) as well as cluster B personality disorders, inevitably increases the prevalence of hyperthymics in bipolar II disorder. The close relationship between hyperthymia and the manic component of an affective illness was probably missed because patients with a clear predominance of manic phases were mixed with those developing only occasional manic and many depressive episodes. This did not allow a proper differentiation between the types of temperament and their relationships to different courses of an affective disorder.

4.2.2 Prospective Field and High-Risk Studies

According to Akiskal's concept, one would expect not only a high frequency of affective disorders, but also of cluster B personality disorders in the families of patients with bipolar affective disorder. This, however, is not the case according to two independent studies (Coryell and Zimmerman 1989; Maier et al. 1995). Coryell and Zimmerman (1989) found only a relative increase of personality disorders from cluster C, in particular obsessive-compulsive personality disorder, in comparison with a group of healthy controls. This fits with the increased rigidity, as measured by the MPT in the study by Maier et al. (1995) of individuals at risk for bipolar disorder. Clearly, features of the melancholic type are not only predominant in the families of patients with unipolar major depression, but also in relatives of individuals with bipolar disorder (see also the high-risk studies carried out by Hecht et al. 1998 and Lauer et al. 1997). This is consistent with the fact that, in this group, the risk of developing a unipolar depression is clearly higher than that of developing a bipolar affective illness (Maier et al. 1995), particularly a bipolar I disorder with a predominantly manic course, in which we would expect premorbid personality features of the manic type to be increased.

Nevertheless, there are indications of features of the manic type in individuals with a high familial risk of bipolar affective disorder (Nurnberger et al. 1988). They show higher scores than controls for hypomanic traits and sensation seeking, particularly so on a scale of disinhibition with respect to drug taking and sexual behavior. However, the possibility cannot be ruled out that subclinical symptoms of a cyclothymic disorder were measured. Yet these would have been regarded by Akiskal (1992, 1996) as part of the subjects' temperament.

Prospective field studies provide interesting information about the development of distinctive premorbid personality features in individuals who later suffer from a bipolar disorder. According to one study, these characteristics clearly do not increase continuously from childhood onward. It appears that it is the other way round, at least in 7-year-old boys. In comparison with probands who remained healthy, they presented similar signs of hostility and motor restlessness as probands who later developed schizophrenia. However, the features were less marked and were only present at this relatively early age (Crow et al. 1995).

This might explain why a study of young Swiss men eligible for military service who later developed a bipolar affective disorder found no significant differences from the norm on any of the scales of the Freiburg Personality Inventory (Clayton et al. 1994), in contrast to others who developed unipolar depression or schizophrenia. The age of the subjects in any psychological study may therefore considerably influence the results. This of course also applies to high-risk studies. Over short time periods, however, the psychometric profiles of individuals at risk of developing an affective disorder remain relatively constant, at least at the age of full maturity, as far as they are not influenced by a mental state disorder which develops during the follow-up interval (Lauer et al. 1998).

All in all, the findings described here from prospective field and high-risk studies are compatible with the concept that features of the manic and the melancholic types are combined premorbidly in bipolar patients or cancel each other out. Thus features of one or the other type of premorbid personality only become evident in those who later develop either a predominantly depressive or an overwhelmingly manic course of the illness (von Zerssen 1996a). The combination of traits of the contrasting affective types may have a positive effect on occupational success, as can be inferred from studies of the employment status of patients with unipolar and bipolar affective disorders (Kraus 1991) and their relatives (Coryell et al. 1989). The results of earlier studies on this topic are confirmed by these findings.

4.3 Personality and the Course of Affective Disorders

A considerable number of studies within the time-frame considered in this chapter have examined the extent to which personality traits can predict the course of illness in patients with affective disorders, and especially whether they predict the response to particular therapeutic measures. It has been repeatedly found that high scores on measures of neuroticism and related concepts, such as harm avoidance, as well as comorbidity with anxiety and personality disorders (which themselves

are typically associated with high scores on such measures; Davidson et al. 1985), worsen the prognosis of an affective disorder (Alnæs and Torgersen 1990).

These factors also worsen the response to therapeutic measures, including the following: acute treatment with antidepressants (Black et al. 1988; Carpenter et al. 1995; Joyce et al. 1994; Möller et al. 1987; Peselow et al. 1992; Shea et al. 1990), even if they only account for a small amount of the variance in outcome (Nelson and Cloninger 1997); psychological interventions (Shea et al. 1990); and the response of seasonal affective disorder to light therapy (Reichborn-Kjennerud and Lingjærde 1996). According to one study, this also applies to the long-term effectiveness of prophylactic lithium therapy (Abou-Saleh and Coppen 1990). The assumption that the compliance of patients, particularly with long-term therapy such as lithium prophylaxis, is dependent on personality factors needs to be tested in controlled therapeutic trials. It can be predicted that those who predominantly have features of the melancholic type of personality would be more compliant than those who tend toward the manic type. It remains unclear whether the usual personality questionnaires are sufficiently valid for this type of investigation. To date, they have not produced many convincing results in relation to the long-term prognosis of depressive disorders (Philipp and Maier 1988; Wittchen et al. 1988; exception: Alnæs and Torgersen 1990). On the other hand, Kröber et al. (1998), who had administered an adjective list to an informant, found that what they called "syntonia" (which probably corresponds to our relaxed, easy-going type; see Sect. 2) had a significant long-term protective effect in terms of relapse of a bipolar manic depressive disorder. These protective, or so-called salutogenic, aspects of personality variables should be given more attention alongside the pathogenic aspects in future research. Until now, there have only been a few isolated examples of studies looking at protective factors in the genesis of depression (e.g. Söldner 1994; see von Zerssen 1996b).

An American multicenter study (Akiskal et al. 1995) investigated whether it is possible, on the basis of personality traits, to predict which patients who have hitherto suffered from major depression will develop manic or hypomanic episodes in the long term. The results were surprising in light of the previously mentioned cross-sectional studies of the relationships between personality and the different courses of an affective disorder. It was not possible, on the basis of the postmorbid assessment of personality traits, to distinguish those who later developed a bipolar I disorder and those who continued to suffer from unipolar depression. On the other hand, those who changed from a unipolar depressive disorder to a bipolar II disorder had been more emotionally labile at index investigation. The authors interpreted this as a confirmation of Akiskal's concept of a close relationship between a cyclothymic temperament and bipolar II disorder. There is clearly an urgent need for research to clarify these nosological problems that are relevant to the prognosis of affective disorders.

5 Interpretation of the Relationships Between Personality and Affective Disorders

The following interpretations of the relationships between personality and affective disorders have been proposed (see Akiskal 1988; Angst 1988; Klerman and Hirschfeld 1988; von Zerssen 1996a):

1. Personality traits are risk factors for or protective factors against an affective disorder.
   a) They act as moderator variables, which either increase (Joffe and Regan 1991) or decrease (Kröber et al. 1998) the risk of developing an affective disorder in conjunction with preexisting vulnerability factors.
   b) They are a direct expression of vulnerability or lack of vulnerability (resilience). Only the pathogenic, i.e. vulnerability aspect will be discussed below. In this context, personality traits may be:
   - Nonspecific indicators of vulnerability whose existence does not allow any clear conclusion about the actual illness-specific vulnerability factors and therefore about the type of disorder to be expected (depressive, manic, or nonaffective; Angst and Ernst 1996; see von Zerssen 1996a in relation to "neurotoid" personality traits in patients with affective disorder).
   - Relatively specific indicators which allow such conclusions (see Akiskal 1996 in relation to the temperaments he derived from Kraepelin's work; von Zerssen 1996a in relation to traits of the manic and melancholic types in patients with different courses of an affective disorder).
   - The actual predisposing factor, in other words the actual illness-specific vulnerability factor itself, i.e. an indispensable precondition without which the affective disorder cannot develop, even if additional precipitating factors may be required. This corresponds approximately to the interpretation by Kraus (1991, 1996) of the role of the features of the melancholic type of personality in melancholia.
2. Personality traits are pathoplastic factors once the illness has developed. Thus personality traits may have nothing to do with the origin of the illness, but may influence features of the disorder and its consequences.

a) They color the symptomatology of the disorder in a personality-specific way, e.g. an anankastic depression in patients with a primary anankastic personality (Videbech 1975).
b) They modify the course of the illness, e.g. personality disorders worsening the prognosis (see Sect. 4.3).
c) They influence recovery, including issues such as compliance and rehabilitation (see Sect. 4.3).

3. Personality traits are an expression of the illness itself:
a) They represent preclinical "formes frustes," as is almost certainly the case in cyclothymic fluctuations seen prior to typical bipolar affective disorder (see Akiskal 1992, 1996; for case descriptions, see Waters 1979).
b) They are a sequela to the disorder in terms of a personality change brought about by the illness, whose nature is not recognized and is then falsely attributed to the patients' premorbid personality. This is accepted by several authors in relation to some of the postmorbid measures of particular personality traits in patients with depression, e.g. emotional dependency, low self-confidence, and low extraversion (see e.g. Akiskal 1996; Angst and Ernst 1996; Hirschfeld and Shea 1992).

These different possibilities are not mutually exclusive. They may be combined either in one type of disorder or in an individual case. Thus some characteristics that are present premorbidly, including autonomic lability and other dysthymic traits such as anxiety and depression proneness, become stronger postmorbidly and at the same time influence the course of the illness unfavorably (see Sect. 4). In any case, such dysthymic traits can often be traced back to childhood in patients with depression (Söldner 1994). When they reach the level of a personality disorder, or when a personality disorder develops in addition, there is a clear predisposition to an early onset of the affective disorder (for cluster B personality disorders, see Black et al. 1988; Brodaty et al. 1991; for clusters C and A, see Parker et al. 1998).

Whether dysthymic traits in patients who later develop a depressive illness are regarded as personality characteristics or as symptoms of a chronic disorder with an early onset that have become habitual, is probably merely a question of semantics (Lauer et al. 1998) and will not be elaborated on further here. It should, however, be emphasized that, according to the author's own observations, which are in agreement with those of Waters (1979), cyclothymic fluctuations in individuals who later develop a full-blown bipolar disorder can usually be differentiated from the individuals' primary personality, even by the patients themselves. This is much less often the case with the hypomanic traits as described as the manic type of personality. In these cases, it appears that the traits are more likely to represent a direct expression of their vulnerability to a predominantly manic course of a bipolar disorder (von Zerssen 1996a).

The role of genetic and environmental factors in the genesis of premorbid personality on one hand and affective disorders on the other can only be hinted at with reference to relevant studies (e.g. Kendler et al. 1993b). Psychoanalytically inclined authors direct their theoretical interpretations essentially toward the influence of early life experiences within the family (Söldner 1994). The considerable influence of genetic factors even on the development of attitudes and (e.g. religious) beliefs has, however, been convincingly demonstrated in recent twin studies (Waller et al. 1990). Molecular genetic studies have revealed associations between polymorphisms of individual genes (or, to be precise, their promoter regions) and personality dimensions as measured by means of questionnaires. Thus there are suggestions of a relationship between the dopaminergic system and extraversion, particularly novelty seeking (Benjamin et al. 1996), and between the serotonergic system and neuroticism (Lesch et al. 1996). A relationship between the monoamine system and extraversion, in particular sensation seeking, had already been postulated by Zuckerman (1985).

Cloninger (1987) suggested a relationship between three of his hypothetical dimensions of temperament and the activity of particular neurotransmitter systems. However, the only relationship that has been convincingly demonstrated so far is that between reduced serotonergic activity, as measured in cerebrospinal fluid, and increased impulsivity, which predisposes to aggressive and autoaggressive behavior, the latter in the form of suicide using "hard" methods (Coccaro 1989; Virkkunen et al. 1989). On the basis of cerebrospinal fluid and psychometric variables, a relationship has been suggested between the dopaminergic system and extraversion in patients with depression (King et al. 1986). There are many open questions left and areas in which research could be further expanded. These include the question of how individual predisposition and social relationships interact in the development of personality on one hand and mental state disorders on the other.

6 References

Abou-Saleh MT, Coppen A (1984) Classification of depressive illness. Clinico-psychological correlates. J Affect Disord 6: 53–66

Abou-Saleh MT, Coppen AJ (1990) Predictors of long-term outcome of mood disorder on prophylactic lithium. Lithium 1: 27–35

Akiskal HS (1988) Cyclothymic and related disorders. In: Georgotas A, Cancro R (eds) Depression and mania. Elsevier, New York, pp 86–95

Akiskal HS (1989) Validating affective personality types. In: Robins LR, Barrett J (eds) The validity of psychiatric diagnosis. Raven, New York, pp 217–227

Akiskal HS (1992) Delineating irritable and hyperthymic variants of the cyclothymic temperament. J Pers Disord 6: 326–342

Akiskal HS (1996) The temperamental foundations of affective disorders. In: Mundt C, Goldstein MJ, Hahlweg K, Fiedler P (eds) Interpersonal factors in the origin and course of affective disorders. Gaskell, London, pp 3–30

Akiskal HS, Mallya G (1987) Criteria for the "soft" bipolar spectrum: treatment implications. Psychopharmacol Bull 23: 68–73

Akiskal HS, Maser JD, Zeller PJ et al (1995) Switching from 'unipolar' to bipolar II. An 11-year prospective study of clinical and temperamental predictors in 559 patients. Arch Gen Psychiatry 52: 114–123

Akiskal HS, Placidi GF, Maremmani I et al (1998) TEMPS-I: delineating the most discriminant traits of the cyclothymic, depressive, hyperthymic and irritable temperaments in a nonpatient population. J Affect Disord 51: 7–19

Akiskal HS, Perugi G, Hantouche E, Haykal RM, Manning S, Connor P (in press). The affective temperament scales of Memphis, Pisa, Paris and San Diego: progress towards a self-rated auto-questionnaire version (TEMPS-A). J Affect Disord

Alnæs R, Torgersen S (1990) Basic character inventory personality traits among patients with major depression, anxiety disorders and mixed conditions. Eur Arch Psychiatr Neurol Sci 239: 303–308

Alnæs R, Torgersen S (1997) Personality and personality disorders predict development and relapses of major depression. Acta Psychiatr Scand 95: 336–342

Angst J (1988) Prämorbide Persönlichkeit – Methodische Probleme. In: Janzarik W (ed) Persönlichkeit und Psychose. Enke, Stuttgart, pp 72–81

Angst J, Clayton PJ (1998) Personality, smoking and suicide: a prospective study. J Affect Disord 51: 55–62

Angst J, Ernst C (1996) Prämorbide und postmorbide Persönlichkeit bei affektiv Erkrankten. In: Gross G, Huber G, Morgner J (eds) Persönlichkeit – Persönlichkeitsstörung – Psychose. Schattauer, Munich, pp 119–132

Angst J, Scharfetter C (1990) Schizoaffektive Psychosen – ein nosologisches Ärgernis. In: Lungershausen E, Kaschka WP, Witkowski RJ (eds) Affektive Psychosen. Schattauer, Stuttgart, pp 23–31

APA (1994) Diagnostic and statistical manual of mental disorders, 4th edn (DSM-IV). American Psychiatric Association, Washington DC

Assion HJ, Müller H, Möller HJ (1998) Selbst- und Fremdrating mit dem Münchner Persönlichkeitstest (MPT) bei Patienten mit und ohne Persönlichkeitsstörung. In: Stieglitz RD, Fähndrich E, Möller HJ (eds) Syndromale Diagnostik psychischer Störungen. Hogrefe, Göttingen, pp 201–203

Bagby RM, Cox BJ, Schuller DR, Levitt AJ, Swinson RP, Joffe RT (1992) Diagnostic specificity of the dependent and self-critical personality dimensions in major depression. J Affect Disord 26: 59–63

Bagby RM, Schuller DR, Levitt AJ, Joffe RT, Harkness KL (1996a) Seasonal and non-seasonal depression and the five-factor model of personality. J Affect Disord 38: 89–95

Bagby RM, Young LT, Schuller DR et al (1996b) Bipolar disorder, unipolar depression and the Five-Factor Model of personality. J Affect Disord 41: 25–32

Beck AT (1983) Cognitive therapy of depression: new perspectives. In: Clayton PJ, Barrett JE (eds) Treatment of depression. Raven, New York, pp 265–284

Benjamin J, Lin L, Patterson C, Greenberg BD, Murphy DL, Hamer DH (1996) Population and familial association between the D4 dopamine receptor gene and measures of novelty seeking. Nat Genet 12: 81–84

Black DW, Bell S, Hulbert J, Nasrallah A (1988) The importance of axis II in patients with major depression. A controlled study. J Affect Disord 14: 115–122

Boyce P, Parker G, Barnett B, Cooney M, Smith F (1991) Personality as a vulnerability factor to depression. Br J Psychiatry 159: 106–114

Boyce P, Hickie I, Parker G, Mitchell P, Wilhelm K, Brodaty H (1993) Specificity of interpersonal sensitivity to non-melancholic depression. J Affect Disord 27: 101–105

Brodaty H, Peters K, Boyce P, Hickie I, Parker G, Mitchell P, Wilhelm K (1991) Age and depression. J Affect Disord 23: 137–149

Bronisch T, Hecht H (1989) Validity of adjustment disorder, comparison with major depression. J Affect Disord 17: 229–236

Bronisch T, Hecht H (1990) Major depression with and without a coexisting anxiety disorder: social dysfunction, social integration, and personality features. J Affect Disord 20: 151–157

Bronisch T, Wolfersdorf M (1996) Persönlichkeit – Persönlichkeitsstörungen und suizidales Verhalten. Roderer, Regensburg

Carpenter D, Clarkin JF, Glick ID, Wilner PJ (1995) Personality pathology among married adults with bipolar disorder. J Affect Disord 34: 269–274

Carton S, Morand P, Bungenera C, Jouvent R (1995) Sensation-seeking and emotional disturbances in depression: relationships and evolution. J Affect Disord 34: 219–225

Cassano GB, Musetti L, Perugi G, Soriani A, Mignani V, McNair DM, Akiskal HS (1988) A proposed new approach to the clinical subclassification of depressive illness. Pharmacopsychiatry 21: 19–23

Cassano GB, Akiskal HS, Savino M, Musetti L, Perugi G (1992) Proposed subtypes of bipolar II and related disorders: with hypomanic episodes (or cyclothymia) and with hyperthymic temperament. J Affect Disord 26: 127–140

Clayton PJ, Ernst C, Angst J (1994) Premorbid personality traits of men who develop unipolar or bipolar disorders. Eur Arch Psychiatry Clin Neurosci 243: 340–346

Cloninger CR (1987) A systematic method for clinical description and classification of personality variants. A proposal. Arch Gen Psychiatry 44: 573–588

Cloninger CR (1994) Temperament and personality. Curr Opin Neurobiol 4: 266–273

Cloninger CR, Svrakic DM, Przybeck TR (1993) A psychobiological model of temperament and character. Arch Gen Psychiatry 50: 975–990

Cloninger CR, Bayon C, Svrakic DM (1998) Measurement of temperament and character in mood disorders: a model of fundamental states as personality types. J Affect Disord 51: 21–32

Coccaro EF (1989) Serotonergic studies in patients with affective and personality disorders. Arch Gen Psychiatry 46: 587–599

Coryell WH, Zimmermann M (1989) Personality disorder in the families of depressed, schizophrenic, and never-ill probands. Am J Psychiatry 146: 496–502

Coryell W, Endicott J, Keller M, Andreasen N, Grove W, Hirschfeld RM, Scheftner W (1989) Bipolar affective disorder and high achievement: a familial association. Am J Psychiatry 146: 983–988

Costa PT, McCrae RR (1992) Revised NEO Personality Inventory (NEOPI-R) and NEO Five Factor Inventory. Professional manual. Psychological Assessment Resources, Odessa/FL

Crow TJ, Done DJ, Sacker A (1995) Childhood precursors of psychosis as clues to its evolutionary origins. Eur Arch Psychiatry Clin Neurosci 245: 61–69

Davidson J, Miller R, Strickland R (1985) Neuroticism and personality disorder in depression. J Affect Disord 8: 177–182

Duncan-Jones P, Fergusson DM, Ormel J, Horwood LJ (1990) A model of stability and change in minor psychiatric symptoms: results from three longitudinal studies. Psychol Med [Suppl] 18: 1–28

Dunner DL (1983) Sub-types of bipolar affective disorder with particular regard to bipolar II. Psychiatr Dev 1: 75–85

Engström G, Alling C, Gustavsson P, Oreland L, Traskman-Bendz L (1997) Clinical characteristics and biological parameters in temperamental clusters of suicide attempters. J Affect Disord 44: 45–55

Ernst C, Angst J, Klesse R, Zuberbühler HU (1996) Unipolar and bipolar disorder: premorbid personality in patients and in community samples. In: Mundt C, Goldstein MJ, Hahlweg K, Fiedler P (eds) Interpersonal factors in the origin and course of affective disorders. Gaskell, London, pp 89–100

Eysenck HJ (1990) Genetic and environmental contributions to individual differences: the three major dimensions of personality. J Pers 58: 245–262

Eysenck HJ (1992) Four ways five facts are *not* basic. Pers Individ Diff 13: 667–673

Fiedler P (1998) Persönlichkeitsstörungen, 4th edn. Psychologie Verlags Union, Weinheim

Furukawa T, Yamada A, Tabuse H, Kawai K, Takahashi K, Nakanishi M, Hamanaka T (1998) Typus melancholicus in light of the five-factor model of personality. Eur Arch Psychiatry Clin Neurosci 248: 64–69

Goodwin FK, Jamison KR (1990) Manic-depressive illness. Oxford University Press, New York

Grove WM, Andreasen NC, Young M, Endicott J, Keller MB, Hirschfeld RM, Reich T (1987) Isolation and characterization of a nuclear depressive syndrome. Psychol Med 17: 471–484

Gunderson JG, Phillips KA (1991) A current view of the interface between borderline personality disorder and depression. Am J Psychiatry 148: 967–975

Hecht H, van Calker D, Spraul G, Bohus M, Wark HJ, Berger M, von Zerssen D (1997) Premorbid personality in patients with uni- and bipolar affective disorders and controls: assessment by the Biographical Personality Interview (BPI). Eur Arch Psychiatry Clin Neurosci 247: 23–30

Hecht H, van Calker D, Berger M, von Zerssen D (1998) Personality in patients with affective disorders and their relatives. J Affect Disord 51: 33–43

Heerlein A, Richter P (1991) Ambiguitätsintoleranz bei affektiven und schizophrenen Störungen. Nervenarzt 62: 269–273

Heerlein A, Santander J, Richter P (1996) Premorbid personality aspects in mood and schizophrenic disorders. Compr Psychiatry 37: 430–434

Hirschfeld RMA, Shea MT (1992) Personality. In: Paykel ES (ed) Handbook of affective disorders, 2nd edn. Churchill Livingstone, Edinburgh, pp 185–194

Hirschfeld RMA, Klerman GL, Lavori P, Keller MB, Griffith P, Coryell W (1989) Premorbid personality assessments of first onset of major depression. Arch Gen Psychiatry 46: 345–350

Hole G (1977) Der Glaube bei Depressiven. Enke, Stuttgart

Janowsky DS, Hong L, Morter S, Silva S, Howe L (1998) Underlying personality characteristics related to affective disorders and suicidality. In: Nomura J (ed) Neurobiology of depression and related disorders. Mie Academic Press, Tsu/Mie, pp 9–30

Joffe RT, Regan JR (1991) Personality and family history of depression in patients with affective illness. J Psychiatr Res 25: 67–71

Joyce PR, Mulder RT, Cloninger CR (1994) Temperament predicts clomipramine and desipramine response in major depression. J Affect Disord 30: 35–46

Kendler KS, Neale MC, Kessler RC, Heath AC, Eaves LJ (1993a) A longitudinal twin study of personality and major depression in women. Arch Gen Psychiatry 50: 853–862

Kendler KS, Kessler RC, Neale MC, Heath AC, Eaves LJ (1993b) The prediction of major depression in women: toward an integrated etiologic model. Am J Psychiatry 150: 1139–1148

King RJ, Mefford IN, Wang C, Murchison A, Caligari EJ, Berger PA (1986) CSF dopamine levels correlate with extraversion in depressed patients. Psychiatry Res 19: 305–310

Klerman GL, Hirschfeld RMA (1988) Personality as a vulnerability factor: with special attention to clinical depression. In: Henderson AS, Burrows GD (eds) Handbook of social psychiatry. Elsevier, Amsterdam, pp 41–53

Kraepelin E (1913) Psychiatrie, 8th edn, vol 3. Klinische Psychiatrie, part 2. Barth, Leipzig (English translation: Manic-depressive insanity and paranoia. Churchill-Livingstone, Edinburgh, 1921)

Kraus A (1988) Ambiguitätsintoleranz als Persönlichkeitsvariable und Strukturmerkmal der Krankheitsphänomene Manisch-Depressiver. In: Janzarik W (ed) Persönlichkeit und Psychose. Enke, Stuttgart, pp 140–149

Kraus A (1991) Neuere psychopathologische Konzepte zur Persönlichkeit Manisch-Depressiver. In: Mundt C, Fiedler P, Lang H, Kraus A (ed) Depressionskonzepte heute: Psychopathologie oder Pathopsychologie? Springer, Berlin Heidelberg New York, pp 42–54

Kraus A (1996) Role performance, identity structure, and psychosis in melancholic manic-depressive patients. In: Mundt C, Goldstein MJ, Hahlweg K, Fiedler P (eds) Interpersonal factors in the origin and course of affective disorders. Gaskell, London, pp 31–47

Kröber HL (1988) Die Persönlichkeit bipolar-manisch depressiv Erkrankender. Nervenarzt 59: 319–329

Kröber HL, Adam R, Scheidt R (1998) Einflüsse auf die Rückfälligkeit bipolar Manisch-Depressiver. Nervenarzt 69: 46–52

Lauer CJ, Bronisch T, Kainz M, Schreiber W, Holsboer F, Krieg JC (1997) Pre-morbid psychometric profile of subjects at high familial risk for affective disorder. Psychol Med 27: 355–362

Lauer CJ, von Zerssen D, Schreiber W, Modell S, Holsboer F, Krieg JC (1998) The pre-morbid psychometric profile is stable over time in subjects at high familial risk for affective disorders. J Affect Disord 51: 45–53

Lesch KP, Bengel D, Heils A et al (1996) Association of anxiety-related traits with polymorphism in the serotonin transporter gene regulatory region. Science 274: 1527–1531

Maier W, Lichtermann D, Minges J, Heun R (1992) Personality traits in subjects at risk for unipolar major depression: a family study perspective. J Affect Disord 24: 153–163

Maier W, Minges J, Lichtermann D, Heun R (1995) Personality disorders and personality variations in relatives of patients with bipolar affective disorders. J Affect Disord 35: 173–181

Marks MN, Wieck A, Checkley SA, Kumar R (1992) Contribution of psychological and social factors to psychotic and non-psychotic relapse after childbirth in women with previous histories of affective disorder. J Affect Disord 24: 253–263

Marneros A (1997) Erweiterter Suizid: Eine blaptophobe Finalität. Z Klin Psychol Psychiatr Psychother 45: 183–195

Marneros A, Deister A, Rohde A (1991) Affektive, schizoaffektive und schizophrene Psychosen. Springer, Berlin Heidelberg New York

Matussek P, Agerer D, Seibt G (1985) Aggression in depressives and psoriatics. Psychother Psychosom 43: 120–125

Matussek P, Luks O, Seibt G (1986) Partner relationships of depressives. Psychopathology 19: 143–156

Merikangas KR, Swendsen JD, Preisig MA, Chazan RZ (1998) Psychopathology and temperament in parents and offspring: results of a family study. J Affect Disord 51: 63–74

Mischel W, Peake PK (1982) Beyond déjà vu in the search for cross-situational consistency. Psychol Rev 89: 730–755

Möller HJ (1992) Die Bedeutung und methodische Problematik der psychiatrischen Persönlichkeitsforschung: der Typus melancholicus und andere Konzepte zur prämorbiden Persönlichkeit von Patienten mit affektiven Psychosen. In: Marneros A, Philipp M (eds) Persönlichkeit und psychische Erkrankung. Springer, Berlin Heidelberg New York, pp 45–65

Möller HJ, von Zerssen D (1987) Prämorbide Persönlichkeit von Patienten mit affektiven Psychosen. In: Kisker KP, Lauter H, Meyer JE, Müller C, Strömgren E (eds) Psychiatrie der Gegenwart, 3rd edn. vol 5: Affektive Psychosen. Springer, Berlin Heidelberg New York, pp 165–179

Möller HJ, Fischer G, von Zerssen D (1987) Prediction of therapeutic response in acute treatment with antidepressants. Results of an empirical study involving 159 endogenous depressive inpatients. Eur Arch Psychiatr Neurol Sci 236: 349–357

Mundt C, Fiedler P, Ernst S, Kohlhoff A (1994) Premorbid personality and observed marital interaction of endogenous depressive patients: first results. Neurol Psychiatry Brain Res 2: 81–86

Mundt C, Backenstrass M, Kronmüller KT, Fiedler P, Kraus A, Stanghellini G (1997) Personality and endogenous/major depression: an empirical approach to typus melancholicus. 2. Validation of typus melancholicus core-properties by personality inventory scales. Psychopathology 30: 130–139

Nelson E, Cloninger CR (1997) Exploring the TPQ as a possible predictor of antidepressant response to nefazodone in a large multi-site study. J Affect Disord 44: 197–200

Nietzel MT, Harris MJ (1990) Relationship of dependency and achievement/autonomy to depression. Clin Psychology Rev 10: 279–297

Nurnberger JI Jr, Hamovit J, Hibbs ED et al (1988) A high-risk study of primary affective disorder: selection of subjects, initial assessment, and 1- to 2-year follow-up. In: Dunner DL, Gershon ES, Barrett JE (eds) Relatives at risk for mental disorder. Raven, New York, pp 161–176

Okumura Y, Kraus A (1996) Zwölf Patientinnen mit erweiterter Selbsttötung – Psychologie, Persönlichkeit, Motivation, Vorgeschichte und psychosoziale Konfliktsituation. Fortschr Neurol Psychiatr 64: 184–191

Ouimette PC, Klein DN, Pepper CM (1996) Personality traits in the first degree relatives of outpatients with depressive disorders. J Affect Disord 39: 43–53

Parker G, Roussos J, Austin MP, Hadzi-Pavlovic D, Wilhelm K, Mitchell P (1998) Disordered personality style: Higher rates in non-melancholic compared to melancholic depression. J Affect Disord 47: 131–140

Parker G, Wilhelm K, Mitchell P, Austin MP, Roussos J, Gladstone G (1999) The influence of anxiety as a risk to early onset major depression. J Affect Disord 52: 11–27

Peselow ED, Fieve RR, DiFiglia C (1992) Personality traits and response to desipramine. J Affect Disord 24: 209–216

Peters UH (1984) Typus melancholicus. In: Freedman AM, Kaplan HI, Sadock BJ (eds) Psychiatrie in Praxis und Klinik, vol 1. Thieme, Stuttgart, pp 338–341

Peters UH (1991) Der Typus melancholicus in Haus und Familie. Vom Typus melancholicus zur Familia melancholica. In: Mundt C, Fiedler P, Lang H, Kraus A (eds) Depressionskonzepte heute: Psychopathologie oder Pathopsychologie? Springer, Berlin Heidelberg New York, pp 55–75

Phares EJ (1988) Introduction to personality, 2nd edn. Scott, Foresman, Glenview/IL

Philipp M, Maier W (1988) Psychopathologische Prädiktion des ambulanten Doxepin-Response: ein Replikationsversuch. Nervenarzt 59: 482–487

Placidi GF, Signoretta S, Liguori A, Gervasi R, Maremmani I, Akiskal HS (1998) The semi-structured affective temperament interview (TEMPS-I). Reliability and psychometric properties in 1010 14–26-year old students. J Affect Disord 47: 1–10

Pössl J, von Zerssen D (1990a) A case history analysis of the "manic type" and the "melancholic type" of premorbid personality in affectively ill patients. Eur Arch Psychiatry Neurol Sci 239: 347–355

Pössl J, von Zerssen D (1990b) Die prämorbide Entwicklung von Patienten mit verschiedenen Psychoseformen. Nervenarzt 61: 541–549

Pukrop R, Herpertz S, Saß H, Steinmeyer EM (1998) Personality and personality disorders. A facet theoretical analysis of the similarity relationships. J Pers Disord 12: 226–246

Reich J, Noyes R Jr, Hirschfeld RM, Coryell W, O'Gorman TW (1987) State and personality in depressed and panic patients. Am J Psychiatry 144: 181–187

Reichborn-Kjennerud T, Lingjærde O (1996) Response to light therapy in seasonal affective disorder: personality disorders and temperament as predictors of outcome. J Affect Disord 41: 101–110

Richter P, Diebold K, Schützwohl M (1993) Zur Persönlichkeit unipolar depressiver und bipolar manisch-depressiver Patienten. Nervenarzt 64: 572–577

Roy A (1990) Personality variables in depressed patients and normal controls. Neuropsychobiology 23: 119–123

Sakado K, Sato T, Uehara T, Sato S, Sakado M, Kumagai K (1997) Evaluating the diagnostic specificity of the Munich Personality Test dimensions in major depression. J Affect Disord 43: 187–194

Sato T, Sakado K, Uehara T, Sato S (1994) Age distribution of the melancholic type of personality (typus melancholicus) in outpatients with major depression: a comparison with a

population without a history of depression. Psychopathology 27: 43–47

Sauer H, Richter P, Saß H (1989) Zur prämorbiden Persönlichkeit von Patienten mit schizoaffektiven Psychosen. In: Marneros A (ed) Schizoaffektive Psychosen. Springer, Berlin Heidelberg New York, pp 109–118

Sauer H, Richter P, Czernik A, Ludwig-Mayerhofer W, Schöchlin C, Greil W, von Zerssen D (1997) Personality differences between patients with major depression and bipolar disorder – the impact of minor symptoms on self-ratings of personality. J Affect Disord 42: 169–177

Schäfer ML (1991) Migräne und Persönlichkeit. Enke, Stuttgart

Schäfer ML (1994) Typus melancholicus as a personality characteristic of migraine patients. Eur Arch Psychiatry Clin Neurosci 243: 328–339

Schützwohl M, Diebold K, Richter P (1992) Zur Persönlichkeit schizophrener und schizoaffektiver Patienten. Schweiz Arch Neurol Psychiatr 143: 541–551

Shea MT, Pilkonis PA, Beckham E, Collins JF, Elkin I, Sotsky SM, Docherty JP (1990) Personality disorders and treatment outcome in the NIMH Treatment of Depression Collaborative Research Program. Am J Psychiatry 147: 711–718

Shea MT, Leon AC, Mueller TI, Solomon DA, Warshaw MG, Keller MB (1996) Does major depression result in lasting personality change? Am J Psychiatry 153: 1404–1410

Söldner ML (1994) Depression aus der Kindheit. Vandenhoeck & Ruprecht, Göttingen

Solomon DA, Shea MT, Leon AC et al (1996) Personality traits in subjects with bipolar I disorder in remission. J Affect Disord 40: 41–48

Sone K, Ueki H (1984) Vergleichende Forschung über die manischen Zustände zwischen der monopolaren Manie und der manisch-depressiven Erkrankung. Z Klin Psychol Psychopathol Psychother 32: 248–259

Steinmeyer EM, Pukrop R, Herpertz S, Saß H (1996) Facettentheoretische Konstruktvalidierung des NEO-Fünf-Faktoren-Inventars (NEO-FFI) und des Sechs-Faktoren-Tests (SFT). In: Möller HJ, Engel RR, Hoff P (eds) Befunderhebung in der Psychiatrie: Lebensqualität, Negativsymptomatik und andere Entwicklungen. Springer, Berlin Heidelberg New York, pp 31–38

Straub R, Wolfersdorf M, Keller F, Hole G (1992) Persönlichkeit, Motivation und Affektivität als modulierende Faktoren suizidalen Verhaltens bei depressiven Frauen. Fortschr Neurol Psychiatr 60: 45–53

Thau K, Lenz G, Rieder N, Kubinger K, Grisar E (1991) Persönlichkeitsprofile bei bipolar schizoaffektiven Psychosen. Ein Vergleich zu bipolar affektiven Psychosen. Nervenarzt 62: 682–688

Tölle R (1987) Persönlichkeit und Melancholie. Nervenarzt 58: 327–339

Tölle R (1988) Beziehungen zwischen Persönlichkeit und Psychose. In: Janzarik W (ed) Persönlichkeit und Psychose. Enke, Stuttgart, pp 82–90

Tölle R, Peikert A, Rieke A (1987) Persönlichkeitsstörungen bei Melancholiekranken. Nervenarzt 58: 227–236

Uluşahin A, Uluğ B (1997) Clinical and personality correlates of outcome in depressive disorders in a Turkish sample. J Affect Disord 42: 1–8

Videbech T (1975) A study of genetic factors, childhood bereavement, and premorbid personality traits in patients with anancastic endogenous depression. Acta Psychiatr Scand 52: 178–222

Virkkunen M, de Jong J, Bartko J, Linnoila M (1989) Psychobiological concomitants of history of suicide attempts among violent offenders and impulsive fire setters. Arch Gen Psychiatry 46: 604–606

von Zerssen D (1988) Der "Typus manicus" als Gegenstück zum "Typus melancholicus" in der prämorbiden Persönlichkeitsstruktur affektpsychotischer Patienten. In: Janzarik W (ed) Persönlichkeit und Psychose. Enke, Stuttgart, pp 150–171

von Zerssen D (1991) Zur prämorbiden Persönlichkeit des Melancholikers. In: Mundt C, Fiedler P, Lang H, Kraus A (eds) Depressionskonzepte heute: Psychopathologie oder Pathopsychologie? Springer, Berlin Heidelberg New York, pp 76–94

von Zerssen D (1992) Der "Typus manicus" – eine Variante der Zyklothymie? In: Marneros A, Philipp M (eds) Persönlichkeit und psychische Störung. Springer, Berlin Heidelberg New York, pp 72–86

von Zerssen D (1994) Persönlichkeitszüge als Vulnerabilitätsindikatoren: Probleme ihrer Erfassung. Fortschr Neurol Psychiatr 62: 1–13

von Zerssen D (1996a) "Melancholic" and "manic" types of personality as premorbid structures in affective disorders. In: Mundt C, Goldstein MJ, Hahlweg K, Fiedler P (eds) Interpersonal factors in the origin and course of affective disorders. Gaskell, London, pp 65–85

von Zerssen D (1996b) Neuere Untersuchungen zur prämorbiden Persönlichkeit bei Patienten mit affektiven Erkrankungen. In: Möller HJ, Deister A (eds) Vulnerabilität für affektive und schizophrene Erkrankungen. Springer, Wien New York, pp 89–102

von Zerssen D (1996c) Forschungen zur prämorbiden Persönlichkeit in der Psychiatrie der deutschsprachigen Länder: Die letzten drei Jahrzehnte. Fortschr Neurol Psychiatr 64: 168–183

von Zerssen D (in press) Diagnostik der prämorbiden Persönlichkeit. In: Stieglitz RD, Baumann U, Freyberger HJ (eds) Psychodiagnostik in Klinischer Psychologie, Psychiatrie, Psychotherapie, 2nd edn. Thieme, Stuttgart

von Zerssen D, Akiskal HS (1998) Personality factors in affective disorders: historical developments and current issues with special reference to the concepts of temperament and character. J Affect Disord 51: 1–5

von Zerssen D, Pössl J (1990) The premorbid personality of patients with different subtypes of an affective illness. Statistical analysis of blind assignment of case history data to clinical diagnoses. J Affect Disord 18: 39–50

von Zerssen D, Pfister H, Koeller DM (1988) The Munich Personality Test (MPT) – a short questionnaire for self-rating and relatives' rating of personality traits: formal properties and clinical potential. Eur Arch Psychiatr Neurol Sci 238: 73–93

von Zerssen D, Pössl J, Gruben S, Tauscher R, Barthelmes H (1994a) An operationalized procedure for the recognition of premorbid personality types in biographical case notes on psychiatric patients. Eur Arch Psychiatry Clin Neurosci 243: 256–272

von Zerssen D, Tauscher R, Pössl J (1994b) The relationship of premorbid personality to subtypes of an affective illness. A replication study by means of an operationalized procedure for the diagnosis of personality structures. J Affect Disord 32: 61–72

von Zerssen D, Barthelmes H, Black C, Breu P, Garczynski E, Hecht H, Pössl J, Wesel E (1996) Das Biographische Persönlichkeits-Interview (BPI) – ein Forschungsinstrument zur Erfassung der prämorbiden Persönlichkeit. In: Möller HJ, Engel RR, Hoff P

(eds) Befunderhebung in der Psychiatrie: Lebensqualität, Negativsymptomatik und andere aktuelle Entwicklungen. Springer, Berlin Heidelberg New York, pp 303–307

von Zerssen D, Asukai N, Tsuda H, Ono Y, Kizaki Y, Cho Y (1997) Personality traits of Japanese patients in remission from an episode of primary unipolar depression. J Affect Disord 44: 145–152

von Zerssen D, Pössl J, Hecht H, Black C, Garczynski E, Barthelmes H (1998a) The Biographical Personality Interview (BPI) – a new approach to the assessment of premorbid personality in psychiatric research. Part I: development of the instrument. J Psychiatr Res 32: 19–25

von Zerssen D, Barthelmes H, Pössl J, Black C, Garczynski E, Wesel E, Hecht H (1998b) The Biographical Personality Interview (BPI) – a new approach to the assessment of premorbid personality in psychiatric research. Part II: psychometric properties. J Psychiatr Res 32: 25–35

Waller NG, Kojetin BA, Bouchard TJJ, Lykken DT, Tellegen A (1990) Genetic and environmental influences on religious interests, attitudes, and values: a study of twins reared apart and together. Psychol Sci 1: 138–142

Waters BGH (1979) Early symptoms of bipolar affective psychosis. Research and clinical implications. Can Psychiatr Assoc J 24: 55–60

WHO (1992) The ICD-10 classification of mental and behavioural disorders. Clinical descriptions and diagnostic guidelines. World Health Organization, Geneva

Wiggins JS (1997) In defense of traits. In: Hogan R, Johnson J, Briggs S (eds) Handbook of personality psychology. Academic Press, San Diego, pp 95–115

Wiggins JS, Trapnell PD (1997) Personality structure: the return of the big five. In: Hogan R, Johnson J, Briggs S (eds) Handbook of personality psychology. Academic Press, San Diego, pp 737–765

Wittchen HU, Lässle R, Bronisch T, Krieg JC, Cording-Tömmel C, von Zerssen D (1988) Zur Prognostik depressiver und Angstsyndrome. In: Wittchen HU, von Zerssen D Verläufe behandelter und unbehandelter Depressionen und Angststörungen. Springer, Berlin Heidelberg New York, pp 211–231

Young LT, Bagby RM, Cooke RG, Parker JD, Levitt AJ, Joffe RT (1995) A comparison of Tridimensional Personality Questionnaire dimensions in bipolar disorder and unipolar depression. Psychiatry Res 58: 139–143

Zimmerman M, Coryell W, Pfohl B, Stangl D (1986) The validity of four definitions of endogenous depression. II. Clinical, demographic, familial, and psychosocial correlates. Arch Gen Psychiatry 43: 234–244

Zuckerman M (1985) Sensation seeking, mania, and monoamines. Neuropsychobiology 13: 121–128

Zuckerman M, Cloninger CR (1996) Relationship between Cloninger's, Zuckerman's, and Eysenck's dimensions of personality. Pers Individ Diff 21: 283–285

CHAPTER

21

G.W. Brown

Role of Life Events in the Causation of Affective Disorders

1 Introduction

Life-event research has arguably proved to be the most effective way of gaining insights into how the external environment helps to generate and perpetuate depressive conditions. An early achievement was to make clear that concepts such as the amount of mere change in activity brought about by an event is largely irrelevant, and it was essential to deal directly with the meaning of events and their emotional impact (Brown and Harris 1978). It has also proved important to pay some attention to ongoing difficulties that can either be brought about by an event (the death of a husband leading to financial problems) or lead to an event (a marital difficulty eventually ending in a separation).

In dealing with meaning, two perspectives have proved productive. The first can be summed up by the statement that we cannot fully know the meaning of most life events until they can be related to the individual's relevant plans and concerns. However, while the mechanisms involved are clearly cognitive, they are not necessarily matters a person is either entirely conscious of or willing or able to report. Concerns "are largely dormant demons: They are dispositions that remain silent as long as conditions conform to the standard, within reasonable limits" (Frijda 1986, p. 336). One way of conceiving of these concerns is in terms of plans and purposes that typically stem from role activity and in that sense are clearly social in nature, e.g. a woman's wish to move from an overcrowded and damp flat to give children "a better start in life". A second perspective concerning meaning assumes that evolutionary-based response patterns that guide us in terms of what to want or to avoid are also likely to be involved. The woman planning to move from her damp flat may therefore experience shame on learning her plans have fallen through if this disappointment results from a friend's mistrust that she will repay the loan that has been planned. At the same time, she may experience increased anxiety if she perceives her child's asthma as likely to worsen again during the coming winter. There are therefore behavioural systems sensitive to a particular range of stimuli. Of course, such responses will be influenced by cultural display rules and individual differences of various kinds, but there is little doubt that such evolutionary-based behavioural systems need to be taken into account if we are to understand the role life events play in the aetiology of common psychiatric disorders such as depression.

The only life-event instrument to deal with both kinds of meaning has been the Life Events and Difficulty Schedule (LEDS) developed in London, and this review will be largely restricted to research based on it. It has been used to estimate *general* as well as various kinds of *specific* threats. It does this by the use of contextual ratings made by the investigator which take into account a person's likely plans and concerns of likely relevance for the event insofar as these can be indirectly assessed from a person's current circumstances and biographical detail. The use of such ratings is also methodologically important, as they go some way to rule out reporting artefacts since such ratings take no account of reported feelings about the event or difficulty or whether or not a disorder followed. This is achieved by means of consensus meetings in which the interviewer holds back from a team of raters such information when describing the event and its immediate context. It is possible to argue that this procedure rules out possible bias and, if anything, will result in a conservative estimate of any aetiological effect. General guidelines for rating severe threat are given in an extensive rating manual containing thousands of examples listed in terms of over 100 event categories (e.g. subject's demotion at work and subject's unplanned pregnancy). Only events having long-term threat in the sense of its presence some 10–14 days after its occurrence are considered for a severe threat rating. The LEDS has been used to study a variety of psychiatric conditions and has been successfully employed in a wide variety of cultural settings. Nothing has emerged from the employment of other life-event instruments that would seriously threaten the LEDS-based findings to be reviewed.

2 Three Key Early Findings

The first use of the LEDS to study depressive conditions occurred in the early 1970s and involved both a patient series seen at the Maudsley Hospital in South London together with a random series of women from the local Camberwell population (Brown and Harris 1978, 1989). All the women were aged between 18 and 65. The Present State Examination (PSE) was used to assess psychiatric state (Wing et al. 1974). The initial threshold of caseness for depression was settled by two psychiatrists (John Copeland and John Cooper) who had previously worked on the PSE. They went by their experience of affective disorder in out-patient practice (Brown and Harris 1978). Later work operationalised this threshold in terms of core symptoms of depression (Finlay-Jones et al. 1980). This has been found, if anything, to be somewhat higher than that of the computer algorithm (ID/CATEGO) of the PSE (Wing

and Sturt 1978) and the Research Diagnostic Criteria (Spitzer et al. 1978) which were developed somewhat later (Dean et al. 1983).

Three key findings emerged:

1. Most onsets of depression were preceded by a "provoking agent" (either a *severely threatening event* or a *major difficulty* lasting at least 2 years). As already noted, only long-term threat in the sense of being present some 10–14 days after the event raised the risk of a depressive onset, and this time element is built into the definition of a severely threatening event. Events with short-term threat alone did not raise the risk. These results emerged despite the use of contextual ratings of events that are approximate and probabilistic and based on a limited amount of information. Again, as already noted, in rating the severity of threat of life events, the investigator did her best to assess relevant plans and purposes by taking into account current circumstances and any biographical detail that appeared relevant. A woman, for example, experiencing a second miscarriage after persistent attempts to have her first child would probably have the event rated severe, but her first miscarriage shortly after marriage would have been rated contextually upsetting but not severely so.
2. However, despite events and difficulties playing a substantial aetiological role in the sense that the majority of onsets of depression in the general population were preceded by at least one such severe threat-provoking agent, the likelihood of a depressive episode was very low without the additional presence of one or more ongoing vulnerability factors of a psychosocial nature. One key vulnerability factor in the original Camberwell research was the lack of effective confiding in a core relationship, particularly where a partner or boyfriend was involved. While this particular result has to be treated with caution given that the material on which the rating was based was provided by the woman after the onset of depression, subsequent prospective enquiries have confirmed the central importance of such background psychosocial factors (Brown et al. 1990c).
3. There was no clear-cut link between the presence of a provoking agent in the Camberwell patient series and diagnostic type. Provoking agents were somewhat less common before onset in the patient series as a whole than onsets of depression in the general community not receiving psychiatric care and, while somewhat less common before endogenous-type depressive conditions when defined in clinical terms, there was little difference from the experience of those with "neurotic" depressive conditions, a result reported earlier by Paykel et al. (1971) and later in several studies (e.g. Katschnig et al. 1986; Bebbington and McGuffin 1989).

3 Subsequent Findings

In what follows, findings from research that has followed this first LEDS study of depression will be outlined. The original research reports (which will be cited) should be consulted for details concerning measurement and design.

3.1 Meaning of Severe Events and Onset

Table 1 deals with the role of severe events based on prospective enquiry carried out in the inner-city area of Islington in North London. It is a typical result. The research was based on 400 women with at least one child at home. They were largely working class, with one fifth living alone with their children. The women were followed up at 12-month intervals. The table deals with the 303 women who, at the time of first contact, did not qualify as patients with depression and who might therefore develop such an episode in the following year. Twenty-nine of the 32 onsets in the first follow-up year were preceded by at least one severe event involving long-term threat in the prior 6 months, most within some 8 weeks of onset. In terms of specific rather than general threat, research has shown that such events typically involved loss, if this was defined not only by a loss of person but loss of a role or cherished idea about oneself or someone close (Finlay-Jones and Brown 1981).

The relevance of the contextual ratings of long-term threat based on a crude assessment of relevant plans and purposes was confirmed by the use of a direct measure of emotional commitment to various role domains made at the time of first contact and therefore

Table 1. Onset rate among Islington women in terms of provoking agent status

Provoking agent status	Onsets (n)	Onset rate (%)
No provoking agent	2/153	1
Provoking agent		
Major difficulty only	1/20	5
Severe event	29/130	22
Total	32/303	11

$\chi^2 = 36.60$, 2 df, $p < .001$.

established before any event or onset occurring in the following year. Table 2 shows that, where a severe event in the follow-up year "matched" a prior area of high commitment (e.g. a child's delinquency in the context of a woman's high commitment to motherhood), the risk of an onset of depression was considerably increased.

However, although loss is typically present, it may not be the factor of central aetiological importance. Table 3 illustrates this by the use of a hierarchical scheme dealing with the likely meaning of each severe event (or a closely related sequence of such events) occurring in a 2-year period excluding any time spent depressed (Brown et al. 1995). Various types of specific meaning are considered. Events are only rated lower down the scale such as subcategory vii ("other key loss") if they have failed to quantify for a higher rating such as subcategory iii ("humiliation: put down"). The design of the rating scheme was influenced by an evolutionary perspective on depression (Gilbert 1989).

The first three subcategories (i–iii) concern possible types of humiliation. The ratings assume that one consequence of such events was likely to be either a sense of being put down or a marked devaluation of self. The first subcategory, for example, includes any separation from a partner or lover where he either took the initiative or where the woman was forced to leave or break off a relationship because of violence or discovery of infidelity.

Events associated with subcategory iv ("entrapment") had to have failed to meet criteria for the three humiliation subcategories. Such events underlined the fact of being imprisoned in a punishing situation that had gone on for some time. The table goes on to deal with four types of loss (in the absence of humiliation or entrapment) and finally danger, a residual group involving threat of a future loss (Finlay-Jones and Brown 1981).

The table as a whole deals with whether a particular severe event (or sequence) was followed by an onset. It shows that there were large differences in risk by category of event when those occurring to Islington women are considered. If those involving humiliation are combined with those involving entrapment, the risk of depression was three times greater than for other severe events (31% versus 9%). The relatively low risk associated with loss alone, except following a death rated severe, suggests that something more than loss is usually necessary to bring about a depressive onset.

The likely importance of devaluation of self and entrapment is seen in another result. Separations occurring to Islington women associated with humiliation (subcategory i in Table 3) were further divided into whether or not the woman took some initiative in bringing about the separation after learning of an infidelity or the experience of a partner's or lover's violence. When the further subcategory of a woman who clearly took the initiative in the separation (vi in Table 3) is taken onto account, Table 4 shows a clear gradient between the amount of control apparently exercised by the woman and risk of an onset.

Table 2. Onset of depression among the 130 women in Islington with severe event in terms of matching commitment

At least one severe event taking place in an area of marked commitment	Onset (%)
Yes	40 (16/40)
No	14 (13/90)
Total	22 (29/130)

$\chi^2 = 9.01$, 1 df, $p < .01$.

Table 3. Rate of onset by humiliation/entrapment/loss/danger following 377 event sequences during a 2-year period among Islington community women

Event sequence	Provoking onset (%)
1. Humiliation/trapped	31 (41/131)
i. Humiliation: separation	35 (12/34)
ii. Humiliation: other's delinquency	19 (7/36)
iii. Humiliation: put down	38 (12/32)
iv. Trapped	34 (10/29)
2. Loss alone	9 (14/157)
v. Death	29 (7/24)
vi. Separation: subject initiated	11 (2/18)
vii. Other key loss	7 (4/58)
viii. Lesser loss	2 (1/57)
3. ix. Danger alone	3 (3/89)
4. Total	15 (58/377)

3.2 Psychosocial Vulnerability

The second theme of the introduction concerned issues surrounding psychosocial vulnerability. Two background factors measured at the time of the first interview in Islington have proved highly predictive of onset during the following year (Brown et al. 1990a,b):

Table 4. Risk following separation in a core tie by degree of control on woman's part – Islington series

Type of separation/risk	Onset (%)
Other's initiative	53 (9/17)
Woman "forced" to act	25 (4/16)
Woman's initiative	11 (2/18)

$\chi^2 = 7.59$, 2 df, $p < .05$.

- Negative psychosocial (negative evaluation of self or chronic subclinical conditions)
- Negative environmental (negative interaction in the home or – for single mothers – lack of close tie defined by the women at the time of first contact as "very close" and seen fairly often).

The predictive power of the indices can be judged by the fact that, while only 23% of the 303 women at risk at the time of first contact had both risk factors, three quarters of all onsets recorded in the 12-month follow-up period occurred among them (Brown et al. 1990c). The findings concerning the importance of humiliation and entrapment, of course, fit well with these two background risk factors.

Figure 1 takes into account both type of provoking event and vulnerability. (Since self-esteem was measured only at the time of first interview, only the first follow-up year has been considered.) The size of the interactive effect is impressive. A severe event, however threatening, was never enough without a vulnerability factor to provoke depression, nor for that matter was vulnerability on its own. Where loss alone was concerned, risk of onset was negligible unless both of the two risk factors were present. (Subsequent research has indicated that onset can occur without such vulnerability following a severe event, but under these circumstances risk is low.)

4 Endogenous Conditions

So far, only depressive onsets in the general population, almost entirely of a "neurotic" kind, have been considered. The third introductory theme concerned the curious failure to obtain any clear differentiation between the presence of a provoking severe event and diagnostic type. At the same time, there can be no doubt about the presence of genuine "endogenous" conditions. A recent study of a London psychiatric patient series dealing with inpatients and outpatients seen over a defined period of time has thrown some possible fight on this puzzling set of findings. Table 5 shows, as expected, that those with melancholic/psychotic conditions were a good deal less common than those without. However, it also shows that only those who had a melancholic/psychotic diagnosis and a prior episode differed in having less chance of experiencing a severe event before onset. This difference between first and later episodes may well go some way to explain the general lack of a relationship. It might also help to explain the inconsistencies in published results, since the proportion of patients with both these characteristics is bound to vary by type of treatment centre, and traditional patient enquiries that have reported such differences have often been carried out in tertiary centres where the proportion of later episodes is likely to be greater. While numbers involved with a high melancholic/psychotic score are small, findings held for two other patient series (Brown et al. 1994a) and a series from Pittsburgh produced some consistent findings (Frank et al. 1994). Such a fall-off with episode number has been found to relate to the course of bipolar conditions (Post et al. 1986), and kindling and sensitisation phenomenon, as with such conditions, may play some part with the melancholic/psychiatric conditions as defined in the London and Pittsburgh enquiries. It is also of interest that this same London study suggested that, despite intensive interviews, as many as one tenth of patients with a "neurotic" depressive disorder were without a hint of being "provoked" (Brown et al. 1995).

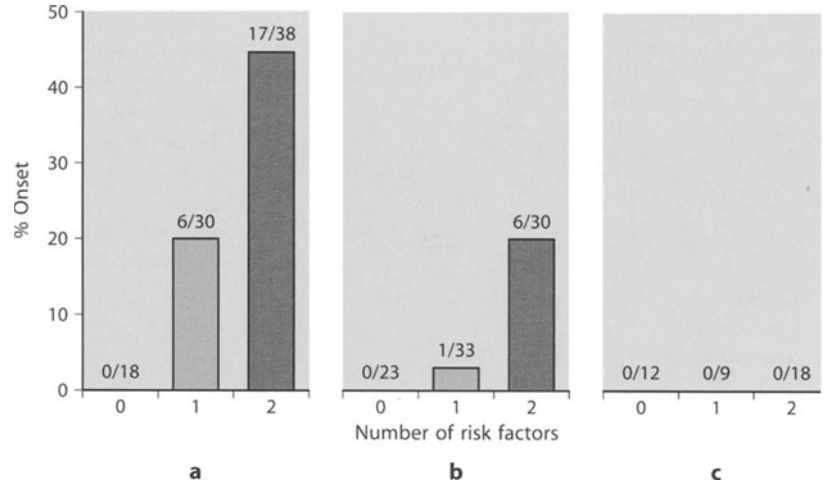


Fig. 1a–c. Rates of onset depression in follow-up year by severe event type and background risk among 130 Islington women. All onsets and severe events (or sequences) taken; only the event nearest onset considered provoking. **a** Any humiliation/entrapment event. **b** Loss event (and not any humiliation/entrapment event). **c** Danger event (and not any humiliation/entrapment or loss event)

Table 5. Ordinal number of adult episodes of depression in North London patients by melancholic/psychotic score and presence of a severe event within 6 months of onset

Melancholic/ psychotic score	Severe event rate (%)					Total excluding first episode	Total
	First episode	Second episode	Third episode	Fourth episode	Fifth episode or more		
Low	71 (25/35)[a]	70 (21/30)	83 (10/12)	60 (3/5)	80 (8/10)	74 (42/57)[b]	73 (67/92)
High	59 (10/17)[a]	15 (2/13)	33 (1/3)	100 (1/1)	0 (0/1)	22 (4/18)[b]	40 (14/35)
Total	67 (35/52)	53 (23/43)	73 (11/15)	67 (4/6)	73 (8/11)	–	64 (81/127)

[a]Not significant.
[b]$x^2 = 13.18$, 1 df, $p < .001$.

5 Outcome

LEDS life-event research has opened up other issues. Perhaps most significant has been the question of outcome. Here, evidence is beginning to emerge that improvement or remission quite often involves the reverse of the psychosocial processes typically leading to the onset of an episode. For example, Table 6 takes account of the presence of positive events in the 20-week period before any significant improvement or remission. Such events on contextual grounds were seen as giving renewed hope about the future. In the Islington series, when episodes lasting 20 weeks or more were considered, 57% of such changes in symptomatology were preceded by such an event compared with an expected proportion of 15% (Brown 1993). No effect emerged for episodes lasting less than 20 weeks. In the patient series, the overall proportion was much the same at 51%; however, those patients recovering or improving on anti-depressant medication had only around half of this positive LEDS experience.

Table 6. Proportion with a positive event among North London patients recovering/improving from depression of at least 20 weeks' duration by medication (with overall Islington series results for comparison)

Patient group	Positive event (%)
North London patients recovering/improving[a]	
Anti-depressant medication	26 (12/47)
Other medication	62 (8/13)
No other medication	48 (28/58)
Total without anti-depressant medication	51 (36/71)
Islington community depressed patients recovering/improving	57 (28/49)

[a]$\chi^2 = 8.20$, 2 dfl, $p < .02$.

One type of "positive" event was defined by its heralding a significant reduction in an ongoing marked difficulty. It is therefore of interest that presence of such an interpersonal difficulty at the point of onset (but no other type) was an important predictor of the chronicity of an episode in both a community and patient series. The experience of a history of childhood abuse or neglect was about as equally effective as a predictor of chronicity. Table 7 gives the percentage of 101 episodes of depression in the Islington community series of women occurring over a number of years in terms of the proportion without improvement/recovery within 12 months by whether they had had either childhood neglect/abuse or an interpersonal difficulty throughout the episode (much the same result was obtained if presence of the difficulty at the point of onset was taken instead) (Brown and Moran 1994; Brown et al. 1994b).

6 Conclusion

There are two ways of looking at the findings that have been reviewed. Firstly, the study of life events has clearly been an effective facilitator, a way of opening up a series of issues concerning depression for study. Once a substantial causal link had been established, a platform was provided for the study of the role of a whole range of other current and biographical experiences. Moreover, as research has progressed, it has pushed back in time to consider the role of early experiences of neglect and abuse which often has event-like characteristics. In more general terms, the study of events has led to the consideration of issues of vulnerability and protection, event production, coping behaviour and issues surrounding chronicity and course of particular episodes. So far, this kind of extension has remained largely within the psychosocial realm, but clearly potential ramifications are much

Table 7. Percentage with chronic course (> 1 year) of 101 episodes of depression among Islington women by the risk factors of interpersonal difficulty and childhood adversity

Risk factor	Chronic course (%)		Odds ratio
	Risk factor present	Risk factor absent	
Childhood adversity	44 (17/39)	16 (10/62)	4.02*
Interpersonal difficulty during course	44 (18/41)	15 (9/60)	4.43*
Childhood adversity *or* interpersonal difficulty during course	44 (24/59)	7 (3/42)	8.91*

*$p < .0001$.

broader. In diagnostic terms, it has raised issues concerning the role of sensitisation and kindling in explaining the apparent lessening of the importance of stress in subsequent onsets of melancholic/psychotic depressive conditions. This is a possibility suggested by research on bipolar conditions (Post et al. 1986).

It has also been possible to isolate a small group of apparently endogenous "neurotic" depressive episodes. Such results call out for collaborative research in both biological and clinical terms. It also suggests that such areas of research are likely to be handicapped if they fail to take account of the now complex psychosocial aetiological models that have been developed around the event–onset link. Particularly significant, for example, for genetic research are findings from disparate cultures that have documented considerable differences in the experience of caseness of depression, e.g. on the one hand at one extreme 2.5% among women between 18 and 65 in a Basque-speaking rural community in a 12-month period (Gaminde et al. 1993) and on the other 30% among women in a black township in Zimbabwe (Broadhead and Abas 1998). Since there were comparable differences in the rates of severe events in the various populations studied, an obvious interpretation is that the differences in the experience of "neurotic" depression had been driven by such psychosocial factors. This argument would hold irrespective of the presence of substantial heritability indices (h^2) within each population (Brown 1996).

The second contribution of such research concerns the depression–event link itself. This is a highly complex issue, because depressogenic life events correlate with a whole range of factors ranging from genetic/personality (Owens and McGuffin 1997) to macrolevel/societal (Brown 1996). However, the findings concerning the role of events involving humiliation or entrapment do suggest that the issue of meaning cannot be entirely divorced from an evolutionary perspective – that in some way an evolutionary-based response pattern in group-living animals closely linked to issues surrounding defeat and exclusion is often involved (Gilbert 1989). However, it does not follow that depression-like states of clinical severity have ever been adaptive. An artificially created life event such as that resulting from the transfer of a dominant marsupial male sugar glider to another group with an accompanying loss of status can create an apparently severe depression-like response (Jones et al. 1995). However, this might largely be a function of captive conditions in which these responses have been documented, and such extreme states may well be largely avoided in natural conditions (M.J. Eales, unpublished work). Nevertheless, this view is fully compatible with such clinically relevant depressive conditions being a complication of essentially non-pathological evolutionary-based submission and appeasement responses to defeat in group-living mammals that are adaptive. Therefore, the high rates of clinically relevant depression that appear to be possible in some human populations may well be a result of our more highly developed cognitive development together perhaps with the event-creating potential of many societies experiencing periods of marked social changes due to factors such as war, industrialisation, urbanisation, changing sexual mores and the like. Such considerations suggest that the study of life events places aetiological research on depression firmly in the courts of social science, biology and clinical medicine, and interdisciplinary collaborative research is likely to prove vital.

7 References

Bebbington PE, McGuffin P (1989) Interactive models of depression. In: Paykel E, Herbst K (eds) Depression: an integrative approach. Heinemann, London, pp 65–80

Broadhead J, Abas M (1998) Life events and difficulties and the onset of depression amongst women in an urban setting in Zimbabwe. Psychol Med 28: 29–38

Brown GW (1993) Life events and affective disorder: replications and limitations. Psychosom Med 55: 248–259

Brown GW (1996) Genetics of depression: a social science perspective. Int Rev Psychiatry 8: 387–401

Brown GW, Harris TO (1978) Social origins of depression. A study of psychiatric disorder in women. Tavistock, London/ Free Press, New York

Brown GW, Harris TO (1989) Life events and illness. Guilford, New York

Brown GW, Moran P (1994) Clinical and psychosocial origins of chronic depressive episodes. 1. A community survey. Br J Psychiatry 165: 447–456

Brown GW, Andrews B, Bifulco A, Veiel H (1990a) Self-esteem and depression. 1. Measurement issues and prediction of onset. Soc Psychiatry Psychiatr Epidemiol 25: 200–209

Brown GW, Bifulco A, Veiel H, Andrews B (1990b) Self-esteem and depression. 2. Social correlates of self-esteem. Soc Psychiatry Psychiatr Epidemiol 25: 225–234

Brown GW, Bifulco A, Andrews B (1990c) Self-esteem and depression. 3. Aetiological issues. Soc Psychiatry Psychiatr Epidemiol 25: 235

Brown GW, Harris TO, Hepworth C (1994a) Life events and 'endogenous' depression: a puzzle re-examined. Arch Gen Psychiatry 51: 525–534

Brown GW, Harris TO, Hepworth C, Robinson R (1994b) Clinical and psychosocial origins of chronic depressive episodes. II. A patient enquiry. Br J Psychiatry 165: 457–465

Brown GW, Harris TO, Hepworth C (1995) Loss, humiliation and entrapment among women developing depression: a patient and non-patient comparison. Psychol Med 25: 7–21

Dean C, Surtees PG, Sashidaran SP (1983) Comparison of research diagnostic systems in an Edinburgh community sample. Br J Psychiatry 142: 247–256

Finlay-Jones R, Brown GW, Duncan-Jones P, Harris TO, Murphy E, Prudo R (1980) Depression and anxiety in the community. Psychol Med 10: 445–454

Finlay-Jones R, Brown GW (1981) Types of stressful life event and the onset of anxiety and depressive disorders. Psychol Med 11: 803–815

Frank E, Anderson B, Reynolds CF, Ritenour A, Kupfer DJ (1994) Life events and research diagnostic criteria endogenous subtype. Arch Gen Psychiatry 51: 519–524

Frijda NH (1986) The emotions: studies in emotion and social interaction. Cambridge University Press, Cambridge

Gaminde I, Uria M, Padro D, Querejeta I, Ozamiz A (1993) Depression in three populations in the Basque country – a comparison with Britain. Soc Psychiatry Psychiatr Epidemiol 28: 243–251

Gilbert P (1989) Human nature and suffering. Erlbaum, Hove

Jones IH, Stoddart DM, Mallick J(1995) Towards a sociobiological model of depression: a marsupial model (petaurus breviceps). Br J Psychiatry 166: 475–479

Katschnig H, Pakesh G, Egger-Zeudener E (1986) Life stress and depressive sub-types: a review of present diagnostic criteria and recent research results. In Katschnig H (ed) Life events and psychiatric disorders: controversial issues. Cambridge University Press, Cambridge, pp 201–245

Owens MJ, McGuffin P (1997). Genetics and psychiatry. Br J Psychiatry 171: 201–202

Paykel ES, Prusoff BA, Klerman GL (1971) The endogenous-neurotic continuum in depression, rater independence and factor distributions. J Psychiatr Res 8: 73–90

Post RM, Rubinow DR, Ballenger JC (1986) Conditioning and sensitisation in the longitudinal course of affective illness. Br J Psychiatry 149: 191–201

Spitzer RL, Endicoll J, Robins E (1978) Research diagnostic criteria: rationale and reliability. Arch Gen Psychiatry 35: 773–782

Wing JK, Sturt E (1978) The PSE-ID-CATEGO system. Supplementary manual. MRC Social Psychiatry Unit, London

Wing JK, Cooper JE, Sartorious N (1974) The measurement and classification of psychiatric symptoms: an instruction for the Present State Examination and CATEGO Programme. Cambridge University Press, Cambridge

CHAPTER

22

M. Bauer, H. Helmchen

General Principles of the Treatment of Depressive and Manic Disorders

Translator: E. Taub

1 Introduction

This chapter describes general guidelines for the treatment of affective disorders. It is thus more specific than Chap. 9 (Vol. 1, Part 2), which concerns the general principles of treatment in psychiatry, and more general than the discussion of specific methods of treatment for the affective disorders found in the following chapters of this volume (Part 1, Chaps. 23–25 and Chap. 31). The aim here is to delineate the role of various forms of treatment in the overall treatment plan, with an emphasis on common, overarching principles.

Depressive disorders are among the more common mental illnesses, with a point prevalence of 4%–8% and a lifetime prevalence of 10%–20% (Kessler et al. 1994; Angst 1995). They are characterized not only by high prevalence, but also by a pronounced tendency to recur and by a chronic course with considerable morbidity and mortality: 40%–80% of patients with unipolar depression suffer a recurrence within 2 years; the likelihood of a recurrence increases with the number of previous depressive episodes and the severity of the current episode; patients with more severe depression have a risk of approximately 15% of dying by suicide (Keller et al. 1986; Frank et al. 1990; Klerman and Weissman 1992; Prien and Kocsis 1995) and also have a higher mortality rate from cardiovascular causes (Glassman and Shapiro 1998; see also Chap. 14, Vol. 2, Part 2). Chronic treatment, involving prophylaxis against recurrence and suicide, is therefore of the greatest importance, in addition to acute treatment (Kupfer 1991; Kupfer et al. 1992; APA 1993).

A complicating factor is that the great majority of depressive patients present first not to the psychiatrist, but to the general practitioner or internist, and, therefore, not all of these patients are correctly diagnosed and properly treated (Keller 1988; Linden et al. 1996; Hirschfeld et al. 1997). In summary, depressive disorders are common, tend to recur, and are associated with considerable mortality.

As is true of other psychiatric illnesses, the etiology and pathogenesis of depressive disorders have not yet been elucidated. At present, it is assumed that their origin is multifactorial, involving not only genetic and biological, but also psychological and social determining factors (Akiskal 1995). The modern treatment of depression must thus take biological, psychological, and social aspects into account (APA 1993). A number of recent studies have shown that a diagnostic distinction between neurotic and endogenous depression makes little sense with respect to either etiopathogenesis or treatment. The two leading diagnostic classification systems have therefore dispensed with this dichotomy; they now list only the diagnoses of the depressive episode (in ICD-10; WHO 1992) or of major depression (in DSM-IV; APA 1994a). Primary distinctions with implications for therapy are now drawn according to degree of severity (mild, moderate, severe), type of disease course (single episode, recurrent), and, especially, the characterization of disease manifestations (e.g. psychotic manifestations, seasonal pattern) (see Chaps. 14, 15, Vol. 3, Part 1; Chaps. 2, 3, Vol. 1, Part 2).

Bipolar disorders are recurrent affective disorders with both depressive and manic episodes. Bipolar I disorder is characterized by one or more manic or mixed episodes, which are usually associated with episodes of major depression. Bipolar II disorder is characterized by one or more episodes of major depression and at least one hypomanic episode (APA 1994a).

Manic disorders are much rarer than depressive disorders. Epidemiologic studies have revealed a 1-year prevalence rate of 0.6%. The lifetime prevalence of bipolar I disorder in the general population has been determined to be between 0.4% and 1.6% (Robins and Regier 1991). In contrast to depressive disorders, which are more common in women, bipolar disorders occur with approximately the same frequency in both sexes. Bipolar disorders are characterized by a high rate of recurrence and thus also by a high morbidity, a high risk of suicide (in 10%–15% of cases of bipolar I disorder; APA 1994a), and many associated social and psychological problems (see Chap. 15, Vol. 3, Part 1).

2 Diagnosis and Indications for Treatment

A correct diagnosis and prognosis, leading to the establishment of proper indications for treatment, is a precondition for the successful therapy of a depressive or manic episode. Careful assessment of the patient's clinical findings, past psychiatric and medical history, and psychosocial stress factors is required. A specific family history with respect to affective disorders must be obtained, and an understanding of the patient's familial and cultural environment is also essential.

Potential organic causes of depressive or manic manifestations must be ruled out by means of a neurological and general medical examination and, when indicated, by appropriate additional diagnostic techniques, including electroencephalography (EEG), biochemical laboratory tests, and radiological and neuroradiological studies, i.e. computed tomography (CT) or magnetic resonance imaging (MRI) of the

head. A medication history is indispensable, as is screening for the covert use of medications and illicit substances when this is suspected (e.g. benzodiazepines, amphetamines, barbiturates, cannabis), both to rule out the possibility of a pharmacogenically induced affective disorder (Müller-Oerlinghausen 1997) and also because these substances may interact with antidepressant or antimanic medications. In particular, if the first manic episode occurs after age 40, a medical disease or substance ingestion must always be ruled out as the cause.

The current nosologic diagnostic classification of the affective disorders is based on two operationalized classification systems, ICD-10 of the World Health Organization (WHO 1992) and DSM-IV of the American Psychiatric Association (APA 1994a); in the German-speaking countries, the diagnostic categories of ICD-10 are in widespread use in the clinical field. The diagnosis of a depressive or manic episode is made when a particular pattern of manifestations listed in the diagnostic manual is present (WHO 1992). Moreover, the syndromal diagnostic evaluation of current clinical findings and the assessment of the severity of depression or mania are particularly important for specific aspects of decision-making relating to therapy. Finally, the possible presence of suicidality and of somatic and social complications of depression must be taken into account. In the case of mania, particular attention must be paid to the extent of self-endangerment (e.g. by large financial expenditures or unrestrained professional activities) and uncritical endangerment of others (e.g. in road traffic).

Treatment is indicated in all reliably diagnosed cases of depressive or manic disorder. Once the diagnosis has been made, a comprehensive and multidimensional treatment plan should be drawn up, taking account not only of the current clinical findings (e.g. the presence or absence of psychotic manifestations, inhibition or agitation in depression, dysphoric-irritative or mixed affective manifestations in mania), the severity of illness, and possible suicidality, but also of the patient's conceptions and preferences regarding specific forms of treatment. At this point, the question also arises as to whether inpatient treatment in a specialized facility is necessary or whether the treatment can be carried out in an outpatient setting.

The indications for inpatient treatment of depressive or manic disorders are summarized in Table 1.

If the patient refuses to be admitted to an inpatient facility, he or she can be committed involuntarily to a locked psychiatric facility if there is acute self-endangerment secondary to depression, according to currently applicable law, i.e. the PsychKG in the German federal states or the federal Care Act (*Betreuungsgesetz*, BtG) (see also Chap. 15, Vol. 1, Part 2). In cases of severe mania, financial or social self-endangerment or uncritical endangerment of others may also necessitate judicial commitment.

Table 1. Indications for inpatient treatment

Depressive episode	Manic episode
Suicidality	Presence of delusions or psychotic manifestations
Presence of delusions or psychotic manifestations	Financial or social self-endangerment
Lack of a favorable social environment	Uncritical endangerment of others
Family conflict constellations that tend to maintain depression	Intractability
Inability of the patient to care for him- or herself	
Intractability	

3 Goals of Treatment

The initial goal of treatment is the complete eradication of affective manifestations (complete remission), followed by the prevention of recurrences and chronification. Because there may be considerable overlap between the therapeutic measures used to achieve these different goals, attention must be paid to course-specific aspects of the illness even during the initial treatment of acute episodes. At the same time, it must be borne in mind that acute treatment and recurrence prevention are two different forms of therapy, each with its own indications.

The short-term, intermediate, and long-term goals of the treatment of a depressive or manic episode are given in detail in Table 2. After the acute affective episode has worn off, the goal of further treatment is to reduce the probability of recurrence (see Table 2).

4 Aspects of Course with Implications for Treatment

The typical course and the correspondingly structured treatment of a depressive illness (episode and recurrence tendency) may be represented graphically according to the model developed in the 1990s in Pittsburgh (USA) by Kupfer and colleagues (Kupfer 1991; Fig. 1). Three stages of treatment were described, corresponding to three stages of the illness:

Table 2. Goals of therapy for depressive and manic episodes and for the tendency toward recurrence

	Depressive episode	Manic episode	Tendency toward recurrence
Short-term goals (hours to days)	Acute alleviation of anxiety, agitation, and insomnia	Control of psychomotor agitation, logorrhea, and aggression	
	Prevention of suicidal behavior	Prolongation of sleep duration	
Intermediate goals (days to weeks)	Improvement of mood, motivation, and cognitive ability	Reduction of grandiose ideas and excessive self-esteem	
	Removal of psychosocial stress factors	Reduction of excessive activity in social, occupational, and sexual areas	
Long-term goals (weeks to months)	Prevention of rapid relapse in the vulnerable period after remission	Prevention of rapid relapse in the vulnerable period after remission	
	Prevention of chronification and resistance to treatment	Prevention of resistance to treatment	
	Regaining of social competence with reintegration into family, occupation, and society	Elimination of unfavorable consequences of excessive activity and reintegration into family, occupation, and society	
Very long-term goals (years)			Prevention of recurrence
			Prevention of rapid cycling, particularly in bipolar disorders

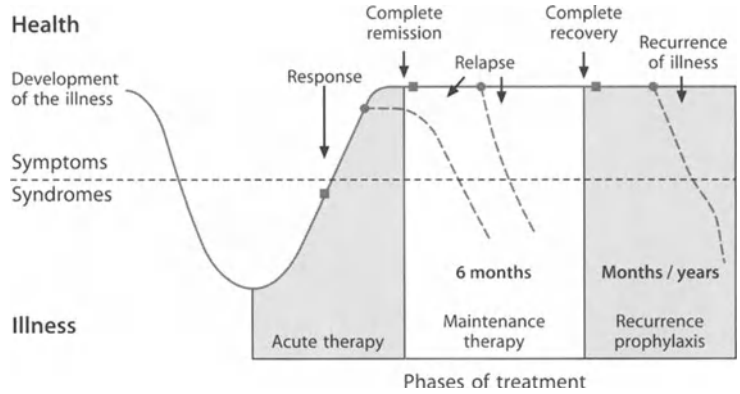


Fig. 1. Long-term course of a depressive illness. (After Kupfer 1991)

1. Acute therapy
2. Maintenance therapy
3. Prophylactic therapy against recurrence

If the premorbid level of functioning has been fully reestablished by acute therapy, we speak of a complete remission. The phase of acute therapy is followed by that of maintenance therapy, which lasts approximately 6 months. Antidepressant pharmacotherapy for the stabilization of remission should be carried out for 6 months with the same antidepressant used for acute therapy, at the same dose. This is also true of the successful augmentative use of lithium "add-on" treatment, as was recently shown in a 6-month,

placebo-controlled study (Bauer et al. 1999). If the depressive syndrome returns during the phase of maintenance therapy, we speak of a relapse. If no relapse occurs during maintenance therapy, a slow tapering (>3 months) of the antidepressant medication is recommended. Nonetheless, if this results in a symptomatic deterioration, the original medication should be given at the original dose for at least another 6 months before a cautious attempt at discontinuing medication is made again.

If the patient has been asymptomatic for approximately 6 months, it is presumed that complete recovery has taken place. This presumption remains to be confirmed by stable, continued freedom from symptoms (complete remission) after the cessation of medication. Moreover, the concept of recovery applies only to individual episodes of the illness, and not to their tendency to recur. A recovery in the latter sense can be spoken of only if the patient remains free of recurrences after the cessation of prophylactic therapy against recurrence.

Prophylactic therapy against recurrence in either mono- or bipolar disorders should be taken into consideration depending on the previous course of illness. A recurrence of disease is said to occur when an affective episode appears during the phase of prophylactic therapy, i.e. after a completely asymptomatic state has been achieved for some length of time (Kupfer 1991). Even though no definite recommendation can be given as to when prophylactic therapy should be initiated, it is clear that it is indicated in certain situations associated with a high risk of recurrence, as listed in Table 3.

For the depressive disorders, a high risk of recurrence is present in patients who have had two episodes within 5 years and in patients who have had more than three episodes overall. Bipolar disorders have a much higher recurrence rate than unipolar depressive disorders. Furthermore, there are a number of other risk factors for the recurrence of affective episodes (Brunello et al. 1995; Greil and Kleindienst 1997; Table 3).

Table 3. Risk factors for recurrence

Depressive disorders	Manic disorders
Two episodes within 5 years	Bipolar course of illness
More than three episodes	Two episodes within 3–4 years
"Double depression" (i.e. depression and dysthymia occurring simultaneously)	Presence of manic or bipolar disorder in first-degree relatives
Residual manifestations in the phase of maintenance therapy	Early age of onset of illness
Simultaneous presence of substance abuse or anxiety disorder	
Depressive episode in first-degree relatives	
First depressive episode before age 30	
Severe depressive episode with suicidality	
Direct transition from depressive episode to manic episode	

Even if only one of these features is present, it is recommended that acute therapy be followed by prophylactic therapy against recurrence, or at least that such therapy be considered, so that the type of maintenance therapy can be chosen partly with a view toward future prophylactic therapy. The medication of choice for recurrence prophylaxis in unipolar depression is either lithium or the antidepressant with which remission was achieved in the index phase (APA 1993). Lithium is also indicated for recurrence prophylaxis in bipolar disorders or unipolar mania, in which it may also be given – particularly in the presence of schizoaffective features – in combination with carbamazepine or, as recently proposed, valproate (see Chap. 23, Vol. 3, Part 1).

5 Spectrum of Options for Treatment of the Acute Episode

The origin of depressive illnesses is currently assumed to be multifactorial, and their treatment is therefore fundamentally multidimensionally oriented. The different forms of treatment available can be combined with one another in different ways; in particular, pharmacotherapy can be combined with supportive or specific psychotherapy. The established forms of treatment for depressive and manic illnesses are listed in Table 4.

The applicability of individual forms of treatment varies over the course of the disorder; in the acute phase, biological and supportive therapies are most important. Furthermore, the severity of illness also determines the choice of therapy. Mild depressive episodes, especially of a reactive type or in the setting of an adaptive disorder, can usually be treated adequately by supportive conversation with the physician or by conflict-centered short-term psychotherapy. Moderate and severe depressive illness, however, is treated primarily by medical therapy with antidepressants.

Nonetheless, according to current concepts, the so-called disorder-specific forms of psychotherapy, such as cognitive therapy, as described by Beck (1967), and interpersonal psychotherapy (IPT), as described by Klerman et al. (1984), play an increasingly important role in the treatment of depression, even in cases of

Table 4. Established methods of treatment for depressive and manic episodes

Depressive episodes	Manic episodes
Pharmacotherapy with antidepressants[a]	Pharmacotherapy with antimanic agents[a]
Supportive therapeutic conversation (supportive psychotherapy)[b]	Electroconvulsive therapy for intractable cases
Specific psychotherapy, e.g. behavior therapy, cognitive therapy, and interpersonal psychotherapy[b]	Accompanying therapeutic conversation (information about the nature and course of the illness, explanation of treatment options, provision of a feeling for and concept of the illness)
Other methods of somatic therapy, e.g. sleep deprivation treatment, phototherapy, and electroconvulsive therapy[c]	Sociotherapeutic measures, e.g. protection from the negative social and financial consequences of the illness
Sociotherapeutic measures, e.g. provision of social activities, protection from the negative social and financial consequences of the illness	Inclusion of family members
Inclusion of family members	Agreement on treatment if a tendency toward recurrence is present

[a]See Chap. 23, Vol. 3, Part 1.
[b]See Chap. 25, Vol. 3, Part 1.
[c]See Chaps. 24 and 31, Vol. 3, Part 1.

moderate severity, in both acute and long-term therapy (see Chap. 25, Vol. 3, Part 1).

The most important component of the treatment of mania is pharmacotherapy with antimanic agents, including neuroleptics, prophylactic agents against recurrence, and benzodiazepines. The recurrence-prophylactic agents lithium, carbamazepine, and valproate are indicated particularly in mild or moderately severe forms of the illness, as they have less serious side effects than the neuroleptics (see Chaps. 15, 23, Vol. 3, Part 1). The accompanying discussion with the physician should inform the patient about the illness, its course, and its treatment. Specific psychotherapeutic measures have no value in the treatment of acute mania. The environment of manic patients should be kept as devoid of stimulation as possible. The established modes of treatment for manic illnesses are listed in Table 4.

6 Medical and Psychotherapeutic Guidance of Depressive Patients

The basis of any treatment for depression is the therapeutic conversation between the patient and the physician, in which the physician must demonstrate empathy, acceptance, and understanding. Reliability and ensuring that the same therapist always treats the patient are especially important factors, because depressive patients are often very sensitive to change and loss. Furthermore, the physician may exert a supportive effect by communicating to the patient that the illness is treatable. Before the treatment is begun, the patient must be informed about the illness and its projected therapy; in particular, the patient must be informed of the latency of effectiveness of the antidepressants and their typical side effects. Thorough explanation of the treatment plan or, even better, an agreement between patient and physician concerning the treatment plan can improve the extent of the patient's cooperation and compliance in taking antidepressants.

It has been found useful for the physician to see the patient frequently at the beginning of pharmacological therapy and to evaluate the therapeutic effect together with the patient, perhaps with the participation of family members as well, over the course of treatment. The basic principles of the medical therapeutic conversation with depressed patients can be delineated as follows:

- Basic therapeutic behavior (provision of empathy, acceptance, and understanding)
- Thorough information provided to the patient about the nature, course, and prognosis of the illness
- Explanation of the multidimensional model of the cause of depression
- Explanation to the patient of the goals and temporal course of treatment
- Creation of a treatment plan
- Explanation of the latency of effectiveness and side effects of antidepressants
- Bringing family members into the discussion
- Common evaluation of the effect of treatment

7 Pharmacotherapy with Antidepressants

Since the introduction of the first tricyclic antidepressant, imipramine, in 1957, many new types of antidepressants have been developed; nearly 30 different antidepressants have been used clinically in Germany up to the present. Newer antidepressants were developed with the goal of finding substances that were as effective against depression as the tricyclic agents, but with fewer side effects, a more rapid onset of effect, and lesser toxicity (Rudorfer and Potter 1989; Leonard 1995). The classes of substances currently available differ little in their antidepressant effect; there is no convincing evidence that any class of antidepressants is more effective, or has a more rapid onset of effect, than the others. Nonetheless, there are important differences in their neurochemical activity profiles and thus also in their side effects. The newer antidepressants of the 1980s and 1990s generally have less serious side effects than the "classical" substance groups of the tricyclic agents and the irreversible monoamine oxidase (MAO) inhibitors (see Chap. 23, Vol. 3, Part 1).

Differential therapeutic considerations concerning the initial choice of an antidepressant are essentially based on various clinical features of, and specific previous experiences in, the individual patient. The antidepressant should be chosen according to the following criteria (Bauer and Berghöfer 1997):

- The clinical features of depression in the individual patient (e.g. a sedating antidepressant for an anxious-agitated depressive syndrome, or an activating and less sedating antidepressant for an inhibited, psychomotorically slowed depressive syndrome)
- Psychiatric comorbidity (e.g. serotonergic agents and MAO inhibitors for patients with comorbid obsessive-compulsive and panic disorders)
- Specific previous experiences (e.g. nonresponse vs. response, intolerance, allergic reactions) with individual antidepressant substances or substance classes
- The range of side effects of the antidepressant
- Accompanying medical and neurological illnesses
- Further specific features of the patient, e.g. age, particularly with regard to somatic comorbidity (Lebowitz et al. 1997); pregnancy (see Chap. 13, Vol. 2, Part 1)

Proper selection of an antidepressant requires a knowledge of the typical side effects of the individual substances, which, in turn, are determined by their pharmacological activity. The neurochemical effects of the antidepressants account for their clinical effectiveness against depression and also, in part, for their unwanted side effects (Leonard 1993; Richelson 1994). The presence of accompanying medical or neurological illness necessitates the choice of an antidepressant whose pharmacologic profile will have a beneficial, or at least not harmful, effect on the accompanying illness. An example of this is the use of substances such as mianserin or selective serotonin reuptake inhibitors (SSRI) in the presence of heart disease, as they have a much weaker effect than the tricyclic agents on the cardiac conductive system, for example (ECG monitoring for potential adverse effects is mandatory).

The use of substances with a strong anticholinergic effect should be considered with special caution in patients over age 70. Older patients are more sensitive to the central anticholinergic effects of these medications and may develop disorientation, confusion, or even delirium as a result. The newer antidepressants, which have fewer side effects (e.g. SSRI, nefazodone, venlafaxine), are preferred to the tricyclic agents in this age-group. The cardiovascular and renal changes usually associated with advanced age lead to an intensification of side effects as compared to those found in younger individuals (Lebowitz et al. 1997). Antidepressants should therefore be given to older patients in slowly increasing doses and in lower overall doses than those given to younger patients (Bezchlibnyk-Butler and Jeffries 1996).

Intoxications and delirious syndromes are considered absolute contraindications for the use of all antidepressants. Relative contraindications for the use of tri- and tetracylcic antidepressants include congestive glaucoma, cardiac dysrhythmia, bladder dysfunction, severe hepatic or renal injury, epilepsy, and blood dyscrasias. There are no relative contraindications for monotherapy with SSRI or reversible inhibitors of monoamine oxidase. If SSRI are used in combination with other substances having serotonergic effects, such as lithium, clomipramine, and especially MAO inhibitors, the possible appearance of a serotonin excess syndrome should be watched for. The use of antidepressants during pregnancy and nursing requires an especially careful assessment of risks and benefits, in view of the possibility of adverse effects on the fetus or infant (Altshuler et al. 1996).

Tri- and tetracyclic agents should be given in slowly increasing doses in order to minimize their anticholinergic effects and possible drug-induced orthostatic hypotension. The typically effective daily dose (150 mg of a tri- or tetracyclic antidepressant) should be reached in 4–7 days; monitoring of blood levels usually plays no role in the introductory phase of antidepressant medication. For antidepressants of the SSRI type, the initial dose (daily dose of 20 mg) usually has an

adequate antidepressant effect. If it does not, and if the medication is well tolerated, an increase of the dose can be attempted before switching to another medication or to combination therapy, if necessary. Inadequate dosing is a common reason for inadequate response to antidepressants (so-called pseudoresistance; Keller 1988). Under stable conditions, the dose may be raised as early as 2 weeks after the initiation of treatment (see Table 6).

Regular laboratory monitoring must be carried out both before and during antidepressant pharmacotherapy, because such treatment may, in rare cases, cause changes in a number of organ systems. The performance of blood tests, including a complete blood count with differential leukocyte count, hepatic enzymes, creatinine, electrolytes, and thyroid hormones, is recommended before and during therapy, so that any changes in these values with respect to the baseline may be detected. An ECG and blood pressure measurements, as well as an EEG in patients known to have epilepsy, should also be performed before treatment is begun. The frequency of monitoring over the course of treatment must be adjusted individually according to patient characteristics (age, accompanying illnesses) and the type of therapy administered (type of antidepressant, dose). At the beginning of treatment, the complete blood count and hepatic enzymes should be monitored biweekly, then monthly; the ECG should also be repeated at 2-week intervals.

8 Resistance to Treatment and Treatment Optimization

Clinical studies and observations in clinical practice have shown that approximately 30%–40% of depressive patients in either the inpatient or the outpatient setting fail to respond adequately to an initial 4- to 6-week treatment with an antidepressant (Möller 1997). As many as 10%–15% of patients fail to improve sufficiently even after several different treatments have been attempted (Nierenberg and Amsterdam 1990). A review of the literature reveals that 12%–15% of depressive patients are not yet asymptomatic 2 years after the onset of the illness, and the illness takes a chronic course in these patients (Scott 1988).

The lack of success of antidepressant or antimanic therapy is often due not to the illness itself, but to suboptimal treatment. Thus several possible means of optimizing the treatment should be considered before switching to another agent (Table 5).

9 Experimental Methods and Combination Therapy

If resistance to therapy remains intractable despite attempts at treatment optimization of the types

Table 5. Methods of treatment optimization

General treatment optimization	Treatment optimization for the depressive episode	Treatment optimization for the manic episode
Assured compliance (blood level monitoring)	Adequate duration of antidepressant medication (4–6 weeks)	Adequate duration of antimanic medication (usually more than 4 weeks)
Addition of neuroleptic agents in the presence of psychotic (delusional) manifestations	Adequate dose (tricyclic agents >150 mg/day, SSRI >20 mg/day)	Elevation of serum levels of phase-prophylactic agents (lithium, carbamazepine, valproate)
Treatment of psychiatric comorbidity (e.g. alcohol and substance abuse, including nicotine and caffeine abuse)	Serotonergic antidepressants (clomipramine, SSRI) for depression with marked obsessive-compulsive manifestations	Combination therapies
Diagnosis and treatment of somatic comorbidity (e.g. intercurrent somatic illness)	Trial of a high dose of the antidepressant (tricyclic agents 200–300 mg/day, SSRI 30–60 mg/day)	
Determination of negative drug interactions (e.g. with medications used to treat general medical illnesses)	Search for psychosocial stress factors that may maintain depression	
Exclusion of medications used to treat general medical illnesses as disease-inducing or disease-maintaining factors		

SSRI, selective serotonin reuptake inhibitors.

outlined, the use of augmentative or combination therapy, or of experimental therapies, should be considered. Experimental therapies for depressive disorders include the administration of thyroxine at high doses (Bauer et al. 1998a,b), psychostimulants, estrogens, MAO inhibitors at high doses, inhibitors of cortisol synthesis, and reserpine (Nolen et al. 1994; Heinz 1997; Nelson 1997). Among augmentative therapies, the addition of lithium to the conventional antidepressants has been found useful (Bauer and Döpfmer 1999; Bauer et al. 1999). Combination therapy with various antidepressants has been tried with variable success, as has so-called pindolol augmentation, a combination of the β-blocker pindolol with a SSRI (Nelson 1997). On the other hand, the combination of a neuroleptic with an antidepressant has proved useful in the treatment of depression with psychotic features (so-called double-rein therapy).

The usefulness of a combination of pharmaco- and psychotherapy in the treatment of depressive disorders is not yet fully clear, because studies on this question have yielded conflicting results (Schramm and Berger 1998). In most studies, this combination was found to be only moderately superior to either form of therapy alone. It should be pointed out, however, that combination therapy was associated with fewer terminations of treatment and with a greater acceptance of treatment than either form of therapy alone (Weissman and Klerman 1990).

Forms of treatment for refractory mania that should be considered experimental include the use of a double or triple combination of various phase-prophylactic agents; atypical neuroleptics (e.g. clozapine, olanzapine, risperidone); newer anticonvulsants (e.g. lamotrigine, gabapentin); augmentation with calcium-channel blockers; and electroconvulsive therapy (Post et al. 1997).

10 Sequential Treatment Strategies (Stepwise Approaches)

Failure of treatment is often attributable to the indiscriminate application and poor monitoring of antidepressant therapy. Clinical decision-making is a topic of major importance in medical research today, as the different modes of application of a therapy clearly account for a considerable portion of the variation in its clinical effectiveness. The unsuccessful treatment of depression is a clear example of the negative consequences of the lack of a clear treatment strategy, particularly with respect to the use of medications.

It thus seems reasonable to pay attention to the procedural elements of quality assurance while providing treatment and to apply them consistently. These elements include the standardized assessment of clinical findings at predetermined, regular intervals, using defined criteria for treatment success, and the decision either to continue effective treatment or to abandon ineffective treatment in favor of another form of treatment (Helmchen 1990). Within this formal framework, the large variety of possible treatments for depression can be applied in a reasonable and mutually compatible sequence.

Stepwise approaches to treatment have the following formal characteristics:

1. An established and rationally justified sequence of therapeutic steps
2. Regular monitoring of therapy
   - At predetermined intervals (biweekly)
   - In the form of standardized assessment of clinical findings, with the aid of established diagnostic assessment scales
3. After the success or failure or therapy has been assessed by means of well-defined criteria, decision between the following:
   - Continuation of antidepressant therapy (if the response to therapy is complete or partial)
   - Change of antidepressant therapy (if unsuccessful)
4. Pharmacologic monitoring (serum levels of antidepressants and lithium, compliance monitoring)
5. Consistent application of the individual components of therapy

In order to provide such a rationally based sequence of treatment steps, a number of so-called sequential treatment strategies (stepwise approaches or flowcharts) have been proposed, i.e. multidimensional decision trees that lay out a sequential problem-solving approach incorporating many individual medical therapeutic decisions. Such strategies are a special form of the heuristic approach to making decisions and taking action in providing treatment; they reduce the complexity of decision-making for the treating physician (Linden 1994). The treating physician may, nonetheless, choose to deviate from a treatment plan of this type, when indicated, to take account of particular features of the individual patient.

An operationalized, consistently performed, and well-monitored stepwise approach to antidepressant therapy is considered an essential component of the prevention or surmounting of resistance to therapy (Helmchen 1974, 1990), although this has not yet been confirmed empirically. Such confirmation can only be obtained from controlled studies and is indispensable for quality assurance in the somatotherapy of depression.

Table 6. Stepwise approach to the somatotherapy of depressive disorders[a]

Therapeutic and diagnostic steps[b]	Description	Durations
1. Discontinuation	Tapering of previously given, unsuccessful medication and undesired (multi-)medication	0–3 days
2. Baseline diagnosis and therapeutic sleep deprivation	Diagnostic classification, exclusion of organic causes, pharmacological "washout"	4–7 days
3. Antidepressant monotherapy	Tricyclic agent or SSRI (see text for initial selection of antidepressant)	2 weeks
4. High-dose antidepressant monotherapy	Increase of dose according to individual tolerance	2 weeks
5. Antidepressant and lithium (lithium augmentation)	Addition of lithium, reduction of the antidepressant to the dose of step 3; reduction of the target serum lithium level to 0.5–0.8 m*M*/l	4 weeks
6. Lithium monotherapy and diagnostic reassessment	Discontinuation of the antidepressant of step 5; reconsideration ("reassessment") of the diagnosis, additional laboratory testing, imaging procedures, neuropsychological testing	1–2 weeks
7. Lithium and MAO inhibitors	Addition of an MAO inhibitor (tranylcypromine)	2 weeks
8. High-dose lithium and MAO inhibitors	Increase of dose according to individual tolerance	2 weeks
9. Medication-free step	Discontinuation of all psychotropic medications; preparation for electroconvulsive therapy (anesthesiological consultation)	1 week
10. Electroconvulsive therapy	Three sessions of electroconvulsive therapy per week	2–4 weeks
Special case for the treatment of delusional (psychotic) depression: double-rein therapy	Additional administration of a neuroleptic agent (haloperidol or olanzapine, 5–15 mg/day)	

MAO, monoamine oxidase; SSRI, selective serotonin reputake inhibitor.
[a]At the in- and outpatient psychiatric clinics of the Free University of Berlin (1997).
[b]Proceed to the next step if no response.

A number of stepwise approaches to the treatment of depressive disorders have been published in recent years (Helmchen 1990; Linden et al. 1994; Kasper 1997; Nelson 1997). Table 6 shows one such that has been in use for 10 years (Linden et al. 1994; Berghöfer et al. 1997). There is as yet no comparable experience with stepwise approaches for the treatment of manic disorders or the prevention of recurrences in the affective disorders, although treatment algorithms for both of these situations in the event of nonresponse to an initial trial of monotherapy have been published recently (Post et al. 1997; Goodwin 1997; Bauer and Ströhle 1999).

11 Treatment-Resistant Tendency Toward Recurrence and Rapid Cycling

Lithium is the agent of first choice in the prophylactic treatment of bipolar disorders, even though one third of all patients have a tendency toward recurrence despite adequate lithium therapy. The anticonvulsants carbamazepine and valproate currently enable more specific phase prophylaxis directed against individual types of disease course and clinical syndromes, but a sizable fraction of patients (perhaps 10%–20%) nonetheless remain refractory to treatment for the prevention of recurrence (Bauer and Ströhle 1999).

A particular form of resistance to phase prophylaxis is found in patients with "rapid cycling," a malignant type of disease course in which more than four affective episodes occur over a 12-month period; this type of course, as a rule, does not respond either to lithium or to other phase-prophylactic agents. The affective episodes arise in any combination or sequence and meet the criteria for a manic, mixed, or hypomanic episode or for an episode of major depression; they are not separated by a 2-month period of remission, nor are they typified by a transition to an episode of opposite polarity (APA 1994a).

Rapid cycling occurs in approximately 10%–20% of patients with bipolar disorders, of which 70%–90% are women. In addition to female sex, other factors that have been repeatedly found to be associated with a higher risk of rapid cycling include the use of tricyclic antidepressants and the presence of either overt or subclinical hypothyroidism (in as many as 50% of cases) (Bauer and Whybrow 1991). The treatment history of patients with rapid cycling is often charac-

terized by diverse, polypharmaceutic, and mostly unsuccessful attempts at treatment. No standard concept of treatment has yet been developed for this subgroup of the affective disorders, in which as many as ten or more cycles may be experienced per year ("ultra-rapid cycling"; APA 1994b).

The treatment of rapid cycling does not differ fundamentally from that of nonresponse to phase prophylaxis, except in that the use of antidepressants should be avoided in patients with rapid cycling for the reasons discussed. A number of authors have shown that a considerable percentage (72%–82%) of patients with rapid cycling do not respond to lithium treatment. Due to favorable observations made in individual cases, a question that is increasingly coming under discussion is whether patients with bipolar rapid cycling might not be better treated initially with valproate or carbamazepine.

Among experimental methods, adjuvant high-dose thyroxine treatment has been found to be of value in bipolar patients with a prophylaxis-resistant tendency toward recurrence, both with and without rapid cycling (Bauer and Whybrow 1990; Baumgartner et al. 1994; Bauer et al. 1998a; see also Chaps. 15, 23, Vol. 3, Part 1).

12 References

Akiskal HS (1995) Mood disorders: introduction and overview. In: Kaplan HI, Sadock BJ (eds) Comprehensive textbook of psychiatry/VI, vol 1. Williams and Wilkins, Baltimore, pp 1067–1079

Altshuler LL, Cohen L, Szuba MP, Burt VK, Gitlin M, Mintz J (1996) Pharmacologic management of psychiatric illness during pregnancy: dilemmas and guidelines. Am J Psychiatry 153: 592–606

Angst J (1995) The epidemiology of depressive disorders. Eur Neuropsychopharmacol Suppl: 95–98

**APA (1993) Practice guideline for major depressive disorder in adults. Am J Psychiatry 150[Suppl 4]: 1–26

APA (1994a) Diagnostic and statistical manual of mental disorders, 4th revision (DSM-IV). American Psychiatric Press, Washington DC

**APA (1994b) Practice guideline for the treatment of patients with bipolar disorder. Am J Psychiatry 151[Suppl 12]: 1–36

Bauer M, Berghöfer A (1997) Leitlinien und praktische Durchführung der Pharmakotherapie mit Antidepressiva. In: Bauer M, Berghöfer A (eds) Therapieresistente Depressionen. Springer, Berlin Heidelberg New York, pp 170–184

**Bauer M, Döpfmer S (1999) Lithium augmentation in treatment-resistant depression – a meta-analysis of placebo-controlled studies. J Clin Psychopharmacol 19: 427–434

Bauer M, Ströhle A (1999) Neue Behandlungsstrategien bei prophylaxeresistenten Bipolaren Störungen. Nervenarzt 70: 587–599

*Bauer MS, Whybrow PC (1990) Rapid cycling bipolar affective disorders. II. Treatment of refractory rapid cycling with high-dose levothyroxine: a preliminary study. Arch Gen Psychiatry 47: 435–440

*Bauer MS, Whybrow PC (1991) Rapid cycling bipolar disorder: clinical features, treatment, and etiology. In: Amsterdam JD (ed) Advances in neuropsychiatry and psychopharmacology. 2. Refractory depression. Raven, New York, pp 191–208

Bauer M, Hellweg R, Baumgartner A (1998a) Hochdosierte Thyroxinbehandlung bei therapie- und prophylaxeresistenten Patienten mit affektiven Psychosen. Nervenarzt 69: 1019–1022

*Bauer M, Hellweg R, Gräf KJ, Baumgartner A (1998b) Treatment of refractory depression with high-dose thyroxine. Neuropsychopharmacology 18: 444–455

Bauer M, Bschor T, Kunz D, Berghöfer A, Ströhle A, Müller-Oerlinghausen (1999) For how long should lithium augmentation be continued? Results of a placebo-controlled trial in unipolar depression. Presented at the Lithium Conference '99. International Society for Lithium Research. Lexington/KY, USA, 7–11 May 1999

Baumgartner A, Bauer M, Hellweg R (1994) Treatment of intractable non-rapid cycling bipolar affective disorder with high-dose thyroxine: an open clinical trial. Neuropsychopharmacology 10: 183–189

*Beck AT (1967) Depression: clinical, experimental and theoretical aspects. Harper and Row, New York

Berghöfer A, Müller EB, Bauer M, Linden M, Mackert A, Müller-Oerlinghausen B, Helmchen H (1997) Sequentielle Behandlungsstrategien zur Vermeidung und Überwindung von Therapieresistenz bei depressiven Erkrankungen. In: Bauer M, Berghöfer A (eds) Therapieresistente Depressionen. Springer, Berlin Heidelberg New York, pp 235–243

Bezchlibnyk-Butler KZ, Jeffries JJ (1996) Clinical handbook of psychotropic drugs, 6th edn. Hogrefe and Huber, Seattle

*Brunello N, Burrows GD, Jönsson CPB et al (1995) Critical issues in the treatment of affective disorders. Depression 3: 187–198

**Frank E, Kupfer DJ, Perel JM et al (1990) Three-year outcomes for maintenance therapies in recurrent depression. Arch Gen Psychiatry 47: 1093–1099

Glassman AH, Shapiro PA (1998) Depression and the course of coronary artery disease. Am J Psychiatry 155: 4–11

Goodwin GM (1997) Treatment of bipolar depressive mood disorders: algorithms for pharmacotherapy. Int J Psychiatry Clin Pract 1: S9–S12

Greil W, Kleindienst N (1997) Rezidivprophylaxe affektiver Störungen mit Lithium. In: Müller-Oerlinghausen B, Greil W, Berghöfer A (eds) Die Lithiumtherapie – Nutzen, Risiken, Alternativen. Springer, Berlin Heidelberg New York, pp 190–218

Heinz A (1997) Experimentelle Behandlungsansätze und Zukunftsperspektiven bei therapieresistenten Depressionen. In: Bauer M, Berghöfer A (eds) Therapieresistente Depressionen. Springer, Berlin Heidelberg New York, pp 235–243

Helmchen H (1974) Symptomatology of therapy-resistant depressions. Pharmakopsychiatry 7: 145 –155

*Helmchen H (1990) Gestuftes Vorgehen bei Resistenz gegen Antidepressiva-Therapie. In: Möller HJ (ed) Therapieresistenz unter Antidepressiva-Behandlung. Springer, Berlin Heidelberg New York, pp 237–250

*Hirschfeld RMA, Keller MB, Panico S et al (1997) The national depressive and manic-depressive association consensus statement on the undertreatment of depression. JAMA 277: 333–340

Kasper S (1997) Treatment of unipolar major depression: algorithms for pharmacotherapy. Int J Psychiatry Clin Pract 1: S5–S7

*Keller MB (1988) Undertreatment of major depression. Psychopharmacol Bull 24: 75–80

*Keller MB, Lavori PW, Rice J, Coryell W, Hirschfeld RMA (1986) The persistent risk of chronicity in recurrent episodes of nonbipolar major depressive disorder: a prospective follow-up. Am J Psychiatry 143: 24–28

**Kessler RC, McGonagle KA, Zhao S et al (1994) Lifetime and 12-month prevalence of DSM-III-R psychiatric disorders in the United States. Arch Gen Psychiatry 51: 8–19

*Klerman GL, Weissman MM (1992) The course, morbidity, and costs of depression. Arch Gen Psychiatry 49: 831–834

Klerman GL, Weissman MM, Rounsaville BJ, Chevron ES (1984) Interpersonal psychotherapy of depression. Basic Books, New York

**Kupfer DJ (1991) Long-term treatment of depression. J Clin Psychiatry 52[Suppl 5]: 28–34

*Kupfer DJ, Frank E, Perel JM et al (1992) Five-year outcome for maintenance therapies in recurrent depression. Arch Gen Psychiatry 49: 769–773

Lebowitz BD, Pearson JL, Schneider LS et al (1997) Diagnosis and treatment of depression in late life. Consensus statement update. JAMA 278: 1186–1190

Leonard BE (1993) The comparative pharmacology of new antidepressants. J Clin Psychiatry 54[Suppl 8]: 3–15

Leonard BE (1995) Mechanisms of action of antidepressants. CNS Drugs 4[Suppl 1]: 1–12

Linden M (1994) Therapeutic standards in psychopharmacology and medical decision making. Pharmacopsychiatry 27: 41–45

Linden M, Helmchen H, Mackert A, Müller-Oerlinghausen B (1994) Structure and feasibility of a standardized stepwise drug treatment regimen (SSTR) for depressed patients. Pharmacopsychiatry 27: 51–53

Linden M, Maier W, Achberger M, Herr R, Helmchen H, Benkert O (1996) Psychische Erkrankungen und ihre Behandlung in Allgemeinarztpraxen in Deutschland. Nervenarzt 67: 205–215

*Möller HJ (1997) Therapieresistenz unter Antidepressiva: Definition, Epidemiologie und Risikofaktoren. In: Bauer M, Berghöfer A (eds) Therapieresistente Depressionen. Springer, Berlin Heidelberg New York, pp 3–15

Müller-Oerlinghausen B (1997) Depression als unerwünschte Arzneimittelwirkung. In: Bauer M, Berghöfer A (eds) Therapieresistente Depressionen. Springer, Berlin Heidelberg New York, pp 57–64

Nelson JC (1997) Augmentation strategies for treatment of unipolar major depression. In: Rush AJ (ed) Mood disorders. Systematic medication management. Mod Probl Pharmacopsychiatry 25: 34–55

**Nierenberg AA, Amsterdam JD (1990) Treatment-resistant depression: definition and treatment approaches. J Clin Psychiatry 51[Suppl 6]: 39–47

*Nolen WA, Zohar J, Roose SP, Amsterdam JD (eds) (1994) Refractory depression: current strategies and future directions, Wiley, Chichester

Post RM, Denicoff KD, Frye MA, Leverich GS (1997) Algorithms for bipolar mania. In: Rush AJ (ed) Mood disorders. Systematic medication management. Mod Probl Pharmacopsychiatry 25: 114–145

Prien RF, Kocsis JH (1995) Long-term treatment of mood disorders. In: Floyd EB, Kupfer DJ (eds) Psychopharmacology: the fourth generation of progress. Raven, New York, pp 1067–1079

Richelson E (1994) The pharmacology of antidepressants at the synapse: Focus on newer compounds. J Clin Psychiatry 55[Suppl 9]: 34–39

**Robins LN, Regier DA (1991) Psychiatric disorders in America: the epidemiological catchment area study. MacMillan, New York

Rudorfer MV, Potter WZ (1989) The new generation of antidepressants. In: Extein IL (ed) Treatment of tricyclic-resistant depression. American Psychiatric Press, Washington, pp 83–134

Schramm E, Berger M (1998) Störungsspezifische Psychotherapie bei Depression. Münch Med Wochenschr 140: 306–313

Scott J (1988) Chronic depression. Br J Psychiatry 153: 287–297

Weissman MM, Klerman GL (1990) Interpersonal psychotherapy and its derivates in the treatment of depression. In: Manning DW, Frances AJ (eds) Combined pharmacotherapy and psychotherapy for depression. American Psychiatric Press, Washington DC

WHO (1992) Tenth revision of the International Classification of Diseases, Chapter V (F): Mental and behavioural disorders. WHO, Geneva

CHAPTER

23

S.S. Shergill, C.L.E. Katona

Pharmacotherapy of Affective Disorders

1 Overview

The pharmacological treatment of affective disorders has revolutionised the course of the illness with marked beneficial effects on the patient's social role and a reduction of the costs to society at large. The use of pharmacotherapy significantly improves the chance that a depressed patient will recover over a period of 3–6 weeks. However, pharmacotherapy itself cannot be seen outside the context of a therapeutic relationship and shared understanding of the illness. A therapeutic alliance is not only useful in minimising non-compliance, but also facilitates a thorough explanation of likely adverse effects of any drug treatment given and the anticipated time taken to respond to treatment, as well as the relevance and likely duration of drug treatment within an overall treatment plan.

Although there have been effective treatments for depression for over 40 years, it is only relatively recently that there has been a wide and growing choice of antidepressants acting on differing receptor subtypes with varying specificity. It will take some time, with more experience through widespread use, before the optimal niche for each of these different drugs is found. Meanwhile, the newer drugs in general are safer and have a more favourable side-effect profile than previously available drugs. Only a decade or so ago, for example, all available antidepressants were highly dangerous in overdose. A further factor to consider is the increasing range of indications being identified for many of the newer antidepressants. Some, for example, have been shown to be useful in obsessive-compulsive, eating or anxiety disorders.

The range of available treatments for bipolar affective disorders (BAD), especially in the treatment and prophylaxis of mania, have not expanded to nearly the same extent as treatments of depressive illness. Increasing evidence has, however, emerged for the efficacy of anticonvulsant medication such as carbamazepine and sodium valproate, both in the treatment of acute mania and as mood stabilisers in the prophylaxis of bipolar disorder. There has also been an increased interest in drugs such as verapamil (a calcium channel blocker), clonazepam (a benzodiazepine also used in the treatment of epilepsy) and the atypical antipsychotics such as clozapine, olanzapine and risperidone.

The following routine investigations are recommended before and during treatment, with suggested tests for different drugs:

- Physical examination
- Thyroid function tests
- Full blood count (FBC)
- Urea, electrolytes (U+E) and general chemistry screen
- Urine toxicology for substance abuse
- Electrocardiography (ECG) in patients over 40 years
- Pregnancy test if relevant
- Serum levels of lithium; monitor thyroid function, urinalysis and U+E
- Serum levels of valproate; monitor U+E, FBC and liver function tests
- Serum levels of carbamazepine; monitor FBC and liver function tests
- Serum levels of the following tricyclics: amitriptyline, imipramine, desipramine and nortriptyline

2 Depressive Disorder

2.1 Available Drugs

The serendipitous discovery of the antidepressant effect of imipramine in 1958 (Kuhn 1958) led to intensive testing of the group of drugs known as the tricyclics in depressive illness. The almost simultaneous and equally serendipitous observation of mood elevation in tuberculous patients treated with iproniazid (Bloch et al. 1954) led to the development of the monoamine oxidase (MAO) inhibitor (MAOI) family of antidepressants. These focused research interest on changes in the activity of the biogenic amine neurotransmitters noradrenaline and 5-hydroxytryptamine (5-HT, serotonin) as mediators in the treatment of depression. Despite early disagreements about the relative contribution of these two receptor systems, it is now generally accepted that there is a complex interrelationship between them and also the dopaminergic system (Moller and Volz 1996). In the light of this, the possibility of manipulating one system without compensatory changes in the others is very small. Models of antidepressant action based simply on changes in availability of one neurotransmitter are clearly overly simplistic.

The research emphasis on amine neurotransmitters has, however, led to the development of the closely related tetracyclic drugs, which together with the tricyclics and MAOI comprise the classic antidepressants. Subsequent progress has resulted in the availability of the serotonin-specific re-uptake inhibitors (SSRI), other atypical antidepressants such as bupropion (used more widely in the United States than in Europe), trazadone and nefazadone, the specific and reversible MAO type A inhibitor moclobemide, and more recently introduced drugs such as mirtazapine, reboxetine and venlafaxine.

These newer classes of antidepressant have varied specificity for the various neuroreceptors, shown in Table 1, (modified from Richelson 1996). This is reflected in their differing side-effect profiles. At the present time, all of these antidepressant drugs, despite their different modes of action, still show similar delays in response (around 2–3 weeks) and no improvement in overall efficacy to that found with classic antidepressants (Cohen 1997). However, they do demonstrate clear advantages in adverse effect profile (reflecting both safety and side-effects) and in particular tend to show significantly decreased levels of cardiotoxicity and sedation.

2.1.1 Tricyclic and Tetracyclic Drugs

The group of tricyclic and tetracyclic drugs includes amitriptyline, clomipramine, desipramine, imipramine, lofepramine, nortriptyline and trimipramine. These are indicated for the treatment and prophylaxis of major depression, the treatment of secondary depression (in association with dementia, Parkinson's disease, etc.) and the depressed phase of BAD. Other approved indications for specific drugs include clomipramine for obsessive-compulsive disorder and imipramine for childhood enuresis.

The short-term effects of these drugs are to reduce the re-uptake of noradrenaline and 5-HT. This has provided one of the cornerstones of the monoamine hypothesis of mood disorders. In addition, they block the muscarinic subtype of acetylcholine receptors and histamine receptors. Longer-term effects (with a time course similar to that of onset of antidepressant action) include down-regulation of β-adrenergic receptors and (possibly) a similar effect on 5-HT receptors.

These drugs are administered orally, and there is a large first-pass effect. They are highly plasma protein bound, with predictable interactions with other similarly bound drugs, and are metabolised primarily in the liver. They demonstrate linear pharmacokinetics, with an increase in dose reflected in the plasma concentration. They have long elimination half-lives (between 10 and 70 h) compatible with once-daily dosage once the steady state has been achieved (usually after approximately 5 days).

The side-effect profile within this group of drugs varies and is dependent on the affinity for the various neuroreceptors (see Table 1); the clinical effects of receptor action are shown below. The effects of acetylcholine muscarinic receptor antagonism include:

- Dry mucous membranes, mouth, eyes, etc.
- Blurred vision
- Constipation
- Urinary retention
- Excess sweating
- Confusion, disorientation, delirium, delusions and hallucinations

The histamine H_1 and α_1-adrenoceptor antagonism effects include:

1. Cognitive effects:
   - Drowsiness
   - Weakness, fatigue
   - Nightmares, agitation, restlessness and insomnia
   - Confusion, disorientation
2. Neurological effects:
   - Fine tremor
   - Akathisia
   - Tardive dyskinesia
   - Seizures
   - Paraesthesia
   - Parkinsonism, dystonia and gait disturbance
3. Cardiovascular effects (exacerbated by muscarinic and 5-HT_2 antagonism):
   - Tachycardia
   - Orthostatic hypotension
   - Increased conduction times
   - Arrhythmias, syncope
4. Gastrointestinal (GI) effects (exacerbated by 5-HT uptake inhibition and muscarinic antagonism):
   - Anorexia, nausea, vomiting and diarrhoea
   - Glossitis and peculiar taste
5. Sexual effects (exacerbated by increased dopaminergic D_2 activity and 5-HT_2 antagonism):
   - Decreased libido and impotence
   - Testicular swelling and retrograde ejaculation
   - Breast engorgement and enlargement
   - Anorgasmia
   - Amenorrhea
   - Galactorrhoea
6. Others:
   - Weight gain (common)
   - Blood dyscrasia (rare)
   - Jaundice, hepatitis and urticaria (rare)
   - Photosensitivity and skin pigmentation (rare)

The therapeutic window is narrow; toxic effects (as seen in overdose) largely reflect anticholinergic and antihistaminic effects culminating in central nervous system depression, altered consciousness, respiratory slowing and onset of seizures. The cardiac effects are the most hazardous; the severity of the toxicity can be monitored electrocardiographically in terms of prolongation of the QRS complex.

Tricyclic and tetracyclic antidepressants have not been shown to be teratogenic and may reasonably safely be used during pregnancy. As with all other drugs, their use in pregnant women should be avoided,

Table 1. Effects of antidepressants on neurotransmitters/receptors (dissociation constants)

	Amitriptyline	Clomipramine	Imipramine	Nortriptyline	Trazadone	Nefazadone	Bupropion	Venlafaxine	Fluoxetine	Sertraline	Paroxetine	Mirtazapine	Reboxetine
Noradrenaline re-uptake inhibition	10–100	10–100	10–100	1–10	1000–10,000	100–1000	1000–10,000	100–1000	100–1000	100–1000	10–100		1–10
5-HT re-uptake inhibition	10–100	1–10	10–100	100–1000	100–1000	100–1000		10–100	10–100	1–10	0.1–1		
DA re-uptake inhibition	1000–10,000	1000–10,000	1000–10,000	1000–10,000		1000–10,000	100–1000	1000–10,000	1000–10,000	100–1000	1000–10,000		
5-HT_1 receptor antagonism block	100–1000	1000–10,000	1000–10,000	100–1000	10–100	10–100							
5-HT_2 receptor antagonism block	10–100	10–100	10–100	10–100	1–10	10–100			100–1000	1000–10,000		1–10	
ACh block	10–100	10–100	10–100	100–1000					1000–10,000	100–1000	100–1000	1000–10,000	1000–10,000
H_1 block	1–10	10–100	10–100	10–100	100–1000		1000–10,000		1000–10,000			0.1–1	1000–10,000
α_1-adrenoceptor block	10–100	10–100	10–100	10–100	10–100	100–1000	1000–10,000		1000–10,000	100–1000	1000–10,000	100–1000	1000–10,000
α_2-adrenoceptor block	100–1000	1000–10,000	1000–10,000	1000–10,000	100–1000	1000–10,000				1000–10,000	1000–10,000	10–100	1000–10,000
D_2 block	1000–10,000	100–1000	1000–10,000	1000–10,000	1000–10,000	100–1000			1000–10,000			1000–10,000	

5-HT, hydroxytryptamine; DA, dopamine; ACh, acetylcholine; H_1, histaminergic; D_2, dopamine 2 receptor.

if possible, especially during the first trimester. They are also secreted in small amounts into breast milk (1%), although this is thought unlikely to represent much hazard to the baby. They should be used cautiously in patients with hepatic or renal disease.

All tricyclic and tetracyclic antidepressants are equally efficacious in the treatment of depressive disorder. However, the choice of drug and dosage regime must be tailored individually, taking into related effects such as sedation, propensity to hypotension and risk of overdose. More sedative antidepressants (such as amitriptyline or trimipramine) may be particularly suitable for agitated patients with poor sleep. Similarly, less sedating antidepressants, such as lofepramine, imipramine and nortryptiline, are indicated for more withdrawn patients. The less cardiotoxic drugs such as lofepramine should be used for patients at higher risk of taking an overdose and in the treatment of elderly patients. Clomipramine is the treatment of choice in patients presenting with concomitant obsessive-compulsive symptoms.

Treatment should be preceded by an ECG in patients over 40 years and routine physical examination, blood count and serum electrolytes and liver function tests. The dose of antidepressant should be increased gradually with a small initial starting dose and incremental raise every few days up to the therapeutic dose. In the case of amitriptyline, the dose should be raised to 150 mg and sustained for 2–3 weeks. In the absence of clinical improvement, the dosage should be increased towards 300 mg if side-effects permit. Recommended doses are usually between 150 and 250 mg a day, and the use of lower doses has been criticised (Quitkin 1985). The response rate to routine doses is approximately 70%. Common side-effects, such as sedation, postural hypotension, dry mouth, blurred vision, constipation and urinary retention, can be ameliorated by initially spreading the dose through the day, although this can be changed to once-daily dosage, for convenience, once the therapeutic dose is reached.

2.1.2 Monoamine Oxidase Inhibitors

The MAOI iproniazid was developed in the early 1950s from the finding that a closely related drug, isoniazid, caused euphoria in patients treated for tuberculosis. This group of drugs (which includes phenelzine, isocarboxazid and tranylcypramine) irreversibly inhibits both subtypes of MAO: MAO-A, which is responsible for degrading noradrenaline and 5-HT, and MAO-B, which is specific for phenylethylamine. Tranylcypramine has an additional direct effect on inhibiting re-uptake of noradrenaline and 5-HT.

Both MAO-A and MAO-B are involved in the metabolism of dopamine. Their efficacy added further support in the 1960s to the monoamine theory of depression and also led to the use of other MAOI compounds; they were subsequently shown to be effective in the treatment of depression.

MAO is widely distributed throughout the body, in the central nervous system, sympathetic terminals, liver, platelets and within the gastrointestinal system, where it is responsible for the metabolism of dietary tyramine. Inhibition of MAO results in passage of tyramine directly into the circulation and a resultant hypertensive crisis. Phenelzine, isocarboxazid and tranylcypramine are irreversibly bound and non-specific, acting on both types of MAO. Reports of potentially fatal interactions with tyramine-rich foods, such as cheese, in patients taking MAOI led to reduction in their usage and realisation of the need to avoid tyramine-containing foods if taking them.

The MAOI have similar indications to the tricyclic and tetracyclic antidepressants. They may be better than the tricyclics in the treatment of "atypical" depression (characterised by hypersomnia, hyperphagia, anxiety and without prominent somatic symptoms; Cohen 1997). They have also been used for the treatment of panic disorder with agoraphobia, post-traumatic stress disorder and social phobia.

MAOI are well absorbed from oral administration. Phenelzine and isocarboxazid are metabolised by acetylation, which is dependant on the patient's acetylation status. Approximately half of the European and a higher proportion of the Asian population are slow acetylators, which can lead to higher than expected serum levels and an increase in adverse effects, even at moderate doses. The half-life of these drugs tends to be around 2–6 h, and they require multiple doses during the day.

The tyramine reaction is an important feature of the side-effect profile of these drugs, and patients need to be warned about the need to avoid ingestion of large doses of tyramine-containing food. Foods to be avoided include the following:

- Alcohol (particularly Chianti and beer)
- Broad beans
- Cheese (especially mature cheeses; cream cheese does not interact)
- Liver (especially beef and chicken)
- Smoked or pickled fish or poultry
- Soups or pate (commercial preparations)
- Yeasts and meat extracts
- Green banana skins

The following should only be eaten in moderation:

- Avocados
- Aubergines
- Raisins
- Soy sauce

- Sour cream
- Spinach
- Yoghurt

In addition, the possibility of interactions with other medications should also be pointed out. The following drugs are contraindicated:

- Anaesthetics (particularly administered with adrenaline)
- Bronchodilators
- Adrenergic antihypertensives (methyldopa, guanethidine)
- L-Dopa or tryptophan
- Opiates
- Cold/flu preparations (containing sympathomimetics, especially dextromethorphan)
- Sympathomimetics (including illicit drugs containing cocaine and amphetamines)
- Serotonin specific re-uptake inhibitors (including clomipramine, trazadone and tricyclics)

The following drugs should be used with caution:

- Antihistamines
- Codeine
- Hydralazine
- Propranolol
- Tricyclic and tetracyclic drugs

Note: The half-life of any contraindicated drugs must be considered before changing to an MAOI, and an adequate washout time must be allowed before starting treatment. The most common antidepressant combinations are phenelzine with amitriptyline or trimipramine. The tricyclic is started before the phenelzine or both are commenced together at a low dose. The traditional MAOI inhibit the enzyme irreversibly; even after stopping the medication, its effects will therefore continue for about 2 weeks until adequate monoamine oxidase is formed. Thus a washout period is required before any contraindicated drugs can be commenced.

The commonest adverse effects are orthostatic hypotension, weight gain, sexual dysfunction and insomnia. Very occasionally, patients present with paraesthesia, secondary to pyridoxine deficiency. Phenelzine and isocarboxazid have been associated with hepatoxicity, and treatment requires monitoring of liver function tests. MAOI are less cardiotoxic and less likely to lower the seizure threshold in patients than tricyclics. They are contraindicated in pregnancy and pass into breast milk. They need to be used with caution in patients with renal, hepatic or seizure disorders.

Moclobemide is a newer drug which is relatively specific for MAO-A and is reversibly bound (an example of a novel class of drugs known as reversible inhibitors of MAO-A, RIMA); this enables it to be displaced by tyramine and makes it far less likely to precipitate a hypertensive crisis. Selegiline is a specific inhibitor of MAO-B used predominantly in Parkinson's disease which may also (in high and non-MAO-B-selective dosage) be effective in refractory depression (Sunderland et al. 1994).

There is no clear evidence to suggest that there are significant differences in the efficacy of any one MAOI over another. The side-effect profile dictates treatment choice. Moclobemide is least likely to precipitate a tyramine reaction, although being the most recently introduced MAOI, there is little long-term data on its effectiveness in a prophylactic or maintenance role. Tranylcypramine is a more "activating" antidepressant and has relatively low hepatotoxicity. Phenelzine is advocated as the safest of the three traditional MAOI, but, like isocarboxazid, it may be hepatotoxic. All MAOI require a slow build-up of dose initially and administration two or three times a day.

2.1.3 Serotonin-Specific Re-uptake Inhibitors

The serotonin specific re-uptake inhibitors (SSRI) are the fruits of the renewed interest in the early 1970 in the role of the 5-HT system in depressive illness. The family of SSRI now includes fluoxetine, paroxetine, fluvoxamine, sertraline and citalopram.

The main indication for treatment with the SSRI is major depression, within which their use (both for acute treatment and in prophylaxis) is similar to that of the tricyclics. They are also indicated for the treatment of obsessive–compulsive disorder, panic disorder and bulimia (fluoxetine). Outside their approved indications, they shown efficacy in dysthymia, agoraphobia and impulsive behaviour.

The exact mechanism of the antidepressant effect of the SSRI remains unclear. They inhibit serotonin re-uptake without direct effects on adrenergic or dopaminergic re-uptake. The result is a down-regulation of 5-HT receptors, e.g. 5-HT_2. This has ill-defined indirect effects on the other neurotransmitter systems. There is some variability between SSRI in their in vitro specific affinity for serotonin and their potency in blocking its re-uptake, but this does not appear to have any clinical significance. The absence of significant anticholinergic, antihistaminergic and anti-adrenergic effects is responsible for their relatively benign side-effects. The commonest adverse effects with SSRI treatment in the central nervous system are headaches, anxiety, insomnia and weight loss; in the gastrointestinal system, nausea, diarrhoea and anorexia are relatively often seen. Sweating is also sometimes reported. The initial reports of increased suicidal acts in patients being treated with fluoxetine have not been

supported. Less common side-effects include disturbances of sexual function (anorgasmia, delayed ejaculation and impotence) and rashes. The SSRI have very low toxicity in overdose.

The SSRI are rapidly absorbed orally, undergoing little first-pass metabolism. They are highly plasma protein bound and will displace other drugs from protein binding, serving to elevate their plasma level; fluvoxamine has the lowest plasma protein-binding affinity. They all undergo hepatic metabolism and to some extent affect the cytochrome P450-metabolising system. This may result in an increase in levels of other drugs metabolised by this route, including antipsychotics such as clozapine.

The SSRI have many clinically significant drug interactions; there are potentially fatal interactions with concomitant L-tryptophan and MAOI treatment. Caution should be exercised with administration of benzodiazepines, antipsychotics, other antidepressants and lithium. Cimetidine has been shown to increase levels of some SSRI.

Treatment with the SSRI can usually be initiated with the therapeutic dose, although some SSRI, e.g. fluvoxamine, paroxetine and sertraline (see the pharmaceutical data sheets provided with the respective drug by the manufacturers), require upward titration if effects are not seen within 4 weeks. They have relatively long elimination half-lives, allowing once-daily dosage. This is particularly the case for fluoxetine and its active metabolite norfluoxetine (70 and 330 h, respectively) and can cause difficulties if a change of medication becomes necessary, as interactions will occur up to 5 weeks after discontinuing treatment. Higher doses of SSRI have been suggested in the treatment of concomitant or "stand-alone" obsessive-compulsive disorder (paroxetine) and bulimia (fluoxetine). Their low level of side-effects makes them suitable for treatment of elderly patients. There have been reports of SSRI withdrawal phenomena (particularly with paroxetine); these include symptoms of anxiety, malaise, akathisia and dyskinesia. They should be distinguished from depressive relapse associated with poor compliance. If they occur during planned cessation, it is suggested that the drug should be re-introduced and withdrawn more gradually.

The ease of administration (standard, single daily dose) of the SSRI and their lack of severe adverse effects has led to their increasingly widespread use and may even have contributed to the shift of care of depressed patients from hospitals into the primary care services. Fluoxetine has had an enormous effect on the public awareness of antidepressant medication and has been requested and prescribed for symptoms falling short of significant depressive episodes, including sensitivity to criticism and low self-esteem (Barondes 1994). The relative expense of the SSRI compared to the tricyclics for essentially equivalent efficacy has resulted in considerable debate as to the financial implications of antidepressant prescription in general. Economic analyses are discussed later in this chapter.

2.1.4 Trazadone and Nefazadone

Trazadone and nefazadone are structurally unrelated to the tricyclics, tetracyclics, SSRI or MAOI but provide equally effective treatment of depressive illness. Nefazadone, a very recently introduced drug, is structurally related to trazadone, but has a different pharmacological profile. They have very little in the way of anticholinergic side-effects and differ from each other in their degree of sedation, with trazadone but not nefazadone having a marked sedative effect.

These drugs are indicated for the treatment of major depression. Trazadone acts as a specific inhibitor of 5-HT re-uptake and is also a post-synaptic 5-HT_2 antagonist. The list in Sect. 1 reflects its additional anti-adrenergic and antihistaminergic effects. Nefazadone also inhibits 5-HT re-uptake and acts as an antagonist at the 5-HT_2 post-synaptic receptors. It has less blocking effect on histaminergic and α-adrenergic receptors than trazadone.

Trazadone and nefazadone are both readily absorbed from the gastrointestinal tract with hepatic metabolism. Trazadone has a longer half-life (6–11 h) than nefazadone (2–4 h). The commonest side-effects with trazadone are sedation, orthostatic hypotension, dizziness and nausea. It is safe in overdose and has the rare side-effect of priapism. Adverse effects with nefazadone are nausea, headache, drowsiness and dizziness. It is relatively free of adverse sexual effects, and cardiotoxicity and orthostatic hypotension are also uncommon. Both drugs are contraindicated in the treatment of pregnant or breast-feeding women.

Nefazadone has been shown to reduce depression-associated anxiety and to improve sleep patterns. It may be particularly useful in patients experiencing intolerable sexual side-effects from other antidepressants. Unlike trazadone, it does however entail the relative inconvenience of twice-daily dosage. Trazadone has similar indications but is more sedative, benefiting patients with prominent insomnia.

2.1.5 Other Antidepressant Drugs

Bupropion

Bupropion is a unicyclic antidepressant drug, originally formulated in 1966 but not used for the treatment of depression until 1985. It was suspended from use in

the United States when it was associated with increased incidence of seizures in treated patients. This adverse effect was subsequently realised to occur at similar levels to other antidepressants, provided therapeutic doses were not exceeded. It was reinstated in 1989.

It is indicated for the treatment of depression. Its mode of action is unclear; it has low-grade re-uptake activity at noradrenergic receptors and possibly dopaminergic receptors. It is well absorbed orally, metabolised via the liver and has a half-life of approximately 12 h.

Bupropion has the rare benefit of being devoid of any pharmacologically distinct activity outside the central nervous system. It is virtually devoid of side-effects in terms of orthostatic hypotension, weight gain or sexual dysfunction. The most commonly reported adverse effects are headache, insomnia and nausea. The risk of seizures has been demonstrated to be dose related and comparable to treatment with tricyclic antidepressants at therapeutic doses. Risk of seizures does, however, rise significantly if high doses are given. Bupropion is contraindicated in patients with a history of brain trauma, seizures or other neurological disease; it should also not used in pregnant or breast-feeding women.

Mirtazapine

Mirtazapine is a recently introduced antidepressant with unique pharmacology. It is marketed as a noradrenergic and specific serotonergic antidepressant (NaSSA). It antagonises presynaptic α_2-adrenoceptors and thereby enhances serotonergic and noradrenergic transmission. It also acts to antagonise post-synaptic 5-HT_2 and 5-HT_3 receptors. This acts to increase the release of 5-HT to act on 5-HT_1 receptors, which may be relevant to its action in improving sleep and anxiety. It has a high affinity for histamine H_1 receptors but little effect on cholinergic or α_1-adrenoceptors.

Mirtazapine is indicated for the treatment of depression. It is extensively metabolised hepatically with a half-life of 20–40 h, permitting once-daily dosage. Prominent adverse effects are sedation, weight gain and dry mouth. It offers the advantage (in comparison with the SSRI) of less nausea and sexual dysfunction with more sedation and some weight gain, which may be of some benefit in suitable patients.

Reboxetine

Reboxetine was also recently introduced and is a selective inhibitor of noradrenaline re-uptake. It has minimal affinity for the serotonergic and dopaminergic systems and does not have major actions on muscarinic, histaminergic or adrenergic receptors. It is indicated for the treatment of major depression. It undergoes hepatic metabolism and requires twice-daily dosage. It is highly protein bound, with a potential interaction with imipramine, chlorpromazine, propranolol, methadone and other similarly bound drugs. The commonest adverse effects are dry mouth, constipation, insomnia and sweating; less common effects include tachycardia, vertigo, urinary hesitancy and impotence.

Reboxetine may be a useful drug for those patients intolerant of SSRI or tricyclic antidepressants. It may also be useful in patients with Parkinson's disease who are particularly sensitive to SSRI-induced extrapyramidal side-effects.

Venlafaxine

Venlafaxine is a bicyclic agent marketed as a serotonin and noradrenaline re-uptake inhibitor (SNRI). It is similar in action to other drugs soon to be released onto the market such as milnacipran. Venlafaxine is a potent inhibitor of 5-HT re-uptake and also causes some inhibition of noradrenaline re-uptake. It has a similar although weaker effect on dopaminergic re-uptake. It has negligible effects on muscarinic, α-adrenergic and histaminergic receptors.

Venlafaxine is indicated for the treatment of depression, and it has been claimed that it is effective in resistant depression. It is hepatically metabolised into another active compound (*O*-desmethylvenlafaxine, ODV). Venlafaxine and ODV possess half-lives of 5 and 11 h, respectively. The routine preparation required twice-daily administration, although a once-daily slow-release preparation is now available. The major route for excretion is via the urine, so doses need to be lower in patients suffering concomitant renal or hepatic disease. The commonest side-effect is nausea; this has been reduced by the introduction of the slow-release preparation. Other adverse effects are headaches, sweating, anxiety and sexual dysfunction. Venlafaxine should be used with caution in patients with hypertension and avoided in pregnancy and while breast-feeding. Abrupt withdrawal has been associated with insomnia, nausea, dizziness and nervousness, and it is recommended that the dose is tapered prior to withdrawal. No dose adjustment is suggested for treatment in elderly patients.

Venlafaxine appears to be a promising drug with a good effect on depressive illness. The use of a slow-release preparation has decreased the otherwise important disadvantage of a high incidence of initial nausea, which, with the original preparation, affected 25%–37% of treated patients. The claims of faster onset of action and effects in refractory depression require further confirmation.

2.2 Choice of Drug

As all antidepressants seem to be equally efficacious in the treatment of depression, other characteristics serve

to distinguish them from each other. While there is little disagreement that the newer antidepressants have a safer side-effect profile, there has been a debate whether the newer antidepressant drugs are as effective in severe depression. Perry (1996) reviewed the literature from controlled trials using the Hamilton Depression Rating Scale changes and suggested that tricyclic antidepressants are more effective in severe endogenous depression; Nierenberg (1994) reviewed the literature comparing SSRI and tricyclic antidepressants in the treatment of severe depression and concluded that the SSRI were equally effective. The latter view is probably a fair reflection of contemporary opinion. The previous dichotomy between psychotic and neurotic depressions is largely unhelpful in guiding treatment, and current thinking views psychotic depression as being at the extreme end of a spectrum. Adjunctive neuroleptic medication is often required in the treatment of severe depression accompanied by psychotic symptoms; this can be withdrawn once the psychotic symptoms are in remission and treatment with antidepressants is continued.

The choice of antidepressant can be guided by the receptor and consequent side-effect profile; in the elderly, significant antagonism of the α_1-adrenergic receptors should be avoided in order to prevent orthostatic hypotension and subsequent falls. Similarly, antidepressants with significant antimuscarinic activity should also be avoided in order to reduce the greater sensitivity to the associated effects of dry mouth, increased urinary retention and constipation. Toxicity in overdose would be a consideration in patients with prominent suicidal ideation treated as outpatients, and the less cardiotoxic drugs would be the drugs of choice. In a similar vein, attempts are made to identify features of the somatic presentation which can be ameliorated using the side-effect profile in an advantageous manner. This would, for example, suggest the prescription of more sedative medication for patients with prominent insomnia and agitation.

Interactions with other medications taken by the patient are a major consideration. It would be inadvisable to treat a patient with a prominent 5-HT re-uptake inhibitor if he or she is being treated with selegeline for Parkinson's disease, for fear of precipitating a serotonergic syndrome as a result of enhanced serotonergic activity. The serotonin syndrome consists of at least three of the following symptoms: confusion, hypomania, agitation, myoclonus, hyperreflexia, diaphoresis, shivering, tremor, diarrhoea, incoordination and fever (Sternbach 1991). Interactions between the newer antidepressants involving their effects on metabolism of other drugs, particularly via the cytochrome P450 system, have been reviewed by Nemeroff et al. (1996). They found that each of the newer antidepressants inhibited a different cluster of cytochrome P450 enzymes. Fluvoxamine inhibited the cytochrome P450 $1A_2$ and caused interactions with theophylline and clozapine, among others. Fluoxetine, sertraline and paroxetine inhibited the cytochrome P450 2D6 enzyme and resulted in elevated levels of desipramine and nortriptyline. Fluoxetine, sertraline and fluvoxamine inhibited the cytochrome P450 2C enzyme and elevated levels of phenytoin and diazepam, among others; these and nefazadone inhibited the cytochrome P450 3A4 enzyme and increased levels of co-administered carbamazepine, terfenadine, alprazolam and other benzodiazepines. Overall, these findings emphasise the need to check known metabolic interactions and to phenotype patients who present as very sensitive to particular drugs.

2.3 Pharmacoeconomics of Antidepressant Drug Prescription

The importance of evaluating the financial aspects of clinical treatments is increasingly being recognised; drug treatments are no exception. An economic evaluation simply comparing costs of providing individual drug treatments for depression show an enormous financial advantage in prescribing older drugs such as the tricyclic antidepressants. This rather simplistic system, currently used by many agencies, has been heavily criticised for ignoring other fundamental treatment variables. These other considerations can be summarised as follows (after Burke et al. 1994):

1. Cost of providing treatment
   a) Medication price
   b) Professional services
   c) Medical tests
   d) Special dosing requirements
2. Likelihood of treatment success
   a) Efficacy (does it work?) vs. effectiveness (likelihood of success)
   b) Ease of dosing
   c) Tolerability and compliance
3. Toxicity
   a) Therapeutic index of medication
   b) Cost of treating adverse effects
   c) Lethality in overdose
4. Potential for drug interactions
5. Interpatient variability
6. Maintenance and prophylaxis

Several factors have been identified as relevant to service utilisation. Time to onset of drug action is clearly important, and this may possibly be reduced with venlafaxine (Guelfi et al. 1995). Dose titration may be labour intensive and cause delays in the case of the tricyclic antidepressants, for example. Some antidepressants require medical tests both prior to initiation

and during treatment monitoring, such as ECG in patients over 40 treated with tricyclics. Drugs may vary in likelihood of response – this can be affected by differences between "real-life" and clinical trial populations and the difficulty in attaining therapeutic dosages of drugs without the intensive support available during a drug study. Drug toxicity may be critical; economic aspects of this include the costs of treating adverse effects directly and the indirect medical costs, e.g. treatment of falls secondary to orthostatic hypotension after treatment with antidepressants with marked adrenergic blockade. Similar considerations apply to treatment of overdosage, factors such as number of days of bed usage being higher for the more toxic antidepressant medications. Tolerability will presumably affect compliance and treatment success and is largely dependent on the adverse side-effect profile. Economic factors are not static across treatment populations; factors such as age and gender may influence liability to particular adverse effects and need to be considered in guiding the optimal choice of drug.

Economic analyses seek to answer the question of which is the most cost-effective treatment for depression using different economic models (Stoddart and Drummond 1984) with varying degrees of complexity. There are two major approaches: partial and complete economic analyses. Partial economic analysis describes either the costs or the outcomes of a treatment. The full analysis examines costs and outcomes together. The interest in treatment of depression requires the use of the full analysis, as both outcome and cost are of concern. There are several different types, and the reader is referred to Chap. 16 (Vol. 1, Part 1) for a fuller discussion of pharmacoeconomics in general. Despite the limitations of individual methods of economic analysis, they provide valuable information on the economic impact of the major depressive disorders. A few examples of these types of analyses applied to treatment of depression are illustrated below.

Health Maintenance Organisations (HMO) in the United States can generate data for a partial economic analysis with relative ease, as they are aware of the direct costs not only for drugs but also for clinic visits and hospitalisation. Comparison of different SSRI (Sclar et al. 1995) illustrates the importance of dose titration even among these drugs. These authors found that dose titration with SSRI correlated with the per capita costs of treatment of depression. Thus patients prescribed paroxetine or sertraline (requiring upward titration in 28% and 40% of patients) were found to result in greater costs to the HMO than fluoxetine (16%). The authors emphasise that this is not an assessment of the cost-effectiveness of these treatments, but only a measure of direct expenditure. The debate over the economic benefits of using tricyclics or SSRI continues apace with research supporting both points of view (Jonsson and Bebbington 1993; Hotopf et al. 1996; Sclar et al. 1994). The discussion and correspondence following publication of these studies has highlighted the difficulties inherent in economic analyses. Priest (1996) reviewed the cost-effectiveness of venlafaxine and emphasised the fact that direct antidepressant drug costs only represent approximately 10% of the full economic costs. Using the various clinical treatment and outcome variables listed above, he concluded that venlafaxine offers financial advantages over treatment with both tricyclic antidepressants and SSRI.

2.4 General Clinical Guidelines for Treatment

The treatment of depression can best be conceptualised as occurring in three phases: acute treatment, continuation treatment and maintenance treatment. Acute treatment is given to achieve remission of the depressive symptoms as quickly as possible. Continuation treatment is to continue improvement through to complete recovery (usually 3–6 months), i.e. to prevent relapse. Maintenance treatment is designed to avoid the development of a new depressive episode.

2.4.1 Acute Treatment

There is ample evidence that treatment success is in large part dependent on an adequate dose of drug being given for an adequate time. Achieving this with the tricyclic and tetracyclic drugs was limited by the need to minimise their side-effects. Dosage adequacy with these drugs could best be achieved through gradual upward titration. Some patients may be unable to tolerate optimal doses but may show a partial response. The plan should be to achieve the highest tolerated dose; if this is ineffective, patients should be changed to a different class of antidepressant, and if there is an adequate response at a low dose, they should remain on this dose unless improvement ceases. If the patient does not respond after 4 weeks, a plasma level may be useful to ascertain adequate compliance and any unusual pharmacokinetic actions, allowing change of dose or drug.

There have been inconsistent findings from studies seeking to relate drug dosage, plasma levels and antidepressant effect; this is unsurprising given the ten-fold variation in the plasma level among individuals given an equal dose of tricyclic antidepressant (Asberg 1976). A review by an American Psychiatric Association Task Force (Task Force on the Use of Laboratory Tests in Psychiatry 1985) concluded that there was good evidence of a linear relationship

between plasma level of imipramine and clinical outcome. The task force suggested that the percentage of patients responding favourably increases with increasing plasma levels up to 250 mg/ml; lower levels may result in response in some patients, but higher levels were unlikely to produce further benefit. The review suggested that nortriptyline has a different response curve, with a therapeutic window of good antidepressant effect achieved at 50–150 mg/ml; this level is achieved by the majority of patients if treated with 150 mg a day (Kragh-Sorensen et al. 1973). Surprisingly, the task force found inconsistent evidence of any such window for amitriptyline, despite nortriptyline occurring as its major metabolite. The newer antidepressants provide what is usually a therapeutic dose from the onset of treatment. Dose increase may nonetheless be indicated in the absence of response after 3–4 weeks of treatment.

Intravenous administration of antidepressants for treatment of depressive disorders has been thought to achieve a more rapid onset of action than conventional oral dosing and has increased in usage (Laux 1993). Other advantages are assured compliance and the use of lower medication doses as there is no first-pass effect. There are relatively few controlled trials, but there is some evidence for the above hypothesis with the use of intravenous maprotiline (Gastpar et al. 1986), doxepin (Laux et al. 1989), clomipramine (Pollock et al. 1989) and citalopram (Bouchard et al. 1997). Loss of efficacy when patients are switched to oral medication was not a problem when adequate oral doses were commenced (Adler et al. 1997). Overall, there is still a paucity not only of studies separating out the psychological from the physical benefits but also ones identifying which patients are most likely to accept and respond to intravenous administration of antidepressants.

2.4.2 Continuation Treatment

Continuation treatment is based on the premise that antidepressant (and antimanic) drugs can suppress the symptoms of the illness while leaving an underlying psychopathological process continuing on its natural course (Prien and Kupfer 1986). Thus withdrawal of the drug before the natural cycle of the disease ends will result in a relapse of the illness. Several placebo-controlled discontinuation design trials have demonstrated the need for continuation treatment; these are reviewed by Prien (1992). The effectiveness of SSRI is suggested to be comparable with that of the tricyclics and tetracyclic drugs during this continuation phase. Prien and Kupfer (1986) suggest that the continuation phase should last at least 4 months from the time of remission of symptoms. The results from the studies carried out by Montgomery and Dunbar (1993) and Doogan and Caillard (1992) suggest that patients should receive 6 months of full-dose treatment from remission of symptoms. There is no overall consensus on the dose of treatment during the continuation phase. It would, however, seem prudent to continue full-dose treatment in patients with previous severe episodes of depression and to attempt a lower-dose strategy in patients with less severe episodes, the latter also receiving regular review and rapid re-establishment of the full dose if symptoms reoccur.

2.4.3 Maintenance Treatment

The chronic and recurrent nature of depressive illness is well recognised. Maintenance treatment should be initiated in patients with a history of three or more episodes of major depression, two or more episodes with rapid recurrence, patients with dysthymia and patients at risk of a life-threatening episode or severe disruption to their functioning (Hirschfield and Schatzberg 1994).

Patients treated with maintenance therapy should be given the same dose that was used to treat their initial episode (Kupfer et al. 1992; Hirschfield and Schatzberg 1994). Full-dose maintenance should be continued for 4–5 years or the equivalent of the duration of two episodes of illness. Patients with three or more episodes of depression should be considered for life-long maintenance treatment.

Lithium has been shown to be an effective maintenance treatment (Baldessarini and Tohen 1988) along with both the SSRI (Hirschfield and Schatzberg 1994) and the tricyclics and tetracyclics (Prien 1992).

2.5 Resistant Depression

Up to 30% of patients with major depression fail to improve with antidepressant treatment, and at least 60% may fail to achieve full remission (Roose et al. 1986). Nierenberg and White (1990) reviewed the literature and suggested that the term "resistant depression" be applied to patients having had treatment with a tricyclic or newer antidepressant for at least 6 weeks at an adequate dosage (e.g. 300 mg imipramine/day or equivalent) with blood level monitoring to ensure adequate exposure. The options at that stage are to augment treatment with lithium or thyroid hormone or to change to a different class of drug. The most robust evidence from placebo-controlled trials favours lithium augmentation (e.g. Katona et al. 1995) and triiodothyronine (T3) augmentation (Joffe et al. 1993). Surveys of psychiatric treatment preferences for refractory depression in the

United States (Nierenberg 1991) and Canada (Chaimowitz et al. 1991) and the United Kingdom (Shergill and Katona 1997) show that lithium augmentation is one of the most popular options. It is, however, only chosen by about one third of psychiatrists, and many psychiatrists prefer to change the class of antidepressant as a first-line measure. Augmentation with lithium or T3 offers the advantage that benefit may still accrue from the initial course of antidepressant treatment, even where this alone has failed to achieve an adequate response. Interestingly, Joffe (1998) has shown that response (to lithium or T3) is equally likely in non-responders and partial responders.

L-Tryptophan has also been used to augment antidepressants. Difficulties with a contaminant, causing life-threatening immunological reactions, led to widespread withdrawal of this tryptophan precursor, although it is now available again under more tightly controlled conditions. Subsequently changing medication to a different class (particularly an MAOI) would be the next step, again with a view to attaining an adequate dose for an adequate time. The MAOI can in turn be augmented with lithium, as above, or with tricyclic antidepressants, starting with a low dose to minimise adverse effects (Amsterdam and Hornig Rohan 1996).

SSRI are an increasingly frequent choice for initial antidepressant treatment. The main options following failure to respond to 6 weeks of high-dose treatment are to change to a tricyclic or tetracyclic antidepressant and follow the suggestions above or to augment the SSRI. Augmentation strategies are reviewed by Amsterdam and Hornig Rohan (1996), who recommend adding lithium (following the suggestions above) or a tricyclic antidepressant.

The place of the newer antidepressants in refractory depression is not clear at present. There is, however, a suggestion from an uncontrolled study (Nierenberg et al. 1994) that venlafaxine may have some benefit in resistant depression.

3 Bipolar Affective Disorder

3.1 Available Drugs

No new range of drugs for the treatment of mania in BAD has emerged to parallel the range available in depression. The main recent advance has been in the demonstration of efficacy of two anticonvulsants (carbamazepine and valproate) both in the treatment of manic phases and in the prophylaxis of manic and depressive phases. Areas that require attention are the treatment of bipolar depression, rapid-cycling disorder and alternatives to lithium for long-term maintenance (Prien and Rush 1996); the current optimal management of these is described below.

The short-term treatment of mania still requires the additional use of sedative medication in the form of antipsychotics such as haloperidol or of relatively short-acting benzodiazepines such as lorazepam or clonazepam. Other drugs available for second-line use in acute mania include the anticonvulsants clonazepam and lamotrigine, calcium channel blockers such as verapamil, and atypical antipsychotics such as clozapine and risperidone.

3.1.1 Lithium

Lithium salts were used for the treatment of gout some 100 years before the discovery (by Cade in 1949) of their therapeutic effect in the treatment of mania. Subsequent work between 1960 and 1970 suggested that lithium was also useful in the prophylaxis of BAD (Gelenberg et al. 1989). Within BAD, lithium is indicated for the treatment of acute manic episodes, prophylaxis of both manic and depressive episodes and treatment of depressive episodes.

Lithium is completely absorbed from the alimentary tract and achieves peak levels after 90 min with a standard preparation and 4 h for sustained-release preparations. It is not protein bound and is excreted unmetabolised via the kidneys. It cannot cross the blood–brain barrier and has a half-life of around 20 h. Equilibrium is reached after approximately 5 days of treatment. The mode of action of lithium is not entirely clear, but a putative mechanism may be through its action on inhibition of the enzyme inisotol-1-phosphatase, which reduces cellular responses linked to the phosphatidyl inositol second-messenger system. Lithium has also been shown to stimulate both sodium- and magnesium-dependent adenosine triphosphatase in cell membranes and to alter the distribution of calcium and magnesium across excitable membranes, serving to stabilise the cell membranes. Furthermore, both noradrenaline and 5-HT show an increased turnover and metabolism in response to lithium treatment in the short term, but a decrease in the longer term. Lithium also increases activity at the 5-HT uptake sites on platelets.

The commonest adverse effects of lithium are seen on the thyroid, heart and kidneys. Thirst, polyuria and gastrointestinal upset are relatively common. A fine tremor, diabetes insipidus, cardiac arrhythmias, weight gain and hypothyroidism may also be induced. Recommended precautions are assessment of renal function, thyroid status and baseline ECG before commencing treatment and regular monitoring (3-monthly for the first 6–12 months and then

6-monthly) of renal and thyroid function. Lithium impairs coronary sinus function and is contraindicated in sick sinus syndrome. Its use in pregnancy and while breast-feeding is not recommended; there have been reports of congenital malformations following exposure to lithium in the first trimester of pregnancy, although the extent of this has been disputed (Schou 1990). Lithium is excreted in clinically significant amounts in breast milk, and its use in breast-feeding mothers is contraindicated. There are case reports of renal failure associated with the use of lithium, but a review of the evidence by Schou (1997) concluded that there were no consistent findings to support a causal relationship.

Initiation of lithium requires low doses initially with blood level monitoring every 5 days until the therapeutic range is achieved; thereafter, it should be checked every 3 months. There is a well-documented narrow therapeutic range of plasma lithium levels (0.6–1.0 mM/l, blood levels being taken 12 h after the previous dose) within which the balance between efficacy and toxicity is optimal. Lower levels are subtherapeutic, and levels greater than 1.2 mM/l can cause toxicity (Gelenberg et al. 1989; Schou 1997). The common symptoms suggestive of lithium toxicity are coarse tremor, dysarthria and ataxia. Lithium toxicity should be regarded as a medical emergency and requires rehydration and, if severe, even haemodialysis. Antipsychotic drugs enhance the neurotoxic effects of lithium treatment, and their dosage should be carefully monitored. Most non-steroidal anti-inflammatory drugs, diuretics and angiotensin-converting enzyme inhibitors can increase lithium concentrations and, if used concurrently, mandate particularly close lithium level monitoring.

3.1.2 Carbamazepine

Carbamazepine is similar in structure to the tricyclic antidepressant imipramine. Though used primarily in neurological practice (especially temporal lobe epilepsy and trigeminal neuralgia), it is being used increasingly in BAD. It is indicated in prophylaxis and may be particularly useful in preventing relapse in rapid-cycling BAD. It has also been found to be effective in the treatment of acute mania (Post et al. 1987).

The mode of action of carbamazepine responsible for its antimanic effects is not fully understood. The efficacy of carbamazepine (and valproate, see below) is consistent with the "kindling" theory of BAD. Kindling refers to the electrophysiological observation that repeated subthreshold stimulations of a neuron may result in an action potential. The kindling model of BAD relies on the phenomenon that episodes tend to increase in frequency and severity. Within this model, BAD is seen as a form of limbic epilepsy, modifiable by anticonvulsants. There are, however, no consistent electroencephalographic (EEG) abnormalities in patients suffering from BAD responding to carbamazepine. Carbamazepine may act as an agonist on peripheral benzodiazepine receptors located on the γ-aminobutyric acid (GABA)-A receptor complex. This may enhance the inhibitory effects of GABA in many brain areas. Peripheral benzodiazepine receptors also act to regulate the calcium channels. This may be relevant in the light of the beneficial effects in BAD of calcium channel blocking drugs such as verapamil.

There is erratic absorption of carbamazepine from the gastrointestinal system, with increased absorption if taken with food. Peak blood levels are reached in 4–8 h, and the half-life varies from 12 to 17 h, depending on the induction of hepatic enzymes. Steady state is reached in 2–4 days. Carbamazepine is metabolised in the liver and excreted via the kidneys.

Carbamazepine commonly induces leucopenia (10% in the first 2 months), although the white cell count usually remains above 3000. Very occasionally, it can also cause agranulocytosis and aplastic anaemia. Patients need to be warned to beware of any symptoms of fever, bruising or easy bleeding. A pruritic rash can occur in 10%–15% of patients within a few weeks of commencing treatment. This can occasionally be the precursor of more severe dermatological conditions.

The most common side-effects of carbamazepine are nausea and gastric disturbance; these can be minimised by gradual upward dose titration. Adverse effects on the central nervous system (such as dizziness, ataxia and sedation) can be minimised by the same strategy. Carbamazepine is not recommended for use in pregnant or breast-feeding women. It is an inducer of catabolic hepatic enzymes and may thus affect the metabolism of multiple other drugs.

The starting dose of carbamazepine is usually 200 mg b.d. It should be taken with meals, and the dose should gradually be increased to between 600 and 1000 mg a day. The dose should be increased at approximately 200 mg every 2–4 days if possible to minimise adverse effects. It can be used alone or with an antipsychotic drug to treat manic episodes. A 3-week trial at therapeutic doses (trough blood levels, 8–12 mg/l) should be enough to assess its effect in the treatment of acute mania. Failure to respond should be followed by addition of lithium or valproate rather than by immediate withdrawal of the carbamazepine.

3.1.3 Sodium Valproate

Sodium valproate was used originally as an antiepileptic in France and is indicated for the treatment of BAD where lithium and carbamazepine are contraindicated or have failed. It is used in acute mania

(Bowden et al. 1994), especially mixed affective states, as well as in the prophylaxis of BAD. Sodium valproate is thought to have its effects by decreasing catabolism of the inhibitory peptide neurotransmitter GABA.

Sodium valproate is converted into valproic acid in the stomach and completely absorbed. It has a half-life of 8–16 h and is metabolised in the liver. Adverse effects are common in the gastrointestinal system; these include nausea (25%), vomiting (5%) and diarrhoea. It can also cause sedation and more rarely ataxia and tremor. Weight gain is common, and alopecia is reported in 5%–10% of patients. Hepatic enzymes can be elevated but rapidly return to normal levels after discontinuation of the drug. Rare cases of severe hepatotoxicity have been reported. Sodium valproate should not be used in pregnant or breast-feeding women. It can interact with other anticonvulsants and will raise levels of tricyclics and warfarin.

Sodium valproate should be given at a starting dose of 200 mg t.d.s.; this should be gradually increased to achieve trough plasma levels between 50 and 100 mg/l. Baseline and regular checks should be made of liver enzymes and the full blood count.

3.1.4 Clonazepam

The benzodiazepine clonazepam is used in the "second-line" treatment of BAD where lithium and/or other anticonvulsants have failed. It is used successfully both in the treatment of acute manic episodes (Chouinard et al. 1983) and as adjunctive therapy, with lithium, in the prophylaxis of BAD. In common with carbamazepine, it acts on GABA-A receptors to facilitate GABA-induced reductions in neuronal firing.

Clonazepam is rapidly absorbed orally, has no major metabolites and has a half-life of 34 h. It should be prescribed once daily, usually at night, with an initial dose of 0.5–1 mg, increasing gradually to 4–8 mg/day over several weeks. As it can cause abnormalities of liver enzymes, routine liver function tests are required. Common adverse effects are drowsiness, fatigue and dizziness; rare effects are irritability and aggression.

3.1.5 Other Drugs

Verapamil

Verapamil is a calcium channel blocker and is mainly used for the treatment of hypertension, angina and certain cardiac arrhythmias. However, it is also useful in the "second-line" treatment of acute mania and for BAD prophylaxis, where lithium and anticonvulsants have been ineffective, and in the treatment of mania (Garza-Trevino et al. 1992). It is thought to act by inhibiting calcium-dependent intracellular protein kinases.

Verapamil is well absorbed, is extensively first-pass metabolised by the liver and has a half life of 5–12 h. Its main adverse effects (as might be expected) are bradycardia, hypotension and atrioventricular heart block. Gastrointestinal complaints are also common. Verapamil can enhance the neurotoxicity of lithium and carbamazepine. It can be fatal if co-prescribed with beta blockers. It should not be used in pregnant or breast-feeding women. Verapamil should be initiated at 40 mg t.d.s. and increased to 120 mg t.d.s. over a period of weeks. Concomitant monitoring of blood pressure, pulse and ECG is required.

Clozapine

Clozapine is an example of an atypical antipsychotic drug which has been used in the treatment of patients with severe bipolar illnesses. In this context, it should be regarded as a "third-line" drug in patients with acute mania refractory to lithium, antipsychotics and second-line drugs (Calabrese et al. 1996). It has been suggested to have mood-stabilising properties in addition to its antimanic effect (reviewed by McElroy et al. 1996). Clozapine has actions at various receptors, and it is unclear which is essential for its action. It has low D_2 receptor antagonism and is a more potent antagonist of D_1, D_4, 5-HT_2 and α_1-adrenergic receptors. Although clozapine is free from extrapyramidal side-effects, it is associated with seizures and agranulocytosis and is currently licensed only for the treatment of resistant schizophrenia.

Lamotrigine

Lamotrigine is an anticonvulsant drug licensed for monotherapy of certain types of epilepsy. It acts by blocking the release of the excitatory neurotransmitters glutamate and aspartate through blocking sodium channels and stabilising presynaptic neuronal membranes. There are case reports and open studies suggesting that it has a beneficial effect in the treatment of BAD, usually in treatment-refractory patients and as an adjunct to more conventional pharmacological approaches (Sporn and Sachs 1997). The adverse effects associated with its usage are headache, nausea, vomiting and diplopia. Confusional states and rashes have also been reported. A gradual increase in starting dose is recommended to reduce adverse effects.

Gabapentin

Gabapentin is an anti-epileptic drug thought to potentiate intracellular GABA, with some evidence of effectiveness in the treatment of bipolar disorder (Schaffer and Schaffer 1997).

Risperidone

Risperidone is an atypical antipsychotic which has shown some efficacy in the treatment of both acute mania and psychotic depression (McElroy et al. 1996).

It has not been as extensively studied as clozapine, and there have been reports of risperidone-associated precipitation of manic episodes, especially in the absence of mood-stabilising medication. Its mode of action is predominantly via 5-HT_2 and D_2 receptor blockade.

Olanzapine

Olanzapine, another atypical antipsychotic agent, has been shown to result in a significant improvement in the symptoms of acute mania when compared to placebo (Sanger et al. 1998).

3.2 Choice of Drug

3.2.1 Acute Treatment

Manic Episodes

Antipsychotics are commonly used in addition to lithium or another antimanic agent to gain rapid control over symptoms in mania. In studies comparing the treatment of mania with lithium and with antipsychotics (reviewed by Gelenberg and Hopkins 1996), lithium appeared better at stabilising mood and ideation, while antipsychotics were more effective in reducing time to clinical response and particularly in reducing hyperactivity acutely. Approximately 80% of patients with acute mania will respond to treatment with lithium, although this may take up to 2 weeks to take effect (Bowden 1996). In the initial treatment of acute mania, the treatment of choice is lithium, which may be given alone or adjunctively at this stage; if it is used alone, management can be problematic in the first few days because, unlike antipsychotics or benzodiazepines, lithium does not have sedative-related effects on aggressive or disinhibited behaviour. Adjunctive medication can be added according to clinical need: high- or medium-potency antipsychotics to control psychosis and induce sleep and sedation or benzodiazepines for sedation in the absence of psychosis.

In conclusion, lithium or valproate are the treatments of choice in mania patients, except for highly aroused patients in whom more rapid control is necessary. In these patients, the choice is to add a neuroleptic, usually with a higher potency such as haloperidol in low doses (5–15 mg/day), or to add a benzodiazepine such as clonazepam or lorazepam to achieve an acute sedative effect. Antipsychotics can be used on their own if patients are intolerant or unwilling to comply with lithium treatment or if there is a chance that they may be pregnant. Valproate is licensed for use in the treatment of acute mania (Bowden et al. 1994) in the United States. Other individual drugs used to treat mania, with evidence of their efficacy, include carbamazepine (Post et al. 1987), clonazepam (Chouinard et al. 1983) and verapamil (Garza-Trevino et al. 1992).

Depressive Episodes Within Bipolar Affective Disorder

Lithium can be used in the acute treatment of the depressive phase of a bipolar illness, but antidepressant drugs (reviewed earlier in this chapter) are increasingly used, additionally, for this indication. There is a danger that antidepressants will precipitate rapid cycling and mania (Wehr and Goodwin 1987). The evidence for this particularly implicates tricyclics, especially in the presence of hypothyroidism (Bauer and Whybrow 1990). Overall, there is insufficient data on the newer antidepressants to assess their risk in this area. However, a meta-analysis comparing SSRI and tricyclics suggested that SSRI may be both effective and less likely to induce a treatment-emergent mania (Peet 1994). However, the length of follow-up in the studies analysed was short and most were treatment efficacy studies with no standardised criteria to assess mania. There is preliminary data that bupropion may also be effective in treating depressive episodes within BAD (Wright et al. 1985).

Overall, mood stabilisers are preferable to treatment with antidepressants in the treatment of mild depressive episodes in BAD (Expert Consensus Guidelines 1996). Non-psychotic but moderate to severe depression will require treatment with an antidepressant in addition to the mood stabiliser; a psychotic depression may require further addition of an antipsychotic drug (Expert Consensus Guidelines 1996). Sachs (1996) reviewed the treatment of treatment-resistant bipolar depression, and a modified version of his suggested treatment algorithm is shown in Fig. 1.

3.2.2 Maintenance Treatment

The decision to commence prophylactic or maintenance treatment is based on the assessment of several factors: the severity of previous episodes of illness, the risk of adverse effects from treatment, the likelihood of a recurrence in the near future and patients' willingness to commit themselves to treatment. Once the acute manic episode has resolved, maintenance treatment with lithium has been demonstrated to decrease the relapse rate of BAD from approximately 80% (with placebo) to 35% (Prien 1992). Best current clinical practice is to introduce lithium maintenance after the *second* episode of illness in BAD, which may be thought of as the first relapse (Expert Consensus Guidelines 1996). There is evidence that maintenance lithium should be continued for at least 2 years from relapse, as subsequent relapses may occur when lithium is discontinued and treatment with lithium may not be as effective in subsequent episodes. It is suggested that, if

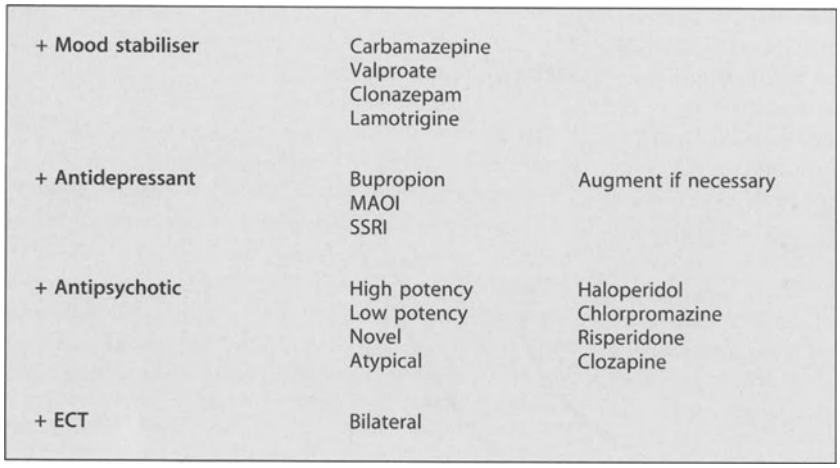


Fig. 1. Lithium treatment algorithm for bipolar depression. The options on the left-hand side should be initiated to control acute symptoms and the next choice on the right-hand side added or used as a substitute if cycling is unchanged for three cycles. MAOI, monoamine oxidase inhibitor; SSRI, serotonin-specific re-uptake inhibitor; ECT, electroconvulsive therapy

a further relapse occurs during lithium maintenance, the lithium should not be discontinued but augmented with carbamazepine or valproate.

Antipsychotics are not a treatment of choice for prophylaxis of BAD despite their common use as an adjunct to lithium (Gelenberg and Hopkins 1996). Antipsychotic treatment is associated with multiple side-effects with chronic use, including tardive dyskinesia, although there is little evidence of the long-term effects of using the newer antipsychotics. Their more favourable side-effect profile may lead to their increased usage in this area. It is generally advised that antipsychotics be used for as short a time as possible before mood-stabilising drugs are effective. They should be reserved for patients intolerant to or non-compliant with mood stabilisers (McElroy et al. 1996).

Lithium is the drug of choice in the prophylaxis of BAD in most patients. Maintenance treatment with lithium is most effective for patients with a history of previous good response, good inter-episode functioning and a history of bipolar illness in the family. General consensus suggests that treatment should be maintained for at least 2 years, and ideally for 5 years, before an attempt to discontinue treatment should be considered. Rapid-cycling illness, where there have been more than four episodes in the year, is less responsive to lithium treatment, and in such patients carbamazepine is probably more effective (Post et al. 1987). However, the results of a meta-analysis of selected studies highlighted the paucity of high-quality controlled studies and contended that there was insufficient data to support this view (Dardennes et al. 1995). A more recent review has supported the continued use of carbamazepine (Post et al. 1997). Similar arguments can be marshalled for the use of sodium valproate in the prophylactic treatment of BAD. Here, too, there are a number of clinical studies suggesting efficacy, but also a lack of good-quality controlled studies (McElroy et al. 1992). Valproate has been shown to benefit some patients with inadequate response to lithium and carbamazepine when used alone, as an adjunct to lithium or with both lithium and carbamazepine as triple therapy (Denicoff et al. 1997). These authors report the following response rates from their prospective randomised study: 33% with lithium, 43% with carbamazepine, 50% with carbamazepine and lithium, 60% with valproate and lithium and 62% with all three.

4 Other Mood Disorders

4.1 Atypical Depression

Atypical depression has been included in DSM-IV as a specifier of major depressive episodes and dysthymia; in ICD-10 it is classed under "other depressive episodes". Lam and Stewart (1996) concluded their review of atypical depression by supporting its diagnostic validity in terms both of clinical description and differential treatment response. In particular, atypical depression responds better to MAOI than to tricyclic antidepressants. Moclobemide also appears to be effective in the treatment of atypical depression and may be superior to fluoxetine in this context (Lonnqvist et al. 1994).

4.2 Dysthymia and Double Depression

The clinical significance of chronic mild depression (dysthymia) is well recognised, but until recently it was thought to be a form of personality type. This would

suggest that psychotherapy, rather than pharmacotherapy, was the treatment of choice. Howland (1990), however, reviewed the literature on the pharmacotherapy of dysthymia and concluded not only that was pharmacotherapy effective, but also that MAOI may be superior to tricyclic antidepressants. Subsequent studies, reviewed by the World Psychiatric Association Dysthymia Working Group (1995), have confirmed the effectiveness of a wide range of different antidepressants in dysthymia, including fluoxetine, imipramine, phenelzine and moclobemide. Long-term maintenance treatment is likely to be required. Sansone and Sansone (1996) recommend the use of SSRI in prophylaxis and suggest that such treatment should be continued for at least 2–3 years.

In patients with "double depression" (where dysthymia is complicated by discrete episodes of major depression), treatment of the major depressive episode also results in improvement of the dysthymia (Akiskal 1994). Continuation treatment should be given for at least 2 years. Where treatment fails, options for refractory depression (reviewed above) should be considered.

4.3 Cyclothymia

Cyclothymia, characterised by episodes of mild depression and hypomania, has a fluctuating course from onset in juvenile or early adult years and confers susceptibility to rapid cycling upon exposure to tricyclic antidepressants (Howland and Thase 1993). The treatment of choice is lithium or other mood stabilisers such as carbamazepine or valproate. Akiskal (1994) reviewed the treatment of depressive episode in cyclothymia, and suggested that such patients should be treated with bupropion, MAOI and low-dose SSRI in conjunction with mood stabilisers.

4.4 Recurrent Brief Depressive Disorder

Recurrent brief depressive disorder (RBDD) is characterised by repeated, relatively brief (less than 2 weeks) episodes of depressive symptoms, otherwise similar in range and severity to those found in major depression. There are limited data on the specific treatment of this disorder, but adoption of the general guidelines for antidepressant treatment is suggested. The value of maintenance treatment with mood stabilisers in RBDD is not yet known.

4.5 Seasonal Affective Disorder

Seasonal affective disorder is characterised by dysphoria and depressive symptoms with onset in winter. The depressive symptoms are classically "atypical" with prominent hypersomnia, hyperphagia and increased lethargy. Phototherapy has been shown to be beneficial in this condition. Pharmacological treatments with antidepressant drugs, such as tricyclic and tetracyclic drugs, MAOI, SSRI and lithium, at standard doses have been used to good effect (Oren and Rosenthal 1992).

5 References

Adler L, Hajak G, Lehmann K et al (1997) On the problems of switching from intravenous to oral administration in drug treatment of endogenous depression Pharmacopsychiatry 30: 62–69

Akiskal HS (1994) Dysthymic and cyclothymic depressions: therapeutic considerations. J Clin Psychiatry 55[Suppl 4]: 46–52

*Amsterdam JD, Hornig Rohan M (1996) Treatment algorithms in treatment-resistant depression Psychiatr Clin North Am 19: 371–386

Asberg M (1976) Treatment of depression with tricyclic drugs – pharmacokinetic and pharmacodynamic aspects. Pharmacopsychiatry Neuropsychopharmacol 9: 18–26

Baldessarini RJ, Tohen M (1988) Is there a long-term protective effect of mood-altering agents in unipolar depressive disorder? Psychopharmacol Ser 5: 130–139

Barondes SH (1994) Thinking about Prozac. Science 263(5150): 1102–1103

Bauer M, Whybrow P (1990) Rapid cycling bipolar affective disorder: treatment of refractory rapid cycling with high dose levothyroxine: a preliminary study. Arch Gen Psychiatry 47: 435–440

Bloch RG, Dooneief AS, Buchberg AS et al (1954) The clinical effects of isoniazid and iproniazid in the treatment of pulmonary tuberculosis. Ann Intern Med 40: 881–900

Bouchard JM, Strub N, Nil R (1997) Citalopram and viloxazine in the treatment of depression by slow drop infusion. A double-blind comparative trial. J Affect Disord 46(1): 51–58

*Bowden CL (1996) Dosing strategies and time course of response to antimanic drugs. J Clin Psychiatry 57[Suppl 13]: 4–12

Bowden CL, Brugger AM, Swann AC et al (1994) Efficacy of divalproex versus lithium and placebo in the treatment of mania. JAMA 271: 918–924

Burke MJ, Silkey B, Preskorn SH (1994) Pharmacoeconomic considerations when evaluating treatment options for major depressive disorder. J Clin Psychiatry 55[Suppl A]: 42–52

Calabrese JR, Kimmel SE, Woyshville MJ et al (1996) Clozapine for treatment-refractory mania Am J Psychiatry 153(6): 759–764

Chaimowitz GA, Links PS, Padgett RW, Carr AC (1991) Treatment-resistant depression: a survey of practice habits of Canadian psychiatrists. Can J Psychiatry 36: 353–356

Chouinard G, Young SN, Annable L (1983) Antimanic effects of clonazepam. Biol Psychiatry 18: 451–466

Cohen LJ (1997) Rational drug use in the treatment of depression. Pharmacotherapy 17: 45–61

Dardennes R, Even C, Bange F et al (1995) Comparison of carbamazepine and lithium in the prophylaxis of bipolar disorders. A meta-analysis. Br J Psychiatry 166: 378–381

Denicoff KD, Earlian E, Smith-Jackson RN et al (1997) Valproate prophylaxis in a prospective clinical trial of refractory bipolar disorder. Am J Psychiatry 154: 1456–1458

Doogan DP, Caillard V (1992) Sertraline in the prevention of depression. Br J Psychiatry 160: 217–222

*Expert Consensus Guidelines (1996) Guidelines for the treatment of bipolar affective disorder. J Clin Psychiatry 57[Suppl 12A]: 7–42

Garza-Trevino ES, Overall JE, Hollister LE (1992) Verapamil versus lithium in acute mania. Am J Psychiatry 149: 121–127

Gastpar M, Gilsdorf U, Baumann P (1986) Comparison of oral and intravenous treatment of depressive states: preliminary results of a WHO collaborative study. Clin Neuropharmacol 9: 434–436

*Gelenberg AJ, Hopkins HS (1996) Antipsychotics in bipolar disorder. J Clin Psychiatry 57[Suppl 9]: 49–52

Gelenberg AJ, Carroll JA, Baudhuin MG et al (1989) The meaning of serum lithium levels in maintenance therapy of mood disorders: a review of the literature. J Clin Psychiatry 50 [Suppl 12]: 17–22

Guelfi JD, White C, Hackett D et al (1995) Effectiveness of venlafaxine in patients hospitalised for major depression and melancholia. J Clin Psychiatry 56: 450–458

Hirschfeld RM, Schatzberg AF (1994) Long-term management of depression. Am J Med 97: 33s–38s

Hotopf M, Lewis G, Normand C et al (1996) Are SSRIs a cost-effective alternative to tricyclics? Br J Psychiatry 168: 404–409

Howland RH (1990) Pharmacotherapy of dysthymia: a review. J Clin Psychopharmacol 11: 83–92

Howland RH, Thase ME (1993) A comprehensive review of cyclothymic disorder. J Nerv Ment Dis 181: 485–493

Joffe (1998) The use of thyroid supplements to augment antidepressant medication. J Clin Psychiatry 59[Suppl 5]: 26–29

Joffe RT, Singer W, Levitt AJ et al (1993) A placebo-controlled comparison of lithium and triiodothyronine augmentation of tricyclic antidepressants in unipolar refractory depression. Arch Gen Psychiatry 50(5): 387–393

Jonsson B, Bebbington PE (1993) What price depression? The cost of depression and the cost-effectiveness of pharmacological treatment Br J Psychiatry 164: 665–673

Katona CLE, Abou-Saleh MT, Harrison DA et al (1995) Placebo-controlled trial of lithium augmentation of fluoxetine and lofepramine. Br J Psychiatry 166(1): 80–86

Kragh-Sorensen P, Asberg M, Eggert-Hansen C (1973) Plasma-nortryptiline levels in endogenous depression. Lancet 1: 113–115

Kuhn R (1958) The treatment of depressive states with G22355 (imipramine hydrochloride). Am J Psychiatry 115: 459–464

*Kupfer DJ, Frank E, Perel JM et al (1992) Five year outcome for maintenance therapies in recurrent depression Arch Gen Psychiatry 49: 769–773

Lam RW, Stewart JN (1996) The validity of atypical depression in DSM-IV. Compr Psychiatry 37: 375–383

Laux G (1993) Antidepressive Infusionstherapie. In: Riederer P, Law G (eds) Neuropsychopharmaka, vol 3. Springer, Berlin Heidelberg New York, pp 257–268

Laux G, Konig W, Lesch K et al (1989) Intravenose versus orale Behandlung endogen depressiver Patienten mit Doxepin – eine Doppelblindstudie mit Plasmaspiegelbestimmungen. Wien Med Wochenschr 22: 525–529

Lonnqvist J, Sivho S, Syvalahti E et al (1994) Moclobemide and flouxetine in atypical depression: a double blind trial. J Affect Disord 32(3): 169–177

*McElroy SL, Keck PE, Pope HG et al (1992) Valproate in the treatment of bipolar disorder: literature review and clinical guidelines. J Clin Psychopharmacol 12[Suppl 1]: 425–525

McElroy SL, Keck PE, Strakowski SM (1996) Mania, psychosis and antipsychotics. J Clin Psychiatry 57[Suppl 3]: 14–26

Moller HJ, Volz HP (1996) Drug treatment of depression in the 1990s. An overview of achievements and future possibilities. Drugs 52: 625–638

Montgomery SA, Dunbar G (1993) Paroxetine is better than placebo in relapse prevention and the prophylaxis of recurrent depression. Int Clin Psychopharmacol 8: 189–195

*Nemeroff CB, DeVane CL, Pollock BG (1996) Newer antidepressants and the cytochrome P450 system. Am J Psychiatry 153: 311–320

Nierenberg AA (1991) Treatment choice after one antidepressant fails: a survey of Northeastern psychiatrists. J Clin Psychiatry 52: 383–385

Nierenberg AA (1994) The treatment of severe depression: is there an efficacy gap between SSRI and TCA antidepressant generations? J Clin Psychiatry 55[Suppl A]: 55–59

Nierenberg AA, White K (1990) What next? A review of pharmacologic strategies for treatment resistant depression. Psychopharmacol Bull 26: 429–460

Nierenberg AA, Feighner JP, Rudolph R et al (1994) Venlafaxine for treatment resistant unipolar depression. J Clin Psychopharmacol 14: 419–423

Oren DA, Rosenthal NE (1992) Seasonal affective disorders. In: Paykel ES (ed) Handbook of affective disorders. Churchill Livingstone, Edinburgh

*Peet M (1994) Induction of mania with selective serotonin reuptake inhibitors and tricyclic antidepressants. Br J Psychiatry 164: 549–550

Perry PJ (1996) Pharmacotherapy for major depression with melancholic features: relative efficacy of tricyclic versus selective serotonin reuptake inhibitor antidepressants. J Affect Disord 39: 1–6

Pollock B, Perel JM, Nathan R et al (1989) Acute antidepressant effect following pulse loading with intravenous and oral clomipramine. Arch Gen Psychiatry 46: 29–35

Post RM, Uhde TW, Roy-Byrne PP et al (1987) Correlates of antimanic response to carbamazepine. Psychiatry Res 21: 71–83

Post RM, Denicoff KD, Frye MA et al (1997) Re-evaluating carbamazepine prophylaxis in bipolar disorder. Br J Psychiatry 170: 202–204

*Prien RF (1992) Maintenance treatment. In: Paykel ES (ed) Handbook of affective disorders. Churchill Livingstone, Edinburgh, pp 419–435

Prien RF, Kupfer DJ (1986) Continuation drug therapy for major depressive disorder: how long should it be maintained? Am J Psychiatry 143: 18–23

Prien RF, Rush AJ (1996) National Institute of Mental Health Workshop Report on the treatment of bipolar disorder. Biol Psychiatry 40: 215–220

Priest RG (1996) Cost-effectiveness of venlafaxine for the treatment of major depression in hospitalized patients. Clin Ther 18: 347–358

Quitkin FM (1985) The importance of dosage in prescribing antidepressants. Br J Psychiatry 147: 593–597

*Richelson E (1996) Synaptic effects of antidepressants. J Clin Psychopharmacol 16: 1s–7s

Roose SP, Glassman AH, Walsh BT et al (1986) Tricyclic nonresponders: phenomenology and treatment. Am J Psychiatry 143: 345–348

*Sachs GS (1996) Treatment-resistant bipolar depression. Psychiatr Clin North Am 19: 215–236

Sanger T, Tohen M, Tollefson G et al (1998) Olanzapine vs placebo in the treatment of acute mania. Schizophr Res 29(1, 2): 152

Sansone RA, Sansone LA (1996) Dysthymic disorder: the depression that never quits. Postgrad Med 99: 233–234

Schaffer C, Schaffer L (1997) Gabapentin in the treatment of bipolar disorder. Am J Psychiatry 154: 291–292

Schou M (1990) Lithium treatment during pregnancy, delivery and lactation: an update. J Clin Psychiatry 51: 410–413

*Schou M (1997) Forty years of lithium treatment. Arch Gen Psychiatry 54: 9–13

Sclar DA, Robison LM, Skaer TL et al (1994) Antidepressant pharmacotherapy: economic outcomes in a health maintenance organization. Clin Ther 16: 715–730

Sclar DA, Robison LM, Skaer TL et al (1995) Antidepressant pharmacotherapy: economic evaluation of fluoxetine, paroxetine and sertraline in a health maintenance organization. J Int Med Res 23: 395–412

Shergill SS, Katona CLE (1997) Pharmacological choices after one antidepressant fails. J Affect Disord 43: 19–25

Sporn J, Sachs G (1997) The anticonvulsant lamotrigine in treatment-resistant manic-depressive illness. J Clin Psychopharmacol 17: 185–189

Sternbach H (1991) The serotonin syndrome. Am J Psychiatry 148: 705–713

*Stoddart G, Drummond M (1984) How to read clinical journals. VII. To understand an economic evaluation. Can Med Assoc J 130: 1428–1433

Sunderland T, Cohen RM, Molchan SE et al (1994) High-dose selegiline in treatment-resistant older depressive patients. Arch Gen Psychiatry 51: 607–615

Task Force on the Use of Laboratory Tests in Psychiatry (1985) Tricyclic antidepressants – blood level measurements and clinical outcome: an APA Task Force report. Am J Psychiatry 142: 155–62

Wehr TA, Goodwin FK (1987) Can antidepressants cause mania and worsen the course of affective illness? Am J Psychiatry 144: 1403–1411

WPA Dysthymia Working Group (1995) Dysthymia in clinical practice. Br J Psychiatry 166: 174–183

Wright G, Galloway L, Kim J et al (1985) Bupropion in the long-term treatment of cyclic mood disorders: mood stabilising effects. J Clin Psychiatry 46: 22–25

CHAPTER

24

U. Voderholzer, M. Berger

Other Methods of Somatic Therapy for Depression

Translator: E. Taub

1 Introduction

Alongside pharmacotherapy with antidepressants, a number of other somatic therapies, including sleep deprivation, electroconvulsive therapy, phototherapy, and, most recently, repetitive transcranial magnetic stimulation can be used to treat depressive illnesses. These therapies are usually given in combination with antidepressants; the purposes of such treatment include giving the patient a degree of relief for several weeks before the effect of antidepressant medication sets in, potentiating the effect of medication once it is established, or, in case of medically refractory depression, providing a different means of brightening the patient's mood.

2 Sleep Deprivation Therapy

The systematic introduction of sleep deprivation for the treatment of depressive illnesses was based not on theoretical considerations, but on case reports from the 1960s. In 1969, Schulte, a psychiatrist in Tübingen (Germany), reported the following: "There are, indeed, melancholic patients who, having deliberately deprived themselves of nighttime sleep, are fresher and more productive the next morning than if they had slept undisturbed" (Schulte 1969, p. 415). This striking improvement of depressive manifestations after a night of wakefulness was unexpected, as it belied the general assumption among physicians, as well as laymen, that sleep has a restorative function, whether in normal individuals or in depressive patients.

2.1 Total Sleep Deprivation for One Night

Schulte's observations were confirmed in the initial study by Pflug and Tölle (1971). Total sleep deprivation for one night has an impressive antidepressant effect in the majority of patients so treated. An improvement of mood after sleep deprivation is often noted by patients as early as the second half of the night spent awake, but it may, in other cases, set in only in the course of the following day.

The effects of sleep deprivation therapy have since been examined in numerous studies. Wu and Bunney (1990), in a meta-analysis of some 60 studies involving a total of more than 1700 patients, concluded that total sleep deprivation for one night leads to a major improvement of mood in the following day in 59% of depressive patients. Patients with the melancholic subtype of depression had a response rate of 75%, while non-melancholic depressed patients had a response rate of only 48%. Wu and Bunney's meta-analysis also revealed, however, that 83% of patients treated with sleep deprivation alone fell back into a depressed mood after sleeping through the following night, in contrast to only 59% of patients who were simultaneously treated with medication.

It has long been known that sleep deprivation may induce a hypomanic mood or a manic episode in patients suffering from bipolar disorders. Nonetheless, in a recently published study involving more than 200 such patients, a transition from depression to hypomania or mania occurred in only 5% after treatment with sleep deprivation (Colombo et al. 1999).

2.1.1 Predictors of Efficacy

Factors predicting a positive response to sleep deprivation include the following:

- "Melancholic" subtype
- Bipolar disorder
- Hyperarousal before sleep deprivation
- Daily fluctuation with morning low
- High variability of mood changes
- Elevated metabolism in the limbic system

Other factors that have no effect on the sleep deprivation response include the following:

- Age
- Sex
- Number of depressive episodes
- Duration of episodes
- Number of hospitalizations
- Severity of depression
- Previous treatments
- Patient expectations

According to the meta-analysis by Wu and Bunney (1990), as already mentioned, 75% of patients with melancholic or endogenous depression respond to sleep deprivation, in contrast to only 48% of depressive patients not meeting these criteria (patients with "non-endogenous" depression). Earlier studies detected no difference in the efficacy of sleep deprivation between patients with bipolar and those with unipolar depression (Elsenga and Van den Hoofdakker 1987), but a more favorable response in bipolar patients has been reported in some more recent studies (Szuba et al. 1991; Barbini et al. 1998). The response rate seems not to depend on whether the patient is suffering from the first episode of depression or from one of many in a recurrent depressive disorder.

Similarly, the patient's age and sex, the duration and severity of the present episode, and the nature of previous treatment have no predictive value with respect to the effect of sleep deprivation therapy (Kuhs and Tölle 1991). Moreover, the effect of a previous attempt using sleep deprivation therapy in the individual patient cannot be used to predict the effect of a new trial (Gordijn et al. 1995): one quarter of all sleep deprivation treatments are successful even if a previous attempt failed to brighten the patient's mood. Only 9% of patients fail to respond to each of three separate attempts using sleep deprivation therapy.

The presence of typical daily fluctuations, with a morning low and a brightening of mood in the evening, has a high positive predictive value. On the other hand, patients with daily fluctuations of reversed polarity (mood low in the evening) and patients without daily fluctuations do not respond as well to sleep deprivation (Reinink et al. 1990; Haug 1992). Carefully performed studies have revealed that it is not the type of fluctuation immediately before sleep deprivation, but rather the variability of mood fluctuations over the preceding days, both over the course of each day and from day to day, that is a good predictor for the efficacy of sleep deprivation (Gordijn et al. 1995, 1998).

Studies of possible neurobiological predictors have mostly yielded conflicting results (for a review, see Kasper and Möller 1996). The response to sleep deprivation was found to be positively associated with "hyperarousal" and with a high level of vigilance (Bouhuys et al. 1989, 1995). Studies using single photon emission computed tomography (SPECT) and positron emission tomography (PET) have shown that patients with an elevated metabolic rate in the limbic system respond more favorably both to sleep deprivation and to pharmacotherapy given immediately thereafter (Ebert et al. 1991; Wu et al. 1992).

2.1.2 Influence of Medication on the Effect of Sleep Deprivation

Because its effect is usually no more than transient, sleep deprivation is not an adequate therapy when used alone; thus the question of its interaction with antidepressants is raised. It seems that the latter do not improve the rate of response to sleep deprivation, but they do reduce the risk of relapse after the following night of sleep. As mentioned above, the meta-analysis performed by Wu and Bunney (1990), involving more than 1700 patients treated with sleep deprivation, revealed that 83% of patients who had received no medication, but only 59% of patients treated with antidepressants, fell back into a depressive mood after a night of sleep. These data disprove the widespread assumption that successful sleep deprivation therapy is regularly followed by a relapse into depression. Simultaneous lithium therapy also reduces the risk of relapse after successful sleep deprivation (Grube and Hartwich 1990; Szuba et al. 1994; Benedetti et al. 1999).

Similar results have been reported for the combination of sleep deprivation and phototherapy. Although the application of bright light during sleep withdrawal does not potentiate its effect (Wehr et al. 1985; Van den Burg et al. 1990), phototherapy after sleep deprivation can reduce the probability of relapse (Neumeister et al. 1996).

A question of major clinical interest is whether repeated sleep deprivation therapy in combination with the administration of antidepressants is superior to the use of antidepressants alone over a treatment period of 4 weeks. A number of open, uncontrolled studies in patients refractory to drug treatment have yielded evidence that sleep depression may improve or accelerate the effect of antidepressants (Leibenluft and Wehr 1992; Van den Hoofdakker et al. 1994; Benedetti et al. 1997). In a controlled study, Kuhs and colleagues (1996) treated depressive patients, in randomized fashion, with amitryptiline either alone or in combination with six partial sleep deprivations over the course of 4 weeks. The combination of sleep deprivation with an antidepressant led to a significantly greater reduction of Hamilton depression scores than did treatment with an antidepressant alone.

2.2 Partial Sleep Deprivation in the Second Half of the Night

Sleep deprivation limited to the second half of the night is an alternative to total sleep deprivation that has a comparable antidepressant effect (Schilgen and Tölle 1980). This form of treatment involves waking up the patient, after several hours of sleep, at 1 or 2 A.M. and keeping him or her awake for the rest of the night and the following day. Sleep deprivation therapy of this type is suitable for patients who consider themselves incapable of undergoing total sleep deprivation or are used to going to bed early. The procedure can also be repeated several nights in a row, possibly with so-called recovery nights in between, to prevent a relapse into depression. In contrast, sleep deprivation in the first half of the night seems to be much less effective (Goetze and Tölle 1981). The superiority of partial sleep deprivation in the second half of the night as compared to the first half of the night was demonstrated in a crossover study by Sack et al. (1988).

General recommendations for the practical application of sleep deprivation in depression are as follows:

1. Sleep withdrawal (total or partial in the second half of the night, starting at 1 to 2 A.M.) should generally be used as *adjuvant* therapy:

a) In combination with the initial administration of antidepressant medication to counteract depression during the latency period of drug effectiveness.
b) In the case of resistance to pharmacological antidepressant treatment.
c) As a differential diagnostic instrument to distinguish between senile depression with pseudodementia and incipient dementing illnesses. Pseudodemented depressive patients often experience a marked, but transient, improvement after sleep withdrawal. Such a finding makes the diagnosis of an incipient dementing illness very unlikely, even if the affected individuals themselves complain of a loss of cognitive ability.

2. After successful sleep deprivation, either a series of sleep deprivation treatments or a treatment with shifting of the phases of sleep is recommended, depending on the preference of the patient, to maintain the therapeutic effect.

2.3 Theories on the Mechanism of Action of Total Sleep Deprivation

First of all, it must be noted that the effect of sleep deprivation in depressive patients can hardly be explained as a psychological or placebo effect. The improvement of mood by sleep deprivation and the relapse into depression after the following night of sleep are contrary to the general expectation of depressive patients. This was further confirmed in the studies conducted by Buddeberg and Dittrich (1978), which revealed the absence of a correlation between the patients' expectations and the objectively measurable effect. It must therefore be assumed that the mechanism of action of sleep deprivation is neurobiological, although there is no generally accepted explanatory model to date. This subject has been reviewed by Van den Hoofdakker (1997) and Kasper and Möller (1996).

A number of hypotheses are based on models of normal and disordered sleep regulation. According to the two-process model of sleep–wake cycle regulation, a homeostatic process S and a circadian process C are the major factors controlling the sleep–wake rhythm. Borbély and Wirtz-Justice (1982) postulated that depression may be associated with a deficiency of process S, which both increases the need to sleep and brightens mood. The therapeutic effect of sleep deprivation is attributed to a transient elevation of factor S. A further hypothesis is based on the disinhibition of rapid eye movement (REM) sleep in depressive patients: sleep deprivation is said to counteract depression primarily through the avoidance of REM sleep. In accordance with this hypothesis, Vogel et al. (1980) achieved an antidepressant effect in depressive patients by selectively suppressing REM sleep. REM sleep is associated with an inhibition of catecholaminergic and a stimulation of cholinergic neurotransmission, i.e. it intensifies the presumed imbalance of these neurotransmitter systems in depression (Berger and Riemann 1993).

Other theories favor a neurochemical explanation, by analogy to the mechanisms of action of the antidepressants. Thus, for example, a serotonergic mechanism has been proposed. This hypothesis is supported by studies in which the administration of serotonergic antidepressants such as clomipramine or fluoxetine was found to stabilize the effect of sleep deprivation. The serotonin system seems to play an important role in the mechanisms of action of lithium and phototherapy, and the administration of either of these was also found to stabilize the effect of sleep deprivation. F. Benedetti and colleagues (personal communication) recently reported that patients who have a functional polymorphism of the serotonin transport gene, and, therefore, a better therapeutic response to selective serotonin reuptake inhibitors (Smeraldi et al. 1998) also have a better response to sleep deprivation therapy.

Findings from neurochemical studies and experimental studies in animals suggest that the effect of sleep deprivation is comparable to that of the amphetamines (Ebert and Berger 1998). Imaging studies have revealed a reduction of limbic metabolism after the acute administration of stimulants (Volkow et al. 1997). As a good response to sleep deprivation was found to be associated with a reduction of limbic metabolism in other studies (Ebert et al. 1991; Wu et al. 1992), these findings are consistent with the hypothesis that sleep deprivation, like the psychostimulants, has a therapeutic effect by provoking neurotransmitter release.

2.4 Sleep Deprivation and Shifting of Sleep Phases

As early as the 1970s and 80s, single cases were reported in which brief daytime naps destroyed the beneficial effect of sleep deprivation therapy. Thus, in clinical practice, efforts are taken to ensure that the patient does not fall asleep after sleep deprivation.

The phenomenon of relapse into depression after brief episodes of sleep was studied systematically by our research group. We found that its occurrence does not depend on the length of the naps or the appearance of a particular stage of sleep, but the time of day does

have an effect: morning naps led to a relapse more frequently than afternoon naps (Wiegand et al. 1993; Riemann et al. 1993). This observation, combined with the fact that sleep deprivation is more effective in the second half of the night than in the first half, led to the hypothesis that there is a critical phase, beginning in the early morning hours and lasting until the early afternoon, in which sleep has a negative effect on mood in depressive patients. Avoidance of sleep in this critical temporal period seems to have an antidepressant effect.

A new method of treatment based on this hypothesis involves shifting the patient's sleeping hours into the noncritical period lasting from the late afternoon until the first half of the night.

The treatment begins with total or partial sleep deprivation for one night. The next day, the patients go to bed at 5 P.M. and are awakened 7 h later, i.e. at midnight. The day after that, the sleeping hours are advanced by 1 h, i.e. the patients go to bed at 6 P.M. and wake up at 1 A.M. Postponement of sleeping hours by 1 h per day continues for another 5 days until the patients have a normal sleep rhythm again, going to bed at 11 P.M. and waking up at 6 A.M. When this procedure is followed, sleep in the second half of the night is totally eliminated for a few days and reduced for a few days thereafter.

Studies to date have shown that the therapeutic benefit persists in two thirds of patients responding to sleep deprivation (Vollmann and Berger 1993; Riemann et al. 1996; Berger et al. 1997; Albert et al. 1998). A comparison of patients who did and did not take medication in addition to sleep deprivation therapy suggested that its therapeutic effect is independent of concomitant antidepressant treatment (Berger et al. 1997).

In an initial randomized and controlled study of 40 depressive patients, the effect of sleep deprivation followed by a shift of sleep phases to the late afternoon and first half of the night ($n = 20$) was compared with that of a control condition, in which sleep deprivation was followed by a phase shift in the opposite direction (from 2 A.M. to 9 A.M. in the first night, and so forth; $n = 20$). A total of 75% of patients in the first group, but only 40% in the second group, experienced a reduction of Hamilton depression scores by at least 30% at the end of treatment (Riemann et al., in press).

Sleep deprivation and the shifting of sleep phases may be used at the beginning of inpatient treatment for depression to carry patients over the latency period before the onset of efficacy of antidepressants, i.e. to brighten the patients' mood several weeks earlier than otherwise possible, with a beneficial effect in two thirds of all patients. As this mode of treatment alone is not expected to have a lasting effect, pharmacological maintenance therapy should be started simultaneously and continued for 6 months; phase prophylaxis should also be given, when indicated.

3 Phototherapy

Phototherapy has been used since the 1980s as a further method of somatic treatment, primarily in patients with seasonally associated depression. Patients generally sit for 1–2 h, in the morning or evening, with their eyes open in front of a lamp from which bright, white light at an intensity of 2500–10,000 lux (but no ultraviolet light) falls directly on the retina. The treatment is usually given for 14 days.

The development of phototherapy for the treatment of depressive illnesses was originally stimulated by the case of a patient with bipolar illness who had personally observed that his depressive and manic episodes varied with the seasons. During a depressive phase beginning in the winter months of 1980–1981, he received his first treatment with bright, white light, which led to a remission of his depressive manifestations (Lewy et al. 1982). This positive experience in a single case was the starting point for numerous studies over the following years of phototherapy at various light intensities. Exposure to light was carried out either in the morning or in the evening, or both.

The same group of clinical researchers that first used phototherapy in depressive patients defined, in the early 1980s, the diagnostic criteria for a new subtype of depressive illness called seasonal affective disorder (SAD; Rosenthal et al. 1984): depressive episodes beginning in the autumn or winter and remitting spontaneously in the following spring or summer, in at least 2 consecutive years. A further requirement for a positive diagnosis is that no relevant psychosocial factors should be present that vary seasonally themselves and might thus explain the seasonal pattern of the affective illness. The clinical manifestations of SAD differ from those of depression of the usual kind by the more frequent occurrence of atypical features such as severe loss of energy, hypersomnia, carbohydrate craving, and weight gain.

In the DSM-IV classification, seasonal depression can be used as a supplementary coding to designate the course of recurrent episodes of an affective disorder. The required criteria for seasonal depression include a pattern of seasonal onset of the episodes over 2 years and a clear predominance of seasonal depressive episodes over nonseasonal episodes in the long term.

Most studies of phototherapy to date have been performed on patients with SAD. Light intensities of 2500 or 10,000 lux were used. The success rates in the individual studies varied from 30% to 70%. The

Table 1. Rates of success of treatment in several studies of phototherapy (after Terman et al. 1989 and the Groningen studies of Meesters et al. 1995, Van den Hoofdakker and Gordijn 1997)

Light exposure	Patients (*n*)	HAMD baseline	HAMD after treatment	HAMD improvement (%)
Morning only[a]	172	17.8	8.1	54
Morning only[b]	16	18.1	8.9	51
Morning only[c]	14	16.9	4.7	72
Afternoon only[a]	34	21.2	12.4	42
Afternoon only[c]	15	15.9	8.4	49
Evening only[a]	143	18.0	10.1	44
Evening only[b]	11	15.8	7.1	50
Evening only[c]	12	17.5	5.5	68
Morning and evening[a]	136	21.1	9.2	56
Evening/morning[c]	14	16.2	8.2	49
Morning/evening[c]	13	19.0	6.5	67
Dim light (control condition)	77	23.4	20.0	15

HAMD, Hamilton depression scores.
[a]After Terman et al. (1989).
[b]Groningen studies: data from 1989/1990.
[c]Groningen studies: data from 1990/1991, 1991/1992, 1992/1993; treatment with 10,000 lux. The other studies used 2500 lux.

percentage of improvement in various studies is given in Table 1. In contrast to studies of antidepressant drugs, the duration of treatment was only 14 days. Positive effects usually appeared a few days after the beginning of treatment.

These studies showed that the time of day at which the patient was exposed to light seemed to make no difference, despite original predictions on the basis of chronobiological hypotheses.

Unlike patients with seasonal depression, patients with typical nonseasonal depression seem to respond poorly or not at all to phototherapy; the rate of response among such patients is much lower than in those with SAD. Phototherapy therefore cannot be recommended as a standard treatment for nonseasonal depression. It should be pointed out, however, that there are fewer published studies of phototherapy in nonseasonal depression than in seasonal depression, and many studies do not satisfy strict methodological criteria (for a review, see Van den Hoofdakker and Gordijn 1997).

3.1 Critical Remarks

Many authors have correctly remarked that the positive effect of phototherapy on seasonal depression may be due to a placebo effect (Eastman 1990); it is known that the placebo effect plays a major role in studies of outpatient therapy in patients with mild to moderately severe depression. In many of the controlled studies of phototherapy, dim light at intensities of 50–500 lux was used as a control condition, a range in which artificial room light also falls. Patients can easily distinguish between dim light and bright, white light, so the difference in the effectiveness of the two intensities might be attributable to different expectations of therapeutic success on the patients' part.

Furthermore, it must be observed that the intensity of natural daylight out of doors in typical European and North American longitudes exceeds 1000 lux even on cloudy and rainy days and may be as high as 100,000 lux, while most phototherapy studies employed an intensity of 2500 lux, usually for 2 h. As the effect of phototherapy seems not to depend on the time of day, the patient might simply take a walk outside for an equivalent length of time instead of sitting in front of a lamp. Wirz-Justice et al. (1996) studied the effect of "natural phototherapy" of this type in seasonal depressives, compared with a control condition in which patients received a small dose of artificial light, which was expected to have no therapeutic effect. Natural phototherapy led to a significant improvement of mood compared to the control condition after 1 week of treatment.

3.2 Mechanism of Action

There is still no generally recognized hypothesis for the mechanism of action of phototherapy (Leonhardt and

Wirz-Justice 1995). Because of the known effect of light on the circadian system, which is mediated by the retina, the retino-hypothalamic tract, and the suprachiasmatic nucleus (the anatomical substrate of the biological clock), it was originally supposed that phototherapy might act by normalizing a disturbance of the circadian rhythm. Nonetheless, no such disturbance could be found in patients with SAD. Furthermore, the effect of phototherapy on disturbances of the sleep–wake cycle is known to depend on the time of day at which treatment is given, while the effect of phototherapy in SAD is independent of the time of day (Wirz-Justice et al. 1993). These facts do not support the hypothesis that the antidepressant action of light is based on an effect on the circadian system.

Other theories postulate an effect on neurotransmitter systems, e.g. the serotonergic system (Neumeister et al. 1997). The dose-dependence of the phototherapy effect and its independence of the time of day are both compatible with this hypothesis.

3.3 Practical Application

At the present time, phototherapy can be recommended as a single therapy against depression (i.e. not in combination with medications) only in milder episodes of seasonal depression. It can also be used in combination with medications in other forms of depression. Furthermore, the use of phototherapy to stabilize the effect of sleep deprivation and to treat sleep disorders in senile depression is also reasonable. The therapeutic effect can be expected to begin 3–4 days after the start of treatment. Recommendations for the application of phototherapy are as follows:

- The patient should look directly at a bright light with open eyes.
- 2500 lux for 2 h or 10,000 lux for 40 min.
- Every day for 2 weeks, in the morning or evening (whichever is more convenient for the patient).
- Maintenance of the proper distance to avoid overdose (according to the instructions of the manufacturer).

Side effects of phototherapy include occasional irritation of the eyes and headache. It should be used with caution in the case of: ophthalmologic illnesses (retinopathies, glaucoma, cataract); use of medications that increase ocular sensitivity to light (e.g. lithium, fluoxetine, propranolol).

4 Repetitive Transcranial Magnetic Stimulation

Repetitive transcranial magnetic stimulation (rTMS) is a new form of treatment in which a powerful magnetic field is used to cause electrical stimulation of brain areas. The possible antidepressant effect of this method is now being investigated in numerous studies; no firm conclusion regarding its clinical usefulness can be drawn at present. If the antidepressant effect of rTMS is confirmed, it will constitute a good alternative to electroconvulsive therapy (ECT), because, unlike ECT, it does not require general anesthesia.

TMS was originally developed for diagnostic use in neurology. In contrast to transcranial electrical stimulation (TES), TMS is painless and noninvasive. It is performed by externally applying a strong magnetic field (1.5–2 T) to the head, which then, penetrating the skull and subarachnoid space, induces an electrical current in the underlying brain structures. Stimulation of the motor cortex in this way with a sufficiently high current, for example, results in neuronal depolarization, which, if the motor pathway is intact, evokes a motor response detectable as electromyographic (EMG) activity in the corresponding muscles. In contrast to single-stimulus TMS, rTMS (repetitive or rapid-rate transcranial magnetic stimulation) employs high-frequency stimulus sequences (frequency >1 Hz, more than two consecutive stimuli, constant stimulus interval).

The initial case report by Höflich et al. (1993) was followed in the ensuing years by studies from a number of different research groups, involving a relatively small number of cases (for a review, see Haag et al. 1997), providing evidence for a possible antidepressant effect. According to the study by Pascual-Leone et al. (1996a,b), high-frequency left prefrontal stimulation was particularly associated with a therapeutic effect.

In a crossover study of 12 patients (George et al. 1997), a sham (placebo) treatment was used as a control. The patients treated with rTMS responded with a significant improvement of mood, while sham treatment resulted in a worsening of Hamilton depression scores. Figiel et al. (1998) treated 50 patients suffering from treatment-resistant depression with rTMS and obtained a response rate of 42%. A study of patients with mania (Grisaru et al. 1998) yielded evidence that stimulation of the right prefrontal cortex is an effective treatment for this illness. This contrasts with the above-mentioned finding that left prefrontal stimulation is effective against depression. In a more recent controlled study (Zwanzger et al. 1999), rTMS at an intensity of 90% and 100% of the motor threshold was compared to placebo treatment; rTMS at 100% intensity was found to be more effective than either of

the other two conditions. No randomized, controlled trial has yet been published comparing the efficacy of TMS and ECT.

4.1 Possible Mechanism of Action

Experimental studies in animals have revealed evidence that TMS and ECT exert similar effects on the behavioral level and lead to comparable biochemical changes in the central nervous system (Belmaker and Grisaru 1998). Multiple research groups have shown that rTMS produces a significant reduction of immobility in Porsolt's Forced Swimming Test (Zyss et al. 1997, 1999; Belmaker and Grisaru 1998). Thus animal models, too, provide evidence supporting an antidepressant effect of rTMS.

In rTMS, unlike ECT, the objective is not to induce an epileptic seizure. Seizures occur rarely during rTMS as an unwanted side effect. As for ECT, the induction of seizures is not necessarily its actual mechanism of action, but it is, apparently, a necessary precondition for a therapeutic benefit. Thus skepticism regarding the efficacy of rTMS as compared to ECT seems appropriate.

4.2 Indications, Side Effects, and Safety

Because there is a small chance that rTMS may induce an epileptic seizure, patients should undergo an EEG before they are treated, and those with evidence of elevated cerebral excitability should be excluded from treatment (Brandt et al. 1997; Wassermann 1998). Sleep deprivation should be avoided before rTMS. Caution is also indicated regarding the use of medications that significantly lower the seizure threshold. There is no evidence that rTMS leads to memory disturbances, as ECT does, nor does it require general anesthesia.

The potential side effects of rTMS are as follows:

- Focal and secondarily generalized epileptic seizures
- Change in auditory threshold (acoustic artifact up to 120 dB)
- Brief unpleasant sensation under the stimulating coil
- Headache
- Migraine attack in patients with known migraine (rare)
- Transient tinnitus

Important safety considerations in the performance of rTMS include the following:

- EEG before initial treatment
- Avoidance of sleep deprivation
- Caution in the presence of medications that lower the seizure threshold
- Hearing protection during TMS
- Not to be used in patients with metallic foreign bodies in the brain or cardiac pacemakers because of the strong magnetic field
- Not to be used in patients with tinnitus

In summary, no recommendation can be given at present for the routine use of rTMS as a treatment for depression. However, there is good reason to believe that rTMS might be a component of the spectrum of antidepressant therapies in the future, in particular as a simpler alternative to ECT, with fewer side effects, in the treatment of medically refractory depression.

5 References

Albert R, Merz A, Schubert J, Ebert D (1998) Schlafentzug und anschließende Schlafphasenvorverlagerung stabilisiert den positiven Schlafentzugseffekt bei depressiven Episoden. Nervenarzt 69: 66–69

Barbini B, Colombo C, Benedetti F, Campori E, Bellodi L, Smeraldi E (1998) The unipolar-bipolar dichotomy and the response to sleep deprivation. Psychiatry Res 79: 43–50

Belmaker RH, Grisaru N (1998) Magnetic stimulation of the brain in animal depression models responsive to ECS. J ECT 14: 194–205

Benedetti F, Barbini B, Lucca A, Campori E, Colombo C, Smeraldi E (1997) Sleep deprivation hastens the antidepressant action of fluoxetine. Eur Arch Psychiatry Clin Neurosci 247: 100–103

Benedetti F, Colombo C, Barbini B, Campori E, Smeraldi E (1999) Ongoing lithium treatment prevents relapse after total sleep deprivation. J Clin Psychopharmacol 19: 240–245

*Berger M, Riemann D (1993) REM sleep in depression – an overview. J Sleep Res 2: 211–223

Berger M, Vollmann J, Hohagen F, König A, Lohner H, Voderholzer U, Riemann D (1997) Sleep deprivation combined with consecutive sleep phase advance as a fast-acting therapy in depression: an open pilot trial in medicated and unmedicated patients. Am J Psychiatry 154: 870–872

Borbély AA, Wirz-Justice A (1982) Sleep, sleep deprivation and depression. Hum Neurobiol 1: 205–210

Bouhuys AL, Beersma DGM, Van den Hoofdakker RH (1989) Observed behavior as a predictor of the response to sleep deprivation in depressed patients. Psychiatry Res 28: 47–61

Bouhuys AL, Van den Burg W, Van den Hoofdakker RH (1995) The relationship between tiredness prior to sleep deprivation and the antidepressant response to sleep deprivation in depression. Biol Psychiatry 37: 457–161

Brandt SA, Ploner CJ, Meyer BU (1997) Repetitive transkranielle Magnetstimulation. Möglichkeiten, Grenzen und Sicherheitsaspekte. Nervenarzt 68: 778–784

Broocks A, Bandelow B, Pekrun G et al. (1998) A comparison of aerobic exercise, clomipramine and plazebo in the treatment of panic disorder. Am J Psychiatry 155: 603–609

Buddeberg C, Dittrich A (1978) Psychologische Aspekte des Schlafentzugs. Arch Psychiatr Nervenkr 225: 249–261

Colombo C, Benedetti F, Barbini B, Campori E, Smeraldi E (1999) Rate of switch from depression into mania after therapeutic sleep deprivation in bipolar depression. Psychiatry Res 86: 267–270

Eastman CI (1990) What the placebo literature can tell us about light therapy for SAD. Psychopharmacol Bull 4: 495–504

Ebert D, Berger M (1998) Neurobiological similarities in antidepressant sleep deprivation and psychostimulant use: a psychostimulant theory of antidepressant sleep deprivation. Psychopharmacology 140: 1–10

Ebert D, Feistel H, Barocka A (1991) Effects of sleep deprivation on the limbic system and the frontal lobes in affective disorders: a study with Tcggm HMPAO SPECT. Psychiatry Res Neuroimaging 40: 247–251

Elsenga S, Van den Hoofdakker RH (1987) Response to total sleep deprivation and clomipramine in endogenous depression. J Psychiatr Res 21: 151–161

Elsenga S, Van den Hoofdakker RH, Dols LCW (1990) Early and late partial sleep deprivation in depression. In: Stefanis C, Soldatos C, Rabavilas A (eds) Psychiatry: a world perspective, vol 2. Excerpta Medica, Amsterdam, pp 374–379

Figiel GS, Epstein C, McDonald WM, Amazon-Leece J, Figiel L, Saldivia A, Glover S (1998) The use of rapid-rate transcranial magnetic stimulation (rTMS) in refractory depressed patients. J Neuropsychiatry Clin Neurosci 10: 20–25

George MS, Wassermann EM, Kimbrell TA et al. (1997) Mood improvement following daily left prefrontal repetitive transcranial magnetic stimulation in patients with depression: a placebo-controlled crossover trial. Am J Psychiatry 154: 1752–1756

Goetze U, Tölle R (1981) Antidepressive Wirkung des partiellen Schlafentzuges während der 1. Hälfte der Nacht. Psychiatr Clin 14: 129–149

*Gordijn MCM, Beersma DGM, Bouhuys AL, Korte HJ, Van den Hoofdakker RH (1995) A longitudinal study of sleep deprivation responses in depression: the variability is highly related to diurnal mood variability. Acta Neuropsychiatr 7: 58–60

Gordijn MCM, Beersma DGM, Bouhuys AL, Van den Hoofdakker RH (1998) Mood variability and sleep deprivation effect as predictors of therapeutic response in depression. In: Beersma DGM, Van Bemmel AL, Folgering H, Hofman WF, Ruigt GSF (eds) Sleep-wake research in the Netherlands, vol 9, pp 41–44

Grisaru N, Chudakov B, Yaroslavsky Y, Belmaker RH (1998) Transcranial magnetic stimulation in mania: a controlled study. Am J Psychiatry 155: 1608–1610

Grube M, Hartwich P (1990) Maintenance of antidepressant effect of sleep deprivation with the help of lithium. Eur Arch Psychiatr Neurol Sci 240: 60–61

Haag C, Padberg F, Möller HJ (1997) Transkranielle Magnetstimulation (TMS). Ein Diagnostikum aus der Neurologie als Therapeutikum in der Psychiatrie? Nervenarzt 68: 274–278

Haug HJ (1992) Prediction of sleep deprivation outcome by diurnal variation of mood. Biol Psychiatry 31: 271–278

Höflich G, Kaspers S, Hufnagel A, Ruhrmann S, Möller HJ (1993) Application of transcranial magnetic stimulation in treatment of drug-resistant depression – a report of two cases. Hum Psychopharmacol 8: 361–365

Kasper S, Möller HJ (eds) (1996) Therapeutischer Schlafentzug: Klinik und Wirkmechanismen. Springer, Berlin Heidelberg New York

*Kuhs H, Färber D, Borgstädt S, Mrosek S, Tölle R (1996) Amitriptyline in combination with repeated late sleep deprivation versus amitriptyline alone in major depression. A randomised study. J Affect Disord 37: 31–41

Kuhs H, Tölle R (1991) Sleep deprivation therapy. Biol Psychiatry 29: 1129–1148

Leibenluft E, Wehr TA (1992) Is sleep deprivation useful in the treatment of depression? Am J Psychiatry 149: 159–168

Leonhardt G, Wirz-Justice A (1995) Bisherige Erfahrungen und praktische Anwendung mit Lichttherapie. In: Zulley J, Wirz-Justice A (eds) Lichttherapie. Roderer, Regensburg, pp 53–62

Lewy AJ, Kern HE, Rosenthal NE et al. (1982) Bright artificial light treatment of a manic-depressive patient with a seasonal mood cycle. Am J Psychiatry 139: 1496–1498

Meesters Y, Jansen JHC, Beersma DGM et al. (1995) Light therapy for seasonal affective disorder. The effects of timing. Br J Psychiatry 166: 607–612

Neumeister A, Goessler R, Lucht M, Kapitany T, Bamas C, Kasper S (1996) Bright light stabilizes the antidepressant effect of sleep deprivation. Biol Psychiatry 39: 16–21

Neumeister A, Praschak-Rieder N, Heßelmann B, Rao ML, Glück J, Kasper S (1997) Effects of tryptophan depletion on drug-free patients with seasonal affective disorder during a stable response to bright light therapy. Arch Gen Psychiatry 54: 133–138

Pascual-Leone A, Catala MD, Pascual AP (1996a) Lateralized effect of rapid-rate transcranial magnetic stimulation of the prefrontal cortex in mood. Neurology 46: 499–502

Pascual-Leone A, Rubio B, Pallardo F, Catala MD (1996b) Rapid-rate transcranial magnetic stimulation of left dorsolateral prefrontal cortex in drug-resistant depression. Lancet 348: 233–237

Pflug G, Tölle R (1971) Therapie endogener Depressionen durch Schlafentzug. Nervenarzt 42: 117–124

Reinink E, Bouhuys A, Wirz-Justice A, Van den Hoofdakker RH (1990) Prediction of the antidepressant response to total sleep deprivation by diurnal variation of mood. Psychiatry Res 32: 113–124

Riemann D, Wiegand M, Lauer CH, Berger M (1993) Naps after total sleep deprivation in depressed patients. Are they depressogenic? Psychiatry Res 94: 109–120

Riemann D, Hohagen F, König A, Schwarz B, Gomille J, Voderholzer U, Berger M (1996) Advanced vs. normal sleep timing: effects of depressed mood after response to sleep deprivation in patients with a major depressive disorder. J Affect Disord 37: 121–128

Riemann D, Hohagen F, König A et al. How to preserve the antidepressant effect of sleep deprivation: a comparison of sleep phase advance and sleep phase delay. Eur Arch Psychiatr Clin Neurosci (in press)

Rosenthal NE, Sack DA, Gillin JC et al. (1984) Seasonal affective disorder: a description of the syndrome and preliminary findings with light therapy. Arch Gen Psychiatry 41: 72–80

Sack DA, Duncan W, Rosenthal NE, Mendelson WE, Wehr TA (1988) The timing and duration of sleep in partial sleep deprivation therapy of depression. Acta Psychiatr Scand 77: 219–224

Schilgen B, Tölle R (1980) Partial sleep deprivation as therapy for depression. Arch Gen Psychiatry 37: 267–271

Schulte W (1969) Klinische Erfahrungen über das Herausgeraten aus der melancholischen Phase. In: Hippius H, Sehlbach H (eds) Das depressive Syndrom. Urban and Schwarzenberg, Munich, pp 415–420

Smeraldi E, Zanardi R, Benedetti F, Di Bella D, Perez J, Catalano M (1998) Polymorphism within the promotor of the serotonin transporter gene and antidepressant efficacy of fluvoxamine. Mol Psychiatry 3: 508–511

Szuba MP, Baxter LR Jr, Fairbanks LA, Guze BH, Schwartz JM (1991) Effect of partial sleep deprivation on the diurnal variation of mood and motor activity in major depression. Biol Psychiatry 30: 817–819

Szuba MP, Baxter LR Jr, Altshuler LL, Allen EM, Guze BH, Schwartz JM, Liston EH (1994) Lithium sustains the acute antidepressant effects of sleep deprivation: preliminary findings from a controlled study. Psychiatry Res 51: 283–295

*Terman M, Terman JS, Quitkin FM et al. (1989) Light therapy for seasonal affective disorder. A review of efficacy. Neuropsychopharmacology 2: 1–22

Van den Burg W, Bouhuys AL, Van den Hoofdakker RH, Beersma DGM (1990) Sleep deprivation in bright and dim light: antidepressant effects on major depressive disorder. J Affect Disord 19: 109–117

*Van den Hoofdakker RH (1997) Total sleep deprivation: clinical and theoretical aspects. In: Honig A, van Praag HM (eds) Depression: Neurobiological, psychopathological and therapeutic advances. Wiley, Chichester, pp 563–589

Van den Hoofdakker RH, Gordijn MCM (1997) Will light brighten the future of the depressive patient? Acta Neuropsychiatr 9: 71–76

Van den Hoofdakker RH, Gordijn MCM, Kasper S (1994) Sleep deprivation in refractory depression. In: Nolen WA, Zohar J, Roose SP, Amsterdam JD (eds) Refractory depression. Current strategies and future directions. Wiley, Chichester, pp 129–142

*Vogel GW, Vogel F, Mcabee RS, Thurmond AJ (1980) Improvement of depression by REM sleep deprivation; new findings and a theory. Arch Gen Psychiatry 37: 247–253

Volkow N, Wang G, Fowler J, Logan J, Lieberman J, Pappas N (1997) Effects of methylphenidate on regional brain glucose metabolism in humans. Am J Psychiatry 154: 50–55

Vollmann J, Berger M (1993) Sleep deprivation with consecutive sleep phase advance therapy in patients with major depression: a pilot study. Biol Psychiatry 33: 54–57

Wassermann E (1998) Risk and safety of repetitive transcranial magnetic stimulation: report and suggested guidelines from the International Workshop on the Safety of Repetitive Transcranial Magnetic Stimulation. Electroencephalogr Clin Neurophysiol 108: 1–16

Wehr TA, Goodwin FK (1981) Biological rhythms and psychiatry. In: Arieti S, Brodie HKH (eds) American handbook of psychiatry, vol 7, 2nd edn. Basic, New York, pp 46–74

Wehr TA, Rosenthal NE, Sach DA, Gillin JC (1985) Antidepressant effects of sleep deprivation in bright and dim light. Acta Psychiatr Scand 72: 161–165

Wiegand M, Riemann D, Schreiber W, Lauer CJ, Berger M (1993) Effect of morning and afternoon naps on mood after total sleep deprivation in patients with major depression. Biol Psychiatry 33: 467–476

Wirz-Justice A, Graw P, Kräuchi K et al. (1993) Light therapy in seasonal affective disorder is independent of time of day or circadian phase. Arch Gen Psychiatry 50: 929–937

Wirz-Justice A, Graw P, Kräuchi K, Sarrafzadeh A, English J, Arendt J, Sand L (1995) Natural light treatment of seasonal affective disorder. J Affect Disord 37: 109–120

Wu JC, Bunney WE (1990) The biological basis of an antidepressant response to sleep deprivation and relapse: review and hypothesis. Am J Psychiatry 147: 14–21

Wu JC, Gillin JC, Buchsbaum MS, Hershey T, Johnson JC, Bunney WE Jr (1992) Effects of sleep deprivation on brain metabolism of depressed patients. Am J Psychiatry 194: 538–543

Zwanzger P, Thoma H, Mikhaiel P, Kathmann N, Hampel H, Möller HJ, Padberg F (1999) Transcranial magnetic stimulation in major depression. Abstract presented at the World Congress of Psychiatry, Hamburg, 6–11 August 1999

Zyss T, Gorka Z, Kowalska M, Vetulani J (1997) Preliminary comparison of behavioral and biochemical effects of chronic transcranial magnetic stimulation and electroconvulsive shock in the rat. Biol Psychiatry 42: 920–924

Zyss T, Mamczarz J, Vetulani J (1999) The influence of rapid-rate transcranial magnetic stimulation (rTMS) parameters on rTMS effects in Porsolt's forced swimming test. Int J Neuropsychopharmacol 2: 31–34

CHAPTER

25

E. Frank, M.E. Thase, C. Spanier,
J.M. Cyranowski, L. Siegel

Psychotherapy of Affective Disorders

This work was supported in part by grants MH49115, MH29618, MH30915, and MH41884 from the National Institute of Mental Health. Portions of this chapter have appeared previously in Frank and Spanier C (1995), Thase (1995), and Thase and Beck (1993).

1 Introduction

It is not mere coincidence that roughly contemporaneous with the development of the Research Diagnostic Criteria (Spitzer et al. 1978) there emerged in the United States a series of short-term psychotherapeutic approaches to the treatment of major depression. The move toward more specific categorization of nonpsychotic disorders undoubtedly facilitated the theoretical work necessary for the development of what have come to be known as the "depression-specific" psychotherapies, but also provided the atmosphere in which these theories could be developed into practical interventions that then could be tested empirically. Although this chapter is entitled "Psychotherapy of Affective Disorders," its major focus will be on the psychotherapy of unipolar depression and, in particular, on the treatment of acute major depressive episodes in adults. We emphasize this area for the simple reason that it is the area of the most consistent theoretical and empirical work.

This chapter will also focus on individual treatments as opposed to marital, family, or group interventions, again because the majority of data available are for individual interventions. The most prominent and empirically supported among these individual interventions have been interpersonal psychotherapy (IPT) as developed by Klerman et al. (1984), cognitive therapy (CT) as developed by Beck et al. (1979), and a variety behavioral therapies that were studied in the 1970s and early 1980s and that are finding a new place in the limelight as the field attempts to meet the demand for ever-briefer interventions. Recently, Shapiro et al. (1991, 1994, 1995) and others in the United Kingdom have examined the efficacy of psychodynamic-interpersonal psychotherapy (IP; Shapiro et al. 1994, 1995; Guthrie et al. 1998, 1999). Although the clinical use of more traditional psychodynamic psychotherapies in depressed patients is common, the multiple extant forms of this therapeutic approach (often provided without manual-based guidelines) and the lack of a coherent body of controlled research on the specific use of this therapy for depressed patients with rationally derived depression outcomes precluded further inclusion of psychodynamic psychotherapies in the current chapter.

As will become apparent as this chapter proceeds, there are a number of important commonalities among the depression-specific psychotherapies. Each was originally conceptualized as a time-limited intervention. They are all highly focused on the patient's depression and on the problems believed to be relevant to the onset and maintenance of the depressive episode. They are all present-oriented and highly pragmatic in their approach. While some provide much more explicit guidelines for therapists' behavior from beginning to end of treatment, and even from beginning to end of session, all share an emphasis on structuring the therapy to a greater or lesser extent. Each of these therapies, in addition, is likely to seem coherent and relevant to the depressed patient. Finally, and perhaps most importantly, each of the depression-specific psychotherapies provides a clear message of hope for the depressed patient.

2 Interpersonal Psychotherapy

2.1 Overview and Description

IPT examines the relationship between depression and problems in the interpersonal realm, making no attribution as to whether the interpersonal problems apparent in the context of depression represent causes of, or have been caused by, the depressive episode. IPT was initially designed to be practiced as a weekly, face-to-face, present-oriented, and relatively short-term therapy "that emphasizes the current interpersonal relations of the depressed patient while acknowledging the role of genetic, biochemical, developmental, and personality factors in the causation of and vulnerability to depression" (Klerman et al. 1984, p. 5). The techniques of IPT, while incorporating both psychodynamically oriented therapies (e.g. exploration, clarification of affect) and cognitive-behavioral therapies (CBT, e.g. behavior change techniques, reality-testing of perceptions), were, in fact, developed for the management of the four basic problem areas which distinguish IPT: (a) unresolved grief, (b) role transitions, (c) interpersonal role disputes (often marital disputes), and (d) interpersonal deficits. Accordingly, IPT is not considered to be unique in terms of its techniques, although the strategies used in IPT (e.g. relating symptom onset to overt or covert disputes with significant others with whom the patient is currently involved) *are* considered to be distinctive (Klerman et al. 1984).

The major goals of IPT are achieved by ascertaining with the patient which of the four types of problems described above was associated with the onset of the current episode of depression and, subsequently, by working with the patient to renegotiate interpersonal difficulties associated with the primary problem area. Rooted in the interpersonal school of psychoanalysis founded by Meyer (1957) and articulated by Sullivan (1953), and informed by Bowlby's attachment theory (Bowlby 1982), IPT recognizes the profound impact of

early developmental experiences and unconscious mental processes (i.e. intrapsychic wishes and conflicts) on later patterns of interpersonal relationships; however, during the actual practice of IPT, the clinician focuses on improving the patient's current social roles and interpersonal relations (Klerman et al. 1984). Rather than analyzing and reconstructing intrapsychic or cognitive events from the past, IPT focuses on improving current interpersonal relations. This emphasis on *current* relational dilemmas is based on the assumption that early childhood experiences will be reflected in current patterns of interpersonal relating and in the patient's current social roles. In the original conceptualization of IPT, it was assumed that interventions focusing on the current interpersonal context of a patient's life would both facilitate recovery from the acute episode as well as provide protection against reemergence of symptoms (Klerman and Weissman 1993; Klerman et al. 1984) as long as treatment was ongoing, and possibly even after treatment discontinuation. The first of these assumptions has been verified empirically, as will be elaborated below. The second assumption, of continued prophylactic benefit after treatment has been discontinued, has received inconsistent empirical support and only for continued improvement in social functioning, not for prevention of recurrence.

The therapeutic stance in IPT is one of warmth, support, and empathy. In other words, the role of the therapist in IPT is one of active patient advocate and not neutral commentator as in psychoanalysis, nor one of providing unconditional positive regard as in Rogerian client-centered psychotherapy. Accordingly, since transference is not facilitated, the therapeutic relationship is conceptualized to be based in reality, as are the patient's interpersonal perceptions of problems outside of therapy.

2.2 Maintenance Interpersonal Psychotherapy

The adaptation of IPT to a maintenance treatment was based extensively on IPT as developed by Klerman et al. (1984). While maintenance interpersonal psychotherapy (IPT-M; Frank 1991) preserves the four distinctive problem areas of IPT and implements the strategies (e.g. exploration of feelings associated with loss of a role) and techniques of IPT (e.g. elicitation of feelings, nonjudgmental exploration of the affective quality of relationships, facilitating the development of satisfying and adaptive interpersonal behaviors), it differs from IPT in terms of its goals and timing. The primary goal of IPT-M is to prevent recurrence, i.e. to sustain wellness in patients who are fully remitted; in contrast, the goal of IPT is to bring about remission from an acute episode of depression. Moreover, IPT-M is designed to treat patients in a maintenance format for several years. For example, patients were treated for up to 3 years with IPT-M in the Maintenance Therapies in Recurrent Depression (MTRD) study for which it was originally developed (Frank et al. 1990). Accordingly, the number of problem areas that are typically the focus of IPT-M is greater than in IPT for the acute treatment of depression.

Since the goal of IPT-M is prevention, the therapist is mindful of the early development of interpersonal problems similar to those associated with the onset of the patient's most recent episode as well as earlier episodes of depression (Frank 1991). The emphasis in IPT-M is on augmenting the patient's strengths and helping the patient assume much of the responsibility for the prevention of future episodes. Thus the patient is encouraged to be watchful for the appearance of early somatic and cognitive symptoms characteristic of prior depressive episodes. If symptoms are reported, preventative strategies are planned and implemented by the therapist and patient to avoid the onset of a new episode. Consistent with an interpersonal approach to the treatment of depression, these strategies are primarily interpersonal in emphasis and designed to improve the patient's mood and adaptive functioning (e.g. increasing supportive social contacts outside of a chronically conflictual marriage, attending to one's own emotional or practical needs, and focusing less on the needs of others). IPT-M, in contrast to weekly IPT for acute treatment, was scheduled once per month in the MTRD protocol (Frank et al. 1990). This modification in the timing of IPT sessions was justified in accordance with the nature of the study population, which consisted of asymptomatic patients with a history of recurrent episodes of depression.

The problem areas that become the focus of treatment in IPT-M usually involve a combination of role transitions, role disputes, and interpersonal deficits and less frequently grief issues, unless a significant loss-related event occurs during maintenance treatment (Frank et al. 1993). Identification of the problem area in IPT-M evolves over the course of acute treatment, in which a number of interpersonal themes surface but are not the focus. These themes, particularly ingrained interpersonal patterns that continue to be problematic for the patient during remission or as a result of remission, are noted by the therapist and explored in IPT-M. While the techniques used in IPT-M are similar to IPT, the time-frame of IPT-M places a greater emphasis on the future (e.g. developing adaptive coping strategies in preparation for an upcoming and potentially stressful interpersonal event).

2.3 Evidence for the Efficacy of Interpersonal Psychotherapy

A relatively small number of researchers have evaluated interpersonal psychotherapy as short-term (DiMascio et al. 1979; Elkin et al. 1989; Schulberg et al. 1996; M.M. Weissman et al. 1979) and long-term (Frank et al. 1990, 1991a; Klerman et al. 1974; Reynolds et al. 1999a; M.M. Weissman et al. 1974) treatments for depression. Although the absolute number of studies is small, they have consistently demonstrated IPT to be efficacious, not only in alleviating depressive symptoms, but also in extending the well interval in patients with recurrent illness.

2.3.1 Interpersonal Psychotherapy as an Acute Treatment for Depression

Three controlled trials have been completed which demonstrate the efficacy of IPT as an acute treatment for major depression. The first study, the New Haven-Boston Collaborative Study of the Treatment of Acute Depression, begun in 1973, was a four-cell, 16-week randomized trial of 81 depressed patients involving weekly IPT and amitriptyline either alone or in combination. The control treatment consisted of non-scheduled psychotherapy. Although there were no significant differences between IPT and medication in reduction of symptoms at the end of treatment, both treatments were more effective than the control treatment, and combined treatment was more effective than either treatment alone (DiMascio et al. 1979; M.M. Weissman et al. 1979). A reexamination of these data revealed that the advantage of the combination of IPT and amitriptyline was limited to the subgroup of patients with more severe endogenous depressions (see Thase, in press). This study also found that patients who had received IPT, either alone or in combination with pharmacotherapy, developed significantly better psychosocial functioning at 1-year follow-up than either treatment control patients or patients who received pharmacotherapy alone (M.M. Weissman et al. 1981). The researchers noted, however, that across the four treatment groups, many patients relapsed by the 1-year follow-up, thus indicating that higher social functioning did not appear to be of significant prophylactic benefit and suggesting that 16 weeks of treatment may be inadequate to maintain recovery from an acute episode of depression.

The second acute treatment study, the landmark National Institute of Mental Health (NIMH) Treatment of Depression Collaborative Research Program (NIMH-TDCRP, Elkin et al. 1989), also strongly supported the efficacy of IPT for the acute treatment of depression in outpatients. A total of 250 unipolar, nonpsychotic depressed outpatients were studied at three sites and were randomly assigned to 16 weeks of either imipramine plus clinical management therapy (a "minimally supportive therapy" provided by a psychiatrist involving regular meetings consisting of support, encouragement, and at times direct advice), IPT, CBT (Beck et al. 1979), or placebo treatment (i.e. pill-placebo plus clinical management, PLA-CM). In contrast to the Boston-New Haven study, a combined treatment condition consisting of medication and psychotherapy was not included. The primary analyses showed that, at termination of the 16-week treatment period, patients in *all* treatments, including PLA-CM, showed significant reduction in depressive symptoms and improved functioning. When aggregating across the multiple outcome measures, the three active treatments generally performed better than the placebo and clinical management; however, these differences typically did not reach conventional levels of statistical significance after adjustments for multiple comparisons. In addition, there was no evidence in the primary analyses that the efficacy of the psychotherapies was significantly different from tricyclic antidepressant treatment. Secondary analyses, in which patients were dichotomized according to baseline severity of depression, revealed no significant differences among the four treatments in the less severely depressed subjects, but consistently significant differences among the treatments for the group of patients who were more severely depressed. Less severely ill patients (baseline Hamilton Depression Score, HRSD < 20) improved with all treatments, including PLA-CM, while the more severely depressed patients (HRSD $\geqslant 20$) in PLA-CM did poorly. For the severely ill patients, IPT was comparable to imipramine in efficacy and both were superior to PLA-CM. IPT was not significantly more effective than CBT in this more severely depressed group, although there were suggestive trends. In a subsequent reanalysis stratifying the sample on the presence or absence of atypical depression, Stewart et al. (1998) found that IPT was equally effective among both typical and atypical patients.

In an 18-month naturalistic follow-up of the course of depressive symptoms (with follow-up data collected at 6, 12, and 18 months), Shea et al. (1992) found that the percentage of patients who both met a stringent criterion of recovery during the acute treatment and who did not relapse over the 18-month follow-up period did not differ significantly among the four treatments and was low, ranging from 19% to 30%. Moreover, the rate of relapse among patients who had recovered across the four groups was high, ranging from 30% to 50%. Similar to the results obtained by M.M. Weissman et al. (1981), these findings indicate

that the recurrence rate for major depression is high, despite effective acute treatment, and that 16 weeks of treatment is insufficient in many instances to maintain lasting remission. Thus it appears that short-term psychotherapy does not have enduring effects in many cases and thus cannot be considered effective in the long term. It is important to note that when the collaborative study was begun, data from the MTRD study (Frank et al. 1990) that demonstrate the efficacy of maintenance treatment were not available.

As a result of controversy surrounding the approach to data analysis, two reanalyses of the NIMH-TDCRP efficacy data have been reported (Gibbons et al. 1993; Klein and Ross 1993). Although the findings were generally consistent with those originally reported (Elkin et al. 1989), the reanalysis by Klein and Ross (1993) showed more distinct differences in efficacy among treatments, particularly among the more severely depressed patients. They found that medication was superior to psychotherapy, and IPT and CBT were superior to placebo, particularly in the severely ill group. Again, they found trends suggesting that IPT was superior to CBT among patients with severe illness.

In an effort to control for the effects of several problems common to longitudinal psychiatric data, including serial correlation, missing data, and person-specific effects, Gibbons et al. (1993) used random regression models (RRM) to reanalyze the NIMH-TDCRP longitudinal HRSD dataset. Gibbons and colleagues argued that the result of incorporating these so-called random effects was "a clear test of the primary hypotheses of the TDCRP study without compromises..." (p. 749). While their results were in agreement with those published earlier, the use of RRM allowed clearer inferences and more confidence in the major findings of the TDCRP. Results of the RRM analyses revealed significant improvement over all subjects ($p \leqslant q.001$) as well as a significantly faster rate of improvement for imipramine relative to placebo ($p \leqslant q.03$), with clearer differences from placebo in this analysis at 16 weeks than in the original analysis. In contrast to other analyses, no further differences were found among treatment groups at 16 weeks. In other words, none of the analyses provided evidence of significant differences in efficacy between the two psychotherapies (IPT vs. CBT) or between these groups and the placebo group (PLA-CM), or between the two psychotherapies (considered both separately and jointly) and IMI-CM. Inspection of trend lines, however, revealed some ordering of treatments; imipramine was superior to psychotherapy, and IPT and CBT were superior to placebo.

The original analysis coupled with the subsequent reanalyses of the NIMH-TDCRP dataset indicate that IPT is an effective acute treatment for depression, even for more severely depressed patients. In addition, based on HRSD recovery criterion scores (i.e. $\leqslant 6$), there is some evidence for the specific effectiveness of IPT as compared with PLA-CM, especially in terms of recovery rates for the more severely depressed patients.

The third major study (Schulberg et al. 1996) compared IPT ($n = 93$) with pharmacotherapy ($n = 91$; nortriptyline, plasma levels 190–270 n*M*) and a treatment as usual comparison group ($n = 92$). This study, which was conducted in four primary care clinics in the Pittsburgh area, included a 16-week acute-phase and 4-monthly continuation phase sessions. Results clearly favored both well-specified treatments over the usual care condition, and after 4 months remission rates were as follows: IPT, 46%; nortriptyline, 48%; and usual care, 18%. After 8 months of treatment, 72% of the IPT group had remitted, as compared to 67% of the patients treated with nortriptyline and only 20% of those in the usual care condition. Subsequent analysis confirmed that IPT was equally effective in more and less severely depressed patients (Schulberg et al. 1998).

Despite this seemingly consistent support for the short-term efficacy of acute IPT treatment, in one notable exception, Reynolds et al. (1999b) failed to find significant benefit for IPT monotherapy as compared to an attention-placebo control. This study of recently bereaved, older (>49 years) adults compared IPT plus nortriptyline ($n = 16$), IPT plus placebo ($n = 17$), medication clinic plus nortriptyline ($n = 25$), and medication clinic plus placebo ($n = 22$). Response rates following the 16-week trial were as follows: IPT plus nortriptyline, 69%; nortriptyline monotherapy, 56%; IPT monotherapy, 29%; and no-treatment control, 45%. Thus nortriptyline, either alone or in combination with IPT, was superior to both placebo and IPT. Notably, however, the combination of medication and IPT was associated with the lowest attrition rate, and thus the highest treatment completion rate, across all conditions.

2.3.2 Interpersonal Psychotherapy as a Continuation and Maintenance Treatment for Depression

Three randomized controlled trials have been completed that demonstrate the efficacy of IPT as a prophylactic treatment for major depression. In the original study of IPT, the New Haven-Boston Collaborative Study of the Treatment of Acute Depression, Klerman et al. (1974) assessed the effect of IPT in an 8-month six-cell trial in 150 depressed female outpatients who had responded with symptom reduction to 4–6 weeks of amitriptyline therapy. This trial would probably be considered continuation treatment rather

than a maintenance treatment study by today's standards. By *continuation treatment* we refer to the phase of treatment that is provided in the 4–6 months following remission of the depressive episode. A return of depressive symptoms during this treatment phase is typically referred to as a relapse of the previous depressive episode. In contrast, *maintenance treatment* refers to the phase of treatment provided after the patient has shown a full recovery of the depressive episode, i.e. a remission that is sustained for an adequate period of time, such as 4–6 months. Depressive symptoms that appear during this time frame are typically considered to represent the appearance of a new episode of depression or a recurrence. (For a full definition of these and related terms, see Frank et al. 1991b.) Thus, while the goal of continuation treatment is to prevent a relapse of the previous depressive episode, the goal of maintenance treatment is to prevent the recurrence of new depressive episodes or to extend the period of wellness between depressive episodes.

In the New Haven-Boston Collaborative Study, patients were randomly assigned to 8 months of weekly IPT, medication, or their combination, IPT and placebo, placebo alone, or no pill. At the conclusion of treatment, relapse rates were highest for patients receiving no treatment (36%) compared with the other three active treatment groups: medication alone (12%), IPT alone (16.7%), and the combined IPT/medication group (12.5%). Patients receiving IPT demonstrated improved social functioning, although these effects were not apparent for 6–8 months (M.M. Weissman et al. 1974). The investigators concluded that superior outcomes were produced by the combination of IPT with medication. However, the design of the study, in which medication was initially administered alone, restricts generalizations regarding the efficacy of IPT to patients who have already responded positively to amitriptyline.

The second study, the MTRD (Frank et al. 1990; Kupfer et al. 1992), is the longest randomized maintenance trial to date. In this 3-year outcome study in depressed patients who showed a clear history of repeated episodes of depression, we (Frank et al. 1990) contrasted maintenance IPT-M with maintenance pharmacotherapy (imipramine), combination pharmacotherapy/psychotherapy, and a control group in a five-cell design in order to determine whether IPT-M alone or in combination with medication would play a significant role in the prevention of recurrence. IPT-M was designed specifically to maintain recovery and reduce vulnerability to future episodes by improving social adjustment; thus the focus was on the interpersonal and psychosocial context of the *well* state. Accordingly, we hypothesized that the risk of recurrence would be reduced in IPT-M by improving social adjustment; the patient would be helped to cope more effectively with interpersonal and social problems associated with the well state which would, in turn, reduce the number and severity of stressful life-events. This, in turn, would reduce the risk of recurrence.

In this trial, all subjects were treated acutely with a combination of IPT and imipramine and continued on that combination until they had sustained a clear-cut remission (HRSD $\leqslant 7$) for 20 weeks. They were then randomly assigned to one of five maintenance treatments: (1) IPT-M alone, (2) IPT-M with placebo tablet, (3) IPT-M with imipramine, (4) medication clinic visits with imipramine, or (5) medication clinic visits with placebo tablet. In contrast to previous studies, imipramine was maintained at the highest dose of medication ever employed in a maintenance trial (mean dose, >200 mg/day), rather than being tapered from acute treatment levels, while IPT was administered monthly and thus at the lowest dose ever in comparison to other clinical trials (Klerman et al. 1994). We found that active imipramine, maintained at an average dose of 208 mg, provided a prophylactic effect to a larger proportion of patients over a longer period of time than in any previous study of maintenance therapy of recurrent depression (Frank et al. 1990). Survival analyses indicated that patients who received imipramine alone or in combination with IPT displayed the greatest mean survival time (124 weeks and 131 weeks, respectively), although these two groups did not significantly differ. The study also demonstrated that when patients received a combination of pharmacotherapy and IPT for the treatment of their acute episode, those who continued to receive IPT on a monthly basis following drug discontinuation remained well significantly longer than those who did not, thus indicating a clinically meaningful and statistically significant effect for IPT-M, i.e. the IPT-M alone and IPT-M plus placebo groups showed mean survival times of 82 and 74 weeks, respectively, which significantly differed from the 45 weeks displayed by patients with no active treatment. Thus this long-term outcome study clearly established the value of maintenance treatment in the prevention of recurrence in major depression and found IPT-M to have significant prophylactic capacity, even when used at a very low dose in patients at high risk for recurrence.

In a follow-up report from this study and in an effort to further understand the relationship between monthly psychotherapy and longer survival time, we (Frank et al. 1991a) examined the contribution of treatment quality and demographic and clinical variables to long-term survival in the IPT-M alone and IPT-M plus placebo conditions. Treatment quality was defined as the extent to which therapists conformed to specific principles, goals, and techniques of IPT psychotherapy, thus labeled "treatment specificity." While no demo-

graphic or clinical variables were related to long-term survival in the IPT-M conditions in this follow-up report, higher (i.e. above the median) specificity of IPT-M was associated with significantly increased survival time, with a median survival time of almost 2 years (102 weeks; SE, 8 weeks; $p < .001$). In contrast, patient–therapist dyads rated as having low treatment specificity of IPT-M or who, in other words, were unable to focus consistently on interpersonal concerns, had a median survival time of less than 5 months (18 weeks; SE, 4.6 weeks). Patients in the low-specificity dyads received no more protection from the IPT-M treatment assignment than was provided by the medication clinic plus placebo assignment (median survival time, 21 weeks). We concluded that, when patients are able to maintain a consistent focus on interpersonal issues, monthly sessions of IPT provide substantial protection against recurrence of depression.

The third study, the Pittsburgh Study of Maintenance Therapies in Late-Life Depression (MTLD; Reynolds et al. 1999a), extended the findings of Frank et al. (1990) regarding the value of IPT-M to include late-life depressed populations. In this 3-year outcome study, the first randomized, placebo-controlled study of maintenance psychotherapy in late-life depression, elderly patients (i.e. patients older than 59 years) with a clear history of recurrent unipolar depression were treated to full recovery with open treatment of nortriptyline and IPT. Recovered subjects ($n = 107$) were then randomized into one of four maintenance treatments: monthly IPT-M sessions alone (with placebo), nortriptyline alone (with monthly medication clinic), combined IPT and nortriptyline, or a no-treatment control (placebo and medication clinic). Results indicated that all three active treatments were better than placebo in preventing depression recurrence. While the preventive capacity of medication alone and IPT-M alone did not significantly differ, combined IPT and medication treatment showed a trend toward added benefit over either treatment alone in preserving recovery, an effect that was particularly notable in patients 70 years and older (Reynolds et al. 1999a).

Taken together, the results from the MTRD study, the MTLD study, the study by Schulberg and colleagues, the NIMH Collaborative Study, and the studies by Klerman, Weissman, and colleagues provide clear evidence of the efficacy of IPT in the treatment of major depression. The performance of IPT in these clinical trials consistently surpasses minimal treatment or placebo control groups in both acute and maintenance treatment of mid- and late-life outpatients with depression. The recent negative study by Reynolds et al. (1999b) does raise some concern about the utility of IPT alone for treatment of bereavement related depressions in later life.

3 Psychodynamic-Interpersonal Psychotherapy

3.1 Overview and Description

One model of psychotherapy that includes psychodynamic elements, is available in manual-based format, and has a growing empirical track record to support its efficacy in the specific treatment of adult depression is Psychodynamic-Interpersonal Therapy (PI). Previously known as *exploratory therapy*, the model for PI is based on Hobson's conversational model (Hobson 1985; Goldberg et al. 1984). Further developed in manual-based format by Shapiro and Firth (1985), PI utilizes psychodynamic, interpersonal, and experiential approaches to therapy. In line with IPT, PI views interpersonal difficulties as primary in the origins of depression. Yet, in line with psychodynamic and experiential approaches, PI places a greater emphasis on the patient–therapist relationship as a tool for revealing and resolving relevant interpersonal issues. There is, however, less emphasis on the interpretation of transference than in most formal dynamic therapies.

The PI model assumes that the patient's depression arises from or is exacerbated by problems within his or her interpersonal relationships. PI therapists seek to understand, with the patient, the bases of interpersonal problems through the exploration of the patient's feelings and mutual negotiation – where the therapist's views are expressed as tentative statements open to correction, elaboration, and feedback. Therapists take a tentative, encouraging, and supportive approach, helping the patient to link his or her distress to specific interpersonal problems. Finally, the therapeutic relationship itself is utilized to address interpersonal problems and to test solutions in the "here and now" (Guthrie et al. 1998, 1999; Shapiro et al. 1994).

3.2 Evidence for the Efficacy of Psychodynamic-Interpersonal Psychotherapy

As noted above, a manual for PI has been developed by Shapiro and Firth (1985), as have adherence rating scales (Shapiro and Startup 1990, 1993). As part of the Sheffield Psychotherapy Projects (see Shapiro et al. 1991), Shapiro and colleagues have implemented a programmatic series of studies regarding the efficacy of PI therapy for the treatment of adult depression. The first such project, which included both depressed and/or anxious patients, utilized a crossover design to compare PI (then termed exploratory therapy) with CBT (termed prescriptive therapy) and indicated

a marginal superiority of CBT over PI (Shapiro and Firth 1987). More recently, Shapiro et al. (1994) compared the effectiveness of PI and CBT in 117 depressed patients diagnosed with the Diagnostic Interview Schedule (DIS; Eaton and Kessler 1985) and stratified for severity with the Beck Depression Inventory (BDI; Beck et al. 1961), who were randomly assigned to one of four treatment conditions: 8 weeks of PI, 16 weeks of PI, 8 weeks of CBT, or 16 weeks of CBT. At post-treatment, CBT and PI did not significantly differ in effectiveness as assessed by the majority of the protocol's post-treatment outcome measures; however, a moderate effect size was obtained on the BDI, indicating some post-treatment advantage of CBT over PI on this outcome (Shapiro et al. 1994). No overall advantage of 16-session treatment over eight-session treatment was obtained at post-treatment, although severely depressed patients were generally found to improve more after 16 sessions than after eight sessions (Shapiro et al. 1994). Notably, however, results obtained at 1-year follow-up (with data collected for 89% of the original study patients) substantially differed from post-treatment results. Specifically, at 1-year follow-up, the eight-session PI treatment condition appeared significantly less efficacious than the other three treatment conditions across nearly all outcome measures. In addition, there was no longer a measurable benefit of 16-session CBT over eight-session CBT, even in the initially severely depressed patients (Shapiro et al. 1995).

Most recently, Guthrie et al. (1999) examined psychological, social, and economic outcomes for non-psychotic psychiatric patients (75.5% of whom had a depressive illness) who were unresponsive to 6 months of routine mental health treatment and who were subsequently randomized to either 8 weeks of PI or usual psychiatric care. Patients receiving PI treatment displayed significantly greater improvement in psychological distress and social functioning, as well as significant reductions in health care utilization in the 6 months post-treatment, as compared with usual care controls (Guthrie et al. 1999).

4
Cognitive Therapy

4.1
Overview and Description

The theoretical basis of cognitive therapy has been derived from three primary sources: (1) the phenomenological perspective, (2) the structural view of personality and psychotherapy, and (3) contemporary work in cognitive and behavioral psychology (Beck 1976; Beck et al. 1979). Cognitive therapy shares with the phenomenological approach an emphasis on concepts of self and the personal world as key determinants of behavior (Frankl 1985). This contribution dates to the ancient Greek school of Stoic philosophy (Beck 1976). The writings of the post-Freudian analysts such as Adler (1936), Horney (1950), and Sullivan (1953) stressed this viewpoint on personality and psychopathology.

Freud's structural theory that partitioned cognition (thought) into primary and secondary processes and recognized the role of conscious, preconscious, and unconscious mental activity was a second major influence. The psychoanalytic model also postulated the existence of personality constructs and defense mechanisms that, at times of distress or conflict, are central to the etiology of psychopathological reactions. George Kelly's (1955) formulation of personal constructs and Jean Piaget's (1954) studies of schemas (i.e. internalized and hierarchical sets of rules used for problem-solving) in the cognitive development of children and adolescents also helped to shape the cognitive model for psychotherapy.

The fundamental principles of cognitive therapy were outlined by Beck in a series of papers in the early 1960s (Beck 1961, 1963, 1964). Subsequently, a number of research studies solidified the importance of cognitive distortions in various psychopathological states (e.g. Beck 1967, 1976; Braff and Beck 1974; A.N. Weissman 1979; Nelson and Craighead 1977; Rizley 1978; Hollon and Kendall 1980). Another influence was the introduction of Albert Ellis' (1962) system of rational-emotive therapy. Ellis, who, like Beck, had been trained as an analyst, also emphasized the significance of irrational or distorted beliefs in the origin and maintenance of "neurotic" psychopathology and similarly advocated active and direct interventions in psychotherapy of depressed patients.

Beck's model of cognitive therapy continued to evolve during the 1960s and 1970s (Beck 1967, 1970) and was first published in a fully developed form in the text *Cognitive Therapy and the Emotional Disorders* (Beck 1976), which also described the use of behavioral methods such as activity scheduling (Lewinsohn et al. 1982), self-reinforcement and self-monitoring (McLean 1982), and social skills training (Hersen et al. 1984).

The cognitive therapy of depression posits that three types of problems in cognition are involved in the genesis and/or maintenance of depression. The first type of cognitive dysfunction derives from the fact that depressed individuals spend a disproportionate amount of time thinking gloomy or unpleasant thoughts about themselves, their world, and their future. Beck (1976) has referred to the content of thoughts in these three domains – self, world, and future – as the cognitive triad. Cognitions that are

particularly relevant to the depressed patients are those that occur almost instantaneously with the worsening of dysphoric moods. These cognitions are referred to as automatic negative thoughts, and they provide the gateway for the cognitive therapist to understand the depressed patient's phenomenological world.

The second type of cognitive dysfunction in depression involves errors in logic and information processing (Beck 1976; Burns 1980). For most patients, errors in information processing are primarily state dependent, i.e. only apparent when a person is in a state of depression or another dysphoric mood (Coyne and Gotlib 1983; Haaga et al. 1991; Robins and Hayes 1993). Information-processing errors associated with depression include overgeneralization, excessive personalization, selective abstraction, emotional reasoning, and all-or-nothing thinking (for an expanded description, see Burns 1980).

The third type of cognitive dysfunction in depression involves hypothesized or "deeper" cognitive structures, such as dysfunctional attitudes and a depressogenic schema (Beck 1976; Segal 1988; Young and Lindemann 1992). In cognitive therapy, both attitudes and schema are considered to be ultimately accessible through questioning techniques, as illustrated by the use of the Socratic method, or guided discovery (Beck et al. 1979).

Cognitive therapy is more traditionally oriented than behavior therapy in that pathological schemas are conceived of as "unconscious," nonobservable constructs. These depressogenic structures are presumed to result from adverse early experiences (Beck 1976; Segal 1988). In individuals prone to depression, pathological schemas in relevant areas such as excessive interpersonal dependence or perfectionistic expectations for success are proposed to be "silent" during times of a stable romantic relationship or a high vocational attainment (Persons and Miranda 1992). However, they are "activated" in response to specific, matching adversities (e.g. Hammen et al. 1989; Segal et al. 1992).

Early in the course of therapy, particularly with more severely depressed patients, cognitive therapists liberally employ behavioral techniques. For example, daily monitoring of moods and activities is used to increase participation in rewarding behaviors, as well as to help establish functional relationships between changes in moods and accompanying patterns of automatic thoughts. Similarly, stepwise graded task assignments are used to help patients begin to approach and overcome problems that are perceived as overwhelming.

Slowly, and in a manner directly tied to each patient's ability to use more abstract cognitive interventions, the therapy moves toward eliciting and testing the accuracy of automatic thoughts, developing rational alternatives, and identifying and modifying maladaptive schemas. Therapeutic strategies, such as the use of written responses to stereotypic automatic negative thoughts ("coping cards") and a printed, five-column form known as the "Daily Record of Dysfunctional Thoughts," are used to teach patients to begin to challenge their negative cognitions. Patients are also encouraged to keep their thought records as part of a journal or notebook so that a coherent summary of the course of therapy is readily available and can be drawn upon repeatedly. When properly performed, each session ends with a new homework assignment that builds logically on the material covered within that session.

It is important to distinguish more simplistic models of cognitive intervention, such as verbal persuasion or the notion that negative thoughts are "replaced" by more positive thoughts, from the actual process of Beck's model of cognitive therapy. In contrast to persuasion, in which the "expert" directly advocates a "correct position," cognitive therapy emphasizes the use of Socratic questioning in order to guide the depressed person's discovery that logical errors or distorted thinking influences his or her assessments and interpretations. These possibilities are then tested in vivo as experiments. That more positive alternative conclusions typically can be identified and validated is at the heart of cognitive therapy.

4.2 Evidence for the Efficacy of Cognitive Therapy

Beck's model of cognitive therapy is the best studied psychological treatment of major depression (Depression Guideline Panel 1993; Thase 1995). Cognitive therapy has been extensively studied in comparison with waiting-list control conditions and with other forms of psychotherapy, as well as pharmacotherapy. However, despite such intensive study, only two trials have compared cognitive therapy with a placebo-clinical management condition (Elkin et al. 1989; Jarrett et al. 1999).

There is little doubt about the efficacy of cognitive therapy as an acute-phase treatment when compared with waiting-list control conditions (Beach and O'Leary 1992; Neimeyer et al. 1989; Propst et al. 1992; Ross and Scott 1985; Rude 1986; Scott and Stradling 1990; Selmi et al. 1990; Thompson et al. 1987). In the meta-analysis conducted by the Depression Guideline Panel (1993) of the Agency for Health Care Policy and Research (AHCPR), cognitive therapy had an overall efficacy rate of 46.6% (±6.9%), with an advantage of 30% (±22%) when compared with waiting-list control conditions.

With respect to comparisons with pharmacotherapy, results of four studies indicate that cognitive therapy is

superior to low-contact or "treatment-as-usual" medication control conditions, whether cognitive therapy is administered alone (Blackburn et al. 1981) or in combination with treatment as usual (Ross and Scott 1985; Scott and Stradling 1990; Teasdale et al. 1984). Each of these studies had a primary care provider or family practitioner prescribe the pharmacotherapy. Results from studies utilizing more rigorous pharmacotherapy conditions, as are provided in psychiatric outpatient clinics (the psychiatric clinic setting of Blackburn et al. 1981; Elkin et al. 1989; Hollon et al. 1992; Jarrett et al. 1999; McKnight et al. 1992; Murphy et al. 1984), have yielded more consistent evidence of parity. Of note, there is little evidence that the combination of cognitive therapy and pharmacotherapy routinely delivers an added benefit when compared to cognitive therapy alone (e.g. Blackburn et al. 1981; Hollon et al. 1992; Murphy et al. 1984), although the methodological limitations of these smaller outpatient studies may obscure detection of more additive benefits (see Thase, in press).

In a reanalysis of the TDCRP study conducted by Stewart et al. (1998), cognitive therapy did quite well among patients with features of atypical depression, and relatively poorly among those with more "typical" symptom profiles. A recent placebo-controlled study by Jarrett et al. (1999) confirmed the utility of cognitive therapy for patients with atypical depression.

Cognitive therapy generally has been found to have an efficacy comparable to that of other active psychotherapies, including behavioral marital therapy (Beach and O'Leary 1992; Jacobson et al. 1991), individual or group behavior therapy (Gallagher and Thompson 1982; Rude 1986; Shaw 1977; Thompson et al. 1987), IPT (Elkin et al. 1989), brief dynamic therapy (Gallagher and Thompson 1982; Thompson et al. 1987), pastoral counseling (Propst et al. 1992), and nondirective group therapy (Hogg and Deffenbacher 1988). In the study by Jacobson et al. (1991), individual cognitive therapy appeared to be more effective than behavioral marital theory in a subset of patients with satisfactory marriages, whereas the latter treatment produced greater gains on measures of marital satisfaction in maritally distressed couples. Interestingly, a combination of individual cognitive therapy and behavioral marital therapy was no more effective than the individual modalities (Jacobson et al. 1991).

4.3 Cognitive Therapy as a Continuation or Maintenance Form of Treatment

During naturalistic follow-up, patients who were treated with cognitive therapy generally have fared better over 1- or 2-year follow-ups than patients who were treated with treatment-as-usual interventions (Ross and Scott 1985; Scott and Stradling 1990) or those who were typically withdrawn from antidepressant pharmacotherapy (Blackburn et al. 1986; Evans et al. 1992; Kovacs et al. 1981; Simons et al. 1986). However, in the TDCRP study, no appreciable difference in relapse rates was found among patients who responded to cognitive therapy, IPT, imipramine, or placebo (Shea et al. 1992). It should be noted that acute-phase cognitive therapy was not significantly more effective than clinical management and placebo condition in this trial (Elkin et al. 1989) and that, in the absence of evidence of acute-phase efficacy for cognitive therapy, the assumption of prophylaxis may not hold true. Comparability in survival rates among several forms of active psychotherapy has been observed in two other follow-up studies (Gallagher-Thompson et al. 1990; Jacobson et al. 1993).

Cognitive therapy is being increasingly studied as a continuation- or maintenance-phase therapy. In one early controlled trial, treatment with monthly cognitive therapy alone achieved the same level of prophylaxis over 6 months as was achieved by continuation pharmacotherapy (Blackburn et al. 1986). Blackburn and Moore (1997) subsequently replicated this finding in a larger study comparing continuation forms of cognitive therapy and pharmacotherapy. Two other early studies, employing relatively small samples and a type of CBT not entirely consistent with Beck's model of treatment, failed to find any prophylactic value for continued "booster" therapy sessions after termination of acute treatment (Baker and Wilson 1985; Kavanaugh and Wilson 1989). This may be because only the subset of patients with incomplete remissions are at increased risk for relapse after acute-phase cognitive therapy (Thase et al. 1992). More recently, Fava and colleagues (1994, 1996, 1998) have found that a short course of cognitive therapy focused on residual symptoms after acute-phase pharmacotherapy significantly reduced the risk of relapse even after discontinuation of medication.

The results of the NIMH-TDCRP study have raised some concerns about the suitability of cognitive therapy for treatment of more severely depressed patients (see, e.g. American Psychiatric Association 1993). Persons et al. (1996) have challenged this contention and pointed out that the TDCRP results are not consistent with the bulk of evidence comparing cognitive therapy and pharmacotherapy (see also Thase 1995). Nevertheless, results from a series of studies conducted at the University of Pittsburgh do suggest that patients with abnormal electroencephalographic sleep profiles (Thase et al. 1996a) or hypercortisolism (Thase et al. 1996b) are less responsive to cognitive therapy than patients with more normal

neurobiological profiles. A similar association between sleep abnormalities and IPT response also has been observed in two studies (Buysse et al. 1999; Thase et al. 1997a). We suspect that these neurobiological abnormalities may be associated with dysregulation of affect, diminished hedonic capacity, and/or impaired information processing that could, in turn, interfere with productive psychotherapy. Such patients may be better treated with antidepressants, either alone (Thase et al. 1997a) or in combination with psychotherapy (Thase et al. 1997b).

5 Behavioral Therapies

5.1 Overview and Description

The conceptual basis for behavioral approaches to psychotherapy originated from two main sources: (1) increasing attempts to understand human behavior on the basis of experimental learning theory, as evident in the early work of Thorndike (1931) and Skinner (1953), and (2) a general dissatisfaction with intrapsychic theories of psychopathology that were difficult to test in an empirical fashion (see Hoberman 1990). In response to this latter "philosophy of science" critique, behavior therapy is often conceptualized as a general scientific approach to the study of human behavior change, rather than a single theoretical model or set of therapeutic techniques. As described by Kazdin (1982), this methodological approach to behavior therapy is characterized by a focus on the maladaptive behaviors for which the patient seeks treatment, a reliance on empirical findings from general psychology (and, particularly, the psychology of learning), and a focus on current rather than historical determinants of behavior. Moreover, behavioral approaches rely heavily on behavioral assessment techniques, including detailed functional analyses of environmental determinants and consequences of behavior, as well as ongoing monitoring of target behaviors and utilization of assessment data to guide treatment implementation (Kazdin 1982).

Most behavioral theories of depression draw from one or more basic tenets of social learning theory. Social learning theory posits that psychological functioning is best conceptualized as a set of ongoing, reciprocal interactions among personal factors (e.g. cognitive sets or expectancies), behavioral factors, and environmental factors (Bandura 1977). Accordingly, behavioral approaches to depression often view people and their environments as reciprocal determinants of each another.

Skinner (1953) and Ferster (1966) were among the first to conceptualize depression as a behavioral extinction-related phenomenon, resulting from an interruption or reduction in behavior caused by decreased positive reinforcements in the social environment. Lewinsohn and colleagues (e.g. Lewinsohn and Shaw 1969; Lewinsohn 1974; Lewinsohn et al. 1979) further refined this position, describing depression as the result of a lack of response-contingent positive reinforcements in one's social environment. According to Lewinsohn, low response-contingent positive reinforcement may be due to a number of causes, including an inability to obtain positive social reinforcements because of social skills deficits, a lack of available positive reinforcers (or surplus of aversive or stressful experiences) in one's environment, and/or a decreased capacity to enjoy positive experiences (Lewinsohn et al. 1973; Lewinsohn and Gotlib 1995). Taking a cognitive-behavioral approach, Rehm's (1977) self-control model postulates that depression arises from deficits in self-monitoring (e.g. a tendency to attend to negative feedback and the immediate versus delayed outcomes of behavior), self-evaluation (e.g. setting unrealistic or perfectionistic standards for evaluating oneself), and self-reinforcement (characterized by low rates of contingent self-rewards and high rates of self-punishment). In contrast, Nezu and colleagues (Nezu 1987; Nezu et al. 1989) have implicated ineffective problem-solving skills in the onset and maintenance of depressive episodes.

At the heart of behaviorally based treatment approaches to depression is the attempt to alter dysfunctional interactions between the patient and his or her social environment and to increase the frequency and effectiveness of contingent positive reinforcements received by the depressed individual for adaptive, nondepressed behaviors. Toward this end, specific interventions may be aimed at changing aspects of the patient's environment, increasing social skills required for the patient to elicit positive reinforcement from the environment (such as positive social interaction, effective communication, or problem-solving skills), increasing the pleasantness and decreasing the aversiveness of patient–environment interactions, and/or increasing rates of self-reinforcement (Foa et al. 1989; Hoberman 1990).

Behavioral treatments of depression share a number of common strategies (see Hoberman and Lewinsohn 1985; Lewinsohn and Gotlib 1995). These typically include monitoring patients' activity level, moods, and/or thoughts, as well as the use of functional analysis to determine environmental stimuli that may predispose, trigger, reinforce, or maintain depressive (versus nondepressive) responses *for the particular individual.* Patients are aided to identify pleasant or reinforcing events or activities, and to increase these events, and,

similarly, to identify and reduce aversive or stressful events or activities. Patients are commonly encouraged to set small and achievable goals early in treatment and to reinforce themselves for graded therapy successes, in order to increase the patient's use of positive self-reinforcement, to obtain early reductions in depressive affect and, thereby, a sense of hopefulness, and to strengthen the patient's sense of self-efficacy. Social skills training is another common component of behavior therapies and may include interventions designed to increase interpersonal skills and improve social interaction, communication, assertiveness, decision-making, and/or problem-solving abilities. Finally, behavioral approaches to psychotherapy are typically time-limited, commonly designed to be implemented in 4–12 weeks (Lewinsohn and Gotlib 1995).

5.2 Evidence for the Efficacy of Behavioral Therapies

Lewinsohn and colleagues (e.g. Lewinsohn et al. 1980) were among the first to develop and test a predominantly behavioral intervention for the treatment of unipolar depression. This structured, 12-session program was specifically designed to increase patient's engagement in pleasant activities, and reduce the intensity and frequency of aversive activities, and included a variety of behavioral and cognitive intervention strategies such as training in assertiveness, relaxation, self-control, decision-making, problem-solving, communication, and time management. Lewinsohn and his group have demonstrated in a randomized clinical trial that, relative to various control groups, behavioral therapy increased pleasant experiences and reduced aversive experiences, which led to subsequent decreases in depression severity (for a comprehensive review of this work, see Lewinsohn and Gotlib 1995).

Based on the rationale that depressed individuals lack the skills necessary to elicit positive reinforcements from the social environment, a number of investigators have utilized behavioral treatments for depression that are explicitly designed to provide social skills training. Bellack et al. (1980) describe one such behavioral treatment program comprising four major components: (1) social skills training, (2) social perception training, (3) practice (including homework), and (4) self-evaluation and self-reinforcement (see also Becker and Heimberg 1985; Becker et al. 1987). Bellack et al. (1981, 1983) and Hersen et al. (1984) demonstrated that social skills training was as effective as amitriptyline in an acute, 12-week clinical trial. Moreover, treatment effects were maintained with six to eight booster sessions provided over a 6-month follow-up period. Similarly, McLean and Hakstian's (1979) behavioral intervention designed to improve social skills, problem-solving, and self-control was found to be equal or superior to insight-oriented psychotherapy, relaxation therapy, and amitriptyline in a 10-week clinical trial. Furthermore, data at a 27-month follow-up showed behavioral therapy subjects to be less depressed and more socially active and productive than patients in the other treatment conditions, particularly as compared with patients in the relaxation therapy condition (McLean and Hakstian 1990). Thus the efficacy of social skills training for depression has received support through a number of sources.

Rehm's self-control therapy is typically administered in six to 12 group sessions and includes structured interventions designed to remediate deficits in self-monitoring, self-evaluation, and self-reinforcement. Evidence supports the superiority of this intervention to no-treatment controls and nonspecific psychosocial treatments, yet suggests that self-monitoring tactics represent the most essential component of this intervention program (see Rehm 1990; Craighead et al. 1998). Finally, Nezu and Perri (1989) have provided preliminary evidence of the efficacy of a 12-week problem-solving intervention in reducing depressive symptoms at post-treatment and 6-month follow-up, and Arean et al. (1993) have reported the superiority of problem-solving therapy to both waiting-list control and reminiscence therapy in the treatment of older adults with depression.

Recent support for the effectiveness of behavioral therapy is evident in the work of Jacobson and colleagues. In a recent component analysis of Beck's cognitive therapy for depression, Jacobson et al. (1996) found that the behavioral activation component of cognitive therapy was itself as effective as either a partial or full cognitive therapy intervention program. Behavioral activation interventions were limited to those strategies specifically designed to "activate people in their natural environment." These included daily activity monitoring, assessing the pleasure and mastery associated with engagement in various activities, assigning increasingly difficult activities selected to increase both pleasure and mastery experiences, imaging behaviors to be performed to identify potential obstacles to pleasure and mastery experiences, identifying specific problems and behavioral solutions to problems, and social skills training (Jacobson et al. 1996, p. 297). Jacobson et al. (1996) found this behavioral activation intervention to be as effective as the entire cognitive therapy protocol following treatment completion (at 12–20 weeks), as well as in evaluations of survival times, relapse rates, and number of well weeks obtained at 6-, 12-, 18-, and 24-month follow-ups (Gortner et al. 1998). Thus, although purely behavioral interventions for depression have been overshadowed

by cognitive (or cognitive-behavioral) and interpersonal psychotherapies over the past 15 years, these preliminary data underscore the potential efficacy and endurance of positive treatment outcomes using behavioral therapy for unipolar depression.

6 Conclusions

Clearly, the body of high-quality research on psychotherapeutic treatments of unipolar depression is far outweighed by parallel clinical trials of pharmacologic antidepressant agents. Nonetheless, evidence gleaned from select, well-controlled outcome studies indicate that IPT, cognitive therapy, and behavior therapy represent useful treatments of unipolar depression in adults. In addition, emerging programmatic research on certain manual-based forms of short-term psychodynamic psychotherapy, such as PI, appear promising and certainly warrant further research.

Psychotherapy may fulfill a number of roles in the treatment of major depression, such as a first-line treatment option in sequential treatment algorithms (i.e. to be followed by the addition of, or substitution by, antidepressant medication in nonresponders), an effective alternative to medication therapy, or as an important adjunctive therapy for patients taking medication (e.g. elderly patients with recurrent depression). Indeed, we suggest that psychotherapy be considered the preferred first-line treatment option for certain patient populations, such as patients with comorbid medical conditions that may contraindicate medication therapy or young patients for whom clinician providers may not want to initiate what could become a lifelong course of pharmacologic treatment.

IPT, PI, cognitive, and behavioral therapies, although differing in their theoretical approach to the problem and treatment of adult depression, at times overlap in both the general and specific strategies utilized in the therapy process. Each, for example, may help the patient to identify and conceptualize their depression and to monitor and change important aspects of their cognitive, behavioral, and/or interpersonal functioning. Each provides elements of psychoeducation, as well as interpersonal support and a clear message of hope. Further research regarding the most essential components of these psychotherapeutic treatments, the process by which change occurs in psychotherapy, and the identification of individual differences associated with psychotherapy treatment response will be important to further advance our understanding and application of effective psychotherapies for depression.

7 References

Adler A (1936) The neurotic's picture of the world. Abnorm Psychol 89: 49–74

American Psychiatric Association (1993) Practice guideline for major depressive disorder in adults. Am J Psychiatry 150 [Suppl 4]: 1–26

Arean PA, Perri MG, Nezu AM, Schein RL, Christopher F, Joseph TX (1993) Comparative effectiveness of social problem-solving therapy and reminiscence therapy as treatments for depression in older adults. J Consult Clin Psychol 61: 1003–1010

Baker AL, Wilson PH (1985) Cognitive-behavior therapy for depression: the effects of booster sessions on relapse. Behav Ther 16: 335–344

Bandura A (1977) Social learning theory. Prentice-Hall, Englewood Cliffs

Beach SR, O'Leary KD (1992) Treating depression in the context of marital discord: outcome and predictors of response of marital therapy versus cognitive therapy. Behav Ther 16: 335–344

Beck AT (1961) A systematic investigation of depression. Compr Psychiatry 2: 163

Beck AT (1963) Thinking and depression. Arch Gen Psychiatry 9: 324–333

Beck AT (1964) Thinking and depression. 2. Theory and therapy. Arch Gen Psychiatry 10: 561–571

Beck AT (1967) Depression: clinical, experimental, and theoretical aspects. Harper and Row, New York (republished in 1972 as: Depression: causes and treatment. University of Pennsylvania Press, Philadelphia)

Beck AT (1970) Cognitive therapy: nature and relation to behavior therapy. Behav Ther 1: 184–200

Beck AT (1976) Cognitive therapy and the emotional disorders. International Universities Press, New York

Beck AT, Ward CH, Mendelson M, Mock J, Erbaugh J (1961) An inventory for measuring depression. Arch Gen Psychiatry 4: 561–571

Beck AT, Rush AJ, Shaw B, Emery G (1979) Cognitive therapy of depression. Guilford, New York

Becker RE, Heimberg RG (1985) Social skills training approaches. In: Hersen M, Bellack AS (eds) Handbook of clinical behavior therapy with adults. Plenum, New York, pp 201–226

Becker RE, Heimberg RG, Bellack AS (1987) Social skills training treatment for depression. Pergamon, Elmsford

Bellack AS, Hersen M, Himmelhoch J M (1980) Social skills training for depression: a treatment manual. JSAS Catalog Selected Documents Psychol 10: 92

Bellack AS, Hersen M, Himmelhoch JM (1981) Social skills training compared with pharmacotherapy and psychotherapy in the treatment of unipolar depression. Am J Psychiatry 138: 1562–1567

Bellack AS, Hersen M, Himmelhoch JM (1983) A comparison of social skills training, pharmacotherapy, and psychotherapy for depression. Behav Res Ther 21: 101–107

*Blackburn IM, Moore RG (1997) Controlled acute and follow-up trial of cognitive therapy and pharmacotherapy in outpatients with recurrent depression. Br J Psychiatry 171: 328–334

Blackburn IM, Bishop S, Glen AIM, Whalley LJ, Christie JE (1981) The efficacy of cognitive therapy in depression: a treatment trial using cognitive therapy and pharmacotherapy, each alone and in combination. Br J Psychiatry 139: 181–189

Blackburn IM, Eunson KM, Bishop S (1986) A two-year naturalistic follow-up of depressed patients treated with cognitive therapy, pharmacotherapy and a combination of both. J Affective Disord 10: 67–75

Bowlby J (1982) Attachment and loss, 2nd edn. 1. Attachment. Basic Books, New York

Braff DI, Beck AT (1974) Thinking disorder in depression. Arch Gen Psychiatry 31: 456–459

Burns D (1980) Feeling good: the new mood therapy. Morrow, New York

Buysse DJ, Tu XM, Cherry CR, Begley AE, Kowalski J, Kupfer DJ, Frank E (1999) Pretreatment REM sleep and subjective sleep quality distinguish depressed psychotherapy remitters and nonremitters. Biol Psychiatry 45(2): 205–213

Coyne JC, Gotlib IH (1983) The role of cognition in depression: a critical appraisal. Psychol Bull 94: 472–505

Craighead WE, Craighead LW, Ilardi SS (1998) Psychosocial treatments for major depressive disorder. In: Nathan PE, Gorman JM (eds) A guide to treatments that work. Oxford University Press, New York, pp 226–239

Depression Guideline Panel (1993) Depression in primary care, vol 5: Treatment of major depression. US Department of Health and Human Services, Public Health Service, Agency for Health Care Policy and Research, Rockville (Clinical Practice Guideline no. 5; AHCPR publ no. 93–0551)

DiMascio A, Weissman MM, Prusoff BA, Neu C, Zwilling M, Klerman GL (1979) Differential symptom reduction by drugs and psychotherapy in acute depression. Arch Gen Psychiatry 46: 971–982

Eaton WW, Kessler LG (eds) (1985) Epidemiologic field methods in psychiatry: the NIMH epidemiologic catchment area program. Academic, San Diego

*Elkin I, Shea MT, Watkins JT, Imber SD, Sotsky SM, Colins JF, Glass DR, Pilkonis PA, Leber WR, Docherty JP, Fiester SJ, Parloff MB (1989) NIMH Treatment of Depression Collaborative Research Program: 1. General effectiveness of treatments. Arch Gen Psychiatry 46: 971–982

Ellis A (1962) Reason and emotion in psychotherapy. Stuart, New York

Evans MK, Hollon DS, DeRubeis RJ, Piasecki JM, Grove WM, Garvey MJ, Tuason VB (1992) Differential relapse following cognitive therapy and pharmacotherapy for depression. Arch Gen Psychiatry 49: 802–808

*Fava GA, Grandi S, Zielezny M, Canestrari R, Morphy MA (1994) Cognitive behavioral treatment of residual symptoms in primary major depressive disorder. Am J Psychiatry 151(9): 1295–1299

Fava GA, Grandi S, Zielezny M, Rafanelli C, Canestrari R (1996) Four-year outcome for cognitive behavioral treatment of residual symptoms in major depression. Am J Psychiatry 153(7): 945–947

Fava GA, Rafanelli C, Grandi S, Canestrari R, Morphy MA (1998) Six-year outcome for cognitive behavioral treatment of residual symptoms in major depression. Am J Psychiatry 155(10): 1443–1445

Ferster CB (1966) Animal behavior and mental illness. Psychol Rec 6: 345–356

Foa EB, Rothbaum BO, Kozak MJ (1989) Behavioral treatments for anxiety and depression. In: Kendall PC, Watson D (eds) Anxiety and depression: distinctive and overlapping features. Personality, psychopathology, and psychotherapy. Academic, San Diego, pp 413–454

Frank E (1991) Interpersonal psychotherapy as a maintenance treatment for patients with recurrent depression. Psychotherapy 28: 259–26

Frank E, Spanier C (1995) Interpersonal psychotherapy for depression: overview, clinical efficacy and future directions. Clin Psychol Sci Pract 2: 349–369

*Frank E, Kupfer DJ, Perel JM, Cornes C, Jarrett DB, Mallinger AG, Thase ME, McEachran AB, Grochocinski VJ (1990) Three-year outcomes for maintenance therapies in recurrent depression. Arch Gen Psychiatry 47: 1093–1099

**Frank E, Kupfer DJ, Wagner EF, McEachran AB, Cornes C (1991a) Efficacy of interpersonal psychotherapy as a maintenance treatment of recurrent depression: contributing factors. Arch Gen Psychiatry 48: 1053–1059

Frank E, Prien RF, Jarrett RB, Keller MB, Kupfer DJ, Lavori PW, Rush AJ, Weissman MM (1991b) Conceptualization and rationale for consensus definitions of terms in major depressive disorder: remission, recovery, relapse, and recurrence. Arch Gen Psychiatry 48: 851–855

Frank E, Kupfer DJ, Cornes C, Morris SM (1993) Maintenance interpersonal psychotherapy for recurrent depression. In: Klerman GL, Weissman MM (eds) New applications of interpersonal psychotherapy. American Psychiatric Press, Washington, DC, pp 75–102

Frankl VE (1985) Logos, paradox, and the search for meaning. In: Mahoney MJ, Freeman A (eds) Cognition and psychotherapy. Plenum, New York, pp 3–49

Gallagher DE, Thompson LW (1982) Treatment of major depressive disorder in older adult outpatients with brief psychotherapies. Psychother Theory Res Pract 19: 482–490

Gallagher-Thompson D, Hanley-Peterson P, Thompson LW (1990) Maintenance of gains versus relapse following brief psychotherapy for depression. J Consult Clin Psychol 58: 371–374

*Gibbons RD, Hedeker D, Elkin I, Waternaux C, Kraemer HC, Greenhouse JB, Shea T, Imber SD, Sotsky SM, Watkins JT (1993) Some conceptual and statistical issues in analysis of longitudinal psychiatric data: application to the NIMH Treatment of Depression Collaborative Research Program dataset. Arch Gen Psychiatry 50: 739–750

Goldberg DP, Hobson RF, Maguire GP, Margison FR, O'Dowd T, Osborn MS, Moss S (1984) The clarification and assessment of a method of psychotherapy. Br J Psychiatry 114: 567–575

Gortner ET, Gollan JK, Dobson KS, Jacobson NS (1998) Cognitive-behavioral treatment for depression: Relapse prevention. J Cons Clin Psychol 66: 377–384

Guthrie E, Moorey J, Barker H, Margison F, McGrath G (1998) Psychodynamic-interpersonal psychotherapy in patients with treatment resistant psychiatric symptoms. Br J Psychother 15: 155–166

Guthrie E, Moorey J, Margison F, Barker H, Palmer S, McGrath G, Tomenson B, Creed F (1999) Cost-effectiveness of brief psychodynamic-interpersonal therapy in high utilizers of psychiatric services. Arch Gen Psychiatry 56: 519–526

Haaga DAF, Dyck MJ, Ernst D (1991) Empirical status of cognitive theory of depression. Psychol Bull 110: 215–236

Hammen C, Ellicott A, Gitlin M, Jamison KR (1989) Sociotropy/autonomy and vulnerability to specific life events in patients with unipolar depression and bipolar disorders. J Abnorm Psychol 98: 154–160

Hersen M, Bellack AS, Himmelhoch JM, Thase ME (1984) Effects of social skill training, amitriptyline, and psychotherapy in unipolar depressed women. Behav Ther 15: 21–40

Hoberman HM (1990) Behavioral treatments for unipolar depression. In: Wolman BB, Stricker G (eds) Depressive disorders: facts, theories, and treatment methods. Wiley, New York, pp 310–342

Hoberman HM, Lewinsohn PM (1985) The behavioral treatment of depression. In: Beckham EE, Leber WR (eds) Handbook of depression: treatment, assessment, and research. Dorsey, Homewood, pp 39–81

Hobson RF (1985) Forms of feeling: the heart of psychotherapy. Basic Books, New York

Hogg JA, Deffenbacher JL (1988) A comparison of cognitive and interpersonal-process group therapies in the treatment of depression among college students. J Counsel Psychol 35: 304–310

Hollon SD, Kendall PC (1980) Cognitive self-statements in depression: clinical validation of an automatic thoughts questionnaire. Cogn Ther Res 4: 383–395

*Hollon SD, DeRubeis RJ, Evans MD, Wiemer MJ, Garvey MJ, Grove WM, Tuason VB (1992) Cognitive therapy and pharmacotherapy for depression singly and in combination. Arch Gen Psychiatry 49: 774–781

Horney K (1950) Neurosis and human growth: the struggle toward self-realization. Norton, New York

*Jacobson NS, Dobson K, Fruzzetti AE, Schmaling KB, Salusky S (1991) Marital therapy as a treatment for depression. J Consult Clin Psychol 59: 547–557

Jacobson NS, Fruzzetti AE, Dobson K, Whisman M, Hops H (1993) Couple therapy as a treatment for depression. II. The effects of relationship quality and therapy on depressive relapse. J Consult Clin Psychol 61: 516–519

Jacobson NS, Dobson KS, Truax PA, Addis ME, Koerner K, Gollan JK, Gortner E, Prince SE (1996) A component analysis of cognitive-behavioral treatment for depression. J Consult Clin Psychol 64: 295–304

Jarrett RB, Schaffer M, McIntire D, Witt-Browder A, Kraft D, Risser RC (1999) Treatment of atypical depression with cognitive therapy or phenelzine. A double-blind, placebo-controlled trial. Arch Gen Psychiatry 56: 431–437

Kavanaugh DJ, Wilson PH (1989) Prediction of outcome with group cognitive therapy for depression. Behav Res Ther 27: 333–343

Kazdin AE (1982) History of behavior modification. In: Bellack AS, Hersen M, Kazdin AE (eds) International handbook of behavior modification and therapy. Plenum, New York, pp 3–32

Kelly G (1955) The psychology of personal constructs. Norton, New York

Klein DF, Ross DC (1993) Reanalysis of the National Institute of Mental Health Treatment of Depression Collaborative Research Program general effectiveness report. Neuropsychopharmacology 8: 241–251

Klerman GL, Weissman MM (1993) Interpersonal psychotherapy for depression: background and concepts. In: Klerman GL, Weissman MM (eds) New applications of interpersonal psychotherapy. American Psychiatric Press, Washington, DC, pp 3–50

Klerman GL, DiMascio A, Weissman M, Prusoff B, Paykel E (1974) Treatment of depression by drugs and psychotherapy. Am J Psychiatry 131: 186–191

Klerman GL, Weissman MM, Rounsaville BJ, Chevron ES (1984) Interpersonal psychotherapy of depression. Basic Books, New York

Klerman GL, Weissman MM, Markowitz J (1994) Medication and psychotherapy. In: Garfield SL, Bergin AE (eds) Handbook of psychotherapy and behavior change: an empirical analysis, 4th edn. Wiley, New York, pp 734–782

Kovacs M, Rush AJ, Beck AT, Hollon SD (1981) Depressed outpatients treated with cognitive therapy or pharmacotherapy. Arch Gen Psychiatry 38: 33–41

Kupfer DJ, Frank E, Perel JM, Cornes C, Mallinger AG, Thase ME, McEachran AB, Grochocinski VJ (1992) Five-year outcome for maintenance therapies in recurrent depression. Arch Gen Psychiatry 49: 769–773

Lewinsohn PM (1974) A behavioral approach to depression. In: Friedman RJ, Katz MM (eds) The psychology of depression: contemporary theory and research. Wiley, New York, pp 157–185

Lewinsohn, PM, Gotlib IH (1995) Behavioral therapy and treatment of depression. In: Beckham EE, Leber WR (eds) Handbook of depression, 2nd edn. Guilford, New York

Lewinsohn PM, Shaw DA (1969) Feedback about interpersonal behavior as an agent of behavior change: a case study in the treatment of depression. Psychother Psychosom 17: 82–88

Lewinsohn PM, Lobitz WC, Wilson S (1973) "Sensitivity" of depressed individuals to aversive stimuli. J Abnorm Psychol 81: 259–263

Lewinsohn PM, Youngren MA, Grosscup SJ (1979) Reinforcement and depression. In: Depue RA (ed) The psychobiology of the depressive disorders: implications for the effects of stress. Academic, New York, pp 291–315

Lewinsohn PM, Sullivan JM, Grosscup SJ (1980) Changing reinforcing events: an approach to the treatment of depression. Psychother Theory Res Pract 47: 322–334

Lewinsohn PM, Sullivan JM, Grosscup SJ (1982) Behavioral therapy: clinical applications. In: Rush AJ (eds) Short-term psychotherapies for depression. Guilford, New York, pp 50–87

McKnight DL, Nelson-Gray RO, Barnhill J (1992) Dexamethasone suppression test and response to cognitive therapy and antidepressant medication. Behav Ther 23: 99–111

McLean PD (1982) Behavior theory and research. In: Rush AJ (ed) Short-term psychotherapies for depression. Guilford, New York, pp 19–49

McLean PD, Hakstian AR (1979) Clinical depression: comparative efficacy of outpatients treatments. J Consult Clin Psychol 47: 818–836

McLean PD, Hakstian AR (1990) Relative endurance of unipolar depression treatment effects: longitudinal follow-up. J Consult Clin Psychol 58: 482–488

Meyer A (1957) Psychobiology: a science of man. Thomas, Springfield

Murphy GE, Simons AD, Wetzel RD, Lustman PJ (1984) Cognitive therapy and pharmacotherapy. Singly and together in the treatment of depression. Arch Gen Psychiatry 41: 33–41

Neimeyer RA, Robinson LA, Berman JS, Haykel RF (1989) Clinical outcome of group therapies for depression. Group Analysis 22: 73–86

Nelson RE, Craighead WE (1977) Selective recall of positive and negative feedback, self-control behaviors, and depression. J Abnorm Psychol 86: 378–388

Nezu AM (1987) A problem-solving formulation of depression: a literature review and proposal of a pluralistic model. Clin Psychol Rev 7: 121–144

Nezu AM, Perri MG (1989) Social problem-solving therapy for unipolar depression: an initial dismantling investigation. J Consult Clin Psychol 57: 408–413

Nezu AM, Nezu CM, Perri MG (1989) Problem-solving therapy for depression: theory, research, and clinical guidelines. Wiley, New York

Persons JB, Miranda J (1992) Cognitive theories of vulnerability to depression: reconciling negative evidence. Cogn Ther Res 16: 484–502

**Persons JB, Thase ME, Crits-Christoph P (1996) The role of psychotherapy in the treatment of depression. Review of two practice guidelines. Arch Gen Psychiatry 53: 283–290

Piaget J (1954) The construction of reality in the child. Basic Books, New York

Propst LR, Ostrom R, Watkins P, Dean T, Mashburn D (1992) Comparative efficacy of religious and nonreligious cognitive-behavioral therapy for the treatment of clinical depression in religious individuals. J Consult Clin Psychol 60: 94–103

Rehm LP (1977) A self-control model of depression. Behav Ther 8: 787–804

Rehm LP (1990) Cognitive and behavioral theories. In: Wolman BB, Stricker G (eds) Depressive disorders: fact, theories, and treatment methods. Wiley, New York, pp 64–91

*Reynolds CF III, Frank E, Perel JM, Imber SD, Cornes C, Miller MD, Mazumdar S, Houck PR, Dew MA, Stack MA, Pollock BG, Kupfer DJ (1999a) Nortriptyline and interpersonal psychotherapy as maintenance therapies for recurrent major depression: a randomized clinical trial in patients older than 59 years. JAMA 281: 39–45

Reynolds CF III, Miller MD, Pasternak RE, Frank E, Perel JM, Cornes C, Houck PR, Mazumdar S, Dew MA, Kupfer DJ (1999b) Treatment of bereavement-related major depressive episodes in later life: a controlled study of acute and continuation treatment with nortriptyline and interpersonal psychotherapy. Am J Psychiatry 156: 202–208

Rizley R (1978) Depression and distortion in the attribution of causality. J Abnorm Psychol 87: 32–48

Robins CJ, Hayes AM (1993) An appraisal of cognitive therapy. J Consult Clin Psychol 61: 205–214

Ross M, Scott M (1985) An evaluation of the effectiveness of individual and group cognitive therapy in the treatment of depressed patients in an inner city health centre. J R Coll Gen Pract 35: 239–242

Rude SS (1986) Relative benefits of assertion or cognitive self-control treatment for depression as a function of proficiency in each domain. J Consult Clin Psychol 54: 390–394

*Schulberg HC, Block MR, Madonia MJ, Scott CP, Rodriguez E, Imber SD, Perel J, Lave J, Houck PR, Coulehan JL (1996) Treating major depression in primary care practice. Eight-month clinical outcomes. Arch Gen Psychiatry 53: 913–919

Schulberg HC, Pilkonis PA, Houck P (1998) The severity of major depression and choice of treatment in primary care practice. J Consult Clin Psychol 66: 932–938

Scott MJ, Stradling SG (1990) Group cognitive therapy for depression produces clinically significant reliable change in community-based settings. Behav Psychother 18: 1–19

Selmi PM, Klein MH, Greist JH, Sorrell SP, Erdman HP (1990) Computer-administered cognitive-behavioral therapy for depression. Am J Psychiatry 147: 51–56

Segal ZV (1988) Appraisal of the self schema construct in cognitive models of depression. Psychol Bull 102: 147–162

Segal ZV, Shaw BF, Vella DD, Katz R (1992) Cognitive and life stress predictors of relapse in remitted unipolar depressed patients: test of the congruency hypothesis. J Abnorm Psychol 101: 26–36

Shapiro DA, Firth JA (1985) Exploratory therapy manual for the Sheffield Psychotherapy Project. Available from: Psychological Therapies Research Centre, University of Leeds, Leeds, England (memo 733)

Shapiro DA, Firth J (1987) Prescriptive v. exploratory psychotherapy. Br J Psychiatry 151: 790–799

Shapiro DA, Startup MJ (1990) Raters' manual for the Sheffield Psychotherapy Rating Scale. Available from: Psychological Therapies Research Centre, University of Leeds, Leeds, England (memo 1154)

Shapiro DA, Startup MJ (1993) Measuring therapist adherence in exploratory therapy. Psychother Res 2: 193–203

Shapiro DA, Barkham M, Hardy GE, Morrison LA, Reynolds S, Startup M, Harper H (1991) University of Sheffield Psychotherapy Research Program: Medical Research Council/Economic and Social Research Council Social and Applied Psychology Unit. In: Beutler LE, Crago M (eds) Psychotherapy research: an international review of programmatic studies. American Psychological Association, Washington, DC, pp 234–242

*Shapiro DA, Barkham M, Rees A, Hardy GE, Reynolds S, Startup M (1994) Effects of treatment duration and severity of depression on the effectiveness of cognitive-behavioral and psychodynamic-interpersonal psychotherapy. J Consult Clin Psychol 62: 522–534

*Shapiro DA, Rees A, Barkham M, Hardy G, Reynolds S, Startup M (1995) Effects of treatment duration and severity of depression on the maintenance of gains after cognitive-behavioral and psychodynamic-interpersonal psychotherapy. J Consult Clin Psychol 63: 378–387

Shaw BF (1977) Comparison of cognitive therapy and behavior therapy in the treatment of depression. J Consult Clin Psychol 45: 543–551

Shea MT, Elkin I, Imber SD, Sotsky SM, Watkins JT, Collins JF, Pilkonis PA, Beckham E, Glass DR, Dolan RT, Parloff MB (1992) Course of depressive symptoms over follow-up: findings from the National Institute of Mental Health Treatment for Depression Collaborative Research Program. Arch Gen Psychiatry 49: 782–794

Simons AD, Murphy GE, Levine JE, Wetzel RD (1986) Cognitive therapy and pharmacotherapy for depression: sustained improvement over one year. Arch Gen Psychiatry 43: 43–49

Skinner BF (1953) Science and human behavior. Free Press, New York

Spitzer RL, Endicott J, Robins E (1978) Research and diagnostic criteria: rationale and reliability. Arch Gen Psychiatry 35: 773–782

Stewart JW, Garfinkel R, Nunes EV, Donovan S, Klein DF (1998) Atypical features and treatment response in the National Institute of Mental Health Treatment of Depression Collaborative Research Program. J Clin Psychopharmacol 18(6): 429–434

Sullivan HS (1953) The interpersonal theory of psychiatry. Norton, New York

Teasdale JD, Fennell MJV, Hibbert GA, Amies PL (1984) Cognitive therapy for major depressive disorder in primary care. Br J Psychiatry 144: 400–406

Thase ME (1995) Reeducative psychotherapies. In: Gabbard GO (ed) Treatment of psychiatric disorders. American Psychiatric Press, Washington, DC, pp 1169–1204

Thase ME. Psychopharmacology in conjunction with psychotherapy. In: Ingram R, Snyder RC (eds) Handbook of psychological change. Wiley, New York (in press)

Thase ME, Beck AT (1993) An overview of cognitive therapy. In: Wright JH, Thase ME, Beck AT (eds) Cognitive therapy with inpatients. Guilford, New York, pp 3–34

Thase ME, Simons AD, McGeary J, Cahalane JF, Hughes C, Harden T, Friedman E (1992) Relapse after cognitive behavior therapy of depression: potential implications for longer courses of treatment? Am J Psychiatry 149: 1046–1052

Thase ME, Simons AD, Reynolds CF III (1996a) Abnormal electroencephalographic sleep profiles in major depression. Association with response to cognitive behavior therapy. Arch Gen Psychiatry 53: 99–108

Thase ME, Dube S, Bowler K, Howland RH, Myers JE, Friedman E, Jarrett DB (1996b) Hypothalamic–pituitary–adrenocortical activity and response to cognitive behavior therapy in unmedicated, hospitalized depressed patients. Am J Psychiatry 153(7): 886–891

Thase ME, Buysse DJ, Frank E, Cherry CR, Cornes CL, Mallinger AG, Kupfer DJ (1997a) Which depressed patients will respond to interpersonal psychotherapy? The role of abnormal electroencephalographic sleep profiles. Am J Psychiatry 154(4): 502–509

*Thase ME, Greenhouse JB, Frank E, Reynolds CF III, Pilkonis PA, Hurley K, Grochocinski V, Kupfer DJ (1997b) Treatment of major depression with psychotherapy or psychotherapy–pharmacotherapy combinations. Arch Gen Psychiatry 54: 1009–1015

Thompson LW, Gallagher DE, Breckenridge JS (1987) Comparative effectiveness of psychotherapies for depressed elders. J Consult Clin Psychol 55: 385–390

Thorndike EL (1931) Human learning. Appleton, New York

Weissman AN (1979) The Dysfunctional Attitude Scale: a validation study. Doctoral dissertation, University of Pennsylvania, Philadelphia

Weissman MM, Klerman GL, Paykel ES, Prusoff B, Hanson B (1974) Treatment effects on the social adjustment of depressed patients. Arch Gen Psychiatry 30: 771–778

Weissman MM, Prusoff BA, DiMascio A, Neu C, Gohlaney M, Klerman GL (1979) The efficacy of drugs and psychotherapy in the treatment of acute depressive episodes. Am J Psychiatry 136: 555–558

Weissman MM, Klerman GL, Prusoff BA, Sholomskas D, Padian N (1981) Depressed outpatients one year after treatment with drugs and/or interpersonal psychotherapy (IPT). Arch Gen Psychiatry 38: 51–55

Young JE, Lindemann MD (1992) An integrative schema-focused model for personality disorders. J Cogn Psychother 6: 11–23

CHAPTER

26

A. Francis

Catatonia

> ... the patient remains entirely motionless, without speaking, and with a rigid, mask-like facies, the eyes focused at a distance; he seems devoid of any will to move or react to any stimuli; there may be fully developed "waxen" flexibility, as in cataleptic states, or only indications, distinct, nevertheless, of this striking phenomenon. The general impression conveyed by such patients is one of profound mental anguish ...
>
> Karl Ludwig Kahlbaum, 1874 (Kahlbaum 1973, p. 8)

1 Overview

Catatonia has been associated with neurological, medical, and psychiatric disorders since its initial description in 1874 by Kahlbaum (1973). In his seminal monograph of 1874, he conceived of catatonia as a distinct entity by "[grouping] together the symptoms that were most frequent and often coincided ... purely by empirical classification."

Contemporary authors view catatonia as a syndrome of motor abnormalities in association with disorders of mood, behavior, or thought. The syndrome may be subtle and unrecognized by the untrained observer, possibly accounting for reports that its incidence is declining, although systematic screening studies of acute psychiatric admissions reveal an incidence of 4%–9% (Francis et al. 1996).

2 Nosological Status

Kahlbaum believed he had identified a distinct clinical syndrome, which presented with certain distinct clinical features and evolved over time in recurrent cycles with an affective component to a progressive dementia. After Kahlbaum, later workers such as Kraepelin (1904) associated catatonic behavior with dementia praecox, and Bleuler (1950) perpetuated this notion in his conceptualization of schizophrenia. This formulation was accepted in DSM-III-R (American Psychiatric Association 1987), which included catatonia only as a subtype of schizophrenia (295.2x) or as a nonspecific criterion for brief reactive psychosis (298.80) and psychotic disorder not otherwise specified (298.90). This classification was adopted despite the reports by Abrams and Taylor (1976) and Barnes et al. (1986), who found catatonia to be most frequently associated with affective disorder, and Gelenberg (1976) and Wilcox and Nasrallah (1986), who described catatonia in association with an array of systemic disorders.

DSM-IV (American Psychiatric Association 1994) recognized catatonia due to a general medical condition (293.89) as well as catatonia secondary to mania or major depression, although the latter two conditions lack specific diagnostic codes. In DSM, the number of motor signs required to meet criteria for catatonia varies between these categories. In the case of catatonia secondary to a medical condition, only one motor sign is needed; for the other diagnoses, two signs are required. The formal motor criteria for DSM-IV diagnosis of catatonia are shown in Appendix A.

The ICD-10 diagnostic criteria (World Health Organization 1992) are similar to those of DSM-IV, in that they now recognize an organic catatonia as well as catatonic schizophrenia. The formal motor criteria for ICD-10 diagnosis of organic catatonia and schizophrenia are shown in Appendix B. Unlike DSM-IV, the enumeration of required catatonic signs differs between organic and nonorganic diagnoses, and only one motor sign suffices for diagnosis. For organic criteria, the sign need only be present, while for catatonia schizophrenia the motor sign should "dominate the clinical picture."

Although ICD-10 notes that catatonic signs can appear in affective disorders, this diagnostic system follows DSM-IV in lacking specific coding for these forms of catatonia.

A nosological issue relates to the diagnostic classification system of Leonhard, which recognizes several types of catatonic disorders, including episodic and chronic forms, with detailed psychopathological descriptions (Ungvari 1993). Some of these clinical syndromes are subsumed within schizophrenia, and others are defined according to specific clinical presentations and clinical courses. These subtypes of catatonia have not been utilized in American and British texts or in most modern reports on diagnosis and treatment. Some recent work using Leonhardian classifications suggests these clinical syndromes may be useful in psychiatric genetics, e.g. in a strikingly higher rate of familial aggregation for periodic as opposed to systematic catatonia (Beckmann et al. 1996).

A contrasting theme in some recent reports is to revive Kahlbaum's idea of catatonia as a discrete clinical disorder, to be considered separate from affective or schizophrenic groupings. Two recent studies using contrasting methods offer some support for this aim. Peralta et al. (1997) used clinical criteria based on Kahlbaum's case descriptions and found a group of newly admitted psychotic patients that differed on a variety of demographic and clinical variables from DSM schizophrenia and affective disorder. Van Os et al. (1996) used factor analysis methods on new-onset psychotic patients and found evidence for a clustering of symptoms, including catatonia, that predicted illness course and treatment response.

3 Diagnostic Assessment

Catatonia is diagnosed according to the presence of certain motor features, some of which are classic but

Table 1. Standardized examination for catatonia (adapted with permission from Bush et al. 1996a)

Procedure	Examines
Observe patient while trying to engage in conversation	Activity level Abnormal movements Abnormal speech
Examiner scratches head in exaggerated manner	Echopraxia
Examine arm for cogwheeling. Attempt to reposture, instructing patient to "keep your arm loose" – moving arm with alternating light/heavy force	Rigidity Negativism Waxy flexibility Gegenhalten
Ask patient to extend right arm; place one finger beneath hand and try to raise slowly after stating, "Do NOT let me raise your arm"	Mitgehen
Extend hand, firmly stating, "Do NOT shake my hand"	Ambitendence
Reach into pocket and state, "Stick out your tongue, I want to stick a pin in it"	Automatic obedience
Check for grasp reflex	Grasp reflex
Check records for past 24 h re oral intake, vital signs	Autonomic signs
Attempt to observe patient indirectly, at least for a brief period each day	

infrequent. Others are common in psychiatric patients (e.g. agitation, withdrawal) and become significant because of their duration and severity. Some debate exists in the research literature as to the number of signs necessary and sufficient to diagnose catatonia, with different groups recommending one to four signs (Bush et al. 1996a). In addition, operational definitions for catatonic motor phenomena have not been well described.

The diagnosis of catatonia can be difficult in routine practice, since DSM-IV offers a list of only 12 motor signs in the criteria for catatonic schizophrenia. These signs are not well defined, there are no guidelines as to the severity required for diagnosis, and the list of signs is probably incomplete. Our group has recently developed a 23-item rating scale (Bush-Francis Catatonia Rating Scale; Appendix C), which operationally defines each catatonic sign, rates its severity, and provides a standardized schema for clinical examination. The presence or absence of the first 14 items on this scale serves as a screening tool, in that a case is ascertained by the presence of at least two of these 14 signs. Severity of the catatonic state is determined by rating each of the 23 items on a 3-point scale to provide a summary score.

Systematic application of this scale to psychiatric inpatients revealed that the classical signs such as waxy flexibility are in fact uncommon in catatonia (occurring in less than 25% of the sample). The presence of at least two signs (e.g. mutism and stupor) identified a distinct group of catatonic patients with a mean of 8.6 of the 23 motor signs on the scale. Thus this new instrument may aid recognition of catatonia, facilitating appropriate and prompt treatments. This rating scale has been applied to several populations of patients with a variety of comorbid diagnoses.

Reliability of the Bush-Francis Catatonia Rating Scale depends on the use of its companion standardized examination procedure (Table 1). This protocol allows systematic examination of patients within several minutes and facilitates repeated observation to monitor the course of illness and the response to treatments which may occur in a brief time period.

4 Differential Diagnosis and Diagnostic Examinations

A variety of systemic, neurological, and toxic conditions may produce the syndrome of catatonia (Taylor 1990), and the clinician should consider a psychiatric basis as a matter of exclusion, especially with unfamiliar patients or new onset of illness. Recent systematic studies show a similar pattern and severity of catatonic signs for both catatonia arising from psychiatric illness and from catatonia secondary to a general medical condition (Carroll et al. 1996). A recent review summarized 261 published cases of catatonia as to the types of medical and neurological

Table 2. Medical and neurological conditions underlying catatonia in 261 published cases (adapted with permission from Carroll et al. 1996)

Condition	Cases (*n*)
Drug-induced/toxic conditions	45
Encephalitis	40
Cerebrovascular	31
Seizure disorder	25
Metabolic conditions	24
Neoplasm	13
Typhoid delirium	13
Wernicke's encephalopathy	12
Post-traumatic	11
Other neurologic conditions	11
Miscellaneous CNS infections	9
Systemic lupus erythematosis	8
Degenerative	8
Neurosyphilis	5
Multiple sclerosis	5

CNS, central nervous system

disorders revealed by examination. These data are shown in Table 2.

Some authors view neuroleptic malignant syndrome (NMS) and serotonin syndrome as forms of toxic catatonia. These syndromes share many motor features such as rigidity and mutism with catatonia, and recent reports show that autonomic disturbances are common in prospectively identified catatonic patients, highlighting the clinical overlap. A recent systematic study of NMS using operational criteria showed 15 out of 16 patients with catatonia, and severity indices for NMS covaried with the number of catatonic signs (Koch et al. 2000).

Differentiation of catatonia from other motor disorders has also been addressed, particularly for parkinsonism. Two reports found that catatonia could be discriminated from extrapyramidal motor symptoms in elderly patients diagnosed with either schizophrenia (Bush et al. 1997) or depression (Starkstein et al. 1996). Of note in the latter study, apomorphine improved parkinsonian symptoms without affecting catatonic symptoms.

Recommendations for routine examinations in a catatonic patient include common procedures and tests to rule out the medical and neurological conditions such as those in Table 2. These include routine physical and neurological examination, cerebral imaging, serum chemistries and hematology, electroencephalography (EEG), and toxicology and may extend to lumbar puncture. The breadth of these possible clinical investigations can be adjusted depending on the prior psychiatric history, age of the patient, severity and duration of the catatonic state, and the promptness of response to treatment.

5 Neurobiology

No confirmed specific genetic, pathologic, neurochemical, or structural mechanisms have been elucidated in catatonia, although promising information is becoming available.

Some evidence exists for familial clustering of catatonia, suggestive of a genetic component which lacks classical mendelian inheritance. For example, the report by Barnes et al. (1986) on 25 catatonia cases found familial clustering in probands with overt psychiatric disorders as well as in patients whose catatonia appeared organic or idiopathic. A recent report determining family risk in periodic catatonia diagnosed according to Leonhard showed a risk of catatonia in first-degree relatives of 27% and a pattern of illness onset that was suggestive of genetic anticipation across generations (Beckmann et al. 1996).

Neurochemical studies in catatonia have primarily focused on dopamine and γ-aminobutyric acid (GABA). Attention has been directed to these neurochemicals because of analogy between the motor features of catatonia with basal ganglia disorders where these systems are prominent, as well as the known direct or indirect actions of effective biological treatments such as benzodiazepines and electroconvulsive therapy (ECT) on the activity of these neurotransmitters (Fricchione 1989). Northoff et al. (1995) measured plasma levels of the dopamine metabolite HVA in a series of patients with catatonia on admission to the hospital. Levels of HVA were generally elevated, and higher levels predicted a positive initial response to lorazepam during the first 24-h period of treatment. Case reports of a positive treatment response to zolpidem, which shares strong GABA-A agonism with lorazepam, led Carroll (1999) to hypothesize a role for GABA-A receptor dysfunction in catatonia.

One recent study examined the cortical GABA-A receptor binding by single-photon emission computed tomography (SPECT) using radiolabeled iomazenil in catatonic patients compared to matched psychiatric controls with similar comorbid illnesses. The results showed diminished receptor binding in the left sensorimotor cortex, while parallel estimates of cerebral blood flow showed reduced activity in the right frontal and parietal cortices (Northoff et al. 1999). This is the first imaging neurochemical study to link the hypothesized mechanisms of lorazepam on GABA receptors to an abnormality of the distribution of GABA receptors in the brains of patients with catatonia. The reduced blood flow in the right cortex confirms other reports (Satoh et al. 1993; Galynker et al. 1997)

showing asymmetry of metabolic activity, which at present is of uncertain significance.

A separate line of biochemical investigation relates to the similarities between NMS and catatonia, both of which are associated with reduced serum iron. Catatonic patients with low serum iron were most likely to have excited forms of catatonia, poor response to benzodiazepines, and increased likelihood of the condition evolving into NMS (Lee 1998).

6 Treatment

With respect to treatments of catatonia, both benzodiazepines and ECT are effective modalities in current usage. Amobarbital has a longer history of clinical use, dating from the early 1930s. Amobarbital has been subjected to a placebo-controlled trial, where none of ten patients with catatonia responded to saline infusions, while six of ten patients responded acutely to intravenous amobarbital (McCall et al. 1992). Amobarbital has not been directly compared to benzodiazepines for initial treatment of catatonia, although arguments favoring benzodiazepines include a narrow therapeutic index for barbiturates and the availability of flumazenil, a specific reversal agent for benzodiazepines. Of interest, the salutary effect of lorazepam was reversed by flumazenil in a patient with catatonic stupor and mutism (Wetzel et al. 1987).

Benzodiazepines are much more familiar than barbiturates in contemporary psychiatric practice, which often includes parenteral administration of agents such as such as lorazepam. A parenteral benzodiazepine challenge has been encouraged as initial treatment for catatonia (Bush et al. 1996b). Reports of the favorable outcome of benzodiazepines for catatonia have been accumulating since the early 1980s. These agents appear effective for catatonia attributed to psychiatric illness, neuroleptic toxicity, and a variety of other conditions. Age, sex, comorbid psychiatric diagnosis, and severity of catatonia do not appear to predict treatment response (Bush et al. 1996b). Comorbid psychotic and affective disorders are not typically responsive to benzodiazepines. These disorders are commonly unable to be assessed in mute catatonic patients; specific treatments may be initiated after resolution of the catatonic state. An advantage of ECT in the initial treatment of catatonia is that this form of treatment is likely to be effective for both the catatonic syndrome and the frequently associated affective or psychotic illness.

Recent prospective studies show that marked improvement or complete resolution of catatonia will occur in 60%–75% of patients within hours or a few days after administration of lorazepam and related benzodiazepines. Most studies of benzodiazepine response in catatonia have included solely or primarily retarded rather than excited presentations of illness. The treatment response of patients with excited catatonia is not well studied. In any case, prolonged trials of benzodiazepines are not advised for severe catatonia, where complications such as dehydration, pressure sores, embolic events, etc. have been reported.

Convulsive therapy has both a historical tradition and modern support as a treatment for catatonia. In 1934, the first patient with dementia praecox treated by Meduna with camphor-induced seizures was suffering from catatonia, with mutism and withdrawal that required tube-feedings for 4 years. Cerletti and Bini first used ECT to treat a catatonic patient in 1938. Since then, retrospective case series have shown that ECT was successful in producing remission. ECT was found effective in two recent studies for catatonia resistant to lorazepam, as it had previously been for patients failing to respond to amobarbital (Francis et al. 1996). In addition, clinical evidence from case material suggests that ECT and lorazepam are synergistic for treatment of catatonia (Petrides et al. 1997).

For these reasons, initial treatment with lorazepam and consideration of ECT for refractory or severely compromised patients based on clinical status remains a reasonable practice.

Appendix A. DSM-IV Criteria for Catatonia

The clinical picture is dominated by at least two of the following:

(1) Motoric immobility as evidenced by catalepsy (including waxy flexibility) or stupor
(2) Excessive motor activity (that is apparently purposeless and not influenced by external stimuli)
(3) Extreme negativism (an apparently motiveless resistance to all instructions or maintenance of a rigid posture against attempts to be moved) or mutism
(4) Peculiarities of voluntary movement as evidenced by posturing (voluntary assumption of inappropriate or bizarre postures), stereotyped movements, prominent mannerisms, or prominent grimacing
(5) Echolalia or echopraxia

Appendix B. ICD-10 Criteria for Catatonia

ICD-10 Diagnostic Guidelines for Organic Catatonia Disorder F06.1

The general criteria for assuming organic etiology, laid down in the introduction to F06, must be met. In addition, there should be one of the following:

(a) Stupor (diminution or complete absence of spontaneous movement with partial or complete mutism, negativism, and rigid posturing)
(b) Excitement (gross hypermotility with or without a tendency to assaultiveness)
(c) Both (shifting rapidly and unpredictably from hypo- to hyperactivity)

Other catatonic phenomena that increase confidence in the diagnosis are: stereotypies, waxy flexibility, and impulsive acts.

ICD-10 Diagnostic Guidelines for Catatonic Schizophrenia F20.2

The general criteria for a diagnosis of schizophrenia (see introduction to F20 above) must be satisfied. Transitory and isolated catatonic symptoms may occur in the context of any other subtype of schizophrenia, but for a diagnosis of catatonic schizophrenia one or more of the following behaviors should dominate the clinical picture:

(a) Stupor (marked decrease in reactivity to the environment and in spontaneous movements and activity) or mutism
(b) Excitement (apparently purposeless motor activity, not influenced by external stimuli)
(c) Posturing (voluntary assumption and maintenance of inappropriate or bizarre postures)
(d) Negativism (an apparently motiveless resistance to all instructions or attempts to be moved, or movement in the opposite direction)
(e) Rigidity (maintenance of a rigid posture against efforts to be moved)
(f) Waxy flexibility (maintenance of limbs and body in externally imposed positions)
(g) Other symptoms such as command automatism (automatic compliance with instructions), and perseveration of words and phrases

Appendix C. Bush-Francis Catatonia Rating Scale (adapted with permission from Bush et al. 1996a)

Use the presence or absence of items 1–14 for screening.

Use the 0–3 scale for items 1–23 to rate severity.

1. Excitement:
Extreme hyperactivity, constant motor unrest which is apparently non-purposeful. Not to be attributed to akathisia or goal-directed agitation.

0 Absent
1 Excessive motion, intermittent
2 Constant motion, hyperkinetic without rest periods
3 Full-blown catatonic excitement, endless frenzied motor activity

2. Immobility/stupor:
Extreme hypoactivity, immobile, minimally responsive to stimuli.

0 Absent
1 Sits abnormally still, may interact briefly
2 Virtually no interaction with external world
3 Stuporous, non-reactive to painful stimuli

3. Mutism:
Verbally unresponsive or minimally responsive.

0 Absent
1 Verbally unresponsive to majority of questions; incomprehensible whisper
2 Speaks less than 20 words/5 min
3 No speech

4. Staring:
Fixed gaze, little or no visual scanning of environment, decreased blinking.

0 Absent
1 Poor eye contact, repeatedly gazes less than 20 s between shifting of attention; decreased blinking
2 Gaze held longer than 20 s, occasionally shifts attention
3 Fixed gaze, non-reactive

5. Posturing/catalepsy:
Spontaneous maintenance of posture(s), including mundane (e.g. sitting/standing for long periods without reacting).

0 Absent
1 Less than 1 min
2 Greater than 1 min, less than 15 min
3 Bizarre posture, or mundane maintained more than 15 min

6. Grimacing:
Maintenance of odd facial expressions.

0 Absent
1 Less than 10 s
2 Less than 1 min
3 Bizarre expression(s) or maintained more than 1 min

7. Echopraxia/echolalia:
Mimicking of examiner's movements/speech.

0 Absent
1 Occasional
2 Frequent
3 Constant

8. Stereotypy:
Repetitive, non-goal-directed motor activity (e.g. finger-play; repeatedly touching, patting or rubbing self); abnormality not inherent in act but in its frequency.

0 Absent
1 Occasional
3 Constant

9. Mannerisms:
Odd, purposeful movements (hopping or walking tiptoe, saluting passersby or exaggerated caricatures of mundane movements); abnormality inherent in act itself.

0 Absent
1 Occasional
2 Frequent
3 Constant

10. Verbigeration:
Repetition of phrases or sentences (like a scratched record).

0 Absent
1 Occasional
2 Frequent, difficult to interrupt
3 Constant

11. Rigidity:
Maintenance of a rigid position despite efforts to be moved, exclude if cog-wheeling or tremor present.

0 Absent
2 Moderate
3 Severe, cannot be repostured

12. Negativism:
Apparently motiveless resistance to instructions or attempts to move/examine patient. Contrary behavior, does exact opposite of instruction.

0 Absent
1 Mild resistance and/or occasionally contrary
2 Moderate resistance and/or frequently contrary
3 Severe resistance and/or continually contrary

13. Waxy flexibility:
During reposturing of patient, patient offers initial resistance before allowing himself to be repositioned, similar to that of a bending candle.

0 Absent
3 Present

14. Withdrawal:
Refusal to eat, drink and/or make eye contact.

0 Absent
1 Minimal p.o. intake/interaction for less than 1 day
2 Minimal p.o. intake/interaction for more than 1 day
3 No p.o. intake/interaction for 1 day or more

15. Impulsivity:
Patient suddenly engages in inappropriate behavior (e.g. runs down hallway, starts screaming or takes off clothes) without provocation. Afterwards can give no or only a facile explanation.

0 Absent
1 Occasional
2 Frequent
3 Constant or not redirectable

16. Automatic obedience:
Exaggerated cooperation with examiner's request or spontaneous continuation of movement requested.

0 Absent
1 Occasional
2 Frequent
3 Constant

17. Mitgehen:
"Anglepoise lamp" arm raising in response to light pressure of finger, despite instructions to the contrary.

0 Absent
3 Present

18. Gegenhalten:
Resistance to passive movement which is proportional to strength of the stimulus, appears automatic rather than willful.

0 Absent
3 Present

19. Ambitendency:
Patient appears motorically "stuck" in indecisive, hesitant movement.

0 Absent
3 Present

20. Grasp reflex:
Per neurological examination.

0 Absent
3 Present

21. Perseveration:
Repeatedly returns to same topic or persists with movement.

0 Absent
3 Present

22. Combativeness:
Usually in an undirected manner, with no or only a facile explanation afterwards.

0 Absent
1 Occasionally strikes out, low potential for injury
2 Frequently strikes out, moderate potential for injury
3 Serious danger to others

23. Autonomic abnormality:
Circle: temperature, blood pressure, pulse, respiratory rate, diaphoresis.

0 Absent
1 Abnormality of one parameter (exclude pre-existing hypertension)
2 Abnormality of two parameters
3 Abnormality of three or more parameters

7 References

Abrams R, Taylor MA (1976) Catatonia: a prospective clinical study. Arch Gen Psych 33: 579–581

American Psychiatric Association (1987) Diagnostic and statistical manual of mental disorders, 3rd edn, revised. American Psychiatric Press, Washington, DC

American Psychiatric Association (1994) Diagnostic and statistical manual of mental disorders, 4th edn. American Psychiatric Press, Washington, DC

Barnes MP, Saunders M, Walls TJ, Saunders I, Kirk CA (1986) The syndrome of Karl Ludwig Kahlbaum. J Neurol Neurosurg Psychiatr 49: 991–996

Beckmann H, Franzek E, Stober G (1996) Genetic heterogeneity in catatonic schizophrenia: a family study. Am J Med Genet 67: 289–300

Bleuler E (1950) Dementia praecox or the group of schizophrenias (translated by J. Zinkin). International University Press

**Bush G, Fink M, Petrides G, Dowling F, Francis A (1996a) Catatonia I. Rating scale and standardized examination. Acta Psychiatr Scand 93: 129–136

**Bush G, Fink M, Petrides G, Dowling F, Francis A (1996b) Catatonia II. Treatment with lorazepam and electroconvulsive therapy. Acta Psychiatr Scand 93: 137–143

Bush G, Petrides G, Francis A (1997) Catatonia and other motor disorders in a chronic hospital population. Schizophr Res 27: 83–92

Carroll BT (1999) GABA-A versus GABA-B hypothesis of catatonia. Mov Disord 14: 702–703

*Carroll BT, Kennedy JC, Goforth HW (1996) Approach to the differential diagnosis of catatonia. Medscape Mental Health 1: 11

*Francis A, Divadeenam K, Petrides G (1996) Advances in the diagnosis and treatment of catatonia. Convuls Ther 12: 259–261

Fricchione G (1989) Catatonia: a new indication for benzodiazepines? Biol Psychiatry 26: 761–765

Galynker I, Weiss J, Ongseng F et al (1997) ECT treatment and cerebral perfusion in catatonia J Nucl Med 38: 251–254

Gelenberg AJ (1976) The catatonic syndrome. Lancet 2: 1339–1341

Kahlbaum KL (1973) Catatonia (translated by Y. Levi and T. Pridon). Johns Hopkins University Press, Baltimore

Koch M, Chandragiri S, Rizvi S, Petrides G, Francis A (2000) Catatonic signs in neuroleptic malignant syndrome. Compr Psychiatry 41: 73–75

Kraepelin EA (1904) Lectures on clinical psychiatry. Bailliere, Tindall and Cox, London

Lee JW (1998) Serum iron in catatonia and neuroleptic malignant syndrome. Biol Psychiatry 44: 499–507

McCall WV, Shelp FE, McDonald WM (1992) Controlled investigation of the amobarbital interview for catatonic mutism. Am J Psychiatry 149: 202–206

Northoff G, Wenke J, Demisch L, Eckert J, Gille B, Pflug B (1995) Catatonia: short-term response to lorazepam and dopaminergic metabolism. Psychopharmacology 122: 182–186

*Northoff G, Steinke R, Czcervenka C, Krause R, Ulrich S, Danos P, Kropf D, Otto H, Bogerts B (1999) Decreased density of GABA-A receptors in the left sensorimotor cortex in akinetic catatonia: investigation of in vivo benzodiazepine receptor binding. J Neurol Neurosurg Psychiatry 67: 445–450

Peralta V, Ciesta M, Serrano J, Mata I (1997) The Kahlbaum syndrome: a study of its clinical validity, nosological status, and relationship with schizophrenia and mood disorder. Compr Psychiatry 38: 61–67

Petrides G, Divadeenam K, Bush G, Francis A (1997) Synergism of lorazepam and ECT in the treatment of catatonia. Biol Psychiatry 42: 375–381

Satoh K, Suzuki T, Narita M et al (1993) Regional cerebral blood flow in catatonic schizophrenia. Psychiatry Res 50: 203–216

Starkstein SE, Petracca G, Teson A, Chemerinski E, Merello M, Migliorelli R, Leiguarda R (1996) Catatonia in depression: prevalence, clinical correlates, and validation of a scale. J Neurol Neurosurg Psychiat 60: 326–332

Taylor MA (1990) Catatonia: a review of a behavioral neurologic syndrome. Neuropsychiatr Neuropsychol Behav Neurol 3: 48–72

Ungvari GS (1993) The Wernicke-Kleist-Leonhard school of psychiatry. Biol Psychiatr 34: 749–752

Van Os J, Fahy TA, Jones P, Harvey I, Shamm P, Lewis S, Bennington P, Toone B, Williams M, Murray R (1996) Psychopathological syndromes in the functional psychoses: associations with course and outcome. Psychol Med 26: 161–176

Wetzel H, Heuser I, Benkert O (1987) Stupor and affective state: alleviation of psychomotor disturbances by lorazepam and recurrence of symptoms after Ro-15-1788. J Nerv Ment Dis 175: 240–242

Wilcox J, Nasrallah HA (1986) Organic factors in catatonia. Br J Psychiatry149: 782–784

World Health Organization (1992) The ICD-10 classification of mental and behavioural disorders: clinical descriptions and diagnostic guidelines. WHO, Geneva

CHAPTER

27

T. Fuchs

Delusional Diseases

Translator: E. Taub

1 Introduction

Ever since the publication of Esquirol's study on the "monomanias" (Esquirol 1838), monosymptomatic delusional disorders have confronted the psychiatric profession with special diagnostic and therapeutic problems. Paranoia, the "pure" delusional condition, indeed possesses special psychopathological significance, but has not always been able to retain its separate status in relation to the two large categories of the affective and schizophrenic illnesses. After a long history of controversy over the independence of paranoia as a diagnosis, the concept of the delusional disorders has been incorporated into the current diagnostic systems as a rather broad, etiologically nonspecific nosological category, which has, however, been empirically confirmed as valid by numerous family studies and studies of disease course. This review builds on the discussion of paranoid processes by Retterstøl (1987) in the third (German) edition of this text and further presents the findings of more recent research on delusional disorders.

2 Nosology

The following changes with respect to the delusional disorders are evident in DSM-IV (Saß et al. 1996; Kendler et al. 1989) and ICD-10 compared with the earlier diagnostic classification systems:

- *Change of concept.* Kraepelin's etiological conception of paranoia has been replaced by the descriptive designation "persistent delusional disorder" (ICD-10: F22.0). This step was taken not least because of the ambiguity of the term "paranoid," variously used in the past to designate the presence of delusions, a particular delusional content (e.g. persecution), a type of schizophrenia, or a personality disorder. The abandonment of this term also leads to the loss of the concept of paranoia as a specific delusional structure narrowly interwoven with individual development, and to the loss of the corresponding distinction between paranoid and paranoiac processes (Schmidt-Degenhard 1998).
- *Broadening of the category.* The central feature of persistent delusional disorder is the presence of a delusion that is usually systematized and typically confined to a single subject and is not associated with an underlying schizophrenic, affective, or organic illness or with a major impairment of personality. The disorder is thus a monothematic and monosymptomatic delusional psychosis. In DSM-III, the corresponding category was restricted to persecution mania and morbid jealousy, but it has now been extended to include delusions of other, non-bizarre content (megalomania, erotomania, delusions related to the body). The major reason for this innovation was the finding in large-scale, long-term studies in Scandinavia that the particular type of delusional content has no predictive specificity for the outcome of paranoia or schizophrenia (Opjordsmoen and Retterstøl 1987; Retterstøl 1991a,b). The possible types of delusional content thus now include the following:
  a) Delusion of being harmed, persecution mania
  b) Morbid jealousy (Othello syndrome)
  c) Erotomania (Clérambault syndrome)
  d) Megalomania (e.g. inventive mania)
  e) Litigation mania
  f) Hypochondriacal psychosis
  g) Delusional parasitosis
  h) Olfactory delusional disorder
  i) Body-dysmorphic delusional disorder (delusional dysmorphophobia)
  j) Induced delusional disorder (*folie à deux*)
- *Inclusion of hallucinations.* The diagnostic criteria also now include clearly manifest hallucinations, as long as they are not visual or auditory; tactile, olfactory, and other vivid body-associated hallucinations are quite common in some forms of delusional disorder (e.g. delusional parasitosis, olfactory delusional syndrome).
- *Persistence.* The delusional disorder is required to be present for 1 month according to DSM-IV, and 3 months according to ICD-10; when it is present for shorter times, it must be classified as an acute, predominantly delusional psychotic disorder (F23.3). The long-term findings by Opjordsmoen and Retterstøl (1991) and Opjordsmoen (1993) indicate that there is a fairly large subgroup of acute delusional disorders, frequently of reactive origin, which remit within 6 months and do not recur, while patients with delusions lasting more than 6 months can expect a significantly worse course of illness. The authors therefore advocate a division of diagnostic categories according to a 6-month criterion, in analogy to the differentiation made between schizophreniform and schizophrenic illnesses.
- *Residual category.* Finally, ICD-10 lists other or not otherwise specified persistent delusional disorders as a residual category (F22.8, F22.9), as well as induced delusional disorder, the so-called *folie à deux* (F24).

As early as the initial studies by Winokur (1977) and Retterstøl (1966, 1970), it was found that first-degree

relatives of patients with delusional disorders had a frequency of paranoid illnesses only slightly higher than that of the general population. Since then, the nosological independence of the delusional disorders has been confirmed by further family studies (Kendler and Hays 1981; Kendler et al. 1985; Watt 1985), which revealed an elevated incidence of paranoid personality disorders in patients' relatives, but no increased occurrence of schizophrenic or schizotypal disorders, as is found in the families of schizophrenics. For the paranoid psychoses of old age, too, family studies indicate that there is no genetic link to schizophrenia (Howard et al. 1997). The hypothesis of a continuous "paranoid spectrum," ranging from non-paranoid and paranoid schizophrenias to paranoia and paranoid personality disorder (Magaro 1981; Munro 1982), has thus not been confirmed: "Most delusional disorders probably have no link to schizophrenia" (Dilling et al. 1993, p. 103). Despite the frequent occurrence of secondary depressive mood alterations in the course of the delusional disorders, the studies referred to above also provide evidence against their belonging to the class of affective disorders.

3 Etiology

3.1 General Considerations

As already mentioned, the studies available to date fail to provide evidence for any significant role of hereditary factors in the causation of delusional disorders. Neuropathological causes are also usually not found, except in the case of paranoia of old age (see below). According to our present state of knowledge, most cases result from a combination of abnormal personality development, unfavorable environmental conditions, and a triggering situation of conflict; the delusion often bears an understandable relation to these preconditions. Thus a continuous growth of the biographically determined delusional theme is often found through the intermediate stages of an overvalued idea and of a delusion-like reaction to the final stage of an irreversible, chronic delusion.

Two major themes continually recur in studies or descriptions of the primary personality of delusional patients: problems of self-confidence and of self-esteem. Thus most patients have personality disorders, usually of the paranoid or sensitive-narcissistic type (see, e.g. Enoch and Trethowan 1979; Opjordsmoen 1988b). According to Kretschmer (1966), a sensitive character is typified by high demands on oneself and corresponding susceptibility to impairment, with a simultaneous tendency toward introversion and affect retention. This structure is now usually subsumed under the further concept of the narcissistic disorder, which may thus also include delusional patients with rather expansive features. Obsessiveness and rigidity are further features of many delusional patients; these are characteristics that impair the ability to gain distance from oneself and that may promote the development of an overvalued idea.

Most patients can also be presumed to have a fundamental disturbance of self-esteem regulation and insecurity in their social relationships, whether because of genetic personality predispositions or because of developmental impairments in early childhood (Retterstøl 1987). Even in youth, these patients often manifest contact disturbances, inhibitions, affect retention, and tendencies toward distrust and withdrawal; supposed or actual slights are perceived in hypersensitive fashion, and deep-seated feelings of inadequacy are formed, along with resentment of others. Injurious or stigmatizing life experiences and conditions such as physical impairment, forced migration, loss of social class, and minority status may further reinforce such tendencies, even in later periods of life (Tölle 1987; Fuchs 1994a,b, 1998a).

Thus, in comparison to schizophrenic or affective disorders, delusional disorders are particularly prone to appear in individuals of low social and educational status and in immigrants (Kendler 1982). In Winokur's study (1977) of 29 delusional patients, more than half had an intelligence quotient below 90. Further factors favoring the development of delusions include social isolation (e.g. in incarceration psychoses), foreign-language environment, sensory deficits leading to an impairment of communication (Fuchs 1993b), impairment of critical ability and affect processing of organic cerebral causes, and persistent substance abuse (Retterstøl 1966; Munro 1988).

When these conditions are present, the delusional theme may gradually come to the fore or, indeed, appear suddenly. The immediate provoking factor is usually the threatened or actual failure of a central personal concern, such as a sexual or marital conflict, a severe narcissistic insult, shame, or loss of social standing, which can no longer be compensated for. In this situation, the individual abjures personal responsibility by paranoid outward projection ("the others are against me") or by inculpation of his or her own body, nature, or disposition ("my body is against me"). Delusions thus often represent a recognizable, projective, and denying defense against feelings of inferiority and inadequacy, against the feared lack of fulfillment of needs for intimacy and dependency, and even against the individual's own aggressive tendencies, fueled by feelings of resentment and revenge. Projective identification can be seen as a reversal of

shameful and degrading experiences: the supposed malevolence of others is used as a cover for one's own supposed inferiority (blame instead of shame; Morrison 1987).

Kretschmer's conception of the sensitive relational delusional state is still of paradigmatic significance for the psychodynamic interpretation of the origin of delusions, even if this conception is primarily concerned with moral and sexual/ethical conflicts as decisive triggers of the disorder. This type of problem seems to have faded in importance in liberal, pluralistic societies; the sensitive relational delusional state is still very much in evidence in Japan, but has receded into the background in the West and is actually entirely unknown in the United States (Rasmussen 1978). Nonetheless, it may be presumed that the change of social values has redirected the central narcissistic problem of relational delusions onto the theme of physical appearance and attractiveness, so that they now typically manifest themselves as body-dysmorphic disorder or olfactory delusional syndrome. Now as before, the battle for self-esteem seems to be the real theme of most delusional patients, who, by displacing it onto the neighboring battlefield of the delusion, "acquire an unconscious pretext by means of which their potential or supposed defeat in life can be covered up, justified, or indefinitely postponed" (Adler 1927, p. 191).

The functionality of the delusion, seen against this background, consists of a compensation for feelings of inadequacy or rejection (as in erotomania), a conviction of one's own greatness, or an enhancement of one's ability to wage battle against the imaginary opponent, with whom the delusional patient forms a "pseudo-community" (Cameron 1959). To this extent, chronic delusions also imply a self-stabilizing defense against a realistic confrontation with one's own life situation. Thus Roberts (1991) compared 17 patients with systematized delusions to a corresponding group of delusional patients in remission. The actively delusional patients attached a significantly higher degree of meaning to their lives and were less depressive, primarily because of the subjects of their delusions.

3.2 Contributions from Phenomenological Anthropology

Phenomenological anthropological research has also made a major contribution to our understanding of the paranoid disorders by attempting to explain delusions as an abnormality of the fundamental structures of the relationship of human beings to the world (Matussek 1963; Glatzel 1981; Blankenburg 1991, 1992). It is assumed that delusions generally consist not of false ideas about neutral and objectifiable states of affairs, but rather of interpretations and assessments that patients attach to a situation in relation to themselves. Delusional content is thus primarily composed of "second-order realities" (Watzlawick 1988), which cannot be directly and objectively tested. Normal, correct judgments in this area can only be made if the individual is basically familiar with the world and capable of assessing it reliably; in Blankenburg's phrase, the experienced world must possess a natural self-evidence (*natürliche Selbstverständlichkeit*; Blankenburg 1971), with respect both to the individual's relationships with others and to his or her own corporeal nature. Delusional disorders develop precisely in the areas where reliable orientation in our environment depends not on rational knowledge and empirical evidence, but on this type of familiarity with the world and emotional soundness. This basic safeguarding mechanism is lacking in individuals predisposed to paranoia; distrust and control are used in its place.

The paranoid individual is on the alert for possible injuries or malevolence and vigilantly scans his or her environment for signs of threat, deception, or betrayal. His or her social perception is thus directed toward hidden meanings, the possible use of pretenses by others, and the hidden "other side" of their behavior. In ambiguous or limited communication structures (e.g. neighborhood gossip, deafness, foreign-language environment), the uncertainty of social perception rises, and room is left for self-referential interpretations in which the greatest attention is paid to the least conspicuous matters – a fleeting gesture, undertones, or words left unsaid. If the paranoid individual further feels him- or herself to be threatened in a central personal concern, the tension of anxious distrust may become unbearable, until a new consistency is achieved by the construction of a delusional interpretation – the certainty of the contempt, betrayal, or malevolence of others, subjectively experienced as a seeing-through or unmasking of their real nature.

This delusional "insight" reduces the previously overwhelming cognitive-emotional complexity of the situation (Luhmann 1973); the tormenting ambiguity of social perception gives way to a subjective emotional certainty, which, despite its negative meaning, is associated with relief and, not least, with the possibility of taking measures of caution or of opposition. Once constructed, the delusional paradigm progressively establishes itself, both by the self-confirming distorted selection of information and ignoring of contrary evidence and by the patient's self-fulfilling behavior, as when enmity toward the supposedly malevolent others eventually provokes an actual rejection by them.

The essentially ambiguous nature of the second-order realities described above implies that any

subjective assessment of them is necessarily aspect bound and provisional. Social relationships are, therefore, based on constant, mutual reassessment of behavior patterns and their interpretation; to be continued, they required one's own view of the shared situation always to be correctable by that of the other person. It is characteristic of delusional individuals to foreclose the possibility of correction and stand by their own interpretations with apodictic certainty, without any intersubjective modulation, as if they were statements about their own inner states (Spitzer 1989).

In the anthropological literature, this phenomenon is known as the lack of assumption of another perspective (Glatzel 1981; Blankenburg 1991; von Baeyer 1991). A delusional condition can thus be construed as an unsuccessful, surrogate-like encounter with others, which securely shields itself against the possibility of doubt: by refusing to switch perspectives, the delusional individual protects him- or herself from becoming burdened once again by the inner conflicts that have been projected onto others. At the same time, the lack of assumption of another perspective brings the delusional individual into an egocentric position: the notion of being observed or threatened by others arises together with the inability to take the point of view of an uninvolved, neutral third party and thereby eradicate the self-referential, delusional perspective (Fuchs 1994b).

On the linguistic/semantic level, these models are reflected in the theory of concretism, according to which a delusion is to be understood not as a symbolic statement about an intersubjectively constituted state of affairs, but as a non-declarative, as it were "exclamatory" expression by the patient of his or her mental state, in a manner analogous to an interjection such as "Help!" or "Ouch!" (Holm-Hadulla 1982; Spitzer 1989; Mundt 1996). The matter troubling the patient is brought to expression, not as such, but masked, in concretistic fashion (e.g. lack of self-esteem as bodily disfigurement in body-dysmorphic disorder, guilt as sexually transmitted disease in hypochondriacal psychosis). Because the topic can no longer be reformulated in symbolic language, the expression of the delusion is also inaccessible to correction by reasoned argument.

3.3
Cognitive Research on Delusional Disorders

The findings of recent cognitive experimental research on the delusional disorders further add to our understanding of this subject. Patients with delusional disorders have been found to be characterized by abnormal attributive and cognitive styles, which may be demonstrated during the performance of specific tasks. Typically, premature conclusions are drawn from limited, ambiguous, or selectively perceived information; too much certainty is attached to one's own judgment in the assessment of probabilities; there is a tendency to ascribe a special significance to coincidental events; and, finally, unfortunate events tend to be blamed on others (Huq et al. 1988; Bentall and Kaney 1989; Bentall et al. 1991; Garety et al. 1991). This would imply that delusional patients lack the ability to see themselves objectively or to assume an attitude of "healthy skepticism" toward their own judgments and thus cannot put events affecting themselves in an overarching, neutral context – or, as expressed elsewhere (Minkowski 1947; Berner 1978), that the thought processes of these patients are characterized by "the exclusion of coincidence."

Although the tasks performed by the subjects in these studies were not related to the themes of their delusions, it should be borne in mind that these findings need not be interpreted as implying a preexisting vulnerability; they may, conceivably, indicate a feature of the course of the disorder. Nonetheless, these findings help to explain the cumulative origin of uncorrectable convictions: when the capacity for self-questioning and intersubjective correction is absent, social perception enters a vicious circle in which there is a progressively severe distortion of reality. The findings discussed above have now been incorporated into techniques of cognitive therapy in the delusional disorders (see below).

4
Individual Forms

The forms of delusional disorder listed in Sect. 2 may be broken down in terms of their basic areas of reference as follows:

1. Delusions primarily with respect to relationships with others:
   a) Injury, aggression (persecution mania, morbid jealousy, litigation mania)
   b) Self-aggrandizement, grandiosity (megalomania, erotomania)
   c) Shame and rejection (olfactory delusional disorder, body-dysmorphic disorder)
2. Delusions primarily with respect to one's own body:
   a) Delusional parasitosis
   b) Hypochondriacal psychosis

The types of delusional disorder listed under item 1 above have in common that the central experience is often one of shame, inadequacy, or inferiority. This becomes explicit only in subitem (c), while it is made almost unrecognizable by defense mechanisms in the

other two groups – by aggressively toned projections (blame instead of shame) or by self-aggrandizement. The primarily intersubjective reference of these forms of delusional disorder is expressed in the patient's frequent self-referential perceptions and delusional ideas.

Distinct from these forms of delusional disorder are the anxious-hypochondriacal forms, referring to the patient's own body, which are listed in item 2; in these forms, ideas of reference occur only rarely. The patient lacks trust not in other people, but in his or her own body and its functions. Admittedly, the excessive stress placed on the patient's own body is generally secondary to a disturbance of the capacity for interpersonal relationships (Küchenhoff 1985).

This dichotomy implies that not all forms of delusion regarding one's own body should be called hypochondriacal, as is usual in the Anglo-American literature. The hypochondriacal conviction of being ill is hardly the only form of disturbed and alienated corporeality; indeed, in the course of a person's life, very different themes may come to bodily expression, including shame, guilt, inferiority, and fear of death (Fuchs 1992). Body-dysmorphic delusional disorder, a pathological process of the social or "external body," thus has little in common with hypochondriacal psychosis per se. Furthermore, body-dysmorphic delusional disorder usually begins in adolescence, while the forms of delusional disorder primarily relating to one's own body usually begin in middle age or old age (Musalek et al. 1989).

The ensuing description of selected individual forms of delusional disorder is intended as an extension of, and supplement to, earlier treatments of the subject (see especially Retterstøl 1987).

4.1 Morbid Jealousy

Morbid jealousy is, along with persecution mania, one of the commoner delusional disorders (approximately 40% of all cases, according to Winokur 1977; Crowe et al. 1988). It predominantly affects men in late adulthood (Musalek et al. 1989). Only a minority of these patients are alcoholics; the supposed close relationship of this disorder to alcoholism is no longer postulated (Enoch and Trethowan 1979; Soyka et al. 1991). Alcoholic morbid jealousy is no longer counted among the delusional disorders in ICD-10, but rather among the disorders caused by psychotropic substances (F10.5). The differential diagnosis must exclude underlying schizophrenic or organic (e.g. dementing) illnesses.

The delusional theme of jealousy is related exclusively to the patient's partner. The patient is incorrigibly convinced of the partner's unfaithfulness and attempts to obtain evidence for it by constant questioning, checking, spying, and searching of the intimate sphere. This usually leads to severe marital conflicts. Typically, the supposed rival remains a rather obscure figure, and the possibility of "catching the partner in the act" tends to be avoided (Enoch and Trethowan 1979), which reflects the reluctance to confront reality that is inherent in the delusional disorders. Instead, the patient's aggressive impulses are directed entirely against the partner; morbid or delusional jealousy is therefore associated with a particularly high risk of violence, including homicide (Soyka 1992).

Narcissistic, latently insecure, and compulsive-controlling personality traits are thought to be predisposing factors for the development of this type of delusional disorder (Enoch and Trethowan 1979). It is often triggered by persistent partnership conflicts or by failures that are painful for the patient, e.g. in the occupational sphere, which are displaced onto the relationship. Feelings of inferiority and inadequacy, fears of loss (possibly secondary to sexual impotence), and also, not uncommonly, the patient's own repressed tendency toward infidelity are projected onto the partner and typically progress from excessive to delusional jealousy. Morbid jealousy exhibits a typical circular structure of intrusive thoughts, distorted perception, and intense feelings of anxiety or rage; this structure has recently proved accessible to cognitive behavior therapy (Tarrier et al. 1990; Dolan and Bishay 1996).

It can be inferred from the findings of several long-term studies that morbid jealousy has a favorable prognosis in comparison to other delusional disorders and only rarely converts to another type of psychotic illness (Crowe et al. 1988).

4.2 Erotomania

Erotomania (Clérambault syndrome) consists of the conviction that one is loved by another person, generally of higher social status, publicly known or held in high regard, who makes his or her passion known through secret signals or messages. The patients are predominantly women of modest social circumstances, in their fourth to sixth decades, who have not established lasting partnerships. They often pursue the admired person (who is usually married) through innumerable letters, telephone calls, or even public scenes and are not discouraged in these activities even by nearly insuperable obstacles, in accordance with the avoidance of confronting reality that is typical of the delusional disorders (see

Sect. 3.1). These patients succeed in reinterpreting their supposed admirers' acts of rejection, no matter how vigorous they may be, as expressions of love ("tests of devotion" etc.). Legal measures taken by the harassed victim of the delusion lead, not uncommonly, to the patient's involuntary commitment to a psychiatric institution.

The contrast between the generally not very successful, erotically unsatisfied, and lonely existence of the patients and the glamour of their supposed admirers suggests that the essential causative factor of erotomania is narcissistic wish-fulfillment (Segal 1989). In accordance with this hypothesis, these patients usually have an inhibited-sensitive or paranoid primary personality and are often notably physically unattractive (Hollender and Callahan 1975). It should be borne in mind, however, that similar manifestations may appear in the setting of schizophrenia or organic delusional syndromes (Signer and Cummings 1987; El Gaddal 1989). The prognosis of pure erotomania is thought to be rather unfavorable; treatment with neuroleptic medications usually has only a palliative influence on the chronic course of the disorder (Opjordsmoen and Retterstøl 1987). Directed reality confrontation or permanent separation from the loved person may have a beneficial effect (Segal 1989), but in many cases leads merely to a change of the "delusional object."

4.3 Body-Dysmorphic Disorder

Dysmorphophobia denotes the subjective perception of one's own body as ugly or deformed, combined with the conviction that others perceive one's body in the same way. As a rule, the supposed defect is located in a specific, usually exposed part of the body, such as the face, ears, nose, jaw, teeth, hair, etc., but it may also be in the limbs or genitalia. These patients develop ideas of reference, imagine that they are being discriminated against or ridiculed by others, and withdraw increasingly from social contact. They consult plastic surgeons, otorhinolaryngologists, or dentists and importune them for an operation that will change their appearance.

The designation "phobia" indicates the close relationship of this syndrome to the social phobias; in view of these patients' environmental anxieties and disturbed contact behavior, the rejection of this previously used term by DSM-IV and ICD-10 is not necessarily appropriate. The new division into separate categories (F45.2: body-dysmorphic disorder; F22.0: somatoform delusional disorder) is not in line with clinical experience, which suggests a continuum extending from excessive preoccupation with one's own appearance, to a neurotically overvalued idea, to the delusional certainty of one's own deformity (Phillips 1991). The assignment of delusional disorders to the category of the schizophrenias, as was customary in the past (Zaidens 1950; Connolly and Gipson 1978), has no proponents today, even if the dysmorphophobic syndrome, like obsessive-compulsive manifestations, can be conceived of as a defense against further psychotic disintegration. Bizarre delusional ideas about physical changes or experiences of being physically influenced from without are evidence for an underlying schizophrenic illness; moreover, dysmorphophobic convictions may also appear in the context of the affective disorders. The classification of the milder forms of dysmorphophobia, however, is still ambiguous in the current systems of classification.

In psychodynamic terms, the connection to the themes of shame and self-esteem is evident; it is a matter of showing oneself and being seen by others. Excessive observation of oneself and an idealized conception of beauty, combined with feelings of inferiority and contact anxieties, characterize these patients (Küchenhoff 1984). As one would expect, sensitive-narcissistic, insecure, and compulsive personality disturbances are also often found in this patient group. The initial manifestation usually appears, not surprisingly, in adolescence, when the individual's body image, concept of self, and interpersonal relationships all come to a crisis. In this situation, the projection of deep-seated feelings of inadequacy, particularly sexual inhibitions, onto a biological bodily defect may perform a self-protective function (Fuchs 1993a): it frees the individual from the need for confirmation, e.g. in competition for partners, and allows negative aspects of the self to be concretized and externalized in a single part of the body, a *pars pro toto* that becomes accessible to repair by cosmetic surgery.

Body-dysmorphic disorder usually takes a chronic course lasting several years or decades and is associated with an elevated risk of suicide (Phillips 1991). Surgical intervention can be expected to produce an improvement, if at all, only in milder forms of the disorder and should be firmly discouraged in more severe forms: when surgery is performed, patients often continue to complain of the supposed deformity because of unrealistic expectations from surgery, or else the symptom migrates to another part of the body, because the underlying conflict concerning the patient's self-esteem remains unresolved (Hay 1970; Andreasen and Bardach 1977; Strian 1984). A marked affective or compulsive component, however, calls for the application of serotonin reuptake-inhibiting antidepressants, which have been repeatedly reported to yield good results (Hollander et al. 1989; Phillips

1991), while the outcome of neuroleptic therapy is usually unsatisfactory.

4.4 Olfactory Delusional Disorder

Olfactory delusional disorder is characterized by the conviction that the sufferer emits a repellent odor (e.g. of sweat, bad breath, or flatus). The primary experience is mainly in the area of perception, i.e. there are abnormal olfactory sensations which, however, are experienced in an "environmentally dependent" fashion, usually only in the presence of other people. There are corresponding ideas of reference, in which words or actions of others that are insignificant in themselves are interpreted as expressions of disgust and aversion. These patients combat the supposed odor by washing or applying perfume, without success; they usually resort to a more or less total social withdrawal. They not uncommonly develop hypochondriacal explanatory delusions of odor-producing diseases of the skin or gastrointestinal tract (Gattaz and Haas 1982).

Olfactory phobia and olfactory delusional disorder have received the greatest amount of attention in psychopathology in Japan, where Morita (1947) grouped them together with dysmorphophobia, erythrophobia, and eye contact phobia as *taijin-kyôfu* syndrome ("fear of mankind," or anthropophobia). The common features of these disorders are the underlying feeling of shame and inferiority, the manifestation of ideas of reference and contact phobia, and social withdrawal (Yamashita 1993; Kimura 1995). Morita believed that the basis of the syndrome lay in the so-called *shinkeishitsu* personality ("constitutional nervous temperament"), characterized primarily by introversion, hyperreflectivity, neurasthenia, and disturbance of self-esteem, and developed a specific form of treatment known as Morita therapy (Kora 1999).

Because olfactory hallucinations also appear frequently in the settings of depression and schizophrenia, these differential diagnoses must be considered (Malasi et al. 1990). The presence of a pure delusional disorder is supported by the development of the delusion on the substrate of a sensitive-narcissistic personality and an existing relationship conflict, with experiences of inadequacy, insult, or exclusion. The perceived inferiority and self-contempt are concretistically projected by the delusion onto the physical sphere, aided not least by the special relationship of the sense of smell to emotional atmospheric experiences, sympathy, and antipathy. Such delusional conditions have a continuous transition into neurotic developments with delusion-like reactions, which are more accessible to psychotherapeutic intervention (Moesler 1992).

4.5 Delusional Parasitosis

Delusional parasitosis, first described by Ekbom (1938) under its German name *Dermatozoenwahn* ("dermatozoal delusion" or epidermozoophobia), consists of the unshakable certainty of being infested by intra- or subcutaneous parasites, usually accompanied by itch. These patients often subject themselves to extensive grooming rituals and bring small collections of skin particles to dermatological clinics or government health departments requesting that they be studied for the presence of parasites ("match-box sign"). The affected individuals, usually women in their fifth to seventh decade, feel severely tormented and socially stigmatized by their supposed condition. Tactile sensations (itching, paresthesia) are a major feature of this delusional disorder, for which the name "chronic tactile hallucinosis" has also been suggested (Bers and Conrad 1954), but they are not always present. A comparative study revealed that delusional patients with tactile phenomena, compared to non-delusional patients with tactile phenomena of other origin, were more often socially isolated and less frequently married, placed higher hygienic demands on themselves, and more commonly had a history of dermatological and mental illness (Musalek 1991).

The nosology of this delusional syndrome has always been particularly difficult; it has been found in association with practically all types of underlying illness. In the diagnostic study carried out by Musalek et al. (1990), approximately 40%–50% of all cases were attributable to organic delusional disorders, usually secondary to arteriosclerosis; Marneros et al. (1988) found this figure to be as high as 70%. Rarer entities to be considered in the differential diagnosis include cocaine- or amphetamine-induced psychoses ("cocaine bugs"), vitamin B_{12} deficiency, diabetes mellitus, lymphoma, and uremia accompanied by pruritus (Morris 1991). The supposed parasitic infestation may also be a delusional expression of feelings of guilt or punishment in the setting of melancholia, while a paranoid attribution of the infestation to another person (contamination delusion) is indicative of schizophrenia. Thus only a minority of cases can be considered pure delusional parasitosis. These cases are usually in the setting of a compulsive personality structure and social isolation. The partners of these patients relatively frequently develop an induced delusion (*folie à deux*) (10%–20%; Musalek and Kutzer 1990).

The prognosis of delusional parasitosis is relatively favorable: intensive application of pharmacological, psychotherapeutic, and sociotherapeutic measures leads to improvement, or even remission, in two thirds of these patients (Musalek 1991; Trabert 1993).

4.6 Hypochondriacal Psychosis

The concept of monosymptomatic hypochondriacal psychosis (Munro 1988) has been used to subsume several disorders, such as olfactory delusional disorder, delusional parasitosis, and body-dysmorphic delusional disorder, that are treated as distinct in this review. As mentioned above, however, the term should be reserved for the hypochondriacal delusional disorders per se, corresponding to the earlier "paranoia hypochondriaca" (Serieux and Capgras 1909), in which patients imagine that they are suffering from an incurable or fatal illness. Hypochondriacal psychosis must also be distinguished from neurotic hypochondriases, which present as diffuse states of ill-being or neurasthenic exhaustion and in which the fear of suffering from a physical illness does not have the same uncorrectable character (Dilling et al. 1993). The so-called circumscribed hypochondriases may, however, assume a character similar to that of a delusion (Hallen 1970); they are characterized by persistent abnormal sensations and foreign-body sensations in particular areas of the body, e.g. in the mouth, which often lead these patients to consult dentists or otorhinolaryngologists.

The affected patients are distinguished by the persistence with which they demand radical diagnostic and operative procedures, usually from multiple physicians. There are often painful dysesthesias in several organ areas, in the skin, or in the sense organs; when these are especially bizarre, the differential diagnosis of coenesthetic schizophrenia should be considered. On the other hand, the fact that the delusion concerns the patient's own state of health, rather than features of the external environment interpreted with reference to the self, suggests the possibility of a depressive disorder, and hypochondriacal delusions may indeed often persist at attenuated levels after the actual depressive phases have passed. It may also be necessary to exclude underlying disorders such as pernicious anemia, uremia, lead poisoning, endocrine diseases, or cerebrovascular disease by specific diagnostic testing, even though this may reinforce the patient's fixation on a somatic cause.

Hypochondriacal psychosis, like hypochondriacal syndromes in general, is typified by a loss of confidence in the corporeal basis of existence and an attitude toward the environment narrowly centered on bodily complaints. The patients' primary personality often exhibits anancastic features and a tendency toward somatization (Munro 1988). The overemphasis of the body is an expression of a loss of the capacity for interpersonal relationships; against this background, the demanding, hostile, and disempowering behavior of these patients toward physicians is to be seen as a surrogate for interpersonal relationships (Fuchs 1992; Küchenhoff 1985). The content of the delusion is not uncommonly of a sexual nature (delusional syphilis or AIDS; Mahorney and Cavenar 1988); this may indicate a sensitive or erotic-conflictual experiential reaction as the origin of the illness. If the disorder takes a chronic course, as it usually does, the frequently occult depressive mood alterations should be watched for, as these are accompanied by an elevated risk of suicide (Bebbington 1976; Opjordsmoen 1988a).

4.7 Delusions Accompanying Physical Impairment

Delusions as consequences of actual physical impairment must be distinguished from hypochondriacal and body-dysmorphic delusional themes. Gaupp (1942) first described such delusions in the case of a teacher named Hager, who felt that he was no longer taken seriously by others, especially women, after an amputation in the war, developed a progressively severe persecution mania, and finally killed a young woman. According to Tölle (1987, 1993), it is primarily amputations and other invasive operations, deformities, or illnesses that lead to social disadvantage and discrimination and are therefore perceived as shameful deficiencies. The resulting delusion of interpretation (delusion of being harmed), which not uncommonly remains hidden for very long periods, bears an important projective, ego-sparing function and is thus in keeping with the etiological concepts presented in Sect. 3.1 above. Clearly, a psycho-organic disturbance of information processing and critical ability plays an additional role in many cases.

4.8 Delusional Disorders in Old Age

Paranoia beginning at age 60 or above occurs with a prevalence of approximately 1%–2% (Christenson and Blazer 1984). Typical delusional contents include those of injury, threat, or persecution by people near the patient's home, although fantastic or confabulatory delusional contents are not uncommon. The nosological classification of the disorders first described by Kay and Roth (1961) as late paraphrenias remains controversial; the existence of a continuous spectrum of transitional forms, from pure delusional disorders with or without hallucinations to unambiguously schizophrenic phenomena, makes it impossible to differentiate delusional disorders of old age with certainty from late-onset schizophrenia (i.e. arising between the ages of 40 and 60) or to establish a clear

diagnostic distinction between schizophrenic and delusional disorders (Howard et al. 1994; Riecher-Rössler et al. 1995). In ICD-10, senile paranoia is assigned to the delusional disorders (F22.0); the presence of persistent auditory hallucinations constrains a classification as persistent delusional disorder of other type (F22.8), as long as the criteria for schizophrenia are not fulfilled (Dilling et al. 1993).

Independently of these problems of classification, certain risk factors and provoking conditions for the onset of dementia in old age have been repeatedly described:

- Female sex (female-to-male ratio typically 7:1; Almeida et al. 1995).
- Paranoid or schizoid premorbid personality traits (Kay et al. 1976; Fuchs 1998b).
- Social isolation (Naguib and Levy 1987; Almeida et al. 1995), in accordance with Janzarik's (1973) concept of contact deficiency paranoia.
- Sensory impairment, particularly deafness (in as many as 40% of cases), which may favor a paranoid misperception of social situations (Cooper et al. 1974; Fuchs 1993a).
- Discriminating, injurious, or threatening life events such as forced migration from one's homeland (in as many as 50% of cases), birth out of wedlock, physical impairments, etc. (Fuchs 1994a, 1998a).
- Mild cognitive disturbances that often can be detected only by neuropsychological assessment; patients with pure delusional disorder tend to have more severe cognitive deficits than those diagnosed as schizophrenic (Naguib and Levy 1987; Howard et al. 1994).

Studies using computed tomography and magnetic resonance imaging have demonstrated organic brain disturbances in a majority of patients; patients of this type generally belong to the category of delusional disorders without schizophrenic manifestations of the first rank (Förstl et al. 1991; Howard and Levy 1992). It is therefore assumed that at least a large fraction of senile paranoid illnesses are caused in part by neurodegenerative processes. This finding is of comparable significance to that of the mainly temporal and prefrontal cerebral abnormalities often found in younger schizophrenics, which probably reflect a disturbance of central nervous system maturation (Häfner 1997).

The course of the senile paraphrenias is usually chronic, although hospitalization is only rarely required. In the study by Howard and Levy (1992) of 64 patients, one third responded partially, and one quarter completely, to long-term neuroleptic treatment. The most effective form appeared to be a depot medication in a rather low dose (e.g. 14 mg fluphenthixol or 9 mg fluphenazine decanoate every 2 weeks). Nonetheless, even when neuroleptic therapy is effective, patients rarely gain insight into their illness. Most patients require permanent maintenance therapy (for a review, see Eastham and Jeste 1997).

5 Therapy

Building a relationship of trust is the indispensable foundation of successful treatment, regardless of what further measures are taken, and this is especially true for the delusional disorders. The physician must thus avoid a direct confrontation with reality, at least at the beginning, but should also take care not to render additional support to the patient's delusional system by an excess of benevolent understanding or even assent.

5.1 Somatotherapy

There is a serious lack of empirical studies, particularly controlled studies, which is not surprising in view of the deficient insight and compliance of this group of patients. There are, however, a number of relatively frequently replicated observations that may be of heuristic value in individual cases. Neuroleptic therapy is often only moderately successful, but if a therapeutic relationship of trust is present, a treatment should be attempted in all cases, at adequate doses and over a sufficiently long period of time. This may be justified to the patient by explaining that the purpose of medication is to heighten inner stability and to protect against the stresses caused by the (supposed) threats or injuries. Despite such assurances, however, the medication itself or any side effects that may occur are not infrequently viewed by the patient in paranoid fashion.

In a number of studies involving relatively few patients, pimozide at a dose of 2–6 mg was found to be an effective agent (Munro 1984, 1988; Pollock 1982; Kaschka et al. 1991); a combination of clomipramine and pimozide may also be useful in depressively colored delusional syndromes (Chiu et al. 1990). Nonetheless, it has not been demonstrated that these agents are fundamentally superior to other highly or intermediately potent neuroleptics. Pimozide may also have cardiotoxic effects in therapeutic doses, which may restrict its indications, particularly for older patients.

The frequent association of delusional disorders with latent or manifest depressive mood alterations (64% according to Munro 1988; 51% according to Marino et al. 1993) makes it seem reasonable to try

antidepressants in many cases. In particular, in both delusional and non-delusional dysmorphophobia, serotonin reuptake inhibitors have been found to be effective in as many as two thirds of all patients (Philips 1996).

5.2 Psychotherapy

Psychotherapeutic approaches to the treatment of delusions have aroused interest only in recent years (Torch and Bishop 1981; Teusch et al. 1987; Nelki 1988; Mundt 1996). A durable, long-term relationship of trust is basic to any psychotherapeutic treatment of delusional patients. According to Mundt (1996), focusing directly on the delusion should therefore be avoided whenever possible, and explorations in this direction should not be too frequent. Instead, the delusional theme should be circumvented and thereby deactualized, in a manner similar to behavior-therapeutic extinction. If treatment is based on the conception that the delusion is a concretistic manifestation of a central life concern that has come turned badly for the patient (see Sect. 3.2), then the therapist's task is to direct his or her efforts to the concern underlying the delusion, without naming it as such. An interpretation of the delusion that is unacceptable to the patient should thus be avoided. Instead, the promotion of new experiences in emotionally neutral fields, and acceptance of the patient by the therapist, can implicitly and covertly help to satisfy the patient's existential needs (e.g. for contact and esteem).

It is important in both the pharmacological and the psychotherapeutic treatment of delusional disorders to recall their compensatory function, i.e. the safe removal of topics that cannot be otherwise processed by means of projection or grandiosity. Structural deficits often manifest themselves as depressive syndromes provoked by a sense of emptiness and loss of meaning when the delusion remits, particularly after successful treatment. The treatment should therefore be directed toward topics of practical importance for the patient's life, the mobilization of resources and preserved abilities, and the recognition of the patient's coping capacity, so that the stabilizing function of the delusion can be replaced by a gradual strengthening of the ego. Once this process is in place, reality confrontation can slowly begin.

When this stage is reached in patients with delusions, or in milder cases with overvalued ideas, cognitive therapy strategies employing directed experiments in perception and behavior may be beneficial; such strategies are currently being developed and tested in connection with cognitive theories of the origin of delusions (see Sect. 3.2). Although the cognitive approach at first seems contradictory to the principle of not arguing with the patient, it can be carried out using careful strategies of circumvention and has been reported to influence the course of the illness favorably in as many as half of all patients (Chadwick and Lowe 1990; Kuipers et al. 1997).

6 Course

Ever since Kraepelin's time, delusional disorders have been held to be chronic, nearly irreversible diseases: "No case of genuine paranoia ever comes to a cure" (Kraepelin 1899). Yet Retterstøl and Opjordsmoen, in their partly prospective, partly retrospective long-term studies of a total of 334 patients, were able to document a much more favorable prognosis of the paranoid psychoses, including the pure delusional disorders. In the study by Opjordsmoen (1988b) of 41 patients with delusional disorders (according to DSM-III-R) followed up after an average of three decades, 15 (37%) had gone into remission. Patients with delusional disorders also differed significantly from schizophrenics with respect to coping with life: they were more frequently married (76% vs. 47%), more frequently had children (73% vs. 37%), and were more frequently employed (46% vs. 26%). Even among 26 patients with classical paranoia in the narrow sense, Retterstøl found that more than one third had gone into remission (Retterstøl 1991a,b). This accords with the finding by Winokur (1977) of a social cure in one third of patients after several years of follow-up; 60% worked in their occupations, and 53% lived in stable marriages. It is to be hoped that the prognosis of the delusional disorders will improve still further as newer, more effective forms of treatment are developed.

7 References

Adler A (1927) Praxis und Theorie der Individualpsychologie, 3rd edn. Bergmann, Munich

Almeida OP, Howard RJ, Levy R, David AS (1995) Psychotic states arising in late life (late paraphrenia). The role of risk factors. Br J Psychiatry 166: 215–228

Andreasen NC, Bardach J (1977) Dysmorphophobia: symptom or disease? Am J Psychiatry 134: 673–676

Bebbington PE (1976) Monosymptomatic hypochondriasis, abnormal illness behaviour and suicide. Br J Psychiatry 128: 475–478

Bentall RP, Kaney S (1989) Content-specific information processing and persecutory delusions: an investigation using the emotional Stroop test. Br J Med Psychol 62: 355–364

Bentall RP, Kaney S, Dewey ME (1991) Paranoia and social reasoning: an attribution theory analysis. Br J Clin Psychol 30: 13–23

Berner P (1978) Psychopathologische Wahnforschung und psychiatrische Hypothesenbildung. Nervenarzt 49: 147–152

Bers N, Conrad K (1954) Die chronische taktile Halluzinose. Fortschr Neurol Psychiatr 22: 254–270

Blankenburg W (1971) Der Verlust der natürlichen Selbstverständlichkeit. Enke, Stuttgart

Blankenburg W (1991) Perspektivität und Wahn. In: Blankenburg W (ed) Wahn und Perspektivität. Enke, Stuttgart, pp 4–28

Blankenburg W (1992) Analysen der Verselbständigung eines Themas zum Wahn. In: Kaschka WP, Lungershausen E (ed) Paranoide Störungen. Springer, Berlin Heidelberg New York, pp 17–32

Cameron N (1959) The paranoid pseudo-community revisited. Am J Sociol 65: 52–58

Chadwick P, Lowe C (1990) The measurement and modification of delusional beliefs. J Consult Clin Psychol 58: 225–232

Chiu S, McFarlane AH, Dobson N (1990) The treatment of monodelusional psychosis associated with depression. Br J Psychiatry 156: 112–115

Christenson R, Blazer D (1984) Epidemiology of persecutory ideation in an elderly population in the community. Am J Psychiatry 141: 1088–1091

Connolly FH, Gipson M (1978) Dysmorphophobia – a long term study. Br J Psychiatry 132: 568–570

Cooper AF, Kay DWK, Curry AR et al (1974) Hearing loss in paranoid and affective psychoses of the elderly. Lancet ii: 851–854

Crowe RR, Clarkson C, Tsai M, Wilson R (1988) Delusional disorder: jealous and nonjealous types. Eur Arch Psychiatry Neurol Sci 237: 179–183

Dilling H, Mombour W, Schmidt MH (1993) Internationale Klassifikation psychischer Störungen (ICD-10). Huber, Bern

Dolan M, Bishay N (1996) The effectiveness of cognitive therapy in the treatment of non-psychotic morbid jealousy. Br J Psychiatry 168: 588–93

Eastham JH, Jeste DV (1997) Treatment of schizophrenia and delusional disorder in the elderly. Eur Arch Psychiatry Clin Neurosci 147: 209–218

Ekbom KA (1938) Der präsenile Dermatozoenwahn. Acta Psychiatr Neurol Scand 13: 227–259

El Gaddal YY (1989) De Clerambault's syndrome (erotomania) in organic delusional syndrome. Br J Psychiatry 154: 714–176

Enoch MD, Trethowan WH (1979) Uncommon psychiatric syndromes. Wright, Bristol

Esquirol E (1838) Des maladies mentales – considerées sous les rapports medical, hygienique et médico-légal. Baillière, Paris

Förstl H, Howard R, Almeida O (1991) Altersparaphrenie. Psychopathologische und computertomographische Hinweise auf zwei Subtypen. Nervenarzt 62: 274–276

Fuchs T (1992) Der hypochondrische Wahn. Z Klin Psychol Psychopath Psychother 40: 396–410

Fuchs T (1993a) Über einen Fall von "Wachstumswahn". Zur Genese und nosologischen Klassifikation der körperdysmorphen Störung. Nervenarzt 64: 199–203

Fuchs T (1993b) Wahnsyndrome bei sensorischer Beeinträchtigung – Überblick und Modellvorstellungen. Fortschr Neurol Psychiatr 61: 257–266

Fuchs T (1994a) Uprooting and late life psychosis. Eur Arch Psychiatry Clin Neurosci 244: 126–130

Fuchs T (1994b) Die Welt als Innenraum. Kafkas "Bau" als Paradigma paranoider Räumlichkeit. Nervenarzt 65: 470–477

Fuchs T (1998a) Life events in late paraphrenia and depression. Psychopathol 32: 60–69

Fuchs T (1998b) Patterns of relation and premorbid personality in late paraphrenia and depression. Psychopathology 32: 70–80

Garety PA, Hemsley DR, Wessely S (1991) Reasoning in deluded schizophrenic and paranoid subjects: biases on performance of a probalistic reasoning task. J Nerv Ment Dis 179: 194–201

Gattaz WF, Haas S (1982) Eigengeruchshalluzinose und die Geruchstrugwahrnehmungen bei endogenen Psychosen. Fortschr Neurol Psychiatr 50: 67–72

Gaupp R (1942) Zur Lehre von der Paranoia. Der Fall des Volksschullehrers Hager. Z Ges Neurol Psychiatry 174: 762–810

Glatzel J (1981) Die paranoide Eigenbeziehung aus der Perspektive einer interaktionalen Psychopathologie. Nervenarzt 52: 147–152

Häfner H (1997) Late-onset schizophrenia and the delusional disorders in old age. Eur Arch Psychiatry Neurosci 247: 173–175

Hallen O (1970) Über circumscripte Hypochondrien. Nervenarzt 41: 215–220

Hay GG (1970) Dysmorphophobia. Br J Psychiatry 116: 399–406

Hollander E, Liebowitz MR, Winchel R, Klumker A, Klein DF (1989) Treatment of body-dysmorphic disorder with serotonin reuptake blockers. Am J Psychiatry 146: 768–770

Hollender MH, Callahan AS (1975) Erotomania or De Clérambault syndrome. Arch Gen Psychiatry 35: 1265–1267

Holm-Hadulla R (1982) Der Konkretismus. Nervenarzt 53: 524–530

Howard R, Levy R (1992) Which factors affect treatment response in late paraphrenia? Int J Geriatr Psychiatry 7: 667–672

Howard R, Almeida O, Levy R (1994) Phenomenology, demography and diagnosis in late paraphrenia. Psychol Med 24: 397–410

Howard RJ, Graham C, Sham P et al (1997) A controlled family study of late-onset non-affective psychosis (late paraphrenia). Br J Psychiatry 170: 511–514

Huq SF, Garety PA, Hemsley DR (1988) Probabilistic judgements in deluded and non-deluded subjects. Quart J Exp Psychol 40A: 801–812

Janzarik W (1973) Über das Kontaktmangelparanoid des höheren Alters und den Syndromcharakter schizophrenen Krankseins. Nervenarzt 44: 515–526

Kaschka WP, Negele-Anetsberger J, Joraschky P (1991) Treatment outcome in patients with delusional (paranoid) disorder. Eur J Psychiatry 5: 240–253

Kaschka WP, Negele-Anetsberger J, Joraschky P (1992) Behandlungsergebnisse bei Patienten mit paranoiden Störungen. In: Kaschka WP, Lungershausen E (eds) Paranoide Störungen. Springer, Berlin Heidelberg New York, pp 131–144

Kay DWK, Roth M (1961) Environmental and hereditary factors in the schizophrenias of old age ("late paraphrenia") and their bearing on the general problem of causation in schizophrenia. J Ment Sci 107: 49–686

Kay DWK, Cooper AF, Garside RF, Roth M (1976) The differentiation of paranoid from affective psychoses by patients' premorbid characteristics. Br J Psychiatry 129: 207–215

Kendler KS (1982) Demography of paranoid psychosis (delusional disorder). Arch Gen Psychiatry 39: 890–902

Kendler KS, Hays P (1981) Paranoid psychosis (delusional disorder) and schizophrenia. A family study. Arch Gen Psychiatry 38: 547–551

Kendler KS, Masterson CC, Davis KL (1985) Psychiatric illness in first-degree relatives of patients with paranoid psychosis, schizophrenia and medical illness. Br J Psychiatry 147: 524–531

Kendler KS, Spitzer RL, Williams JBW (1989) Psychotic disorders in DSM-III-R. Am J Psychiatry 146: 953–962

Kimura B (1995) Zwischen Mensch und Mensch. Strukturen japanischer Subjektivität. Wissenschaftliche Buchgesellschaft, Darmstadt

Kora T (1999) Die Schinkeischitsu-Neurose. In: Katz L, Watanabe N (eds) Die Morita-Therapie im Gespräch. Psychosozial, Giessen, pp 46–100

Kraepelin E (1899) Psychiatrie, 6th edn. Barth, Leipzig

Kretschmer E (1966) Der sensitive Beziehungswahn, 4th edn. Springer, Berlin Heidelberg New York

Küchenhoff J (1984) Dysmorphophobie. Nervenarzt 55: 122–126

Küchenhoff J (1985) Das hypochondrische Syndrom. Nervenarzt 56: 225–236

Kuipers E, Garety PA, Fowler D et al (1997) London-East Anglia randomized controlled trial of cognitive-behavioural therapy for psychosis. I. Effects of the treatment phase. Br J Psychiatry 171: 319–327

Luhmann N (1973) Vertrauen. Ein Mechanismus zur Reduktion sozialer Komplexität, 2nd edn. Enke, Stuttgart

Magaro PA (1981) The paranoid and the schizophrenic: the case for distinct cognitive style. Schizophr Bull 7: 632–661

Mahorney SL, Cavenar JO (1988) A new and timely delusion: the complaint of having AIDS. Am J Psychiatry 145: 1130–1132

Malasi TH, El-Hilu SR, Mirza IA, El-Islam M, Fakhr (1990) Olfactory delusional syndrome with various aetiologies. Br J Psychiatry 156: 256–260

Marino C, Nobile M, Bellodi L, Smeraldi E (1993) Delusional disorder and mood disorder: can they coexist? Psychopathol 26: 5–61

Marneros A, Deister A, Rohde A (1988) Delusional parasitosis. A comparative study to late-onset schizophrenia and organic mental disorders due to cerebral arteriosclerosis. Psychopathology 12: 167–174

Matussek P (1963) Wahrnehmung, Halluzination und Wahn. In: Gruhle HW, Jung R, Mayer-Gross W, Müller M (eds) Psychiatrie der Gegenwart, vol I/2. Springer, Berlin Göttingen Heidelberg, pp 23–76

Minkowski E (1947) Phénoménologie et analyse existentielle en psychopathologie. Evol Psychiatr 4: 137–185

Moesler TA (1992) Eigengeruchswahn. In: Kaschka WP, Lungershausen E (eds) Paranoide Störungen. Springer, Berlin Heidelberg New York, pp 99–109

Morita S (1947) Shinkeishitsu no hontai oyobi ryoho ("The nature of Schinkeischitsu and its therapy"). Tokyo 1928

Morris M (1991) Delusional infestation. Br J Psychiatry 159: 83–87

Morrison NK (1987) The role of shame in schizophrenia. In: Lewis HB (ed) The role of shame in symptom formation. Erlbaum, London, pp 51–87

Mundt C (1996) Zur Psychotherapie des Wahns. Nervenarzt 67: 515–523

Munro A (1982) Paranoia revisited. Br J Psychiatry 141: 344–349

Munro A (1984) Excellent response of pathologic jealousy to pimozide. Canad Med Ass J 131: 852–853

Munro A (1988) Monosymptomatic hypochondriacal psychosis. Br J Psychiatry 153[Suppl 2]: 37–40

Musalek M (1991) Der Dermatozoenwahn. Thieme, Stuttgart

Musalek M, Kutzer K (1990) The frequency of shared delusions in delusions of infestations. Eur Arch Psychiatry Neurol Sci 239: 263–266

Musalek M, Berner P, Katschnig H (1989) Delusional theme, sex and age. Psychopathology 22: 260–267

Musalek M, Bach M, Passweg V, Jaeger S (1990) The position of delusional parasitosis in psychiatric nosology and classification. Psychopathology 23: 115–124

Naguib M, Levy R (1987) Late paraphrenia: neuropsychological impairment and structural brain abnormalities on computed tomography. Int J Geriatr Psychiatry 2: 83–90

Nelki J (1988) Making sense of a delusion of smell. A psychotherapeutic approach. Br J Med Psychol 61: 267–275

Opjordsmoen S (1988a) Hypochondriacal psychoses: a long-term follow-up. Acta Psychiatr Scand 77: 587–597

Opjordsmoen S (1988b) Long-term course and outcome in delusional disorder. Acta Psychiatr Scand 78: 576–586

Opjordsmoen S (1993) The duration criteria of delusional disorder in modern classification. Psychopathology 26: 85–89

Opjordsmoen S, Retterstøl N (1987) Hypochondriacal delusions in paranoid psychoses: course and outcome compared with other types of delusions. Psychopathology 20: 272–284

Opjordsmoen S, Retterstøl N (1991) Delusional disorder: the predictive validity of the concept. Acta Psychiatr Scand 84: 250–254

Phillips KA (1991) Body dysmorphic disorder: the distress of imagined ugliness. Am J Psychiatry 148: 1138–1149

Phillips KA (1996) Pharmakological treatment of body dysmorphic disorder. Psychopharmacol Bull 32: 597–605

Pollock BG (1982) Successful treatment of pathological jealousy with pimozide. Can J Psychiatry 27: 86–87

Rasmussen S (1978) Sensitive delusion of reference. "Sensitiver Beziehungswahn", some reflections on diagnostic practice. Acta Psychiatr Scand 58: 442–448

Retterstøl N (1966) Paranoid and paranoiac psychoses. Thomas, Springfield/IL

Retterstøl N (1970) Prognosis in paranoid psychosis. Thomas, Springfield/IL

Retterstøl N (1987) Nicht-schizophrene paranoide Entwicklungen und Paranoia. In: Kisker KP, Lauter H, Meyer JE, Müller C, Strömgren E (eds) Psychiatrie der Gegenwart, vol 4. Springer, Berlin Heidelberg New York, pp 211–235

Retterstøl N (1991a) Course and outcome in paranoid disorders. Psychopathology 24: 277–286

Retterstøl N (1991b) Erotomania – erotic self-reference psychosis in old maids. A long term follow-up. Psychopathology 24: 388–397

Riecher-Rössler A, Rössler W, Förstl H (1995) Late onset schizophrenia and late paraphrenia – a history of confusion about terms and concepts. Schizophr Bull 21: 345–354

Roberts GA (1991) Delusional belief and meaning in life: a preferred reality? Br J Psychiatry 159[Suppl 14]: 20–29

Saß H, Wittchen HU, Zaudig M (1996) Diagnostisches und statistisches Manual psychischer Störungen DSM-IV. Hogrefe, Göttingen

Schmidt-Degenhard M (1998) Zur Problemgeschichte und Psychopathologie der Paranoia. Fortschr Neurol Psychiatry 66: 313–325

Segal JH (1989) Erotomania revisited: from Kraepelin to DSM-III-R. Am J Psychiatry 146: 1261–1266

Sérieux P, Capgras J (1909) Les follies raisonnantes. Baillarger, Paris

Signer SF, Cummings JL (1987) De Clerambault's syndrome in organic affective disorder. Two cases. Br J Psychiatry 151: 404–407

Soyka M (1992) Zur Klinik des Eifersuchtswahns. In: Kaschka WP, Lungershausen E (eds) Paranoide Störungen. Springer, Berlin Heidelberg New York, pp 53–63

Soyka M, Naber G, Völcker A (1991) Prevalence of delusional jealousy in different psychiatric disorders. Br J Psychiatry 158: 549–553

Spitzer M (1989) Ein Beitrag zum Wahnproblem. Nervenarzt 60: 95–101

Strian F (1984) Die Dysmorphophobie als Kontraindikation kosmetischer Operationen. Handchir Mikrochir Plast Chir 16: 243–245

Tarrier N, Beckett R, Harwood S, Bishay N (1990) Morbid jealousy: a review and cognitive-behavioral formulation. Br J Psychiatry 157: 319–326

Teusch L, Köhler KH, Finke J (1987) Die Bearbeitung von Wahnphänomenen in der klientenzentrierten Gesprächspsychotherapie. In: Olbrich HM (ed) Halluzination und Wahn. Springer, Berlin Heidelberg New York, pp 168–173

Tölle R (1987) Wahnentwicklung bei körperlich Behinderten. Nervenarzt 58: 759–763

Tölle R (1993) Somatopsychic aspects of paranoia. Psychopathology 26: 127–137

Torch EM, Bishop ER (1981) Delusions of parasitosis: psychotherapeutic engagement. Am J Psychother 35: 101–106

Trabert W (1993) Epidemiologische Aspekte des Dermatozoenwahns. In: Möller HJ, Rohde A (eds) Psychische Krankheiten im Alter. Springer, Berlin Heidelberg New York, pp 180–187

von Baeyer W (1991) Begegnung als humaner Lebensvollzug und seine Störung im Wahn – Zur Psychopathologie der Perspektivenübernahme. In: Blankenburg W (ed) Wahn und Perspektivität. Enke, Stuttgart, pp 29–38

Watt JAG (1985) The relationship of paranoid states to schizophrenia. Am J Psychiatry 142: 1456–1458

Watzlawick P (1988) Münchhausens Zopf oder Psychotherapie und "Wirklichkeit". Huber, Bern

Winokur G (1977) Delusional disorder (paranoia). Compr Psychiatry 18: 511–521

Yamashita I (1993) Taijin-Kyofu or delusional social phobia. Hokkaido University Press, Hokkaido

Zaidens SH (1950) Dermatologic hypochondriasis: a form of schizophrenia. Psychosom Med 12: 250–253

CHAPTER

28

H. Beckmann, E. Franzek

Cycloid Psychoses and Their Differentiation from Affective and Schizophrenic Psychoses

Translator: E. Taub

1 Introduction

An understanding of the causes, prevention, treatment, and rehabilitation of the endogenous psychoses has been a major goal of research for decades, not only in psychology and the allied humanities, but also in biology and the natural sciences. Nonetheless, despite the tremendous progress of basic science research, knowledge in this area has remained sparse. The slow pace of developments in this field is frequently attributed to the genetic heterogeneity and multifactorial etiology of the endogenous psychoses, yet efforts to identify valid subgroupings of these diseases are still being made today.

Leonhard's (1999) concept of the cycloid psychoses can be traced back to the historical beginnings of scientific clinical psychiatry. This group of diseases is well characterized clinically, narrowly circumscribed, and reliably diagnosable. In the following sections, the heuristic significance of the cycloid psychoses will be discussed on the basis of both older and newer research findings.

2 Exceptional Position in Psychiatric Nosology

By subdividing the endogenous psychoses into manic-depressive disorders and dementia praecox, Kraepelin (1898, 1923) seemed at first to have introduced a certain order into the chaos that had previously reigned in the classification of mental illnesses. Psychoses with a favorable long-term prognosis were assigned to the group of manic-depressive disorders, while those with an unfavorable long-term prognosis were termed dementia praecox. Nonetheless, it was often impossible to determine the long-term prognosis of a psychosis from its momentary clinical manifestations, and it was found that psychoses resembling each other in the initial stages could either partly regress or result in more or less severe states of mental deficiency.

Kraepelin's concept of dementia praecox was, therefore, soon replaced by that of schizophrenia, or "splitting insanity" ("*Spaltungsirresein*"). Bleuer introduced a new designation, "the group of schizophrenias," with the following rationale:

It soon became clear ... that many diseases that could not be distinguished, in their psychopathological appearance, from psychoses leading to 'imbecility,' have a good prognosis, as does manic-depressive insanity. A term needed to be devised that would link together the disease forms with like symptomatology, even though some of them end in recovery, others in mental deficiency, and yet others in imbecility (Bleuler 1911, p. 47).

Bleuler's concept of schizophrenia has remained current until the present day, although it met with criticism from the beginning (Gruhle 1932) and has proved to be less than fully adequate for the purposes of either clinical practice or research. In this context, Schneider's famous remark on the diagnosis of schizophrenia seems to express resignation on his own part:

Among the numerous varieties of experience [i.e. disease manifestations] that arise in schizophrenia, there are some that we call symptoms of the first rank, not because we think of them as fundamental disorders, but because they have special importance for this diagnosis, in distinction to either nonpsychotic mental abnormalities or cyclothymia. Their value is, therefore, only with respect to diagnosis. Nothing is thereby said about the theory of schizophrenia. ... Wherever such varieties of experience are determined to exist, and no underlying physical diseases are to be found, we speak in all modesty of schizophrenia (Schneider 1967, p. 135).

Kurt Schneider's "atheoretical" view is also that of the currently used operationalized classification schemes, which are revised at regular intervals in accordance with expert polling and consensus and are intended to produce high interrater reliability, as they indeed do, though not uncommonly at the expense of clinical validity.

Independently of this development, Leonhard (1999) continued to pursue the type of clinically and empirically oriented research initiated by Kahlbaum (1863) and Kraepelin. Building on the earlier work of Wernicke (1900) and Kleist (1926), he created a differentiated nosology of the endogenous psychoses that went far beyond Kraepelin's prognostic dichotomy. Leonhard subdivided the two major groups of disease into further, nosologically independent diseases, on the basis of longitudinal studies, some of which extended over the patients' lifetime (Fig. 1).

As Fig. 1 makes clear, it must have been mainly the cycloid psychoses that led Bleuler to abandon prognostic diagnosis. These "atypical psychoses" fell into the manic-depressive disease group of Kraepelin, but they belong to the schizophrenic group according to Bleuler, Schneider, and the modern operationalized classification schemes (APA 1994; WHO 1991), because patients present with symptoms typical of schizophrenia in the momentary clinical view.

Fig. 1. Prognostic dichotomy of the endogenous psychoses

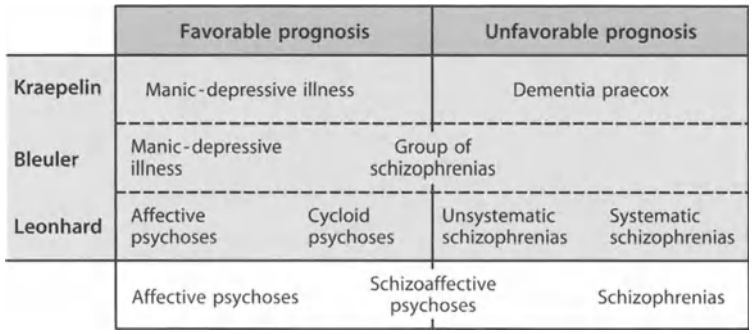


Table 1. Kraepelin's manic-depressive group of illnesses as subdivided by Leonhard (1999)

Phasic monopolar affective psychoses		Bipolar phasic psychoses	Cycloid psychoses
Pure mania	Pure melancholia	Manic-depressive disease	Anxiety-happiness psychosis
Monopolar euphorias	Monopolar depressions	–	Motility psychosis
–	–	–	Confusion psychosis

Table 2. Bleuler's group of schizophrenias as differentiated by Leonhard (1999)

Cycloid psychoses	Unsystematic schizophrenias	Systematic schizophrenias
Anxiety-happiness psychosis	Affective paraphrenia	Systematic paraphrenias
Motility psychosis	Periodic catatonia	Systematic catatonias
Confusion psychosis	Cataphasia	Hebephrenias

Table 1 shows Leonhard's differentiated subdivision of Kraepelin's manic-depressive group of diseases, and Table 2 shows Leonhard's differentiation of Bleuler's "group of schizophrenias."

3 Clinical Aspects

The bipolar cycloid psychoses had already been distinguished from manic-depressive illness and the schizophrenias as a nosologically independent group of diseases by Kleist (1926) and by Fünfgeld (1936), but Leonhard was the first to lay down precise symptomatological criteria for their accurate diagnosis. He distinguished three clinical forms:

- Anxiety-happiness psychosis
- Excited-inhibited confusion psychosis
- Hyperkinetic-akinetic motility psychosis

The characteristic manifestations of these individual subtypes are detailed in Tables 3–5, and three brief case presentations will serve to illustrate them:

1. *Anxiety-happiness psychosis.* The patient first became ill at the age of 27. She was referred to a psychiatric clinic with the diagnosis of "paranoid-hallucinatory psychosis." On admission, she was tense and very mistrustful and had no insight into her illness. She complained of fatigue during the day and derealization phenomena. She stated that she had been feeling happy at times, but very anxious at other times, for 4 weeks. She was discharged in an asymptomatic condition after a 1-month inpatient stay. One year later, she was hospitalized a second time when she suddenly ran around naked, said she was afraid of being poisoned, and told her husband to look her in the eye, because she was God. During the psychiatric examination on admission, she was very lively and alert. She said that she felt the power of God within herself and believed that all people would love each other from now on and that there would be no more war. This time, too, she was discharged in good health, but she soon required rehospitalization because she had developed the delusion that she was both God and "the archangel" and had attempted to prove it by laying her hand on a hot stove. Her mood in the hospital alternated between elevated and anxious, and she had

Table 3. Manifestations of anxiety–happiness psychosis

Anxiety pole	Happiness pole
Severe anxiety with ideas of reference and delusional perceptions	Ecstatic mood with ideas of mission, happiness, and salvation
Anxious, paranoid ideas of being threatened, persecuted, and annihilated	Affect-congruent illusions and hallucinations (often as visions)
Affect-congruent illusions and hallucinations	Ecstatic facial expression, pathetic movements and gestures
Anxiety-colored physical misperceptions	–
Ideas of self-sacrifice for the salvation of others	
Rapid fluctuations between anxiety and happiness	

Table 4. Manifestations of excited–inhibited confusion psychosis

Excited pole	Inhibited pole
Incoherent thoughts with logorrhea	Inhibition of thought with perplexity
Incoherent choice of subject (digressiveness)	Perplexed ideas of significance (delusional perceptions)
Transient misidentification of persons	Delusions of interpretation with perplexity
Delusions of interpretation	Illusions and hallucinations
Transient hallucinations	–
Anxious and ecstatic fluctuations of mood, of much lesser intensity than in anxiety–happiness psychosis	

Table 5. Manifestations of hyperkinetic–akinetic motility psychosis

Hyperkinetic pole	Akinetic pole
Increase of expressive and reactive movements	Loss of expressive and reactive movements
Senseless, purposeless compulsive movement	Flaccid or rigid akinesia
Severe distractibility by environmental stimuli (hyperreactive distractibility, hypermetamorphosis after Wernicke)	Slow, sluggish or absent voluntary movement
Occasionally incoherent speech	Loss of spontaneous speech, up to mutism
Illusions, hallucinations, and delusions of interpretation	
Rapidly alternating affect (elevated–anxious–irritated–depressed)	

delusions of interpretation and poverty delusions. She again recovered completely. The next phase of the illness occurred when she was 36 years old. This time, she expressed multiple ideas of a religious nature, claiming, among other things, that the Pope had appeared to her. She then complained again of severe anxiety, fearing that a world war and the end of the world were approaching. After discharge, she remained healthy and able to manage her household for 6 years. She became ill again at age 42, when she developed severe anxiety and felt threatened and persecuted by a man in her neighborhood. She also heard threatening voices. This time, too, she was discharged in good condition and had good insight into her illness.

2. *Confusion psychosis.* The patient was first hospitalized at age 17 because of a depressive mood disturbance. After being discharged, he passed an examination for a sales position. Three years later, he occasionally became very depressed and spoke extremely little. He was hospitalized a second time at age 23. He was anxious and had severe cognitive inhibition. He had recovered fully by the time of discharge. One year later, he had to be rehospital-

ized because of "thoroughly disordered" thinking associated with a "cheerful and silly" mood. Over the course of the hospital stay, times of excitement, in which he continuously "talked nonsense," alternated with states of perplexed inhibition. He regained mental balance and was discharged. In the ensuing years, he was hospitalized repeatedly. A severe formal cognitive disorder was always at the center of the psychopathology. Frequently, there was logorrhea of incomprehensible and incoherent content. On the other hand, there were also repeated occurrences of severely inhibited, perplexed, mutistic states. His mood was either euphorically elevated or depressed and anxious. Threatening, frightening voices and misidentification of people occurred at varying degrees of severity. He was well between the individual disease phases and was able to assist his elder brother with farm work.

3. *Motility psychosis.* This patient first became ill at age 19. She made "silly, clownish" gestures, stood in place with a facial expression of "quiet rapture," and held her hands, arms, and trunk in "peculiar, bizarre postures." She suffered from auditory hallucinations and sometimes misidentified people. According to her hospital record, she lay in bed moving her arms about constantly and waved her upper body and head back and forth "with a certain degree of grace." She continually repeated words that she heard being spoken around her. Later, she was "silly and in high spirits" and constantly in motion, threw slippers and pillows all over the place, and poured coffee on the floor. After 4 months, she was calmer and was discharged in good condition. Three years later, she became ill again. This time, according to the hospital record, she threw herself back and forth in bed and made noises expressing pleasure. There were accompanying auditory and visual hallucinations. On being discharged 3 months later, she was well again. There were six more inpatient hospitalizations until she reached the age of 60. Psychomotor abnormalities were prominent during all of these. The daily progress notes document, among other things, that she breathed in a mannered fashion and twitched in a comical, clownish way whenever she was touched. She displayed considerable psychomotor agitation, and she sang in a high soprano pitch. She spoke confusedly, with logorrhea, and was in a state of psychomotor excitation; she slammed doors, her face was rigid like a mask, and all of her movements were slowed. She always recovered fully between phases. On an outpatient visit at the age of 64, she was found to be mentally balanced, friendly, and able to modulate her affect. She had been widowed shortly before and lived alone in a well-kept detached house.

All three subtypes of cycloid psychosis are characterized by bipolarity, a phasically remitting course, and a lack of residual mental abnormalities. Leonhard conceded that a certain degree of impairment of internal energy might develop after multiple phases and hospitalizations, but considered this to be a reactive phenomenon. Several more recent studies have confirmed that the long-term prognosis is favorable (Beckmann et al. 1990; Maj 1990; Perris 1974).

These diagnoses are made in accordance with hierarchical clinical conceptions, and each of them has its own characteristic manifestations. If these manifestations are lacking, the diagnosis must be questioned. The diagnosis of a cycloid psychosis according to Leonhard is, in this sense, highly operationalized (as are all of Leonhard's diagnoses). After careful observation for a sufficiently long time, a precise diagnosis of one of the subtypes can usually be made. Leonhard repeatedly stressed that the illness of a given patient belongs to the overall group of cycloid psychoses only insofar as it can be confidently and precisely diagnosed as one of the recognized subtypes and that it is therefore not sufficient to diagnose a case of "cycloid psychosis." Misdiagnoses are particularly likely to occur if a diagnosis is constructed out of incomplete syndromes or if a diagnosis of cycloid psychosis is made only because the course is favorable.

As the above remarks will have made clear, the grouping together of diagnostic criteria into an overall complex called "cycloid psychosis," as was advocated by Brockington et al. (1982) and, in fact, carried out in ICD-10 under the synonymous term "acute polymorphic psychotic disorders," leads to a loss of clinical validity. The cycloid psychoses as originally described by Leonhard are found in all diagnostic categories of the operationalized classification systems, depending on their particular manifestations (Franzek et al. 1996; Franzek and Beckmann 1998a,b) and, despite some overlap, are not identical to any of the defined categories; in particular, they are not identical to the schizophreniform psychoses of DSM-III and DSM-IV or to the acute transient psychotic disorders of ICD-10 (Franzek et al. 1994; Pfuhlmann 1998).

4 Important Considerations in Clinical Practice

Ever since Bleuler's time, the diagnosis of schizophrenia has been associated with an uncertain prognosis. Neuroleptic prophylaxis is therefore recommended for nearly every case of psychosis with schizophrenic symptoms, not limited to acute psychotic episodes (DGPPN 1998). We know from our own clinical experience, however, that neuroleptic phase

prophylaxis often fails to prevent relapses of the cycloid psychoses. Perris (1978, 1986) stated emphatically that neuroleptics are useful for the acute therapy of cycloid psychoses, but he also found that they were significantly less effective than lithium when administered chronically for phase prophylaxis. Indeed, it was the effectiveness of lithium for relapse prevention in affectively colored intermittent psychoses with schizophrenic manifestations that led to the removal of these entities from the broad group of schizophrenias as defined by Bleuler and contributed to the ascendance of the term "schizoaffective psychoses," a vague diagnosis of little clinical utility.

A question for future study is whether low doses of "atypical" neuroleptics of the current generation, which have considerably fewer side effects, might be superior to classical neuroleptic treatment for phase prophylaxis or relapse prevention in the cycloid psychoses. Comparative studies with lithium, carbamazepine, or valproate would be desirable.

Franzek et al. (1994) reported that both malignant neuroleptic syndrome and life-threatening catatonic syndrome tend to occur preferentially in the course of cycloid psychoses, especially motility psychoses. Careful analysis of their own data and of the extensive literature on the subject led these authors to postulate that these two syndromes, when they complicate the course of acute cycloid psychoses, are very likely to be identical. They therefore recommend that no time be wasted on the pointless task of differential diagnosis, that neuroleptics be discontinued as the first step in the treatment of every patient, and that all other necessary therapeutic measures be initiated immediately afterward. The authors warn that the continued administration of highly potent neuroleptics, on the assumption that the patient's condition is the result of pernicious catatonia, will not only fail to improve this syndrome, but will actually make it worse.

Lastly, it is a great relief when a good prognosis can be delivered, and such is the case in postpartum psychotic illnesses, the great majority of which are cycloid psychoses (Pfuhlmann et al. 1998). The psychological comfort that a good prognosis affords the patient's entire family is inestimable.

5 Research Findings

Leonhard himself repeatedly stressed that a hereditary predisposition to disease plays only a small role in the cycloid psychoses. In his family studies, he found that only 4%–5% of patients' first-degree relatives were affected with these conditions – a figure considerably smaller than those for manic-depressive illness (20%–21%) and the unsystematic schizophrenias (13%–22%), which are more clearly hereditary. Only the figure for the systematic schizophrenias was lower (2%–4%) (Leonhard 1999).

New family studies performed to a high methodological standard have since confirmed a number of Leonhard's findings. Affective psychoses were found in 6.6% of the first-degree relatives of patients with cycloid psychoses, while there were no cases of schizophrenic psychoses in these families (Franzek and Beckmann 1998b). More than 20% of the first-degree relatives of patients with periodic catatonia (a clinical subgroup of the unsystematic schizophrenias) were similarly affected, while fewer than 5% of the first-degree relatives of patients with systematic catatonia had that condition (Beckmann et al. 1996; Stöber et al. 1995, 1997).

Another focus of recent investigation is the question of the relative importance of genetic and environmental factors in the causation of the cycloid psychoses. Twin studies, in which pairs of twins with at least one affected member are systematically recruited, are the most appropriate method for addressing this issue, but such studies are rarely performed today because of their strict methodological requirements.

The classical twin-study method is based on a comparison of monozygotic and dizygotic twin pairs and assumes that both kinds of twins are subject to the same environmental influences. According to Galton's rule, which remains valid today, heritability is implied if monozygotic twins are concordant for a particular trait more often than dizygotic twins. Conversely, monozygotic and dizygotic twins should not have significantly different rates of concordance and discordance for nonheritable traits. All phenotypic differences between monozygotic, and therefore genetically identical, twins are assumed to be due to environmental factors. Twin studies must always take possible complications of pregnancy and delivery into account, because the prenatal "environments" of twins, and particularly monozygotic twins, may be markedly different – the well-known twin transfusion syndrome is an extreme example.

Franzek and Beckmann (1998a,b) performed a systematic twin study meeting all of these methodological requirements. They investigated 12 monozygotic and 11 dizygotic twin pairs in which at least one twin suffered from a type of cycloid psychosis. A comparison of the proband-associated concordance rates of monozygotic and dizygotic twins clearly disproved the hypothesis of a primary genetic disposition. A total of 36% of affected monozygotic twins, and 31% of affected dizygotic twins, had an affected partner; these figures are not significantly different. This negative finding is quantitatively reflected in the low calculated heritability index of 0.14 (the maximum

possible value is 1) and the low monozygotic-to-dizygotic ratio of 1.16 (the minimum expected value is 1).

For the unsystematic schizophrenias, however, the findings were entirely different, in that significant differences between monozygotic and dizygotic twin pairs were found. The proband-associated concordance rate was 89% for monozygotic twins and only 25% for dizygotic twins. Accordingly high values were obtained for the heritability index (0.72) and the monozygotic-to-dizygotic ratio (3.56). Thus, according to Galton's rule, there is a major primary genetic predisposition for the unsystematic schizophrenias, but not for the cycloid psychoses. A further, remarkable finding of this twin study was the lack of monozygotic twins with systematic schizophrenia, despite systematic recruitment. The explanations proposed to date for this observation are all speculative, and a further discussion would be beyond the scope of this chapter (Leonhard 1999; Franzek and Beckmann 1998b).

Because genetic factors apparently play little or no role in the causation of the cycloid psychoses, predisposing environmental factors must be sought. Twin studies may give evidence for such factors in the form of environmental differences between affected and unaffected monozygotic twins; this is a situation in which the individual's genetic constitution is held constant, while the environment is the variable under study. The study referred to above involved 12 pairs of monozygotic twins in which the index twin had a cycloid psychosis. Three of these pairs were concordant and nine were discordant, i.e. the co-twin in nine pairs was unaffected. It was found that the index twins of the discordant pairs had had more frequent and more severe perinatal complications than their unaffected co-twins, to a high level of statistical significance ($p < 0.01$). This finding is in accordance with the conclusion of earlier studies of discordant twin pairs, to the effect that "schizophrenic" monozygotic twins had severe perinatal complications more frequently than their healthy co-twins.

Consideration of the case histories of the patients in the latter studies reveals that most of the "schizophrenic" probands were probably suffering from cycloid psychoses, according to Leonhard's definition (Pollin et al. 1965, 1966). It is also known today that perinatal complications probably represent only the last stage of damage sustained over the course of pregnancy (Nelson and Ellenberg 1986; Kuban and Leviton 1994) and that prenatal insults may affect only one of a pair of twins, even if they are monozygotic (Davis et al. 1995). Thus, in many cases, perinatal complications may be an expression of preexisting disorders of fetal development.

Of course, developmental disorders of the central nervous system might occur without any necessary association with severe perinatal complications. Figure 2 shows the magnetic resonance images (MRI) of a pair of 26-year-old male twins whom we recently had the opportunity to study. Twin A had acutely developed a cycloid psychosis (confusion psychosis) that fully resolved after a few weeks of treatment with perazine. His brother, twin B, was psychologically normal and had been healthy his entire life. According to their mother, the twins had been born by cesarean section (birth weights: twin A, 3000 g; twin B, 3100 g). There had been no other complications perinatally,

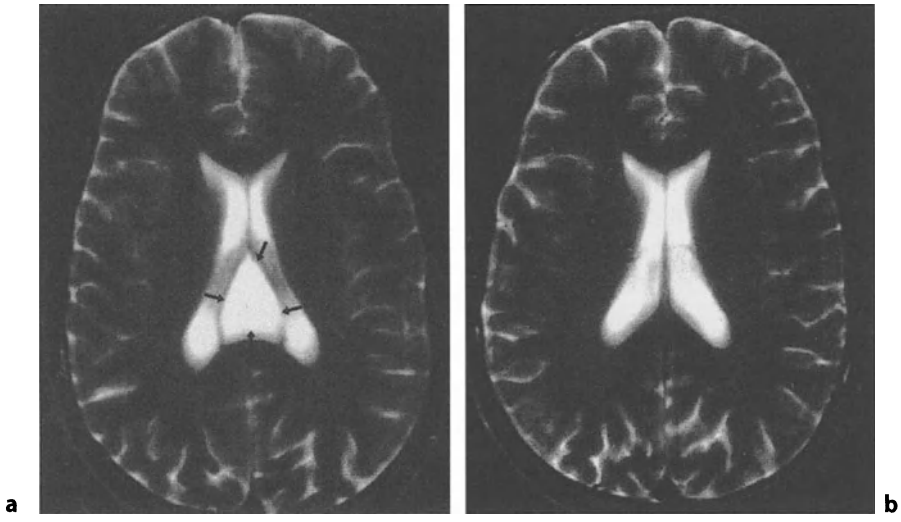


Fig. 2a,b. Magnetic resonance imaging (MRI) at the level of the lateral ventricles in two 26-year-old monozygotic twins. **a** Twin A: cycloid psychosis (confusion psychosis). *Small arrow*, cavum velum interpositum. **b** Twin B: normal

postnatally, or in early childhood development. Both twins passed the *Abitur* (high-school graduation examination qualifying for university) and were studying at the university level at the time of twin A's hospitalization.

As seen in Fig. 2, twin A, who suffered from an acute confusion psychosis, had a septal anomaly in the form of a cavum velum interpositum. This abnormality most likely reflects a developmental anomaly of cerebral maturation. The MRI of the healthy twin, twin B, is wholly unremarkable. For twin A, the septal anomaly itself had had no adverse consequences, neither perinatally nor in later development.

Unfortunately, the more recent studies of discordant twins contain no case histories, but merely speak in global terms of "DSM-III(-R) schizophrenia." We know from our own experience, however, that the index cases in pairs of monozygotic twins discordant for DSM-III(-R) schizophrenia suffer mainly from cycloid psychoses. Discussion of these studies is therefore appropriate here. Reveley et al. (1982) found a considerable difference of ventricular width between monozygotic twins discordant for "schizophrenia": the "schizophrenic" probands had significantly wider ventricles than their own healthy co-twins or than other, healthy, monozygotic twin controls. This is all the more remarkable as ventricular width is known to be largely genetically determined (Bartley et al. 1997).

Similar findings, i.e. wider ventricles in the affected twin, were reported by both Casanova et al. (1990) and Suddath et al. (1990). The latter study is of particular interest, because the neuroradiologist, who was blinded to the psychiatric diagnosis, was able to identify the affected twin in 12 of 15 cases merely by visual inspection of the MRI images for enlargement of the cerebrospinal fluid spaces. In accordance with these findings, Franzek et al. (1996), in a comparison study of psychiatric patients with unspecified abnormalities on cerebral computed tomography (CT) and a matched control group with normal CT scans, found a disproportionately high number of patients with cycloid psychoses in the group with abnormal CT scans. This observation was statistically significant. The abnormalities on CT consisted mainly of ventricular asymmetry and/or enlargement and were said by a neuroradiologist to be the likely result of pre- or perinatal insults.

The question naturally arises as to what the nature of these insults might be. An observation directly relevant to this question has been replicated by independent groups more often than any other in schizophrenia research and has given rise to much speculation: schizophrenics are significantly more likely than the general population to have been born in the winter or spring months (Bradbury and Miller 1985). A detailed discussion of all of the proposed explanations for this phenomenon would be beyond the scope of this chapter. The favored hypothesis at present is that insults arising preferentially in the colder seasons may affect a vulnerable phase of fetal cerebral development in such a way that the individual is predisposed to develop a schizophrenic psychosis in adulthood (Torrey 1987).

It has been repeatedly reported in recent years that excess births in the winter and spring months are found particularly among schizophrenics without schizophrenic relatives (D'Amato et al. 1991; O'Callaghan et al. 1991). Franzek and Beckmann (1992) studied the phenomenon of birth seasonality in a large group of patients diagnosed by Leonhard as having cycloid psychoses and unsystematic and systematic schizophrenias. There was an excess of winter and spring births in the overall group as compared to the general population, as expected. Consideration of the individual diagnostic groups, however, led to the surprising result that the excess was limited to patients with cycloid psychoses and systematic schizophrenias (both types of psychoses that tend not to be familial), while patients with the more familial unsystematic psychoses actually manifested a deficit of births in the months in question (Fig. 3).

The birth deficit among patients with unsystematic schizophrenias was interpreted as implying that individuals whose cerebral development is already impaired by a genetic defect may be less likely to survive if they sustain a further, exogenous insult to the central nervous system (Beckmann and Franzek 1992). The individuals that do not survive do not appear in later birth statistics, of course. On the other hand, if there is no primary genetic defect, the precise moment and anatomical location of the insult are probably what determine whether the individual will be predisposed to developing a cycloid psychosis or a systematic schizophrenia. Both of these hypotheses are entirely consistent with the "neurodevelopmental concept" of schizophrenic psychoses (Jakob and Beckmann 1986); evidence for disturbances of cerebral development has been found in both familial and sporadic forms of schizophrenia (Beckmann and Jakob 1991).

Maternal viral illnesses such as influenza A are considered likely to be a cause of these exogenous insults, as it is now known that they may damage the fetus directly or indirectly through a number of mechanisms (e.g. receptor-mediated endocytosis, axonal retrograde transport, transcytosis, overshoot of the maternal immune response; Franzek and Beckmann 1996). Attention was first drawn to the influenza A virus by the observation by Mednik et al. (1988) that significantly more schizophrenics had been born in the wake of an influenza A epidemic than in the periods immediately before and after. Although this finding has been disputed (Crow and Done 1992), more precise

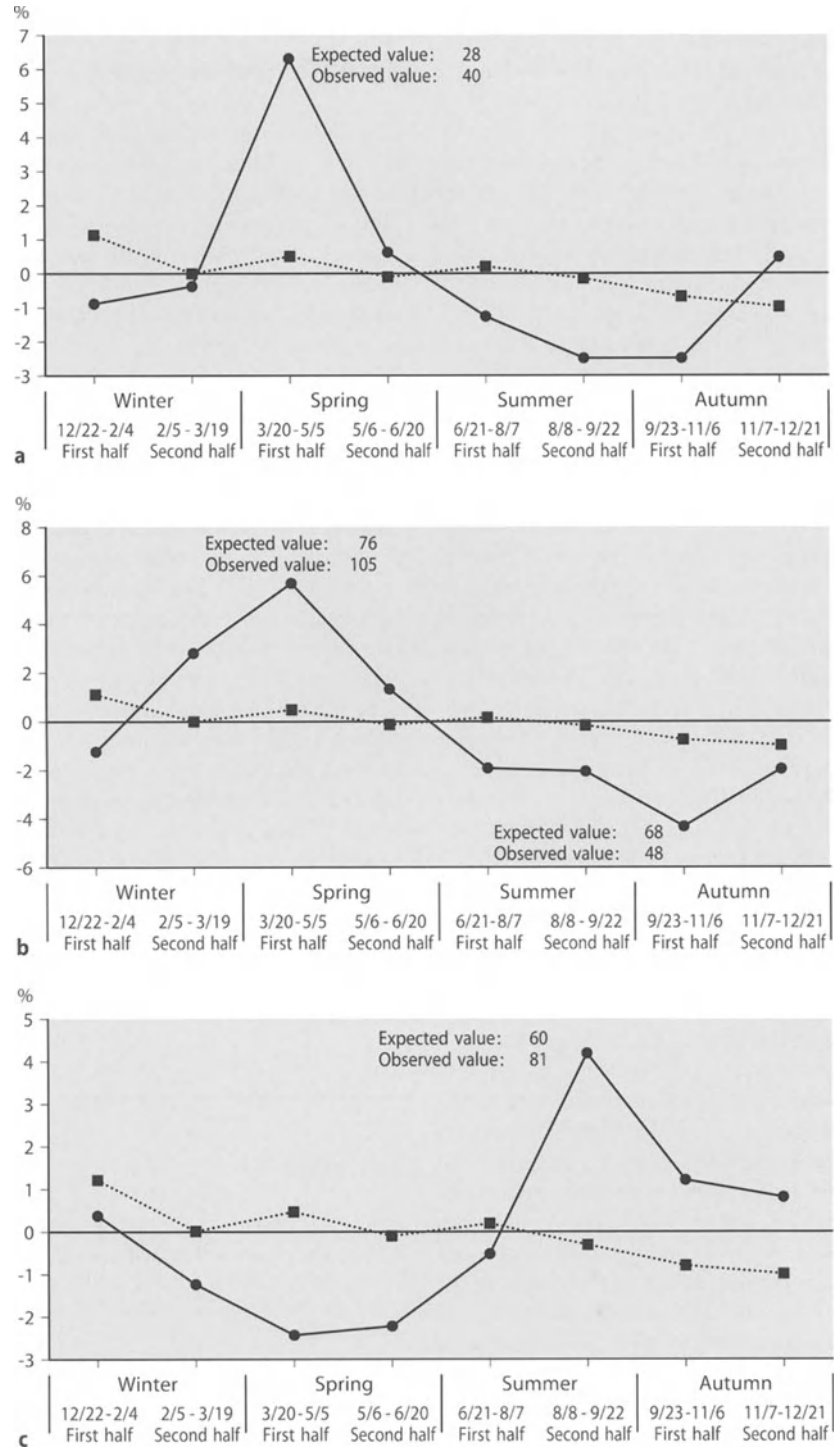


Fig. 3. Seasonal distribution of births in patients with **a** cycloid psychoses, **b** systematic schizophrenias, and **c** unsystematic schizophrenias as compared to the general population. *Dotted lines*, distribution of births in the normal population; *solid lines*, distribution of births among patients with psychosis/schizophrenia

analysis of the data reveals that an increased frequency of schizophrenia was always found among individuals living in densely populated areas. Negative findings, in contrast, were predominantly obtained in thinly or variably populated areas. The effect of an epidemic on the population is known to be more severe wherever people live more closely together, and less severe where there is more space between affected areas; thus the negative reports do not necessarily belie a causal relationship between the influenza A epidemic and the elevated frequency of schizophrenic psychoses.

Stöber et al. (1992, 1997) were the first to document a direct relationship between maternal infectious diseases and the later occurrence of cycloid psychoses and systematic (sporadic) schizophrenias. When questioned retrospectively, the mothers of patients of cycloid psychoses reported that they had flu-like and febrile illnesses in the first trimester of pregnancy more frequently than control mothers. Among mothers of patients with systematic schizophrenia, an elevated occurrence of viral illness as compared to control mothers was found in the second trimester of pregnancy. In contrast, the mothers of patients with manic-depressive illness or any form of unsystematic schizophrenia did not differ significantly from the control mothers. These findings have yet to be confirmed by prospective studies, and they can therefore be interpreted only in a preliminary and hypothetical fashion. It has been proposed that viral (or other) insults affecting the prenatal development of certain parts of the brain are involved in the causation of cycloid psychoses and systematic schizophrenias. The timing, localization, and extent of cerebral damage apparently determine the degree to which the individual then becomes vulnerable to the different forms of psychosis.

Lastly, functional investigative methods such as the measurement of P300 have revealed significant differences between patients with cycloid psychosis and patients with either manic-depressive illness or schizophrenic psychoses (Strik et al. 1996). The topography of P300 in patients with cycloid psychoses did not differ from that of controls, while a displacement to the right was observed in schizophrenics (both systematic and unsystematic). The P300 amplitude, compared to that of control subjects, was lower in schizophrenics, no different in manic-depressives, and elevated in patients with cycloid psychoses. The latter finding probably reflects an elevated state of arousal of the central nervous system. These results are in accordance with the finding by Warkentin et al. (1992) that patients in the acute phase of cycloid psychosis have a significantly elevated mean hemispheric blood flow. Schizophrenics, on the other hand, are known to have regionally decreased cerebral perfusion (hypofrontality).

6 Conclusion and Overview

The division of endogenous psychoses into affective and schizophrenic types, which still prevails today, has failed to bring about a decisive breakthrough either in clinical practice or in the scientific study of these diseases. Psychoses that do not fit neatly into either category are designated as atypical, schizophreniform, schizoaffective, or (more recently) transient polymorphic disorders and are assigned to either the affective or the schizophrenic category, depending on which of the two basic orientations is preferred, that of Kraepelin or that of Bleuler.

Leonhard's classification of the endogenous psychoses, whose roots extend back to the earliest days of psychiatry as a medical specialty, offers a promising alternative for the resolution of this dilemma. His concept of the cycloid psychoses as an independent nosological category distinct from both manic-depressive illness and the schizophrenias has proved especially useful in recent years.

Leonhard's original conception of the cycloid psychoses has been largely confirmed by current research. Indeed, modifications of his concept, especially those toward greater simplicity, have been found to result in a loss of heuristic value. A renewed emphasis on the precise characterization of individual manifestations and syndromes, as in the works of Wernicke, Kraepelin, Kleist, and Leonhard, holds great promise as a means of reviving progress in the scientific understanding of the etiology, genetics, prognosis, and differential treatment of the endogenous psychoses.

7 References

APA (1994) Diagnostic and statistic manual of mental disorders, 4th edn. American Psychiatric Association, Washington DC

Bartley AJ, Jones DW, Weinberger DR (1997) Genetic variability of human brain size and cortical gyral patterns. Brain 120: 257–269

Beckmann H, Franzek E (1992) Deficit of birthrates in winter and spring months in distinct subgroups of mainly genetically determined schizophrenia. Psychopathology 25: 57–64

Beckmann H, Jakob H (1991) Prenatal disturbances of nerve cell migration in the entorhinal region: a common vulnerability factor of functional psychoses? J Neural Transm (Gen Sect) 84: 155–164

Beckmann H, Fritze J, Lanczik M (1990) Prognostic validity of the cycloid psychoses. A prospective follow-up study. Psychopathology 23: 205–211

**Beckmann H, Franzek E, Stöber G (1996) Genetic heterogeneity in catatonic schizophrenia. A family study. Am J Med Genet 67: 289–300

Bleuler E (1911) Dementia praecox oder die Gruppe der Schizophrenien. In: Aschaffenburg G (ed) Handbuch der Psychiatrie. Deuticke, Leipzig, p 47

Bradbury TN, Miller GA (1985) Season of birth in schizophrenia: A review of evidence, methodology and etiology. Psychol Bull 98: 569–594

Brockington IF, Perris C, Kendell RE, Hillier VE, Wainwright (1982) The course and outcome of cycloid psychosis. Psychol Med 12: 97–105

Casanova MF, Sanders RD, Goldberg TE, Bigelow LB, Christison G, Torrey EF, Weinberger DR (1990) Morphometry of the corpus callosum in monozygotic twins discordant for schizophrenia: a magnetic resonance imaging study. J Neurol Neurosurg Psychiatry 53: 416–421

Crow TJ, Done DJ (1992) Prenatal exposure to influenza does not cause schizophrenia. Br J Psychiatry 161: 390–393

D'Amato D, Dalery J, Rochet T, Terra JL, Marie-Cardine M (1991) Saisons de naissannce et psychiatrie. Etude rétrospective d'une population hospitière. Encéphale 17: 67–71

Davis J, Phelps A, Bracha HS (1995) Prenatal development of monozygotic twins and concordance for schizophrenia. Schizophr Bull 21: 13–18

DGPPN (Deutsche Gesellschaft für Psychiatrie, Psychotherapie und Nervenheilkunde) (1998) Praxisleitlinien in Psychiatrie und Psychotherapie. In: Gaebel W, Falkai P (eds) Behandlungsleitlinie Schizophrenie, vol 1. Steinkopf, Darmstadt, pp 1–59

Franzek E, Beckmann H (1992) Season-of-birth effect reveals the existence of etiologically different groups of schizophrenia. Biol Psychiatry 32: 375–378

**Franzek E, Beckmann H (1996) Gene-environment interaction in schizophrenia: season-of-birth effect reveals etiologically different subgroups. Psychopathology 29: 14–26

Franzek E, Beckmann H (1998a) Different genetic background of schizophrenia spectrum psychoses: a twin study. Am J Psychiatry 155: 76–83

Franzek E, Beckmann H (1998b) Psychosen des schizophrenen Spektrums bei Zwillingen. Springer, Heidelberg Berlin New York

Franzek E, Stöber G, Beckmann H (1994) Malignes neuroleptisches und akut lebensbedrohlich katatones Syndrom. Eine identische Komplikation im Verlauf von funktionellen Psychosen. Neuropsychiatrie 8: 151–158

Franzek E, Becker T, Hofmann E, Flöhl W, Stöber G, Beckmann H (1996) Is computerized tomography ventricular abnormality related to cycloid psychosis? Biol Psychiatry 40: 1255–1266

Fünfgeld E (1936) Die Motilitätspsychosen und Verwirrtheiten. Karger, Berlin

Gruhle HW (1932) Theorie der Schizophrenie. In: Bumke O (ed) Handbuch der Geisteskrankheiten, vol 9, part 5. Springer, Berlin, pp 705–714

Jakob H, Beckmann H (1986) Prenatal developmental disturbances in the limbic allocortex in schizophrenia. J Neural Transm 65: 303–326

Kahlbaum K (1863) Die Gruppierung der psychischen Krankheiten und die Einteilung der Seelenstörungen. Kofemann, Danzig

Kleist K (1926) Über zykloide Degenerationspsychosen, besonders Verwirrtheiten und Motilitätspsychosen. Zentralbl Ges Neurol Psychiatr 44: 265–267

Kraepelin E (1898) Psychiatrie, 5th edn. Barth, Leipzig

Kraepelin E, Lange J (1923) Psychiatrie, vol 3: Klinische Psychiatrie, 8th edn. Barth, Leipzig

Kuban KCK, Leviton A (1994) Cerebral palsy. N Engl J Med 330: 188–195

**Leonhard K (1999) Classification of endogenous psychoses and their differentiated etiology, 2nd edn. Springer, Berlin Heidelberg New York

Maj M (1990) Cycloid psychotic disorder: validation of the concept by means of a follow-up and a family study. Psychopathology 23: 196–204

Mednick SA, Machon RA, Huttunen MO, Bonett D (1988) Adult schizophrenia following prenatal exposure to an influenza epidemic. Arch Gen Psychiatry 45: 189–192

Nelson KB, Ellenberg JH (1986) Antecedents of cerebral palsy: multivariate analysis of risk. N Engl J Med 315: 81–86

O'Callaghan E, Gibson T, Colohan HA, Walshe D, Buckley P, Larkin C, Waddington JL (1991) Season of birth in schizophrenia. Evidence for confinement of an excess of winter births to patients without a family history of mental disorder. Br J Psychiatry 158: 764–769

Perris C (1974) A study of cycloid psychoses. Acta Psychiatr Scand Suppl 253: 7–64

Perris C (1978) Morbidity suppressive effect of lithium carbonate in cycloid psychosis. Arch Gen Psychiatry 35: 328–331

Perris C (1986) The case for the independence of cycloid psychotic disorder. In: Marneros A, Tsunag MT (eds) Schizoaffective psychoses. Springer, Berlin Heidelberg New York, pp 272–308

Pfuhlmann B (1998) Nosologie der zykloiden Psychosen. Psycho 24: 338–345

**Pfuhlmann B, Stöber G, Franzek E, Beckmann H (1998) Cycloid psychoses predominate in severe postpartum psychiatric disorders. J Affect Disord 50: 125–134

Pollin W, Stabenau JR, Tupin J (1965) Family studies with identical twins discordant for schizophrenia. Psychiatry 28: 60–78

Pollin W, Stabenau JR, Mosher L, Tupin J (1966) Life history differences in identical twins discordant for schizophrenia. Am J Orthopsychiatry 36: 492–509

Reveley AM, Reveley MA, Clifford CA, Murray RM (1982) Cerebral ventricular size in twins discordant for schizophrenia. Lancet i: 540–541

Schneider K (1967) Klinische Psychopathologie, 8th edn. Thieme, Stuttgart

Stöber G, Franzek E, Beckmann H (1992) The role of maternal infectious diseases during pregnancy in the etiology of schizophrenia in the offspring. Eur Psychiatry 7: 147–152

Stöber G, Franzek E, Lesch KP, Beckmann H (1995) Periodic catatonia: a schizophrenic subtype with dominant inheritance and anticipation. Eur Arch Psychiatry Clin Neurosci 245: 135–141

**Stöber G, Kocher I, Franzek E, Beckmann H (1997) First trimester maternal gestational infections and cycloid psychosis. Acta Psychiatr Scand 96: 319–324

Strik WK, Fallgatter AJ, Stöber G, Franzek E, Beckmann H (1996) Specific P300 features in patients with cycloid psychosis. Acta Psychiatr Scand 94: 471–476

Suddath RL, Christison GW, Torrey EF, Casanova MF, Weinberger DR (1990) Anatomical abnormalities in the brains of monozygotic twins discordant for schizophrenia. N Engl J Med 322: 789–794

Torrey EF (1987) Hypotheses on the seasonality of schizophrenic births. In: Cazullo L, Invernizzi G, Sacchetti E (eds) Etiopathogenetic hypotheses of schizophrenia: the impact of

epidemiological, biochemical and neuromorphological studies. MTP, Lancaster, pp 41–48

Warkentin S, Nilsson A, Karlson S, Risberg J, Franzen G, Gustafson L (1992) Cycloid psychosis: regional cerebral blood flow correlates of a psychotic episode. Acta Psychiatr Scand 85: 23–29

Wernicke C (1900) Grundriß der Psychiatrie in klinischen Vorlesungen. Thieme, Leipzig

WHO (1991) Tenth revision of the International Classification of Diseases, chapter V (F): mental and behavioral disorders. World Health Organization, Geneva

CHAPTER

29

M.T. Tsuang, J.C. Simpson, J.A. Fleming

Schizoaffective Disorder

Preparation of this chapter was supported in part by grants R37-MH43518, U01-MH46318, and R01-DA04604 from the National Institutes of Mental Health and Drug Abuse and by a Veterans Affairs Research Merit Review Grant to Dr. Tsuang.

1 Introduction

When we refer to the diagnosis of "schizoaffective disorder," we think of a disorder which is a mixture of schizophrenic and affective (depressive or manic) symptoms. Efforts to clarify the boundary between schizophrenia and affective disorders have resulted in a number of terms proposed to label such patients, but no single definition has met with widespread acceptance. The inclusion of the term "schizoaffective disorder" in DSM-IV (American Psychiatric Association 1994) has substantially improved the situation and has allowed clinicians and researchers to use consistent definitions. Even with these written criteria, there still remains some controversy as to precisely where schizoaffective disorder falls between the major categories of schizophrenia and affective disorders.

In this chapter, we will survey the widely differing views regarding the diagnosis and treatment of schizoaffective disorder in the light of recent empirical research and refer to interesting findings from phenomenologic, genetic, epidemiologic, and follow-up studies. We will begin with a selective historical review of the concept of schizoaffective disorder.

2 Historical Overview

Confusion over the classification of those patients who share schizophrenic and affective features has existed at least since the time of Kraepelin, who wrote in 1919 that he believed differential diagnosis was most difficult when confronted by a "mingling of morbid symptoms of both psychoses" (Kraepelin 1971). The "acute schizoaffective psychoses," a phrase first employed by Kasanin (1933), expresses the problem succinctly. Kasanin applied this term to a group of patients characterized as having a sudden onset in a state of marked emotional turmoil, distortion of the outside world (including false sensory impressions in some), and recovery following a short-lived psychosis of a few weeks to a few months.

This type of patient, who when viewed cross-sectionally appeared schizophrenic but who fully recovered, attracted increasing attention in the 1960s. Vaillant (1964), relying heavily on the work of Langfeldt and Kant (Kant 1937; Langfeldt 1939, 1956), derived a group of prognostic features which he used to predict remission in schizophrenic patients. Stephens et al. (1966) generated a longer, but analogous list of prognostic features also used to predict recovery in such patients. Both of these investigators were able to predict remission in approximately 80% of their patients by combining various prognostic features, including affective symptoms and affective heredity (or absence of schizophrenic heredity) as predictors of remission. By referring to their patients as having "remitting schizophrenia," however, these authors implied their belief that mixed cases nonetheless remain schizophrenic.

Other investigators have argued that mixed cases are better thought of as variants of affective disorder. Evidence in support of this viewpoint has primarily come from studies of schizoaffective patients with a predominantly manic affective picture (e.g. Abrams and Taylor 1976; Pope et al. 1980; Rosenthal et al. 1980).

The final major three ways in which schizoaffective disorder has been viewed are (1) that it is a heterogeneous mixture of disorders, (2) that it is a distinct disorder separate from schizophrenia and affective illness, and (3) that it represents an artificial categorization of clinical phenomena lying on a continuum between schizophrenia and affective disorder. Family and outcome studies (which are discussed below) have been the primary focus of efforts to evaluate these alternative positions. At this point, we merely note that the continuum hypothesis assumes a dimensional as opposed to a categorical perspective and that, although dimensional nosologic approaches can have certain advantages (e.g. see Widiger and Frances 1985), when applied to psychoses conceptual (and clinical) confusion could arise if the continuum hypotheses is taken to suggest that the psychoses can be divided into a limitless number of categories for the purposes of prognosis and treatment. If, on the other hand, the heterogeneous view of schizoaffective disorder is adopted (e.g. see Levitt and Tsuang 1988), it suggests the usefulness of the strategy of subdivision into a few major homogeneous subtypes.

The latter approach raises the issue of how best to subtype schizoaffective disorder. For example, can schizoaffective disorder be most profitably subtyped into depressive and bipolar types, as in DSM-IV (American Psychiatric Association 1994), or should it initially be divided into schizo-dominant and affect-dominant types, as suggested by M.T. Tsuang et al. (1986), or should both approaches be used as in the Research Diagnostic Criteria (RDC) of Spitzer et al. (1978)? We shall see that empirical evidence from long-term outcome studies appears to support distinctions based upon polarity, but not to the exclusion of other subtyping schemes. Unipolar and bipolar subtypes have also proved useful for clinical purposes, e.g. in guiding cross-sectional pharmacologic treatment, but the evidence from family studies is not conclusive. It has also been pointed out that,

longitudinally, the course of schizoaffective disorder can often be described as polymorphous (Marneros et al. 1986), which could make any dichotomy an oversimplification. Because of these complexities, this remains a very active area of schizoaffective research.

Additional interest in schizoaffective disorder has been stimulated by a recent trend to conceptualize schizophrenia in terms of negative and positive symptoms (Crow 1980; Andreasen et al. 1991). For example, Cuesta and Peralta (1995) showed that, although both positive and negative symptoms were relevant to differential diagnosis between schizophrenia and other psychotic disorders, negative symptoms presented higher significant differences between diagnostic groups (including schizoaffective disorder) than positive symptoms. Such findings indicate the need to distinguish between the depressive component of schizoaffective disorder and the negative syndrome of schizophrenia. This distinction can be difficult cross-sectionally in view of the similarity of the two syndromes. One potential way to disentangle these two overlapping sets of symptoms is to follow patients longitudinally, as negative symptoms are thought to persist over the long term (Pfohl and Winokur 1983; Andreasen et al. 1991), whereas depressive symptoms tend to remit. An additional complication is that negative symptoms themselves are not homogeneous and can result from such diverse causes as drug-induced akinesia, institutionalization, positive symptoms, depression (Andreasen et al. 1991), or as an emotional response to psychosis itself, e.g. postpsychotic depression or demoralization.

In summary, the question that arises is whether schizoaffective disorder shares its unknown etiology with schizophrenia, with affective disorder, or with neither, or whether schizoaffective disorder simply labels a genotypically mixed group made up partly of affective disorder patients and partly of patients with schizophrenia (Coryell 1986).

3 Diagnostic Criteria for Schizoaffective Disorder

The term "schizoaffective disorder" has been widely used, but with considerable differences in meaning and application (Brockington and Leff 1979). However, most definitions of schizoaffective disorder do have in common that they identify patients who concurrently share characteristic and pronounced schizophrenic and affective features and hence do not qualify for typical diagnoses of schizophrenia or affective disorder (Levitt and Tsuang 1990). In particular, such patients have sufficient affective symptoms to exclude an uncomplicated diagnosis of schizophrenia or, alternatively, sufficient schizophrenic features to exclude an uncomplicated diagnosis of affective disorder. It follows that the concept of schizoaffective illness will be directly influenced by how broadly or narrowly schizophrenia and affective psychosis are defined (Kendell 1986).

Despite Kasanin's original description of schizoaffective disorder as characterized by very sudden onset followed by recovery after a few weeks or months (Kasanin 1933), not all subsequent definitions have emphasized the course of the illness. For example, the criteria of Welner et al. (1974), Brockington et al. (1980a,b), Abrams and Taylor (1976), and the proposed clinical criteria of M.T. Tsuang et al. (1986) are primarily or exclusively cross-sectional. However, in view of the complexity of schizoaffective syndromes, involving various mixtures of schizophrenic and affective features over time, it should not be surprising that there has been an increasing emphasis on the use of longitudinal features to define schizoaffective disorder.

The most widely used criteria for studies of schizoaffective disorder have been the RDC of Spitzer et al. (1978), which specify (but do not necessarily require) periods of time when schizophrenia-like symptoms clearly dominate the clinical picture. This concept – the persistence of psychotic symptoms in the absence of affective symptoms – is also important in RDC for distinguishing between "mainly schizophrenic" and "mainly affective" subtypes of schizoaffective disorder, and several major studies have demonstrated the predictive validity of this clinical feature (Brockington et al. 1980a; Himmelhoch et al. 1981; Coryell et al. 1990a,b).

At the same time that the RDC were introduced, an influential review paper by Pope and Lipinski (1978) forcefully argued that "schizophrenic" symptoms occur with some regularity in affective disorders, and by themselves (e.g. cross-sectionally) should not be the basis for differential diagnosis. In consequence of these several developments, there appears to have been a shift in the core concept of schizoaffective disorder from the co-occurrence of psychotic (and in particular, mood-incongruent) symptoms and affective symptoms to one that represents "patients whose psychotic symptoms are not clearly linked to their affective episodes" (Blacker and Tsuang 1992, p. 1476). In 1985, Maj and Perris proposed a set of diagnostic criteria distinguished by their emphasis on course to subclassify schizoaffective disorder. Subsequently, the DSM-III-R criteria required the persistence of delusions or hallucinations in the absence of prominent mood symptoms for at least 2 weeks; this is essentially

a modified version of the RDC criteria for schizoaffective disorder, mainly of the schizophrenic type. This longitudinal criterion was carried over into DSM-IV (American Psychiatric Association 1994), with the additional limiting requirement that this 2-week (or longer) period should occur during the same period of the illness characterized by the concurrent expression of affective or schizophrenic syndromes. Specifically, the essential feature of schizoaffective disorder, as defined by DSM-IV, is an uninterrupted period of illness during which, at some time, there is a major depressive, manic, or mixed episode concurrent with characteristic symptoms that meet the symptom criterion for schizophrenia. In addition, during the same period of illness, there should have been delusions or hallucinations for at least 2 weeks in the absence of prominent mood symptoms. Moreover, the mood symptoms should be present for a substantial portion of the total duration of the illness. Finally, the symptoms must not be due to the direct physiological effects of a substance (e.g. cocaine) or a general medical condition. To meet the DSM-IV criteria for schizoaffective disorder, the essential features must occur within a single uninterrupted period of illness.

One effect of these requirements could be to reemphasize somewhat the cross-sectional picture in diagnosis by, for example, not including in schizoaffective disorder a hypothetical patient who had a mixed clinical picture in one episode followed in a separate episode by 2 weeks of predominantly schizophrenic symptoms. It appears that the prognostic and nosologic significance of longitudinal features will be an active area of research for some time to come.

At present, the criteria most commonly used to define schizoaffective disorder for research purposes include the RDC and the similar DSM-IV criteria. In addition, investigators have often modified these criteria (e.g. Marneros et al. 1992) or proposed their own criteria (e.g. M.T. Tsuang et al. 1986) as a means of studying course, family data, and other potentially important defining characteristics of these patients. Doubtless there will continue to be a variety of ways used to define schizoaffective disorder, and given the controversial status of the illness, this is to be preferred to settling prematurely on any one set of criteria. In fact, it is perhaps best explicitly to label all current diagnostic criteria for schizoaffective disorder as provisional. To emphasize this point (which can tend to be overlooked with repeated use of the phrase "schizoaffective disorder"), it could be helpful to use alternative terms such as "schizoaffective syndrome," "RDC schizoaffective disorder," or "DSM-IV schizoaffective disorder" in appropriate contexts.

4 Epidemiology

In contrast to schizophrenia, which has been the subject of innumerable epidemiologic investigations, relatively little is known about the epidemiology of schizoaffective disorder, e.g. regarding incidence, prevalence, demographics, mortality, and associations with socioeconomic variables and other potential risk factors. Most community surveys have not described rates of schizoaffective disorder. This is probably due largely to the uncertain nosologic status of schizoaffective disorder and to the resulting difficulty in reconciling across studies results that can be greatly affected by how the disorder is defined. To minimize this problem, we have emphasized in this review those studies that used objective diagnostic criteria.

4.1 Incidence

Only a few studies have examined the incidence of schizoaffective disorder. Brockington and Leff (1979) reported that patients meeting three or more definitions of schizoaffective disorder (out of eight definitions examined) comprised 4.5% of first-admission psychiatric patients in the Camberwell catchment area of London in 1973–1974. "Schizomanic" patients (manic patients with schizophrenic or paranoid symptoms) had an incidence rate of 1.7 per 100,000 population per year, which was considerably smaller than the rate of 4 per 100,000 population per year for "schizodepressive" patients. The numbers of schizoaffective patients identified in this sample exceeded the number of manic patients and were roughly half the number of incident schizophrenic cases, indicating that schizoaffective patients comprise a clinically significant population.

Additional evidence on this point can be obtained from an analysis of Epidemiologic Catchment Area Study data, obtained from Diagnostic Interview Schedule (DIS) interviews with randomly selected cases in five U.S. communities (Tien and Eaton 1992). These authors compared incidence rates in three non-overlapping groups of patients with delusions and hallucinations: individuals with the diagnosis of schizophrenia; patients with only delusions and hallucinations; and a "psychotic affective syndrome" group who also had manic or depressive episodes. The latter group is similar to, but not identical with most definitions of schizoaffective disorder, because 59% of the group experienced delusions and hallucinations only when they had a mood disturbance. Despite this limitation, the study is

important because it found that 1-year incidence rates for the psychotic affective group were approximately the same as for schizophrenia, namely, 1.7 per 1,000 population per year versus 2.0 per 1,000 per year for schizophrenia. Even if only 40% of the psychotic affective group were to meet criteria for schizoaffective disorder, this result still exceeds the earlier estimates made by Brockington and Leff (1979) by at least one order of magnitude. In part, the difference can be attributed to the difference between a community sample and a treated sample. However, the difference between the incidence rates for schizophrenia is just as large and not readily explainable.

4.2 Prevalence

Most of the information regarding the prevalence of schizoaffective disorder has been obtained from clinical samples. An exception is Torrey's prevalence study of schizophrenia in a rural area of western Ireland (Torrey 1987). Using key informants, Torrey estimated the 6-month prevalence rate for broadly defined schizophrenia to be 12.6 per 1,000 population. The actual numbers of patients included 21 with DSM-III schizophrenia and 11 with schizoaffective disorder; as with the incidence study by Brockington and Leff (1979), the ratio of schizophrenia to schizoaffective disorder is roughly 2 to 1.

Prevalence estimates of schizoaffective disorder in clinical populations have varied widely, as can be expected given the multiple factors affecting selection for treatment and duration of treatment. For example, Rosenthal et al. (1980) reported that 35% of the manic patients in a lithium clinic were RDC schizoaffective manic. Müller-Oerlinghausen et al. (1992) found that the prevalence of schizoaffective disorder in lithium clinics varied widely from city to city: 7% (Aarhus), 15% (Berlin), 23% (Vienna), and 32% (Hamilton). Junginger et al. (1992) found that 14% of the delusional patients in a chronic and highly selected population of inpatients and outpatients met DSM-III-R criteria for schizoaffective disorder, compared to 60% for schizophrenia, 17% for bipolar disorder, and 4% for major depression. In the Cologne Study on long-term course and outcome, Marneros et al. (1991) applied longitudinal criteria adapted from DSM-III-R to a large sample of patients with major psychoses and found that the number with schizoaffective disorder (28.5%) was about the same as the number with affective disorders (30%), and less than the number with schizophrenia (42%). Taken together, these studies validate current interest in schizoaffective disorder as being a clinically significant population of psychiatric patients.

A number of investigators have compared the prevalence of the manic and depressive subtypes of schizoaffective disorder. Clayton, in an influential review, concluded that the manic type is more frequent (Clayton 1982). A subsequent first-admission study by Berner and Lenz (1986) supports this conclusion in a comparison of subtypes of RDC schizoaffective disorder (12.5% schizoaffective manic vs. 8% schizoaffective depressed). However, a broader sample of first-admission patients in the study by Brockington and Leff (1979) showed substantially more schizodepressive patients (3.7%) than schizomanic patients (1.5%). Two recent studies of consecutive inpatient samples also indicate that the depressive subtype is more prevalent: Kitamura and Suga (1991), using RDC criteria, and Marneros et al. (1991), using modified DSM-III criteria. In view of these inconsistent results, we can hypothesize that selection factors largely determine the relative numbers of schizoaffective manic and schizoaffective depressed patients who are treated. More definitive findings will have to await information from future epidemiologic studies of schizoaffective disorder.

4.3 Demographic Factors

Given the variability in how schizoaffective disorder is defined and the evident influence of selection biases in prevalence studies, it is not surprising that studies of demographic factors such as gender and age at onset have produced widely varying results. Nevertheless, some general conclusions are possible.

With regard to gender, there appear to be as many or more females than males, e.g. 71% in the sample of M.T. Tsuang et al. (1986) and 63% in that of Marneros et al. (1990a). Berner and Lenz (1986) reported that the male to female ratio ranged from 0.3:1 to 1:1 depending on the definition of schizoaffective disorder employed. In their general population study, Tien and Eaton (1992) reported higher 1-year incidence rates for females with DIS psychotic affective syndrome (a heterogeneous group that probably includes schizoaffective disorder) for all age groups examined and computed a relative risk of 6.8 for developing this syndrome in females compared to males. Clayton (1982) hypothesized that the results also depend on polarity, e.g. with approximately equal numbers of males and females in the manic subtype, compared to almost two-thirds female patients in the depressive subtype. Kitamura and Suga (1991), however, found approximately equal numbers of males and females in both schizoaffective manic and schizoaffective depressed subtypes.

Age at onset is one of the basic characteristics that reliably distinguishes major subgroups – e.g. male versus female schizophrenic patients (Goldstein et al. 1989) and bipolar versus unipolar affective disorder (Smeraldi et al. 1983) – and for that reason is also of interest in schizoaffective disorder. First, age at onset appears to be a discriminating factor in comparison with other disorders. For example, in her 1982 review, Clayton concluded that most studies found the average age at onset to be youngest for schizoaffective disorders compared to unipolar and bipolar disorders. Similarly, M.T. Tsuang et al. (1986) reported that the mean age at onset of schizoaffective disorder (29 years) was significantly younger than for manic and depressed groups defined using Washington University criteria (34 and 44 years, respectively); however, there was no difference between schizoaffective disorder and schizophrenia. Marneros et al. (1990a) reported that the median age at onset for schizoaffective disorder (29 years) was younger than the median age for affective disorders (35 years) and older than for schizophrenia (24 years). Berner and Lenz (1986) showed that the age at onset depends largely on the diagnostic criteria employed; for example, DSM-III identified a schizoaffective patient population that was older than the RDC schizoaffective depressed subtype and substantially older than the RDC schizoaffective manic subtype of schizoaffective disorder.

Within schizoaffective disorder, there might also be a substantial effect of gender on the age at onset, namely, that women tend to be older at onset of the disorder (Angst 1986). To the extent that their DIS psychotic affective syndrome includes schizoaffective disorder, the results presented by Tien and Eaton (1992) also support this generalization: 1-year incidence rates for females were substantial for age groups greater than 34 years, whereas incidence rates for males declined sharply after age 34.

Information on marital status is available in a few studies (e.g. Clayton 1982; Marneros et al. 1990b), but confounding by gender makes the available results difficult to interpret or compare across studies.

4.4 Mortality

Several factors combine to make premature mortality a special concern in schizoaffective disorder. Schizoaffective patients share some of the symptoms and other characteristics of schizophrenic and affective disorder patients, who have repeatedly shown to be at increased risk of death, principally (but not exclusively) from suicides and accidents (Simpson 1988). Buda et al. (1988) directly compared the mortality experience of DSM-III schizophrenics with a heterogeneous group that included patients with schizoaffective disorder, schizophreniform disorder, and atypical psychotic disorder using follow-up data beginning with index hospital stays between 1934 and 1945. Excess mortality for the atypical group in this historical sample was observed for specific causes of death including infections, neoplasms, cardiovascular disease, and suicides. Furthermore, suicide occurred in excess among the other psychotic groups compared with schizophrenia and with the general population. Angst et al. (1990) compared suicides in patients with affective disorder and schizoaffective disorder using data from Zurich, Bonn, and New York and found that the suicide risk for schizoaffective patients closely resembled that for patients with affective disorders. In particular, the risk of suicide was constant over a lifetime and did not increase or decrease with age. Marneros et al. (1989a), in a long-term follow-up study of schizoaffective patients, reported suicide attempts over a lifetime in 43% of the unipolar subgroup and in 29% of the bipolar subgroup, usually by drug overdose. Clearly, the suicide risk in these patients is substantial and enduring.

Fortunately, there is evidence that suicide in psychiatric patients is to some extent predictable and preventable, largely in response to standard psychiatric interventions (M.T. Tsuang et al. 1992). Evidence in support of the view was gathered by Müller-Oerlinghausen et al. (1992) in a study of the long-term effect of lithium treatment (average duration of treatment, 7 years) on the mortality of patients with schizoaffective and affective disorders. That study, which included clinic samples from Canada and three European countries, found that mortality was not significantly increased in lithium-treated schizoaffective patients compared with the general population, although the mortality risk in the bipolar and unipolar comparison groups was actually somewhat less than for schizoaffective disorder.

5 Clinical Picture

Most definitions of schizoaffective disorder depend entirely on the clinical picture, and since these definitions vary so greatly, the associated clinical picture will also vary depending on the label (Coryell 1986). Because these patients suffer from psychosis, an associated loss of insight, and often severe psychomotor disturbance, many have problems remembering historical events in their lives. At the time of admission, affective symptoms may be altogether overshad-

owed by the patient's delusional preoccupation, hallucinations, or bizarre behavior. In fact, at this point, such patients often deny affective symptoms that they later recall. The distinction between affective, schizoaffective, and schizophrenic psychosis, therefore, must not depend on the patient interview alone. The clinician should seek knowledgeable informants to learn whether affective symptoms preceded the psychotic ones.

One recent study (Taylor and Amier 1994) attempted to discriminate the psychoses by their classic symptoms. Three groups were statistically compared, including chronic schizophrenia, schizoaffective disorder, and affective disorder. Their analysis discriminated chronic schizophrenia from affective disorder, but schizoaffective disorder overlapped both groups. In addition, the unipolar subtype of schizoaffective disorder resembled chronic schizophrenia, and the bipolar subtype resembled affective disorder. However, these discriminations also substantially overlapped, and among nonaffective positive features, formal thought disorder was best at discrimination. The findings of this study did not fully support the classification system and suggested that the emphasis on hallucinations and delusions is overvalued.

A later study (Kendler et al. 1995) sought to determine whether the DSM-III-R category of schizoaffective disorder differs meaningfully from schizophrenia and affective illness in terms of clinical features and other validators. On the basis of these various clinical validators, they found that schizoaffective disorder defines a syndrome that differs from both schizophrenia and affective illness. The division of schizoaffective disorder into bipolar and depressive subtypes was, however, not validated.

In order to evaluate the basic symptom differences of schizophrenia, schizoaffective disorder, and bipolar disorder, a consecutive series of 72 outpatients was recruited for another study (Ricca et al. 1997), including 28 with a diagnosis of schizophrenia, 29 with bipolar disorder, and 15 with schizoaffective disorder. Data obtained suggest that perception and thought disturbances are the most characteristic experiences of schizophrenic patients in comparison with bipolar patients. The analysis did not highlight a characteristic basic symptom profile of schizoaffective disorder when compared with bipolar affective disorder and schizophrenia. Cognitive features were also compared, by Manschreck et al. (1997), among patients with schizophrenia, schizoaffective disorder, major depression, and normal controls. The results indicate that schizophrenic subjects attain smaller gains in verbal recall when context is increased compared to depressed and normal controls. Schizophrenic and schizoaffective subjects, however, did not differ in recall gain on this task. The authors concluded that schizoaffective subjects could not be distinguished from schizophrenic subjects on this cognitive feature (verbal recall when context is increased).

Blehar et al. (1996) identified clinical variables that increased the diagnostic differentiation among three groups: schizoaffective bipolar disorder, bipolar disorder, and schizophrenia. Based on the clinical diagnostic interview, the schizoaffective bipolar disorder group was more clearly differentiated from the bipolar disorder group than from the schizophrenic group. The primary difference between the schizoaffective bipolar disorder group and the bipolar disorder group was with measures of psychosis, including specific hallucinations and delusions, ratings of the bizarre and fragmented nature of psychosis, catatonia, thought disorder, and comorbid phobias. The schizoaffective bipolar disorder group had higher values on all these variables than the bipolar disorder group. On the other hand, subjects in the schizoaffective bipolar disorder group had a higher frequency of manic episodes and longer average duration of mania than patients in the schizophrenia group.

5.1 Precipitating Factors

One of the defining characteristics of schizoaffective disorder, as initially conceptualized by Kasanin (1933), has been the presence of precipitants such as stressors or major life events. To some extent this notion has been validated empirically. Brockington et al. (1980b) reported that ten of 32 schizomanic patients had obvious recent stress such as childbirth, surgery, head injury, or the disruption of important personal relationships. M.T. Tsuang et al. (1986) found significantly more precipitants of any type in schizoaffective disorder (60%) compared with schizophrenia (11%), mania (27%), or depression (39%). Furthermore, schizoaffective patients had significantly more psychosocial, physical, and postpartum precipitants in most of the comparisons with these other diagnostic groups. In contrast, Marneros et al. (1990b) found equal percentages of schizoaffective and affective disorder patients with life events before onset (51%), compared to only 24% in schizophrenia. Tien and Eaton (1992) found that antecedent alcohol problems, but not daily marijuana abuse, increased the relative risk of DIS psychotic affective syndrome to 5.7.

5.2 Premorbid Factors

Premorbid or predisposing factors can give useful clues to the etiology and early course of the disorder,

but unfortunately little is known about such factors in schizoaffective disorder. Marneros et al. (1990b) reported a significantly higher educational level in schizoaffective patients compared with schizophrenic patients, but no difference between schizoaffective and affective disorder patients. The same pattern was obtained for employment/vocational training. Compared with schizoaffective patients, those patients with schizophrenia were more often in lower social classes; this was partly attributed to the parents' social class, and partly to the schizophrenic patients' downward mobility at the time of onset. In contrast, there was no significant difference between the schizoaffective and affective disorder group. Finally, all three diagnostic groups had similar percentages of patients from broken homes.

6 Genetic and Family Studies

Family studies of schizoaffective disorder have led to inconsistent results, which can be attributed in part to the inherent heterogeneity of this group of patients and to the lack of uniformity in diagnosis. It is of some interest, therefore, to find three studies that used different definitions of schizoaffective disorder yet obtained convergent results.

First, a review by Fowler (1978) summarized several studies showing schizophrenic patients with good prognosis to have a familial risk of affective disorder (20%) between that of unipolar depressive (16%) and manic patients (35%), and to have a familial risk of schizophrenia significantly greater (6%) than manic (<1%) or unipolar depressive patients (<1%). This was interpreted to be consistent with the conclusion that schizophrenia with a good prognosis was associated with elevated risks in relatives for both affective disorder and schizophrenia, and hence most consistent with an interpretation of heterogeneity. Gershon et al. (1988), in contrast, studied schizoaffective disorder as largely defined by RDC, where the emphasis is more on mixing affective and schizophrenic features rather than on a good outcome. The schizoaffective, chronic subtype in this study also revealed elevated risks (compared with controls) for nonaffective psychosis and affective disorder (pooling unipolar and bipolar cases). A third example of yet a different definition of schizoaffective disorder, namely, patients with "atypical psychosis" who for a variety of reasons fail to meet diagnostic criteria for schizophrenia or affective disorder, is found in a study by M.T. Tsuang (1991), where again high rates of schizophrenia and affective disorders were found in relatives of schizoaffective probands. Hence, whether good prognosis, mixing of affective and schizophrenic features, or atypicality is emphasized in the definition of schizoaffective disorder, family studies have found elevated rates of both schizophrenic and affective disorder in relatives of schizoaffective patients, consistent with genetic heterogeneity.

Another possibility is that schizoaffective disorder is a distinct entity. In general, however, family studies have not demonstrated an increased risk of schizoaffective disorder and no other disorders in relatives of schizoaffective patients, which would be the pattern most supportive of schizoaffective disorder as a distinct entity or "third psychosis" (Zerbin-Rudin 1986).

We have also tested the effect of gender on the familial risk for schizophrenia and affective disorders in probands with schizoaffective disorder (Goldstein et al. 1993). The sample consisted of 42 DSM-III schizoaffective probands and 149 first-degree relatives. Findings showed that relatives of females had significantly higher rates of schizophrenia and unipolar disorder than relatives of males. In addition, among relatives, males were at significantly higher risk for schizophrenia spectrum disorders than females. Results were similar when probands were subtyped into "schizoaffective depressed" or "schizoaffective manic" categories. Implications for the taxonomy of schizoaffective disorder suggest a stronger relationship with schizophrenia, although the relationship with affective disorder remains unclear.

Although there have been no molecular genetic studies of schizoaffective disorder, studies of schizophrenia and bipolar disorder have implicated genetic loci that may be involved in both disorders. If so, these may indicate that some susceptibility genes influence a broad range of psychoses, including schizoaffective disorder. For example, Maziade et al. (1997) failed to detect linkage at 6p24–22 in 18 large, multigenerational pedigrees from Eastern Quebec, using either broad or narrow definitions of schizophrenia. They did find suggestive evidence in one large pedigree, however, that the locus was associated with vulnerability to both schizophrenia and bipolar disorder when they utilized a broad phenotypic definition that included schizophrenia, schizoaffective disorder, schizophreniform disorder, bipolar disorder (I and II), and major depression (recurrent).

Moreover, other researchers have also suggested that at least one disorder in the schizophrenia spectrum – schizoaffective disorder – might belong to an affective disorder spectrum as well (Bertelsen and Gottesman 1995). Consistent with this possibility, evidence for linkage was obtained at this locus for bipolar disorder in Old Order Amish pedigrees (Ginns et al. 1996). Similarly, the chromosome 10p region was implicated

for both schizophrenia and bipolar disorder in the NIMH Genetics Initiative pedigrees (Rice et al. 1997; Faraone et al. 1998; Foroud et al. 1998).

Wildenauer et al. (1996) reported suggestive evidence of linkage to a region on chromosome 18p, using a sib-pair analysis. Their findings are interesting, in part because they obtained their highest logarithm of differences (lod) scores when they used a broad phenotypic definition that included relatives of schizophrenic patients with bipolar disorder and major depression, in addition to schizophrenia and schizoaffective disorder. This group recently reported a lod score of 3.1 using this approach and confirmed in principle findings reported by Maziade et al. (1997) at chromosome 6p24–22. Moreover, the chromosome region at 18p has also been implicated in bipolar disorder (Berrettini et al. 1994; Stine et al. 1995).

If schizoaffective disorder is genetically heterogeneous, the next logical question is whether it can be successfully subtyped. M.T. Tsuang (1991) has demonstrated one approach to such subtyping by segregating atypical psychotic patients into those with an increased probability for schizophrenia or affective disorder, and a residual undifferentiated group. Marneros et al. (1989b) have approached the problem longitudinally by showing that schizoaffective disorder can be divided in terms of long-term course into those with bipolar and unipolar types. Others, including Angst et al. (1979) and Gershon et al. (1988), have had difficulty subtyping schizoaffective disorder when using family studies as a validator.

7 Long-Term Course and Outcome

Reasons for examining the long-term course of schizoaffective disorder include the importance of understanding the prognosis as the basis for clinical management and for comparisons of treatment efficacy; the possibility of identifying prognostic factors; and the use of longitudinal information to validate diagnostic concepts and to investigate diagnostic heterogeneity (e.g. by identifying homogeneous subgroups).

As a basic description of the course of the illness, we provide the following illustrative summary adapted from Samson et al. (1988) and subject to the limitations noted elsewhere in this chapter regarding the wide variability in the definition of schizoaffective disorder. Reported rates of recovery from schizoaffective episodes have ranged from 83% reported in a short-term follow-up study (Clayton et al. 1968) to 29% who recovered at any time during a 6-month follow-up period (Coryell et al. 1984). Approximately 20%–30% of schizoaffective patients go on to show a deteriorating course, e.g. one typified by persistent psychosis (Holmboe and Astrup 1957; Brockington et al. 1980a,b; Gross et al. 1986; Coryell et al. 1990a,b; M.T. Tsuang 1990; Grossman et al. 1991). Approximately 10% of patients show diagnostic shifts over time, becoming either more affective or more psychotic in symptom manifestation (Angst 1986). Schizoaffective patients have been reported to spend on average about 20% of their lifetimes hospitalized or in an episode (Angst 1986). Angst (1986) has also reported that the median number of hospitalizations over a 25-year period was between six and seven. Although many definitions of schizoaffective disorder allow for the sequential manifestation of psychotic and affective symptomatology, most patients show concurrent expressions of psychotic and affective symptoms (Marneros et al. 1986).

To put the results of longitudinal studies of schizoaffective disorder in context, it is helpful to keep in mind that, in general, the long-term outcome of schizoaffective disorder is better than in schizophrenia, but worse in comparison to affective disorder (for extensive reviews of this topic, see Harrow and Grossman 1984; Angst 1986; Samson et al. 1988). There is also some evidence that long-term outcome for the manic subtype of schizoaffective disorder is similar to that for mania or bipolar disorder, whereas outcome comparisons of the depressive subtype and major depression reveal substantial differences (Brockington et al. 1980a,b; Clayton 1982). Differential outcomes of other subgroupings of schizoaffective disorder have also been investigated, with much interest focusing on the mainly affective versus mainly schizophrenic distinction in the RDC. Earlier results indicated worse outcomes for the mainly schizophrenic subtype (Levinson and Levitt 1987), but several recent reports have not confirmed this pattern (Coryell et al. 1990a,b; Grossman et al. 1991). For example, Coryell et al. (1990a,b) concluded that chronicity was more predictive of outcome in both depressive and manic subtypes of RDC schizoaffective disorder.

Another study (Tsuang and Coryell 1993) examined the 8-year follow-up of patients with DSM-III-R psychotic depression, schizoaffective disorder, and schizophrenia. The purpose of this study was to investigate the long-term outcome of patients with functional psychoses. The functional status of patients with mood-congruent and mood-incongruent psychotic depression, schizoaffective disorder, and schizophrenia was examined. Patients with psychotic depression had much better outcomes than patients with schizoaffective disorder or schizophrenia. Forty-four percent of patients with psychotic depression were free from psychosis at follow-up in marked contrast to those who had schizoaffective disorder or

schizophrenia, none of whom had recovered. In terms of long-term outcome, patients with schizoaffective disorder could not be distinguished from patients with schizophrenia.

In a German multicenter treatment study (Doering et al. 1998), 354 patients with schizophrenia and schizoaffective disorder were followed for 2 years for the purpose of determining predictors of relapse and rehospitalization. In a statistical analysis of the entire sample, the diagnosis of schizoaffective disorder predicted a worse outcome, in contrast to the well-established view that affective symptoms in schizophrenia are connected with a better outcome. In part, this result might reflect the emphasis on relapse and repeated hospitalizations in the outcome criteria as well as an episodic course with phases that are more frequent, but shorter and more clearly demarcated and with less chronicity and disability than in schizophrenia. In this way, because of their changing psychopathology, patients with the diagnosis of schizoaffective disorder are more likely to meet the criteria for relapse and to need repeated hospitalizations. Nevertheless, outcomes in areas such as social adjustment or employability would usually be better for schizoaffective patients than for patients with schizophrenia.

A major goal of longitudinal studies is to identify prognostic factors, particularly those variables that predict especially good or poor outcomes. The one variable that has emerged with some consistency is the persistence of psychotic symptoms in the absence of affective symptoms, which substantially increases the risk of poor outcomes such as persistent psychosis (Brockington et al. 1980a; Himmelhoch et al. 1981; Maj et al. 1987; Coryell et al. 1990a,b). Other risk factors for a negative outcome include poor premorbid personality (Coryell et al. 1990a,b; del Rio Vega and Ayuso-Gutierrez 1990), premorbid instrumental skills (McGlashan and Williams 1990), chronicity at index assessment (Maj et al. 1987; Coryell et al. 1990a,b), frequency of relapses (del Rio Vega and Ayuso-Gutierrez 1990), and the number of typically schizophrenic symptoms (McGlashan and Williams 1990). Marneros et al. (1993) investigated factors influencing the long-term outcome of schizoaffective disorders in 101 patients with a mean duration of illness of 25.5 years. The most important factors influencing the development of persisting psychological alterations proved to be the absence of pure melancholic episodes during the course, life events at onset, the presence of first-rank schizophrenic symptoms, schizomanic–depressive mixed episodes, and a higher annual frequency of episodes. The most relevant factors influencing the development of negative social consequences due to the illness were found to be an asthenic/low self-confidence premorbid personality and elevated numbers of episodes and cycles. There were only partial similarities with schizophrenia or affective disorder.

To assist clinicians in identifying patients with an increased likelihood of having a schizophrenia-like course and outcome, Coryell et al. (1990a,b) developed the following "poor-outcome prototype": a patient who has had a poor premorbid or adolescent social adjustment and inadequate social adjustment as an adult, a chronic course, and, at some point, persistent psychotic features that dominated the clinical picture. This prototype appeared to be valid in both depressive and manic types of RDC schizoaffective disorder (Coryell et al. 1990a,b). A focus for future research will be to validate such predictors in other patient populations and to refine the concept of a poor prognosis prototype in homogeneous subgroups that are identified as being nosologically important.

8 Treatment Strategies

Studies addressing psychopharmacologic treatment of schizoaffective disorder are made more difficult to generalize from because of the nonuniform way in which schizoaffective disorder has been defined in the literature. As it is our contention that schizoaffective disorder is genetically heterogeneous, subtyping is a natural next step. We have found it useful for psychopharmacologic treatment purposes to divide schizoaffective disorder by polarity (e.g. bipolar vs. unipolar schizoaffective disorder) and to pay especially close attention to interepisode psychotic symptomatology (i.e. to psychotic symptoms that persist in the absence of a full affective syndrome).

When viewed cross-sectionally, schizoaffective disorder, bipolar type (or "schizomania") probably represents primarily a mixture of those patients with bipolar affective disorder and those with schizophrenia in excited states (Clayton 1982). A high index of suspicion for organic/toxic states is also important. If a certain proportion of these patients actually suffer from a form of bipolar disorder, certainly the use of antimanic agents, including neuroleptics, lithium, and anticonvulsants (e.g. carbamazepine and valproate), makes sense. Such agents have been explored alone (Prien et al. 1972; Abrams and Taylor 1976; Brockington et al. 1978; Pope et al. 1980) and in combination (Biederman et al. 1979; Okuma et al. 1989a,b). Biederman et al. (1979), for example, found a modest benefit from the combination of lithium carbonate and haloperidol, which must then be balanced against the additional risk of greater toxicity. When intermorbid psychotic symptoms persist after mixed symptom

episodes subside, it has been our experience that chronic neuroleptic therapy more likely is necessary. Okuma et al. (1989b) reported that carbamazepine, in combination with neuroleptics, was modestly useful in the treatment of excited states in a mixed group of schizophrenic and schizoaffective patients. Potential adverse effects included increased hallucinatory behavior and worsened psychomotor activity, possibly related to the reported inducing by carbamazerpine of its own metabolism and that of concomitant neuroleptics thus lowering neuroleptic blood levels. Valporate, which does not have a tricyclic structure and does not induce the metabolism of hepatically cleared agents (McElroy et al. 1992), would be a possible alternative. Clozapine has recently also been proposed as an additional alternative (McElroy et al. 1991) for patients with schizoaffective disorder, bipolar or depressed type.

Cross-sectional diagnoses of schizoaffective disorder, unipolar type (or "schizodepression") probably primarily represent a mixture of patients with psychotic affective disorder (e.g. delusional depression) and schizophrenic patients with depressive syndromes (Brockington et al. 1980a), together with some patients with bipolar disorder and diverse other conditions (Clayton 1982). Again, combination treatment suggests itself. The use of antidepressants from all classes in combination with neuroleptics is plausible, with tricyclics and monoamine oxidase inhibitors (MAOI) so far being most closely examined (Siris et al. 1978, 1987). It has been suggested that one source of variance in whether these patients respond to the addition of an antidepressant is whether or not patients with current florid psychotic symptoms are excluded (Kramer et al. 1989). In light of additional weight gain seen as a potential side effect of combining phenothiazines and tricyclics (Prusoff et al. 1979), selective serotonin reuptake inhibitors (which do not cause weight gain) might be an attractive alternative, although to our knowledge this alternative has so far not been studied systematically. Augmentation with lithium, analogous to that used with treatment-resistant depressions (de Montigny et al. 1983), has also been suggested for the treatment of refractory psychotic depression, especially that associated with a bipolar course (Nelson and Mazure 1986). Additionally, it has been suggested by M.T. Tsuang et al. (1979) that electroconvulsive therapy may reduce mortality in schizoaffective patients.

In a recent review of controlled treatment studies, Keck et al. (1996) found that in the acute treatment of schizoaffective disorder, bipolar type, typical antipsychotics and lithium were comparable in efficacy except in agitated patients, for whom antipsychotics were superior. The combination of lithium and antipsychotics appeared to be superior to antipsychotics alone in this patient subgroup. Some preliminary data from open trials suggests that the mood stabilizers valproate and carbamazepine and the novel antipsychotics clozapine and resperidone may be promising treatments for schizoaffective disorder. Fenn et al. (1996) also found, in a 5-year naturalistic study of phamacotherapy in schizoaffective disorder, that valproate monotherapy had increased as a percentage of total pharmacotherapies, while lithium monotherapy had declined.

9 Conclusion

We have reviewed schizoaffective disorder, perhaps the most common atypical psychosis, from multiple perspectives. Our necessarily provisional view is that schizoaffective disorder is a genetically heterogeneous condition primarily composed of schizophrenia and affective disorders and perhaps a residual, currently undifferentiated condition. For research purposes, schizoaffective disorder can accordingly be subtyped into schizophrenic, affective, and undifferentiated categories. Other subtyping schemes (e.g. chronic vs. nonchronic, and manic vs. depressive) also have potential value and should not be prematurely excluded at this stage of the investigation. For psychopharmacological treatment purposes, subtyping into bipolar and unipolar categories has proved to be of clinical value. Future research will be needed to resolve the continuing uncertainty regarding diagnosis, treatment, and prognosis in this heterogeneous but important group of psychoses.

10 References

Abrams R, Taylor MA (1976) Mania and schizo-affective disorder manic type: a comparison. Am J Psychiatry 133: 1145–1147

American Psychiatric Association (1994) Diagnostic and statistical manual of mental disorders, 4th edn. American Psychiatric Association, Washington, DC

Andreasen NC., Flaum M, Arndt S et al (1991) Positive and negative symptoms: assessment and validity. In: Marneros A, Andreasen NC, Tsuang MT (eds) Negative versus positive schizophrenia. Springer, Berlin Heidelberg New York, pp 28–51

**Angst J (1986) The course of schizoaffective disorders. In: Marneros A, Tsuang MT (eds) Schizoaffective psychoses. Springer, Berlin Heidelberg New York, pp 63–93

Angst J, Felder W, Lohmeyer B (1979) Are schizoaffective psychoses heterogeneous? II. Results of a genetic investigation. J Affect Disord 1: 155–165

Angst J, Stassen HH, Gross G et al (1990) Suicide in affective and schizoaffective disorders. In: Marneros A, Tsuang MT (eds) Affective and schizoaffective disorders. Springer, Berlin Heidelberg New York, pp 168–185

Berner P, Lenz G (1986) Definitions of schizoaffective psychosis: mutual concordance and relationship to schizophrenia and affective disorder. In: Marneros A, Tsuang MT (eds) Schizoaffective psychoses. Springer, Berlin Heidelberg New York, pp 31–49

Berrettini WH, Ferraro TN, Goldin LR et al (1994) Chromosome 18 DNA markers and manic-depressive illness: evidence for a susceptibility gene. Proc Nat Acad Sci 91: 5918–5921

*Bertelsen A, Gottesman II (1995) Schizoaffective psychoses: genetical clues to classification. Am J Med Genet Neuropsychiatr Genet 60: 7–11

Biederman J, Lerner Y, Belmaker RH (1979) Combination of lithium carbonate and haloperidol in schizo-affective disorder. A controlled study. Arch Gen Psychiatry 36: 327–333

Blacker D, Tsuang MT (1992) Contested boundaries of bipolar disorder and the limits of categorical diagnosis in psychiatry. Am J Psychiatry 149: 1473–1483

*Blehar MC, Faraone SV, Zeller PJ et al (1996) Differentiation of schizoaffective bipolar disorder from bipolar disorder and schizophrenia. Depression 3: 309–315

Brockington IF, Leff JP (1979) Schizo-affective psychosis: definitions and incidence. Psychol Med 9: 91–99

Brockington IF, Kendell RE, Kellet JM et al (1978) Trials of lithium, chlorpromazine and amitriptyline in schizoaffective patients. Br J Psychiatry 133: 162–168

Brockington IF, Kendell RE, Wainwright S (1980a) Depressed patients with schizophrenic or paranoid symptoms. Psychol Med 10: 665–675

Brockington IF, Wainwright S, Kendell RE (1980b) Manic patients with schizophrenic or paranoid symptoms. Psychol Med 10: 73–83

Buda M, Tsuang MT, Fleming JA (1988) Causes of death in DSM-III schizophrenics and other psychotics (atypical group) a comparison with the general population. Arch Gen Psychiatry 45: 283–285

*Clayton PJ (1982) Schizoaffective disorders. J Nerv Ment Dis 170: 646–650

Clayton P, Rodin L, Winokur G (1968) Family history studies. III. Schizoaffective disorder, clinical and genetic factors including a one to two year follow-up. Compr Psychiatry 9: 31–49

*Coryell W (1986) Schizoaffective and schizophreniform disorders. In: Winokur G, Clayton P (eds) The medical basis of psychiatry. Saunders, Philadelphia, pp 102–114

Coryell W, Lavori P, Endicott J (1984) Outcome in schizoaffective, psychotic, and nonpsychotic depression. Arch Gen Psychiatry 41: 787–791

Coryell W, Keller M, Lavori P, Endicott J (1990a) Affective syndromes, psychotic features, and prognosis. I. Depression. Arch Gen Psychiatry 47: 651–657

Coryell W, Keller M, Lavori P, Endicott J (1990b) Affective syndromes, psychotic features, and prognosis. II. Mania. Arch Gen Psychiatry 47: 658–662

Crow TJ (1980) Molecular pathology of schizophrenia: more than one disease process? Br Med J 280: 66–68

Cuesta MJ, Peralta V (1995) Are positive and negative symptoms relevant to cross-sectional diagnosis of schizophrenic and schizoaffective patients? Compr Psychiatry 36: 353–361

de Montigny C, Cournoyer G, Morissette R et al (1983) Lithium carbonate addition in tricyclic antidepressant-resistant unipolar depression: correlations with neurobiologic actions of tricyclic antidepressant drugs and lithium ion on serotonin system. Arch Gen Psychiatry 40: 1327–1334

del Rio Vega JM, Ayuso-Gutierrez JL (1990) Course of schizoaffective psychosis: a retrospective study. Acta Psychiatr Scand 81: 534–537

Doering S, Muller E, Kopcke W et al (1998) Predictors of relapse and rehospitalization in schizophrenia and schizoaffective disorder. Schizophr Bull 24: 87–98

Faraone SV, Matise T, Svrakic D et al (1998) A genome scan of the European-American schizophrenia pedigrees: results from the NIMH Genetics Initiative & Millennium Consortium. Am J Med Genet Neuropsychiatr Genet 81: 290–295

Fenn HH, Robinson D, Luby V et al (1996) Trends in pharmacotherapy of schizoaffective and bipolar affective disorders: a 5-year naturalistic study. Am J Psychiatry 153: 711–713

Foroud T, Castellucio PF, Koller DL et al (1998) Genomewide scan of affected relative pairs using the NIMH Genetics Initiative bipolar affective disorder pedigrees. Am J Med Genet Neuropsychiatr Genet 81: 462

Fowler RC (1978) Remitting schizophrenia as a variant of affective disorder. Schizophr Bull 4: 68–77

Gershon ES, DeLisi LE, Hamovit J et al (1988) A controlled family study of chronic psychoses schizophrenia and schizoaffective disorder. Arch Gen Psychiatry 45: 328–336

Ginns EI, Ott J, Egeland JA et al (1996) A genome-wide search for chromosomal loci linked to bipolar affective disorder in the old order Amish. Nature Genet 12: 431–435

Goldstein JM, Tsuang MT, Faraone SV (1989) Gender and schizophrenia: implications for understanding the heterogeneity of the illness. Psychiatry Res 28: 243–253

Goldstein JM, Faraone SV, Chen WJ, Tsuang MT (1993) The role of gender in understanding the familial transmission of schizoaffective disorder. Br J Psychiatry 163: 763–768

Gross G, Huber G, Armbruster B (1986) Schizoaffective psychoses – long-term prognosis and symptomatology. In: Marneros A Tsuang MT (eds) Schizoaffective psychoses. Springer, Berlin Heidelberg New York, pp 188–203

Grossman LS, Harrow M., Goldberg JF, Fichtner CG (1991) Outcome of schizoaffective disorder at two long-term follow-ups: comparisons with outcome of schizophrenia and affective disorders. Am J Psychiatry 148: 1359–1365

Harrow M, Grossman L (1984) Outcome in schizoaffective disorders: a critical review and reevaluation of the literature. Schizophr Bull 10: 87–108

Himmelhoch JM, Fuchs CZ, May SJ et al (1981) When a schizoaffective diagnosis has meaning. J Nerv Ment Dis 169: 277–282

Holmboe R, Astrup C (1957) A follow-up study of 255 patients with acute schizophrenia and schizophreniform psychoses. Acta Psychiatr Scand Suppl 115: 9–61

Junginger J, Barker S, Coe D (1992) Mood theme and bizarreness of delusions in schizophrenia and modd psychosis. J Abnorm Psychol 101: 287–292

Kant O (1937) Study of a group of recovered schizophrenic patients. Psychiatr Q 15: 262–283

Kasanin J (1933) The acute schizoaffective psychoses. Am J Psychiatry 90: 97–126

Keck PE, McElroy SL, Strakowski SM (1996) New developments in the pharmacologic treatment of schizoaffective disorder. J Clin Psychiatry 57: 41–48

Kendell RE (1986) The relationship of schizoaffective illnesses to schizophrenic and affective disorders. In: Marneros A,

Tsuang MT (eds) Schizoaffective psychoses. Springer, Berlin Heidelberg New York, pp 18–30

**Kendler KS, McGuire M, Gruenberg AM, Walsh D (1995) Examining the validity of DSM-III-R schizoaffective disorder and its putative subtypes in the Roscommon family study. Am J Psychiatry 152: 755–764

Kitamura T, Suga R (1991) Depressive and negative symptoms in major psychiatric disorders. Compr Psychiatry 32: 88–94

Kraepelin E (1971) Dementia praecox and paraphrenia. Facsimile of the 1919 edition. Krieger, Huntington

Kramer MS, Vogel WH, DiJohnson C et al (1989) Antidepressants in 'depressed' schizophrenic inpatients. A controlled trial. Arch Gen Psychiatry 46: 922–928

Langfeldt G (1939) The schizophreniform states. Munksgaard, Copenhagen

Langfeldt G (1956) The prognosis in schizophrenia. Acta Psychiatr Scand Suppl 110: 7–66

Levinson DF, Levitt MEM (1987) Schizoaffective mania reconsidered. Am J Psychiatry 144: 415–425

Levitt JJ, Tsuang MT (1988) The heterogeneity of schizoaffective disorder: implications for treatment. Am J Psychiatry 145: 926–936

*Levitt JJ, Tsuang MT (1990) Atypical psychoses. In: Hyman S, Jennike M (eds) Manual of clinical problems in psychiatry. Little Brown, Boston, pp 45–52

Maj M, Perris C (1985) An approach to the diagnosis and classification of schizoaffective disorders for research purposes. Acta Psychiatr Scand 72: 405–413

Maj M, Starace F, Kemali D (1987) Prediction of outcome by historical, clinical and biological variables in schizoaffective disorder, depressed type. J Psychiatry Res 21: 289–295

Manschreck TC, Maher BA, Beaudette SM, Redmond DA (1997) Context memory in schizoaffective and schizophrenic disorder. Schizophr Res 26: 153–161

Marneros A, Rohde A, Deister A, Risse A (1986) Schizoaffective disorders: the prognostic value of the affective component. In: Marneros A, Tsuang MT (eds) Schizoaffective psychoses. Springer, Berlin Heidelberg New York, pp 155–163

Marneros A, Steinmeyer EM, Deister A et al (1989a) Long-term outcome of schizoaffective and schizophrenic disorders: a comparative study. III. Social consequences. Eur Arch Psychiatry Neurol Sci 238: 135–139

Marneros A, Rohde A, Deister A (1989b) Unipolar and bipolar schizoaffective disroders: a comparative study. II. Long-term course. Eur Arch Psychiatry Clin Neurosci 239: 164–170

Marneros A, Deister A, Rohde A (1990a) Psychopathological and social status of patients with affective, schizophrenic and schizoaffective disorders after long-term course. Acta Psychiatr Scand 82: 352–358

Marneros A, Deister A, Rohde A (1990b) Sociodemographic and premorbid features of schizophrenic, schizoaffective, and affective psychoses. In: Marneros A, Tsuang MT (eds) Affective and schizoaffective disorders. Springer, Berlin Heidelberg New York, pp 130–145

Marneros A, Deister A, Royde A (1991) Stability of diagnoses in affective, schizoaffective and schizophrenic disorders: cross-sectional versus longitudinal diagnosis. Eur Arch Psychiatry Clin Neurosci 241: 187–192

Marneros A, Deister A, Rohde A (1992) Comparison of long-term outcome of schizophrenic, affective and schizoaffective disorders. Br J Psychiatry 161: 44–51

Marneros A, Rohde A, Deister A (1993) Factors influencing the long-term outcome of schizoaffective disorders. Psychopathology 26: 215–224

Maziade M, Bissonnette L, Rouillard E et al (1997) 6p24–22 region and major psychoses in the Eastern Quebec population. Am J Med Genet Neuropsychiatr Genet 74: 311–318

McElroy SL, Dessain EC, Pope HG Jr et al (1991) Clozapine in the treatment of psychotic mood disorders, schizoaffective disorder, and schizophrenia. J Clin Psychiatry 52: 411–414

McElroy SL, Keck PE Jr, Pope HG et al (1992) Valproate in the treatment of bipolar disorder: literature review and clinical guidelines. J Clin Psychopharmacol 12: 42S–52S

**McGlashan TH, Williams PV (1990) Predicting outcome in schizoaffective psychosis. J Nerv Ment Dis 178: 518–520

Müller-Oerlinghausen B, Ahrens B, Grof E et al (1992) The effect of long-term lithium treatment on the mortality of patients with manic-depressive and schizoaffective illness. Acta Psychiatr Scand 86: 218–222

Nelson JC, Mazure CM (1986) Lithium augmentation in psychotic depression refractory to combined drug treatment. Am J Psychiatry 143: 363–366

Okuma T, Yamashita I, Takahashi R et al (1989a) Clinical efficacy of carbamazepine in affective, schizoaffective and schizophrenic disorders. Pharamacopsychiatry 22: 47–53

Okuma T, Yamashita I, Takahashi R et al (1989b) A double-blind study of adjunctive carbamazepine versus placebo on excited states of schizophrenic and schizoaffective disorders. Acta Psychiatr Scand 80: 250–259

Pfohl B, Winokur G (1983) The micropsychopathology of hebephrenic/catatonic schizophrenia. J Nerv Ment Dis 171: 296–300

Pope JG Jr, Lipinski J (1978) Diagnosis in schizophrenia and manic-depressive illness: a reassessment of the specificity of "schizophrenic" symptoms in the light of current research. Arch Gen Psychiatry 35: 811–828

Pope JG Jr, Lipinski JF, Cohen BM, Axelrod DT (1980) Schizoaffective disorder: an invalid diagnosis? A comparison of schizo-affective disorder, schizophrenia and affective disorder. Am J Psychiatry 137: 921–927

Prien RF, Point P, Caffey EMJ, Klett CJ (1972) A comparison of lithium carbonate and chlorpromazine in the treatment of excited schizoaffectives: report of the Veterans Administration and Nation Institute of Mental Health Collaborative Study Group. Arch Gen Psychiatry 27: 182–189

Prusoff BA, Williams DH, Weissman MM, Astrachan BM (1979) Treatment of secondary depression in schizophrenia: a double-blind, placebo-controlled trial of amitriptyline added to perphenazine. Arch Gen Psychiatry 36: 569–575

Ricca V, Galassi F, La Malfa G et al (1997) Assessment of basic symptoms in schizophrenia, schizoaffective and bipolar disorders. Psychopathology 30: 53–58

Rice JP, Goate A, Williams JT et al (1997) Initial genome scan of the NIMH Genetics Initiative bipolar pedigrees: chromosomes 1, 6, 8, 10, and 12. Am J Med Genet Neuropsychiatr Genet 74: 247–253

Rosenthal NE, Rosenthal LN, Stallone F et al (1980) Toward the validation of RDC schizoaffective disorder. Arch Gen Psychiatry 37: 804–810

Samson JA, Simpson JC, Tsuang MT (1988) Outcome studies of schizoaffective disorders. Schizophr Bull 14: 543–554

Simpson JC (1988) Mortality studies in schizophrenia. In: Tsuang MT, Simpson JC (eds) Handbook of schizophrenia, vol 3.

Nosology, epidemiology and genetics of schizophrenia. Elsevier, Amsterdam, pp 245–273

Siris SG, van Kammen DP, Docherty JP (1978) Use of antidepressant drugs in schizophrenia. Arch Gen Psychiatry 35: 1368–1377

Siris SG, Morgan V, Fagerstrom R et al (1987) Adjunctive imipramine in the treatment of postpsychotic depression. Arch Gen Psychiatry 44: 533–539

Smeraldi E, Gasperini M, Macciardi F et al (1983) Factors affecting the distribution of age at onset in patients with affective disorders. J Psychiatr Res 17: 309–317

Spitzer RL, Endicott J, Robins E (1978) Research diagnostic criteria: rationale and reliability. Arch Gen Psychiatry 35: 773–782

Stephens JH, Astrup C, Mangrum JC (1966) Prognostic factors in recovered and deteriorated schizophrenics. Am J Psychiatry 122: 116–1121

Stine OC, Xu J, Koskela R et al (1995) Evidence for linkage of bipolar disorder to chromosome 18 with a parent-of-origin effect. Am J Hum Genet 57: 1384–1394

Taylor MA, Amier N (1994) Are schizophrenia and affective disorder related?: the problem of schizoaffective disorder and the discrimination of the psychoses by signs and symptoms. Compr Psychiatry 35: 420–429

Tien AY, Eaton WW (1992) Psychopathologic precursors and sociodemographic risk factors for the schizophrenia syndrome. Arch Gen Psychiatry 49: 37–46

Torrey EF (1987) Prevalence studies in schizophrenia. Br J Psychiatry 150: 598–608

Tsuang D, Coryell W (1993) An 8-year follow-up of patients with DSM-III-R psychotic depression, schizoaffective disorder, and schizophrenia. Am J Psychiatry 150: 1182–1188

**Tsuang MT (1990) Follow-up studies of schizoaffective disorders: a comparison with affective disorders. In: Marneros A, Tsuang MT (ed) Affective and schizoaffective disorders. Springer, Berlin Heidelberg New York, pp 123–129

Tsuang MT (1991) Morbidity risks of schizophrenia and affective disorders among first-degree relatives of patients with schizoaffective disorders. Br J Psychiatry 158: 165–170

Tsuang MT, Dempsey GM, Fleming JA (1979) Can ECT prevent premature death and suicide in "schizoaffective" patients? J Affect Disord 1: 167–171

*Tsuang MT, Simpson JC, Fleming JA (1986) Diagnostic criteria for subtyping schizoaffective disorder. In: Marneros A, Tsuang MT (eds) Schizoaffective psychoses. Springer, Berlin Heidelberg New York, pp 50–62

Tsuang MT, Simpson JC, Fleming JA (1992) Epidemiology of suicide. Int Rev Psychiatry 4: 117–129

Vaillant GE (1964) Prospective prediction of schizophrenic remission. Arch Gen Psychiatry 11: 509–518

Welner A, Croughan JL, Robins E (1974) The group of schizoaffective and related psychoses – critique, record, follow-up and family studies. I: A persistent enigma. Arch Gen Psychiatry 31: 628–631

Widiger TA, Frances A (1985) The DSM-III personality disorders: perspectives for psychology. Arch Gen Psychiatry 42: 615–623

Wildenauer D, Hallmayer J, Albus M et al (1996) A susceptibility locus for affective and schizophrenic disorder? Psychiatr Genet 6: 152

Zerbin-Rudin E (1986) Schizoaffective and other atypical psychoses: the genetical aspect. In: Marneros A, Tsuang MT (eds) Schizoaffective psychoses. Springer, Berlin Heidelberg New York, pp 225–231

CHAPTER

30

J.E. Cooper, S.P. Singh

Acute and Transient Psychoses

1 Introduction: Disorders but Not Diseases

The traditional binary division of functional psychosis into schizophrenia and manic-depressive psychosis does not account satisfactorily for acute psychotic disorders which have an abrupt onset, short duration, no association with organic states and good outcome. Such disorders are now recognised in both the ICD-10 and the DSM-IV as being separate from the schizophrenic syndromes and the affective psychoses (see the Appendix). This indicates a widespread international agreement that these conditions have a reasonable degree of identity as descriptive clinical states, but in both the classifications this recognition is only as "disorders", i.e. as sets of symptoms and behaviour that are "clinically recognisable" (ICD-10) or "clinically significant" (DSM-IV). There is not yet sufficient knowledge about their aetiology, their genetics or the mechanisms and processes that underlie their symptoms to move them any significant distance along the sliding scale that leads from the modest status of clinical disorder to recognition as a disease entity.

The psychiatric literature is replete with terms for such short-lasting, non-affective and good-outcome psychosis. Floru (1974) collected one hundred such terms, the more common ones being "acute", "atypical", "psychogenic", "reactive", "hysterical" and "third" psychosis. Some of these terms, such as psychogenic psychosis, reactive psychosis and hysterical psychosis, presume an aetiological significance of stress or personality factors, while others, such as atypical psychosis and third psychosis, reflect only a presumption of their distinction from other functional psychosis. Some of the original terms have maintained their regional popularity, such as reactive psychosis in Scandinavia, *bouffée délirante* in France and hysterical psychosis in India. The overlap between concepts is well illustrated in Perris's monograph on cycloid psychoses; he lists 28 different terms proposed by psychiatrists in five different countries that, as he points out, "label quite similar psychotic syndromes" (Perris 1974).

The main justification for the present descriptive separation of the acute and transient psychoses from other functional psychoses does not depend only upon a very acute onset, since the association of an acute onset with a comparatively brief illness has been observed in almost all types of psychoses. For instance, the report on the International Pilot Study of Schizophrenia contains the conclusion that "acute cases have a propensity to recover regardless of the cultural setting" (Sartorius et al. 1978). It is the association of acute onset and favourable outcome with an unusually varied and changing clinical picture that possesses nevertheless a reasonable degree of clinical similarity on recurrence that tips the balance in favour of separation. Other studies provide additional justification (Guinness 1992; Ohaeri 1993; Susser et al. 1995a).

Some of the original authors of these concepts made overt assumptions that these symptoms were caused by underlying pathological processes or mechanisms that would soon be discovered; almost 100 years later, these mechanisms (if they exist at all) are still unknown. Manschreck and Petri (1978) pointed out the several methodological shortcomings of research in this area, especially the inadequacies of definitions, aetiological prejudices, the use of non-standardised assessment methods often under emergency conditions, poor pre-morbid and follow-up assessments and a limited understanding of the interactions between cultural influences and psychopathology. Their extensive review of this complicated topic made it clear that, up until then, the different findings could not be compared in any detail. Cultural issues concerning acute psychoses have been discussed extensively (German 1972; Leff 1981) and are still one of the priority areas for further fieldwork (see also Chap. 14, Vol. 2, Part 1).

An understanding of the present position (illustrated by the criteria to be found in the ICD-10 and DSM-IV; see the Appendix) is best approached by a brief commentary on the historical development of concepts. Further progress is unlikely to occur until large-scale epidemiologically based and prospective studies that use defined criteria and comparable methodology have been carried out. In addition, direct participation by medical and social anthropologists in both the fieldwork and the interpretation of data is necessary to clarify the cultural issues involved. This chapter follows the ICD-10 and DSM-IV in excluding from this group acute psychosis with predominantly depressive or manic symptoms, however abrupt their onset or brief their duration. Two types of psychoses that are usually acute and of brief duration are mentioned in this chapter for completeness, but are described elsewhere in more detail; these are schizo-affective psychoses (see Chap. 29, Vol. 3, Part 1) and puerperal psychoses (see Chap. 13, Vol. 2, Part 1).

2 Historical Development of Concepts

The historical development of concepts will be discussed in four parts: (1) three distinct national traditions (of France, Germany and the Nordic countries), (2) schizo-affective and schizophreniform psychoses, (3) acute psychoses associated with social

changes or with particular cultures and (4) puerperal psychoses.

2.1 Three National Traditions

The three national traditions discussed here overlap in both time and in clinical content. The approximate historical sequence is as follows: (a) the French concept of *bouffée délirante*, (b) the German school of Kleist and Leonhard, emphasising "cycloid psychoses", later developed by Perris, and (c) the Nordic concept of reactive (and psychogenic) psychoses.

2.1.1 Bouffée Délirante

Magnan (in 1886) and Legrain (in 1890) were the first to suggest that a special group of very acute and transient psychotic states could be identified as separate from other and more lasting mental illnesses (Pichot 1982). Their description of *bouffée délirante* has stood the test of time as a clinical state that can be recognised if sought after, but there is very little information about its frequency outside France. Even there, the fairly infrequent use of this diagnosis (around 10% or less in the study of one mental hospital) may be diminishing in favour of more conventional categories such as acute schizophrenia (Johnson-Sabine et al. 1983).

The first descriptions emphasised that 75%–80% of cases had no obvious precipitating cause. In an attempt to convey the abrupt and spectacular nature of the disturbance, the original authors used phrases such as "the delusions may erupt with astonishing speed", "delusions are many and various", "hallucinations are frequent and florid" and "the patient undergoes marked changes of mood, violent oscillations, and 'waves' of delusions" (Ey 1960). Remission of the first episode is described as usually complete within anything from a few days to a few weeks, but recurrence is expected in at least 40%–50% of patients. In addition, several studies of long-term outcome agree that as many as 20% of patients eventually develop obvious chronic schizophrenia.

2.1.2 German School of Kleist and Leonhard and the Concept of Cycloid Psychosis

Kleist (1928) and his pupil Leonhard (1961, 1972) studies over many years some special varieties of acute psychoses and gave them names such as "anxiety-elation psychosis", "confusion psychosis" and "motility psychosis", collectively referred to as "cycloid psychoses" (Chap. 28, Vol. 3, Part 1). Their main features are a sudden onset, a varied and fluctuating clinical picture and rapid resolution; they therefore have many similarities with *bouffée délirante*. Kleist himself made it clear that he regarded these states as nosological entities separate from schizophrenia, since he assumed that "there must be different centres in the brain stem in which these psychoses originate". Most other authorities have not accepted this presumption of a distinct anatomical or physiological basis and have preferred to regard these states as special varieties of either schizophrenia or of manic-depressive psychosis.

The concept of a cycloid psychosis was developed further by Perris in Sweden (Perris 1974), who follows Kleist and Leonhard in regarding it as separate from both the schizophrenic syndrome and the affective psychoses, but on the grounds of familial occurrence and clinical features, rather than a presumption of disturbed brain physiology or structure. More recent attempts to study this concept in the UK (retrospectively re-diagnosed using case notes of patients originally studied for other reasons) have produced only qualified support for the usefulness of this concept as a separate clinical category (Cutting et al. 1978; Brockington et al. 1982a).

2.1.3 Nordic Concept of Reactive (or Psychogenic) Psychosis

Psychogenic psychosis has been a common diagnosis in Scandinavian countries since the turn of the century, used for patients who develop an acute psychosis, usually of short duration with complete recovery, immediately following a severe emotional trauma. Wimmer of Copenhagen started this tradition in 1916, and Faergeman (1963) and others later developed the psychoanalytic and dynamic aspects of the concept. But again, the problems associated with these concepts were evident in Faergeman's follow-up study, in that about half of the 70 patients followed up had clearly developed a schizophrenic or affective illness after 16–19 years. In contrast, Anderson and Laerum (1980) found more recently that only 10% of their cohort of psychogenic psychosis developed schizophrenia over a 10- to 14-year follow-up period. Half the patients who had a recurrence were again diagnosed as having psychogenic psychosis in recurrent episodes. McCabe (1975) applied Jasper's criteria for psychogenic reactions to all patients newly admitted to a Danish hospital. Jasper's criteria include a temporal correlation between onset of illness and a precipitating stress, adequacy of the precipitating stress in psychogenesis and a symbolic meaning such

as defence, wish fulfilment etc. between the contents of the reaction and the stress. Patients thus identified as having a reactive psychosis had an acute onset precipitated by stress, affective symptoms, good premorbid adjustment, preservation of affect, brief duration of illness and absence of family history of schizophrenia. Unfortunately, the application of this logical set of criteria to real patients resulted in many being rejected. McCabe started with 388 patients with a hospital diagnosis of reactive psychosis, of which 250 appeared to fulfil these criteria at the documentary level. Just over 100 of these were examined in the study, only 40 of which fulfilled the criteria above. In other words, his study confirms the puzzling fact that the majority of patients with acute and brief psychoses have illnesses which do not have any understandable relationships with their personality or with their life experiences. Nevertheless, psychogenic and reactive psychoses are still popular concepts in the Scandinavian countries and may be gaining acceptance elsewhere. Stromgren (1989) has argued that reactive psychoses are nosologically distinct and "invariably have a good prognosis". In a recent multicentric study of Nordic psychiatrists, Hansen et al. (1992) found that inter-rater reliability for "reactive psychosis" was as high as for schizophrenia and affective psychosis (but their study was based only upon case summaries, and not upon joint live assessments). A 6-month prospective study of "reactive psychosis" in Chandigarh, North India, found that the diagnosis was stable over time and most patients showed clinical and social recovery over 6 months (Chavan and Kulhara 1988). As compared to Scandinavian patients, Indian patients had more delusions, hallucinations and emotional syndromes.

Other retrospective case studies have emphasised the stability of this diagnosis over time, its relation to stress and its better outcome as compared to schizophrenia. It is likely that the authors of many of these reports also assume that schizophrenia is by definition a comparatively long-standing illness with an unfavourable outcome (Kapur and Pandurangi 1979; Andersen and Laerum 1980; Stephens et al. 1982; Beighley et al. 1992). Reactive psychosis is reported to be similar to hysterical psychosis except for the presence of premorbid histrionic traits in the latter (Modestin and Bachmann 1992) and also similar to the French diagnostic category of *bouffée délirante* (Pichot 1986).

2.2 Schizo-affective and Schizophreniform Psychoses

Schizo-affective disorders are dealt with in Chap. 29 (Vol. 3, Part 1), but deserve a brief mention here because some of the patients to whom this term has been applied in the past are also covered by the other acute psychotic states described in this chapter. The term "schizo-affective" was first used by Kasanin (1933) in the USA to describe patients whose psychosis was characterised by "a very sudden onset in a setting of marked emotional turmoil with a distortion of the outside world ... the psychosis lasts a few weeks to a few months and is followed by a recovery." He also emphasised that these psychoses were usually associated with a severe mental stress. It seems clear that Kasanin regarded these patients as having unusual forms of schizophrenia rather than a different illness, but the resemblance to the clinical states described above is obvious. Schizo-affective has more recently been used with a quite different meaning in both the ICD-10 and DSM-IV to describe illnesses with a mixture of marked schizophrenic and affective symptoms present at the same time, in more or less equal proportions and with no necessary implication of abrupt onset, psychological causation or short duration (although these may well be present).

The term "schizophreniform" has also been subject to a change of meaning over the years. It was originally used by Langfeldt (1939) to describe patients with atypical forms of schizophrenia that had a clear psychogenic cause, an acute onset, a rapid response to treatment and a favourable outcome; Langfeldt was interested in differentiating these illnesses from more typical or "process" types with an insidious onset and poor outcome. The DSM-III (American Psychiatric Association 1980) used it to cover quite different patients who had typical symptoms of schizophrenia but with a duration less than the 6 months required for that diagnosis. There has been a change in the DSM-IV, which brings it back closer to the original meaning of Langfeldt, but it remains separate from the DSM-IV category of "brief psychotic disorder". The term "schizophreniform" has never been used in the ICD system.

2.3 Acute Psychoses Associated with Social Changes or with Particular Cultures

The occurrence of atypical and acute psychiatric disorders, including psychoses, has often been reported in association with potentially stressful social changes such as immigration (Lambo 1965; Leff 1981; Collins et al. 1996). There are also many reports suggesting that people living in non-European cultures are more likely to develop atypical psychoses than Europeans living in Europe (German 1972). This, of course, is using "atypical" as viewed from a European perspective, and it is not surprising that medical anthropologists, and some psychiatrists from non-

European cultures, have objected to the implicit assumption that what European psychiatrists regard as atypical is accepted as such in other parts of the world. This is one of the main issues that show the need for a contribution from social anthropologists to future studies of acute psychoses.

Psychiatric disorders of abrupt onset and often spectacular symptoms that are alleged to occur only in specific cultural settings have been described for many decades and are now featured in the Research Diagnostic Criteria of the ICD-10 (and also in the DSM-IV). Often previously described as the "culture-bound psychoses", there is now increasing agreement that these are not psychoses, however that term is defined. The more they are observed at first hand and in detail, the more apparent it becomes that they are best regarded as culturally influenced (and in a sense socially acceptable) states of anxiety, acute depression, distress, aggression and attention-seeking behaviour (Yap 1974; Leff 1981; see also Chap. 14, Vol. 2, Part 1).

2.4 Other Concepts

The term "hysterical psychoses" was revived briefly in the USA (Hollander and Hirsch 1964; Hirsch and Hollander 1969) in an attempt to emphasise the often individual and reactive clinical picture found in these disorders. The concept is of a disorder with a sudden and dramatic onset, temporally related to a psychosocial stressor in an individual with a "hysterical character". The clinical symptoms included hallucinations, delusions, depersonalisation, transient and circumscribed thought disorder, lack of affective flattening and grossly disruptive behaviour. The disorder lasts between 1 and 3 weeks with full recovery and no residual symptoms. Hysterical psychosis has also been described in India (Kuruvilla and Sitalakshmi 1964) and Switzerland (Modestin and Bachmann 1992).

2.5 Puerperal Psychosis

There has been a long-standing debate about whether puerperal psychosis should be regarded as a separate clinical entity. Because of the association with childbirth, puerperal psychoses have been given much more attention in recent years than other varieties of acute psychoses, and most of the information available is of good quality. The large majority of experts now accept that puerperal psychoses are best regarded as unusually acute variants of either affective psychoses (about 80%–90% of cases) or schizo-affective and schizophrenic psychoses (Brockington et al. 1982b). They are important because they are the most common of the very acute and transient psychoses, occurring after about two births in every 1000. Around 75% of these illnesses become manifest within 16 days of delivery. The most frequent pattern is that the mother becomes perplexed, restless and confused (although not disoriented in the organic sense) a few days after giving birth; over the next week or so, a more obvious affective state develops with manic symptoms being quite common. The optimal treatment and management of women with puerperal psychosis requires skilled assessment by experienced staff working together as a specialised team; it should be possible to combine admission facilities for a small number of mothers and babies with an extensive community and home-visiting service (Oates 1989).

3 Current Nosological Status

The increasingly widespread use of standardised interviewing and rating methods, e.g. in the World Health Organisation studies on schizophrenia and related disorders (Sartorius et al. 1978; Jablensky et al. 1992), has led to an improvement of the quality of research information on some important aspects of psychotic disorders. Only a limited part of this information was relevant to acute and atypical psychoses. There was nevertheless sufficient information to encourage a new and more systematic approach to this subject in the preparation of ICD-10 Chap. V (F). The structure of the final version of F23 – "Acute and Transient Psychoses" – allows the recording of the main clinical syndromes that have typified the national traditions already discussed but without making assumptions about relationships between the syndromes and the presence or absence of either acute stress or typical schizophrenic symptoms (see the Appendix). This point needs emphasis, because of the common (but false) clinical assumptions that a psychological stress can always be identified as the cause of an acute psychosis and that typical schizophrenic symptoms are never present. The 3-month duration criteria in the ICD-10 are supported by the finding of a recent study that 80% of patients with acute psychosis have a mean episode duration of 2 months (Susser et al. 1995b).

The diagnostic criteria for brief psychotic disorder in the DSM-IV (see the Appendix; American Psychiatric Association 1994) are broadly similar to those of F23 in the ICD-10, but the category has fewer subdivisions; it is probably compatible with a condensed form of F23.

4 Epidemiology

There are very few epidemiological studies of acute and brief psychoses in which standardised and reliable methods for the assessment of symptoms and diagnoses have been used. Using the data from some large-scale WHO International Collaborative Studies of psychotic disorders (Jablensky et al. 1992; Varma et al. 1992; Cooper et al. 1990), and also from a separate study in the USA, Susser and colleagues have found evidence to support the separation of acute brief psychoses from both schizophrenia and the affective psychoses. The field work of these studies preceded the publication of both the ICD-10 and DSM-IV, but the criteria put together for "acute non-affective remitting psychoses" are broadly similar to those of the ICD10 and DSM-IV. (Susser and Wanderling 1994; Susser et al. 1995a,b). These studies suggest that there may be large differences in the incidence of acute brief psychosis between different centres, but without the differences in incidence between the sexes that has been found for both schizophrenia and affective psychoses. However, until studies are done which use methods designed especially for the study of acute psychoses, these conclusions must be regarded as provisional.

5 Clinical Features

The clinical features of acute psychotic states have already been described in Sect. 2. The dramatic and often unexpected nature of the onset is the main feature that sets them apart from other psychoses, together with the liability for there to be rapid changes within a few hours from one predominant type of symptom to another. A good pre-morbid adjustment and the frequent absence of a family history of psychiatric disorder can make the development of such a clearly abnormal state very surprising for family and friends. The perplexity and "confusion" that are often prominent in the earliest stages need to be distinguished from the more persistent and stable disorientation of organic states, but with careful observation this distinction is usually not difficult to make. The content of the psychotic symptoms is sometimes understandable in the light of the patient's cultural background and experience of recent stress, but there are no clear links in a large proportion of patients. It is not uncommon to find that a person presenting with acute psychosis also has a mild non-specific pyrexia (Collins et al. 1996) or a history of having taken non-toxic amounts of alcohol, other drugs or local herbal remedies (i.e. amounts or substances that are not in themselves unusual for that person or that social setting). The significance of this is unknown, and in most cases it is likely that the substances were taken in the hope of counteracting the unpleasant early symptoms of psychosis that were already being experienced.

A large proportion of patients with acute psychoses make a full recovery within a few months, and three quarters may be in full remission at 1 year (Varma et al. 1996); others with persistent symptoms will require a change of diagnosis to schizophrenia, affective psychosis or persistent delusional disorder.

6 Management and Treatment

The main aims of the early management are to facilitate a rapid resolution of symptoms and to minimise the disruptive influence of the illness on the patient's personal and family relationships and activities.

6.1 Initial Assessment

Individuals with acute psychoses usually present initially to the medical services before being seen by an experienced psychiatrist, often at inconvenient times in inconvenient places. It is often many hours, if not several days, before detailed and skilled psychiatric assessment is carried out, by which time the most acute (and clinically important) phase has probably passed or has been modified by medication. The initial assessment should include the usual detailed personal history and full evaluation of the mental state. Interviews with close relatives or friends will be needed, since it is likely that the patient will be able to give only information that is heavily influenced by the symptoms, or at best only a partial account. Focus should be on the onset and early course of symptoms, with special attention paid to possible relationships with life events. Enquiry into pre-morbid coping skills, past vulnerability to stress, alcohol and drug history, family history of psychiatric illness and current social situation including family support will allow the development of an individual treatment plan, assuming that laboratory investigations have also been done to ensure that no physical illnesses are present. Careful assessment of cognitive functions is of particular importance, since perplexity in an acute psychosis can mimic the acute confusional state due to organic causes. However, patients in these "functional" acute

psychoses do not display the marked and persistent disorientation of an organic delirium.

In the differential diagnosis, a variety of medical disorders need to be considered, the most important being drug-induced delirium, head injury, epilepsy and drug-induced psychosis. It is likely that a number of laboratory investigations will be required, including a toxicology screen.

6.2 Initial Treatment

Patients will usually need to be admitted to hospital. This allows close observation of the mental state (often rapidly changing) and provides a safe environment in which overactivity and emotional turmoil can be contained. Hospitalisation also makes it possible for the patient to be observed for a drug-free period of several days; this is important, since some acute psychotic states may remit within days without medication. If drug therapy is started at the earliest possible time in every case, spontaneous remission cannot be differentiated from response to medication. Medication should only be started immediately if the symptoms are particularly severe or distressing. If the symptoms persist for more than 3–4 days, medication should be started.

Anti-psychotics are the mainstay of drug treatment of acute psychosis. A high-potency anti-psychotic such as haloperidol may be used, although others such as chlorpromazine have the advantage of sedation. Oral administration is usually sufficient, but in some patients liquid preparations to ensure intake or intramuscular use for more immediate action may be necessary. The anti-psychotics should be given initially in a low dose, with the increases adjusted according to the size and behaviour of the individual. There is some evidence that treatment response may be quicker in acute psychosis than in schizophrenia.

Anti-parkinsonian agents should not be given routinely, since only a comparatively short period of medication may be required. It is best to reserve their use for when comparatively large doses of anti-psychotics are needed and when extrapyramidal signs have appeared. In these circumstances, their use may also improve later compliance with drug therapy.

The short-term use of benzodiazapines in acute psychotic states is well established. They allow a rapid control of agitation, excitement and other behavioural disturbances, and their use may permit a lower dose of anti-psychotics. If excitement, overactivity and euphoria are very prominent (suggesting that the illness might be better viewed as a manic episode), then it is worth remembering that lithium can have a short-term anti-manic therapeutic effect, separate from its use as a long-term prophylactic.

A review of pharmacotherapy of acute psychoses is given in Remington et al. (1998).

6.3 Intermediate Treatment and Management

The early stage of the illness does not provide clues which allow the confident prediction of which patients will have a rapid resolution of symptoms with no further relapse (about one in four) and which patients will have enduring symptoms (again about one in four). Similarly, it is impossible to predict which patients will relapse after recovery from an episode. It is therefore best to adopt a cautious approach to decisions about how long to continue anti-psychotic medication. All patients with an acute psychotic illness of any type, however brief, should continue anti-psychotic medication for at least 6 months. For the last 3 months of this period, a comparatively low dose is justified so long as no symptoms re-appear. If schizophrenic symptoms were present at any time, the anti-psychotic medication should be continued for at least 1 year. The withdrawal of anti-psychotic medication should never be abrupt. The dose should be diminished gradually by stages over 2–3 months with the full knowledge of the patient and family, so that all concerned can watch out for early signs of relapse. A long-term maintenance regime of anti-psychotics (as used for other persistent psychotic disorders such as delusional disorder and the schizophrenic syndrome) is justified for patients who have two or more episodes of acute psychosis within a period of 2 years.

Patients whose symptoms persist beyond the time limits suggested in the diagnostic classifications will require a change in diagnosis; at this stage, they should be subject to a complete clinical review, with the adoption of a new and longer-term programme of management and treatment. The reader may also wish to refer to Chap. 11 (Vol. 3, Part 1) for the treatment of acute schizophrenia, Chap. 23 (Vol. 3, Part 1) for the treatment of acute mania and depression and Chaps. 13–17 (Vol. 2, Part 2) for the treatment of acute psychiatric states accompanying somatic illness.

6.4 Psychosocial Measures

As with all patients admitted to hospital with severe psychiatric disorders, the maintenance of contact between the patient, family and friends should be given a high priority by all members of the therapeutic team. The unexpected and often spectacular behaviour frequently proves to be unusually distressing for all concerned, and both the patient and those most directly

affected will benefit from positive contacts and support from the professional staff. Hospital visiting should be encouraged, even during the acute stage.

In all cases, individual "supportive" psychotherapy with a key worker should be started in the acute phase, and as the acute psychotic symptoms diminish, various psychosocial interventions should be considered. The therapeutic aims are to restore the patient's self-esteem and reduce any feelings of self-blame and inadequacy and fears of further violence or loss of control. It is important to assess the patient's vulnerability to stress and life events which may have triggered the psychosis; this should include a review of the coping mechanisms of the patient and family. It may be possible to improve problem-solving skills and to suggest strategies that may help in the avoidance of future personal crises. A return to normal functioning may be helped by a period in a structured environment such as a day-hospital, and long-term social support or psychotherapy should be considered for those patients with enduring personality or family problems.

In all these measures, the support and co-operation of the family are vital in ensuring the best possible compliance with treatment and to pick up early signs of relapse. Educative discussions by various members of the therapeutic team with both patient and family about the symptoms, the nature of the illness, problems and purposes of medication and likely prognosis and outcome should therefore be a prominent part of the management package.

6.5 Early Intervention and Cognitive Therapy

The early detection of patients developing any type of psychosis has come into prominence recently (Birchwood et al. 1997; Yung et al. 1996), with the aim of facilitating the earliest possible intervention in the psychotic clinical state. The addition of new forms of cognitive therapy for both acute and longer-lasting psychotic symptoms to the general principles of treatment and management described above is often a part of this new approach (Drury et al. 1996a,b; Kuipers et al. 1997). This work is attracting considerable interest and seems likely to develop rapidly in the near future.

Appendix. ICD-10 Criteria for Acute and Transient Psychotic Disorders and DSM-IV Criteria for Brief Psychotic Disorder

ICD-10 Chapter V (F)

(For reasons of space, only the brief description taken from the short glossary is given here. More detailed descriptions are contained in the *Clinical Descriptions and Diagnostic Guidelines* (WHO 1992) and the *Diagnostic Criteria for Research* (WHO 1993) of the ICD-10).

F23 Acute and Transient Psychotic Disorders

A heterogeneous group of disorders characterised by the acute onset of psychotic symptoms such as delusions, hallucinations and perceptual disturbances and by the severe disruption of ordinary behaviour. Acute onset is defined as a crescendo development of a clearly abnormal clinical picture in about 2 weeks or less. For these disorders there is no evidence of organic causation. Perplexity and puzzlement are often present, but disorientation for time, place and person is not persistent enough to justify a diagnosis of organically caused delirium. Complete recovery usually occurs within a few months, often within a few weeks or even days. If the disorder persists, a change in diagnosis will be necessary. The disorder may or may not be associated with acute stress, defined as usually stressful events preceding the onset by 1–2 weeks.

F23.0 Acute Polymorphic Psychotic Disorder Without Symptoms of Schizophrenia

An acute psychotic disorder in which hallucinations, delusions or perceptual disturbances are obvious but markedly variable, changing from day to day or even from hour to hour. Emotional turmoil, with intense transient feelings of happiness or ecstasy, or anxiety and irritability, is also frequently present. The polymorphism and instability are characteristic for the overall clinical picture and the psychotic features do not justify a diagnosis of schizophrenia (F20.-). These disorders often have an abrupt onset, developing rapidly within a few days, and they frequently show a rapid resolution of symptoms with no recurrence. If the symptoms persist the diagnosis should be changed to persistent delusional disorder (F22.-).

Includes:

- *Bouffée délirante* without symptoms of schizophrenia, or unspecified
- Cycloid psychosis without symptoms of schizophrenia, or unspecified

F23.1 Acute Polymorphic Disorder with Symptoms of Schizophrenia

An acute psychotic disorder in which the polymorphic and unstable clinical picture is present, as described in F23.0; despite this instability, however, some symptoms typical of schizophrenia are also in evidence for the majority of the time. If the schizophrenic symptoms persist the diagnosis should be changed to schizophrenia (F20.-).

Includes:

- *Bouffée délirante* with symptoms of schizophrenia
- Cycloid psychosis with symptoms of schizophrenia

F23.2 Acute Schizophrenia-Like Psychotic Disorder

An acute psychotic disorder in which the psychotic symptoms are comparatively stable and justify a diagnosis of schizophrenia but have lasted for less than about one month; the polymorphic unstable features, as described in F23.0, are absent. If the schizophrenic symptoms persist, the diagnosis should be changed to schizophrenia (F20.-).

Includes:

- Acute (undifferentiated) schizophrenia
- Brief schizophreniform disorder/psychosis
- Oneirophrenia
- Schizophrenic reaction

F23.3 Other Acute Predominantly Delusional Psychotic Disorders

Acute psychotic disorders in which comparatively stable delusions or hallucinations are the main clinical features, but do not justify a diagnosis of schizophrenia (F20.-). If the delusions persist, the diagnosis should be changed to persistent delusional disorder (F22.-).

Includes:

- Paranoid reaction
- Psychogenic paranoid psychosis

F23.4 Other Acute and Transient Psychotic Disorders

Any other specified acute psychotic disorders for which there is no evidence of organic causation and which do not justify classification to F23.0–F23.3.

Note: Any other acute psychotic disorders that are not classifiable under any other category in F23 (such as acute psychotic states in which definite delusions or hallucinations occur but persist for only small proportions of the time) should be coded here. States of undifferentiated excitement should also be coded here if more detailed information about the patient's mental state is not available, provided that there is no evidence of an organic cause.

Diagnostic Criteria for Brief Psychotic Disorder of DSM-IV

A. Presence of one (or more) of:
   1. Delusions
   2. Hallucinations
   3. Disorganised speech (e.g. frequent derailment or incoherence)
   4. Grossly disorganised or catatonic behaviour

Note: Do not include a symptom if it is a culturally sanctioned response pattern.

B. Duration of an episode of the disturbance is at least 1 day but less than one month, with eventual full return to level of premorbid level of functioning.

C. The disturbance is not better accounted for by a Mood Disorder With Psychotic Features, Schizoaffective Disorder, or Schizophrenia and is not due to the direct physiological effects of a substance (e.g. a drug of abuse, a medication) or a general medical condition.

Specify if:

- With marked stressor(s) (brief reactive psychosis): if symptoms occur shortly after and apparently in response to events that, singly or together, would be markedly stressful to almost anyone in similar circumstances in the person's culture
- Without marked stressor(s): if psychotic symptoms do *not* occur shortly after, or are not apparently in response to events that, singly or together, would be markedly stressful to almost anyone in similar circumstances in the person's culture
- With postpartum onset: if onset within 4 weeks postpartum

7 References

American Psychiatric Association (1980) Diagnostic and statistical manual of mental disorders, 3rd edn (DSM-III). American Psychiatric Assocation, Washington, DC

American Psychiatric Association (1994) Diagnostic and statistical manual of mental disorders, 4th edn (DSM-IV). American Psychiatric Association, Washington, DC

Anderson J, Laerum H (1980) Psychogenic psychosis: a retrospective study with special reference to clinical course and prognosis. Acta Psychiatr Scand 62: 331–342

Beighley P, Brown G, Thompson J (1992) DSM-III-R brief reactive psychosis among Air Force recruits. J Clin Psychiatry 53: 283–288

Birchwood M, McGorry P, Jackson H (1997) Early intervention in schizophrenia. Br J Psychiatry 170: 2–5

Brockington IF, Perris C, Kendell RE, Hillier VE, Wainwright S (1982a) The course and outcome of cycloid psychosis. Psychol Med 12: 97–105

Brockington IF, Winokur G, Dean C (1982b) Puerperal psychoses. In: Brockington IF, Kumar R (eds) Motherhood and mental illness. Academic, London

Chavan BS, Kulhara P (1988) A clinical study of reactive psychosis. Acta Psychiatr Scand 88: 712–715

Collins P, Wig N, Day R et al (1996) Psychosocial and biological aspects of acute brief psychosis in three developing country sites. Psychiatr Q 67: 177–193

Cooper JE, Jablensky A, Sartorius N (1990) WHO collaborative studies on acute psychoses using the SCAAPS schedule. In: Stefanis CN, Rabavilas AD, Soldatos CR (eds) Psychiatry: a world perspective, vol 1. Elsevier, Amsterdam, pp 185–192

Cutting J, Clare A, Mann A (1978) Cycloid psychosis: an investigation of the disease concept. Psychol Med 8: 637–648

Drury V, Birchwood M, Cochrane R, Macmillan F (1996a) Cognitive therapy and recovery from acute psychosis: a controlled trial. I. Impact on psychotic symptoms. Br J Psychiatry 169: 593–601

Drury V, Birchwood M, Cochrane R, Macmillan F (1996b) Cognitive therapy and recovery from acute psychosis: a controlled trial. II. Impact on recovery time. Br J Psychiatry 169: 602–607

Ey H, Bernard P, Brisset C (1960) Psychoses delirante aigues. In: Manuel de psychiatrie. Masson, Paris, pp 244–254 (Reprinted in: Hirsch SR, Shepherd M (1974) (eds) Themes and variations in European psychiatry. Wright, Bristol)

Faergeman P (1963) Psychogenic psychoses. Butterworth, London

Floru L (1974) Reactive, psychogene und schizophrenie-ähnliche Psychosen. Ein Überblick des Problems. Schweiz Arch Neurol Neurochir Psychiatrie 114: 107–123

German GA (1972) Aspects of clinical psychiatry in sub-Saharan Africa. Br J Psychiatry 121: 461–480

Guinness E (1992) Brief reactive psychosis and the major functional psychoses: descriptive case studies in Africa. Br J Psychiatry Suppl 16: 24–41

Hansen H, Dahl A, Bertelsen A et al (1992) The Nordic concept of reactive psychosis – a multicenter reliability study. Acta Psychiatr Scand 86: 55–59

Hirsch S, Hollander MH (1969) Hysterical psychosis; clarification of the concept. Am J Psychiatry 125: 909–915

Hollander MH, Hirsch SJ (1964) Hysterical psychosis. Am J Psychiatry 120: 1066–1074

Jablensky A, Sartorius N, Ehrenberg G, Anker M, Korten A, Cooper JE, Day R, Bertelsen A (1992) Schizophrenia: manifestations, incidence and course in different cultures. A World Health Organisation ten country study. Psychol Med Monogr Suppl 20: 7–16

Johnson-Sabine EC, Mann AH, Jacoby RJ, Wood KH, Peron-Magnan P, Olie JP, Deniker P (1983) Bouffeé délirante: an examination of its current status. Psychol Med 13: 771–778

Kapur R, Pandurangi A (1979) A comparative study of reactive psychosis and acute psychosis without precipitating stress. Br J Psychiatry 135: 544–550

Kasanin J (1933) The acute schizoaffective psychoses. Am J Psychiatry 13: 97–126

Kleist K (1928) Über zykloide, paranoide und epileptoide Psychosen und über die Frage der Degenerationspsychosen. Reprinted in: Hirsch SR, Shepherd M (1974) (eds) Themes and variations in European psychiatry. Wright, Bristol

Kuipers E, Garety P, Fowler D, Dunn G, Bebbington P, Freemen D, Hadley C (1997) London–East Anglia randomised controlled trial of cognitive-behavioural therapy for psychosis. 1. Effects of the treatment phase. Br J Psychiatry 171: 319–327

Kuruvilla K, Sitalakshmi N (1964) Hysterical psychosis. Ind J Psychiatry 120: 1066–1074

Lambo T (1965) Schizophrenia and borderline states. In: de Reuck AVS, Porter R (eds) Transcultural psychiatry. Ciba Found Symp: 78–94

Langfeldt G (1939) The schizophrenic states. Munksgaard, Copenhagen

Leff J (1981) Psychiatry around the globe. Dekker, New York

Leonhard K (1961) Cycloid psychoses – endogenous psychoses which are neither schizophrenic nor manic-depressive. J Ment Sci 107: 633–648

Leonhard K (1972) Aufteilung der endogenen Psychosen in der Forschungsrichtung von Wernicke und Kleist. In: Kisker KP, Meyer JE, Müller M, Strömgren E (eds) Psychiatrie der Gegenwart, 2nd edn, vol 2, part 1. Springer, Berlin Heidelberg New York

Manschreck TC, Petri M (1978) The atypical psychoses. Culture Med Psychiatry 2: 233–268

McCabe M (1975) Reactive psychosis. Acta Psychiatr Scand Suppl 259: 1–33

Modestin J, Bachmann K (1992) Is the diagnosis of hysterical psychosis justified?: clinical study of hysterical psychosis, reactive/psychogenic psychosis, and schizophrenia. Compr Psychiatry 33: 17–24

Oates M (1989) Management of major mental illness in pregnancy and the puerperium. In: Oates M (ed) Psychological aspects of obstetrics and gynaecology. Baillieres Clin Obstet Gynaecol 3/4: 905–920

Ohaeri J (1993) Long-term outcome of treated schizophrenia in a Nigerian cohort. Retrospective analysis of 7-year follow-ups. J Nerv Ment Dis 181: 514–516

Perris C (1974) A study of cycloid psychoses. Acta Psychiatr Scand Suppl 253: 1–77

Pichot P (1982) The diagnosis and classification of mental disorders in French-speaking countries: background, current views and comparison with other nomenclatures. Psychol Med 12: 475–492

Pichot P (1986) The concept of "bouffeé délirante" with special reference to the Scandinavian concept of reactive psychosis. Psychopathology 19: 35–43

Remington GJ, Bezchlibnyk-Butler K (1998) Current concepts in the pharmacotherapy of acute psychosis. CNS Drugs 9: 191–202

Sartorius N, Jablensky A, Shapiro R (1978) Cross-cultural differences in the short-term prognosis of schizophrenic psychoses. Schizophr Bull 4: 102–113

Stephens JH, Shaffer JW, Carpenter WT (1982) Reactive psychosis. J Nerv Ment Dis 170: 57–663

Stromgren E (1989) The development of the concept of reactive psychoses. Br J Psychiatry 154[Suppl 4]: 47–50

Susser E, Wanderling J (1994) Epidemiology of nonaffective acute remitting psychosis vs. schizophrenia. Arch Gen Psychiatry 51: 294–301

Susser E, Varma VK, Malhotra S, Conover S, Amador XF (1995a) Delineation of acute and transient psychoses in a developing country setting. Br J Psychiatry 167: 216–219

Susser E, Fennig S, Jandorf L, Amador X, Bromet E (1995b) Epidemiology, diagnosis and course of brief psychoses. Am J Psychiatry 152: 1743–1748

Varma VK, Malhotra S, Jiloha RC (1992) Acute non-organic states in India: symptomatology. Ind J Psychiatry 34: 89–101

Varma VK, Malhotra S, Yoo ES et al (1996) Course and outcome of acute non-organic psychotic states in India. Psychiatr Q 67: 195–207

WHO (1992) The ICD-10 classification of mental and behavioural disorders: clinical descriptions and diagnostic guidelines. WHO, Geneva

WHO (1993) The ICD-10 classification of mental and behavioural disorders: diagnostic criteria for research. WHO, Geneva

Yap PK (1974) Comparative psychiatry. University of Toronto Press, Toronto

Yung A, McGorry P (1996) The prodromal phase of first episode psychosis; past and current conceptualisations. Schizophr Bull 22: 353–371

CHAPTER

31

M.S. Nobler, H.A. Sackeim

Electroconvulsive Therapy

1 Introduction

Despite its long history of use as an effective treatment for severe psychiatric illness, electroconvulsive therapy (ECT) is often misunderstood, even by mental health professionals. This is partly due to negative portrayals in the media, which often have distorted crucial facts about patient selection (e.g. ECT being used as punishment for antisocial behavior) or treatment technique (e.g. ECT being administered without anesthesia). However, the lack of knowledge about ECT may also have resulted from the relative lack of controlled research, especially in comparison to the large body of psychopharmacological research that emerged in the 1960s and 1970s. These disparate phenomena, and the efficacy of psychopharmacological agents, contributed to the reduced utilization of ECT from the 1960s to the early 1980s. This trend has reversed, and the past 15 years has seen renewed interest in ECT, reflected both in stabilization of its utilization and increased research, especially in the areas of treatment technique. In addition to highlighting new findings, this chapter will also review central clinical issues in ECT as well as research into mechanisms of action.

The use of ECT also has an impact on health care economics. Recent research has shown that the length of hospitalization and associated financial costs are reduced in depressed inpatients who receive ECT early during hospitalization compared to inpatients who do not receive ECT. This effect is particularly marked in psychotic depression (Olfson et al. 1998).

2 Patient Selection: Indications and Contraindications

2.1 Diagnostic Considerations

One of the most interesting aspects of ECT is that it is an effective treatment for a variety of severe psychiatric and neuropsychiatric syndromes. However, the broad spectrum of action of ECT should never obviate the necessity of careful diagnostic assessment prior to a treatment recommendation.

While introduced as a treatment for severe psychosis, early on it became evident that ECT had profound antidepressant effects. This clinical observation withstood rigorous scientific scrutiny, as both direct comparisons of ECT with antidepressant medications and a series of sham-controlled studies (i.e. general anesthesia without the passage of current) in the United Kingdom demonstrated the efficacy of ECT in major depressive disorder (Sackeim et al. 1995). Indeed, no other somatic treatment has ever been demonstrated to be superior in efficacy to ECT in the treatment of major depression, and ECT soon became the "gold standard" against which new antidepressant medications were compared. Response rates for major depression in contemporary research studies, which often employ stringent criteria, are on the order of 65%–80%. A conservative estimate of the efficacy rates in clinical samples would be even greater, perhaps in the range of 80%–90%.

Though less well studied, both clinical reports and controlled research have demonstrated a marked antimanic effect of ECT (Mukherjee et al. 1994). Indeed, ECT appears to be effective in manic patients who have failed treatment with neuroleptics or lithium. The clinical significance of the role of ECT in mania cannot be overemphasized. In fact, ECT may be a life-saving procedure in cases of manic delirium.

The role of ECT in the treatment of schizophrenia is perhaps both the least understood and the most debated. As mentioned, the first use of ECT was in chronically psychotic, institutionalized patients with essentially no other options for treatment. Partly due to sociocultural forces and partly due to the introduction of neuroleptic medications in the 1950s, the use of ECT for the treatment of schizophrenia receded. Since very little research had been conducted, the guidelines for using ECT in patients with schizophrenia that have evolved over the years are conservative and rather vague. For example, ECT is considered appropriate in patients with schizophrenia where "affective symptomatology is prominent" (American Psychiatric Association 1990, p. 8). At least two lines of research call such conservatism into question (Krueger and Sackeim 1995). First, a reevaluation of the few published studies in this area indicate that the efficacy of ECT in schizophrenia rivaled that of neuroleptics and suggested the particular usefulness of ECT early in the course of illness. Second, there is now a modest body of literature demonstrating that the combination of ECT and neuroleptics may be useful in cases of medication-resistance and is more effective than either ECT or antipsychotic medications alone.

In addition to these major psychiatric syndromes, ECT is useful in the treatment of several syndromes that bridge the gap between psychiatry and neurology (American Psychiatric Association 1990). One such illness is catatonia, which is uniquely responsive to ECT. Perhaps related on a pathophysiological level to catatonia is the neuroleptic malignant syndrome, where ECT has been reported to be of substantial benefit. Among neurological illnesses, Parkinson's disease (PD) has received attention as an indication

for ECT (Kellner et al. 1994). Many patients with PD are prone to develop severe depression. In addition, some patients become increasingly refractory to antiparkinsonian medications, leading to psychomotor stupor, while others develop psychosis secondary to pharmacological treatment. ECT is often dramatically effective in these clinical situations. Independent of its antidepressant and antipsychotic mechanisms in PD, ECT also appears to ameliorate motor symptoms, presumably via dopaminergic pathways. Unfortunately, such benefit can be short-lived.

In sum, while the majority of clinical outcome studies in the modern era involve patients with major depression, ECT should also be considered as a valuable tool in the treatment of mania (especially medication-resistant mania), catatonia, medication-resistant psychosis, severe parkinsonism, and certain rare organic mental states.

2.2 Medical Risk–Benefit Ratio

In addition to the appropriate diagnostic indication, the decision to employ ECT is also based on the overall medical health of the patient. In fact, ECT is often considered because a patient cannot tolerate the side effects of psychotropic medication. While ECT is generally a safe procedure, any potential adverse medical effects (see Sect. 4) must be weighed against the potential benefit of the reduction of severe psychopathology, such as extreme suicidality. All patients referred for ECT must undergo a basic medical evaluation, including physical examination, electrocardiogram, and blood and urine testing. A chest radiograph is also generally recommended. Neuroimaging with computed tomography (CT) or magnetic resonance imaging (MRI) is definitely indicated when there is clinical suspicion of intracranial pathology, but the cost-effectiveness of routine screening with these techniques is questionable.

While there are no *absolute* contraindications for the use of ECT, there are situations that pose increased medical risk (American Psychiatric Association 1990; Abrams 1992). The two most commonly encountered areas are cardiac disease (e.g. ischemic heart disease, severe congestive heart failure, or serious arrhythmia) and central nervous system disease (space-occupying lesions resulting in increased intracranial pressure). Other relative contraindications include unstable vascular malformations, pheochromocytoma, and retinal detachment. If any of these situations are known to exist (or are discovered during the pre-ECT medical evaluation), the next step should be to consult the appropriate medical or surgical specialist and obtain any additional diagnostic tests (e.g. cardiac stress test). Once the evaluation is complete, the patient, his or her family, and the physician must then reassess the potential benefits from ECT (e.g. rapid amelioration of suicidal intent) versus the potential medical risks (e.g. cardiac ischemia). Also central in this decision-making process are the possibilities of adding extra safeguards to ECT, such as modification of the cardiovascular response with intravenous beta blockers. Ultimately, if the decision is made to proceed with ECT, then the medical record and the informed consent process should be clear in indicating all aspects of this discussion.

2.3 Predicting Response

As with other treatment interventions, the ability to predict a likely responder (or nonresponder) to ECT would be of considerable clinical utility. In fact, over the past four decades there has been a significant effort to identify clinical or biological predictors of outcome with ECT (Nobler and Sackeim 1996). The vast majority of this research has involved patients with major depression. Ironically, from a statistical point of view, the uniformly high overall response rates to ECT make it difficult to discern possible outcome predictors. Indeed, some have contended that if the clinical diagnosis of major depression is accurate, then the issue of predicting response to ECT is moot. On the other hand, since ECT is increasingly provided as a treatment of "last resort," failure to respond often leaves patients and their families quite demoralized. Obviously, being able to provide patients with more realistic estimates of likelihood of response would still be valuable.

With these caveats in mind, some general conclusions can be drawn from the literature. In terms of subtypes of depression, despite the common clinical belief that patients with melancholic major depression are especially responsive to ECT, the available evidence suggests that patients without melancholic features are just as likely to have good outcome. In contrast, patients suffering from delusional depression do appear to be particularly responsive to ECT, which is in keeping with clinical experience. Independent of the clinical subtype of melancholic or delusional depression, severity of depression itself does not appear to be predictive of ECT response. While patient age may be statistically associated with response (with older patients doing better), this relation is probably too weak to be of clinical significance. As it turns out, the strongest clinical predictor of ECT response is episode duration, with a briefer length of the index episode of depression being associated with a better ECT outcome (Nobler and Sackeim 1996).

A more recent line of inquiry has focused on the role of proven medication resistance in terms of predicting

ECT response (Prudic et al. 1990, 1996). This approach is in keeping with the shift in referral trends, with ECT being considered after several unsuccessful medication trials. In fact, proven resistance to adequate pharmacotherapy predicts poorer ECT outcome. This finding may, in turn, help explain why both a longer duration of depressive episode and nondelusional (versus delusional) subtype are associated with lower ECT response rates. In the first situation, the longer an episode lasts, the more likely a patient will have received (and failed) at least one adequate medication trial. In the latter, the opposite may be true, as patients with delusional depression are unlikely to have received adequate pharmacotherapy (in this case, combination treatment of antidepressant plus neuroleptic medication). In fact, recent research indicates that very few patients with delusional depression referred for ECT ever receive or tolerate an adequate medication trial (Prudic et al. 1996).

In addition to the search for clinical predictors of ECT response, there has been much effort to identify biological predictors, including measures of endocrine, biochemical, and neurophysiological function (Nobler and Sackeim 1996). Historically, the greatest attention has been given to the dexamethasone suppression test (DST). Unfortunately, the DST has not emerged as a reliable predictor of ECT response. Similarly, the thyrotropin-releasing hormone stimulation test has been of doubtful value. Two promising areas for research include neurophysins and biopterins, but these preliminary findings require independent replication. Recently, much attention has been given to using the electroencephalographic (EEG) manifestations of individual ECT treatments as potential response predictors. For example, the degree of postictal bioelectric EEG suppression immediately following an ECT treatment appears to be related to positive ECT outcome (Nobler et al. 1993; Krystal et al. 1995). While this observation is statistically significant, the magnitude of the association may not be robust enough for direct clinical application. Another promising area for research focuses on correlating changes in cerebral blood flow (CBF) and cerebral metabolic rate (CMR) with clinical outcome (Nobler and Sackeim 1998). Again, preliminary observations await replication.

3 Treatment Technique

3.1 Questioning Basic Premises

From a historical perspective, it must be stressed that ECT was not the first form of convulsive therapy (Fink 1979). In fact, therapeutic seizures were first elicited chemically. For a variety of reasons, the production of seizures by means of passing an electric current soon became the predominant mode of treatment. However, the link between all the forms of convulsive therapy was the elicitation of a generalized seizure. Thus, over the ensuing decades, the elicitation of the seizure at ECT was viewed as both necessary and sufficient for therapeutic action. In other words, the belief was that as long as a seizure of at least adequate duration was produced, the treatment was considered as optimally effective. This was initially supported by studies which indicated that ECT was less effective if subconvulsive electrical stimuli were administered (i.e. stimuli which did not elicit a generalized seizure). Later, a series of classic research studies in the 1960s reinforced this notion, demonstrating that seizures modified by lidocaine (which has anticonvulsant properties and interfered with seizure expression) had diminished therapeutic efficacy, while seizures elicited with very high electrical stimuli led to greater cognitive side effects (Ottosson 1960). The role of electrical stimulus intensity was otherwise overlooked in this epoch, and the belief arose that, while cognitive side effects were due to the passage of current, this was an unavoidable step in the elicitation of the all-important seizure.

The net result of the early research was that practitioners became focused only on ensuring that seizures were elicited at ECT, and not on any particular aspects of the electrical stimulus required to cause the seizure, nor on aspects of the seizures themselves, apart from their duration. Indeed, an influential paper from the 1970s emphasized that longer seizure duration was associated with superior ECT outcome (Maletsky 1978). However, these concepts came under question due to two lines of converging research in the mid-1980s: (1) The discovery that generalized seizures of adequate duration could be elicited by a form of right unilateral (RUL) ECT that lacked therapeutic efficacy (Sackeim et al. 1987) and (2) the finding that speed of response to bilateral (BL) ECT could be influenced by the intensity of electrical stimulation (Robin and de Tissera 1982). These data led to a critical reevaluation of fundamental notions about the primacy of the seizure and its duration and the role of the electrical stimulus.

3.2 Stimulus Dosing and Electrode Placement

Apart from the waveform (sine wave versus brief pulse), overall intensity, and specific parameters (frequency of pulses, pulse width, current, and train duration) of the electrical stimulus, the other main technical feature of ECT is the positioning of the two

electrodes. Over the decades, a wide variety of electrode placements have been employed, with the major distinction being whether both sides (BL) or just one side of the head (unilateral) are stimulated. In practice, BL ECT has become standardized as bifrontotemporal placement of the electrodes, while widely spaced positioning in unilateral ECT (frontotemporal and vertex) on the nondominant hemisphere is the most common form of RUL ECT.

From very early on in the history of ECT, it was recognized that RUL electrode placement resulted in fewer adverse cognitive effects than BL placement, often dramatically so. At the same time, however, there was considerable disagreement as to whether RUL ECT matched BL ECT in antidepressant efficacy. This led to the controversy in ECT, which exists to this day, of whether BL or RUL ECT should be the treatment of choice. Several significant findings in the past decade are now helping to end this controversy. As mentioned, it was demonstrated that a form of RUL ECT could elicit adequate seizures at ECT that lacked therapeutic efficacy in comparison with BL ECT. Specifically, this was RUL ECT delivered at a stimulus intensity that was just above the minimum intensity that was required to elicit a generalized seizure (also known as the seizure threshold). This seizure threshold was empirically determined or "titrated" at the patient's first ECT treatment by first stimulating at very low levels of charge and then restimulating with successively greater amounts of charge until a seizure was produced. A subsequent study found that if the stimulus intensity relative to the initial seizure threshold were increased to two and a half times this threshold, then the efficacy rate of RUL began to approach that of BL ECT (Sackeim et al. 1993). (The duration of the seizures in each of these treatment conditions were not significantly different, indicating that seizure duration lacked any prognostic significance.) New research now indicates that markedly suprathreshold RUL ECT is as effective as BL ECT while retaining advantages with respect to cognitive side effects (McCall et al. 2000; Sackeim et al. 2000). This, and other research (Abrams et al. 1991), essentially demonstrates that there is a dose-response function for RUL ECT by which this form of ECT may be optimized.

In sum, increasing the stimulus intensity of RUL ECT enhances the efficacy of this modality, while increasing the intensity of either BL or RUL ECT appears to quicken the speed of response to both of these modalities. It must be emphasized that what may be critical is not the absolute level of the electrical stimulus, but rather the degree to which it exceeds seizure threshold. Seizure threshold ranges as much as 50-fold among patients, and the only current method of determining this variable with a modicum of precision is by empirical titration (Lisanby et al. 1996). Thus, especially with RUL ECT, inadequate response can occur in patients with high seizure thresholds even when the absolute stimulus intensity is high.

3.3 Frequency and Number of Treatments

Although remission of major depression is occasionally achieved after only two or three treatments, the typical ECT course consists of six to 12 treatments. Of course, the exact number of ECT treatments required for a given patient cannot be predetermined. Rather, ECT is administered until there is remission of target symptoms. If a patient is receiving RUL ECT and has not shown significant improvement after at least eight treatments, switching to BL ECT may be advisable. If a patient has received ten to 15 BL treatments and not responded, the appropriateness of further ECT needs to be seriously reconsidered.

Treatments are typically given two or three times per week, on nonconsecutive days. While ECT three times per week leads to a quicker response, the same level of symptomatic remission can be reached with twice weekly ECT (Lerer et al. 1995). Moreover, less closely spaced treatments generally result in fewer adverse cognitive effects.

4 Adverse Effects

4.1 Medical Considerations

For the vast majority of patients, ECT is an extremely safe procedure. Patients arrive for treatment having had nothing to eat or drink for at least 6 h. An intravenous catheter is placed, and monitors for electrocardiogram, blood pressure, pulse, oxygen saturation, and EEG are routinely set up. Pretreatment with an anticholinergic agent such as atropine is given to protect against bradycardic events. This is followed by general anesthesia with a brief-acting agent such as methohexital.

While most practitioners prefer methohexital, alternatives include thiopental, propofol, etomidate, and ketamine. Once the patient is unconscious, succinylcholine is given to achieve muscle paralysis. The patient is oxygenated by mask with 100% oxygen during this time until spontaneous respiration resumes. The electrical stimulus is applied, and a seizure is elicited, usually lasting for less than 1 min. This entire process lasts for about 5–7 min. Once patients have regained consciousness, are breathing without assistance, and

vital signs have returned to baseline, they are moved to a recovery area. Patients are generally observed in the recovery area for about 30 min or until they are reoriented and can ambulate without assistance.

In the preanesthetic era, there was a high complication rate with ECT, largely due to the effects of unmodified convulsions, with consequent bone fractures and hypoxia. In contemporary practice, the major area of concern is the potential cardiovascular effects of ECT (Zielinski et al. 1993). Generalized seizures are associated with dramatic increase in heart rate and systolic and diastolic blood pressure. Individuals without ischemic heart disease can tolerate such sustained elevation for the brief ictal period. However, the same hemodynamic state may precipitate coronary artery ischemia in vulnerable patients. As mentioned earlier, a careful pre-ECT medical evaluation should identify such high-risk patients, and, in addition to optimizing medication regimens, intravenous agents (such as beta blockers) can be given during ECT to modify the cardiovascular response. Tachyarrhythmias may also occur in the immediate postictal period. Most are transient and benign, but more malignant rhythms can be rapidly treated with lidocaine. Analogous to ischemic cardiac disease, patients with preexisting arrhythmias are much more likely to develop arrhythmias as a consequence of ECT. Independent of cardiovascular risk, the pre-ECT evaluation should be tailored to the medical status of the patient. For example, while spine films are no longer routine, they may be of value in patients with severe osteoporosis or a history of fracture. In each of these instances, a thorough medical evaluation and modifications of anesthetic technique can minimize adverse consequences of ECT.

4.2 Cognitive Effects

Of more practical concern are the cognitive effects of ECT, which constitute the major factor limiting its use (Sackeim 1992). The most common and stereotypical of these are transient postictal disorientation and retrograde and anterograde amnesia. Less common but more worrisome are severe organic mental syndromes. In general, there is rapid recovery of cognitive function just after the ECT course, with return to baseline by several weeks following treatment. There may be, however, permanent spottiness in memory for events prior to, during, and immediately following the ECT course. In very rare instances, a patient may complain of severe, persistent, and pervasive memory loss. Such patients have been difficult to study in a systematic fashion.

Technical aspects of ECT treatment greatly impact on adverse cognitive effects. BL electrode placement is much more likely than RUL placement to result in prolonged postictal disorientation, amnesia, and organic mental syndromes. The type of electrical waveform delivered by the ECT device also affects cognitive outcome. Less efficient stimulus waveforms, such as sine wave, are much more toxic than brief or ultrabrief pulse stimuli. Finally, increasing the electrical stimulus dose relative to seizure threshold appears to worsen short-term cognitive side effects, and, with BL ECT, may increase the risk of severe organicity (Sackeim 1992).

5 Long-Term Outcomes

5.1 Relapse

While ECT is perhaps the most effective somatic antidepressant treatment, it has long been recognized that relapse rates following a successful course of ECT are disappointingly high. Despite the obvious public health burden of this phenomenon, there has been very little research into possible predictors of relapse after ECT. In terms of clinical variables, one study found that medication resistance prior to successful ECT was associated with higher relapse rates following treatment among patients receiving continuation pharmacotherapy (Sackeim et al. 1990). We also lack biological predictors of relapse. The DST has received the most study, but the findings are not at all consistent and lack clinical usefulness (Nobler and Sackeim 1996).

5.2 Continuation Pharmacotherapy

High rates of relapse after ECT with placebo treatment were observed during a series of trials in the 1960s which demonstrated the superiority of continuation pharmacotherapy with tricyclic antidepressants (TCA) or monamine oxidase inhibitors (MAOI). Due to the influence of these studies, the standard recommendation became to place patients on these medications following ECT. However, many of the patients in these trials had received ECT as a first-line treatment, and it is unknown how many would have been resistant to antidepressants. Given that clinical practice and referral patterns have moved ECT back in the line of treatment options (with many patients having failed antidepressant trials prior to ECT) and that proven medication resistance predicts post-ECT relapse, the wisdom of reflexively placing patients on these medications after ECT has been questioned (Sackeim et al. 1990). New attempts at improving continuation phar-

macotherapy following ECT are in progress. Preliminary findings from a recent multicenter trial indicate that the combination of TCA with lithium is superior to TCA alone and to placebo in reducing 6-month relapse rates.

5.3 Continuation Electroconvulsive Therapy

In the field of psychopharmacology, once a patient responds to an acute trial of medication, the compound is usually continued for some minimum period of time (e.g. 6–9 months), depending on the particular illness and the severity of symptoms. Thus, from a logical point of view, it may seem odd that once a patient responds to ECT, we typically discontinue it. Of course, indefinitely continuing ECT two or three times a week would pose the risk of worsening cognition, not to mention the medical risk of multiple repeated general anesthesia. However, ECT practitioners have known for years that many patients are able to well tolerate a transition from an acute ECT course to a course of continuation ECT (C-ECT). In fact, C-ECT may be the only viable treatment option for patients who cannot tolerate (or repeatedly fail) continuation pharmacotherapy and are at high risk for relapse. There is a large body of literature of an uncontrolled nature demonstrating the utility of C-ECT (Sackeim 1994a). For various practical reasons, it has been difficult to conduct prospective, randomized, and controlled studies in this area, although a large multicenter trial is underway in the United States. C-ECT usually begins as once per week treatment, with a gradual lengthening of the space between treatments to roughly once per month, as needed. In addition, continuation treatments are well suited for outpatients and can be minimally intrusive in patients' lives.

6 Mechanisms of Action

Ironically, while ECT may be the most effective somatic treatment in psychiatry, its definitive mechanism of action remains unknown. Instead, we possess a vast body of animal and clinical data as to its effects on neurophysiology, biochemistry, and endocrine physiology, which reflects the myriad of changes that take place in the brain and in the body following a generalized seizure. This has led, in turn, to a variety of theories regarding mechanisms of action. In understanding these various theories, it is important to realize that they may not be mutually exclusive, but rather may explain interrelated phenomena at different hierarchical levels. As this is an extensive topic, the following is intended as a synopsis, and the reader is also referred to recent comprehensive reviews (Fochtmann 1994; Sackeim 1994b).

6.1 Methodological Considerations

When relying on preclinical data, certain caveats must be kept in mind. First, the overwhelming majority of animal data derives from studies in rodents. Despite the problem of interspecies differences, the ability to generalize findings to humans is also hampered by the lack of good animal models for the major psychiatric syndromes. Second, the time course of biological changes relative to electrical stimulation needs to be considered. For instance, neither the effects of a single electroconvulsive shock (ECS) nor the application of multiple ECS in a single day may be relevant to the clinical situation. Third, the majority of preclinical studies compare ECS to standard antidepressant medications. However, while both ECT and antidepressant medications have some degree of overlap in terms of therapeutic applications, ECT clearly has a broader spectrum of action. Thus studies that limit the comparison of ECS to TCA or MAOI may thereby be skewing the interpretations of the findings towards a common mechanism of action.

While clinical studies are immediately relevant, they are constrained at the practical level by the fact that assays of the central nervous system are often indirect, as when peripheral measures of a neurotransmitter metabolite are used as an estimate of central activity. On the theoretical level, the fact that ECT is so highly effective has constricted our ability to tease apart critical factors related to its therapeutic mechanisms. In other words, to demonstrate putative mechanisms of action, the biological effects of an active treatment intervention should be compared among responders and nonresponders to the treatment. Alternatively, the effects of the active treatment could be compared with either an inactive treatment or an attenuated form of active treatment. Since sham ECT studies are generally no longer conducted, the ability to compare the effects of "active" ECT (BL and high-dose RUL ECT) with "weak" ECT (low-dose RUL ECT) has offered some opportunities to test theories regarding mechanisms of action.

6.2 Biochemical Mechanisms

Several decades of research, mostly preclinical, has examined the effects of ECS and ECT on the classical

neurotransmitter systems (Kety 1974; Fochtmann 1994). Much of this work has been derived from animal studies comparing ECS with TCA or MAOI. One of the most consistent findings is that, similar to antidepressant medications, ECS leads to a downregulation of β-adrenergic receptors. This fact has led many to the conclusion that the antidepressant effect of both standard medications and ECT is mediated through noradrenergic pathways. However, human studies have yielded inconsistent results as to changes in noradrenergic function. Furthermore, if such a theory is postulated, it is difficult to explain why ECT would be effective in patients resistant to TCA or MAOI. In contrast to the downregulation of the 5-hydroxytryptamine receptor (5-HT_2) typically seen with antidepressants, ECS leads to an increased density of 5-HT_2 receptors. While human studies have demonstrated an increase of 5-HT metabolites in the cerebrospinal fluid following a course of ECT, other probes of 5-HT function have yielded inconsistent results. Animal studies have consistently demonstrated profound increases in dopaminergic function following acute and chronic ECS. Clinical studies have been less consistent, but also point in the direction of enhanced dopaminergic functioning with ECT. Indeed, this may underlie the antiparkinsonian effect of ECT. Other preclinical work has focused on transmitters such as γ-aminobutyric acid (GABA), adenosine, and the endogenous opioids. These systems may be particularly important in understanding the biochemical basis of the anticonvulsant effects of ECT.

6.3 Endocrine Mechanisms

A wide variety of endocrine changes occur following ECT, and it has been challenging to separate out potential therapeutic mechanisms from what may simply be epiphenomena of seizure activity. For instance, prolactin is released acutely following ECT, and elevations in plasma levels are consistently easy to detect. However, while BL and higher-dosage ECT lead to a greater relative prolactin surge, a recent study found that, when treatment technique is controlled for, prolactin values are not independently associated with clinical outcome (Lisanby et al. 1998). Other endocrine theories have focused on alterations in the hypothalamic–pituitary–adrenal or the thyroid axis. As mentioned, the DST has yielded inconsistent results with respect to clinical outcome. While investigations of the hypothalamic–thyroid axis have also proved inconsistent, the report that thyroid hormone may augment the efficacy of ECT indicates the need for further study in this area (Stern et al. 1991).

6.4 Neurophysiological Mechanisms

At the neurophysiological level, generalized seizures – the repetitive hypersynchronous depolarization of neurons – are accompanied by dramatic changes in blood–brain barrier permeability and marked increases in cerebral blood flow and metabolic rate (Nobler and Sackeim 1998). The postictal period is characterized by decreases in CBF and CMR as well as stereotypical changes in the EEG. Just following a course of ECT, CBF and CMR may remain decreased, while slow-wave activity becomes prominent in the EEG. The degree, anatomic location, and, to some extent, persistence of these effects appear to be related to positive ECT response (Nobler et al. 1994; Sackeim et al. 1996).

Brain imaging studies of major depression have generally indicated that there are baseline reductions in CBF and CMR, both regionally and globally. Thus, at first glance, it seems paradoxical that further reductions in CBF would be involved in therapeutic mechanisms. Studies are underway to help clarify this issue.

6.5 Anticonvulsant Mechanisms

One particular neurophysiological theory regarding the mechanism of action of ECT centers on its anticonvulsant properties (Sackeim 2000). In animals, ECS raises the threshold for subsequent seizures. The seizure threshold also progressively increases in patients during a course of ECT, while seizure duration decreases. Both an increase in seizure threshold and a decrease in seizure duration are the hallmarks of an anticonvulsant effect. Additional theoretical support for an anticonvulsant mechanism is lent by the fact that both anticonvulsant medications and ECT are powerful treatments for mania.

The anticonvulsant theory posits that the therapeutic mechanisms for ECT are tied to the endogenous processes which terminate seizures, and not to the seizures themselves. These processes include the release of endogenous inhibitory or anticonvulsant compounds such as GABA and adenosine. Further, the magnitude of this active inhibition is most likely indexed by acute reductions in CBF and CMR and by

the degree of immediate postictal bioelectric suppression on the EEG.

7 References

Abrams R (1992) Electroconvulsive therapy, 2nd edn. Oxford University Press, New York

Abrams R, Swartz CM, Vedak C (1991) Antidepressant effects of high-dose right unilateral electroconvulsive therapy. Arch Gen Psychiatry 48: 746–748

*American Psychiatric Association (1990) The Practice of ECT: recommendations for treatment, training and privileging. American Psychiatric Press, Washington, DC

Fink M (1979) Convulsive therapy: theory and practice. Raven, New York

Fochtmann LJ (1994) Animal studies of electroconvulsive therapy: foundations for future research. Psychopharmacol Bull 30: 321–444

Kellner CH, Beale MD, Pritchett JT et al (1994) Electroconvulsive therapy and Parkinson's disease: the case for further study. Psychopharmacol Bull 30: 495–500

Kety SS (1974) Biochemical and neurochemical effects of electroconvulsive shock. In: Fink M, Kety S, McGaugh J, Williams TA (eds) Psychobiology of convulsive therapy. Winston, Washington, DC, pp 285–294

Krueger RB, Sackeim HA (1995) Electroconvulsive therapy and schizophrenia. In: Hirsch SR, Weinberger D (eds) Schizophrenia. Blackwell, New York, pp 503–545

Krystal AD, Weiner RD, Coffey CE (1995) The ictal EEG as a marker of adequate stimulus intensity with unilateral ECT. J Neuropsychiatr Clin Neurosci 7: 295–303

*McCall WV, Reboussin DM, Weiner RD et al (2000) Titrated, moderately suprathreshold versus fixed, high dose RUL ECT: Acute antidepressant and cognitive effects. Arch Gen Psychiatry 57: 438–444

Lerer B, Shapira B, Calev A et al (1995) Antidepressant and cognitive effects of twice- versus three-times-weekly ECT. Am J Psychiatry 152: 564–570

Lisanby SH, Devanand DP, Nobler MS et al (1996) Exceptionally high seizure threshold: ECT device limitations. Convulsive Ther 12: 156–164

Lisanby SH, Devanand DP, Prudic J et al (1998) Prolactin response to ECT: effects of electrode placement and stimulus dosage. Biol Psychiatry 43: 146–155

Maletzky BM (1978) Seizure duration and clinical effect in electroconvulsive therapy. Compr Psychiatry 19: 541–550

Mukherjee S, Sackeim HA, Schnur DB (1994) Electroconvulsive therapy of acute manic episodes: a review of 50 years' experience. Am J Psychiatry 151: 169–176

Nobler MS, Sackeim HA (1996) Electroconvulsive therapy: clinical and biological aspects. In: Goodnick P (ed) Predictors of treatment response in mood disorders. American Psychiatric Press, Washington, DC, pp 177–198

Nobler MS, Sackeim HA (1998) Mechanisms of action of electroconvulsive therapy: functional brain imaging studies. Psychiatr Ann 28: 23–29

Nobler MS, Sackeim HA, Solomou M et al (1993) EEG manifestations during ECT: effects of electrode placement and stimulus intensity. Biol Psychiatry 34: 321–330

*Nobler MS, Sackeim HA, Prohovnik I et al (1994) Regional cerebral blood flow in mood disorders. III. Treatment and clinical response. Arch Gen Psychiatry 51: 884–897

Olfson M, Marcus S, Sackeim HA et al (1998) Use of ECT for the inpatient treatment of recurrent major depression. Am J Psychiatry 155: 22–29

Ottosson J (1960) Experimental studies of the mode of action of electroconvulsive therapy. Acta Psychiatr Scand [Suppl] 145: 1–141

Prudic J, Sackeim HA, Devanand DP (1990) Medication resistance and clinical response to electroconvulsive therapy. Psychiatr Res 31: 287–296

**Prudic J, Haskett RF, Mulsant B et al (1996) Resistance to antidepressant medications and short-term clinical response to ECT. Am J Psychiatry 153: 985–92

Robin A, de Tissera S (1982) A double-blind controlled comparison of the therapeutic effects of low and high energy electroconvulsive therapies. Br J Psychiatry 141: 357–366

Sackeim HA (1992) The cognitive effects of electroconvulsive therapy. In: Moos WH, Gamzu ER, Thal LJ (eds) Cognitive disorders: pathophysiology and treatment. Dekker, New York, pp 183–228

Sackeim HA (2000) The anticonvulsant hypothesis of the mechanisms of action of ECT: current status. J ECT 15: 5–26

Sackeim HA (1994a) Central issues regarding the mechanisms of action of electroconvulsive therapy: directions for future research. Psychopharmacol Bull 30: 281–308

Sackeim HA (1994b) Continuation therapy following ECT: directions for future research. Psychopharmacol Bull 30: 501–521

Sackeim HA, Decina P, Kanzler M et al (1987) Effects of electrode placement on the efficacy of titrated, low-dose ECT. Am J Psychiatry 144: 1449–1455

Sackeim HA, Prudic J, Devanand DP et al (1990). The impact of medication resistance and continuation pharmacotherapy on relapse following response to electroconvulsive therapy in major depression. J Clin Psychopharmacol 10: 96–104

**Sackeim HA, Prudic J, Devanand DP et al (1993) Effects of stimulus intensity and electrode placement on the efficacy and cognitive effects of electroconvulsive therapy. N Engl J Med 328: 839–846

**Sackeim HA, Prudic J, Devanand DP et al (2000) A prospective, randomized, double-blind comparison of bilateral and right unilateral ECT at different stimulus intensities. Arch Gen Psychiatry 57: 425–434

Sackeim HA, Devanand DP, Nobler MS (1995) Electroconvulsive therapy. In: Bloom F, Kupfer D (eds) Psychopharmacology: the fourth generation of progress. Raven, New York, pp 1123–1142

Sackeim HA, Luber B, Katzman GP et al (1996) The effects of electroconvulsive therapy on quantitative electroencephalograms. Relationship to clinical outcome. Arch Gen Psychiatry 53: 814–824

Stern RA, Nevels CT, Shelhorse ME et al (1991) Antidepressant and memory effects of combined thyroid hormone treatment and electroconvulsive therapy: preliminary findings. Biol Psychiatry 30: 623–627

Zielinski RJ, Roose SP et al (1993) Cardiovascular complications of ECT in depressed patients with cardiac disease. Am J Psychiatry 150: 904–909

CHAPTER

32

S. Ecker, F. Henn

Psychosurgery

1 Definition

The term "psychosurgery" describes neurosurgical operations performed in order to improve the psychic condition of patients. In other words, brain surgery for psychiatric indications is described as "psychosurgical intervention." The American Psychiatric Association (APA) defines psychosurgery as a "neurosurgical intervention to sever fibers connecting one part of the brain with another or to remove or destroy brain tissue with the intent to modifying or altering severe disturbances of behavior, thought content or mood" (Stone 1989).

The term is not meant to denote "direct surgical intervention into the human psyche," which might suggest a direct correlation between psyche and the connecting fibers of the brain, which is not intended. Therefore, the term psychosurgery might be misleading. In spite of terminological questions, no alternative terms such as "psychiatric surgery," "behavioral surgery," or "psychiatrically indicated neurosurgery" have prevailed thus far. In our opinion, the adoption of the term "psychiatrically indicated stereotactic neurosurgical intervention" would be more adequate to the subject in question.

2 History

The first psychosurgical intervention was performed in 1935 by Egaz Moniz, Professor of Neurology, with the help of the neurosurgeon Almeida Lima. They initially performed a prefrontal leukotomy by injecting pure alcohol into the frontal lobe. After seven operations they changed their technique. They severed the white substance of the frontal lobes with a small needle ("leukotome") (Moniz 1936, 1937). At that time, there were hardly any therapeutic possibilities available to help mentally ill patients. Thus, when Moniz announced that 14 out of 20 severely ill patients (seven schizophrenic patients, 13 patients with affective disorders or phobias) showed significant improvement after the operation (Moniz 1936), considerable enthusiasm was elicited from the medical community. In 1937, the neuropsychiatrist Walter Freemann and the neurosurgeon James Watts applied a modified technique in which they severed all connections to the prefrontal cortex by bilaterally entering the brain through the orbit (Freeman and Watts 1937, 1950). This method was soon commonly adopted, and the following two decades saw a sharp increase in neurosurgical operations on mentally ill patients. Different groups developed new methods with the intent to reduce side effects by limiting the seize of the lesion. In 1949, Professor Moniz received the Nobel Prize for his outstanding work in the field of psychosurgery. It is estimated that, in 1950 alone, approximately 5000 psychiatrically indicated neurosurgical operations were conducted. In 1952 and 1953, approximately 300 reports on neurosurgical interventions were published per year.

In spite of the new techniques, the indication for psychosurgery, and thus the number of operations performed, was significantly reduced by the late 1950s (Cosyns et al. 1994). This was due to an often uncritical application of the new methods with corresponding unsatisfactory results and the development of effective pharmacological treatments (chlorpromazine in 1952, benzodiazepine in 1957, haloperidol in 1958, and tricyclic antidepressants in 1958). From the mid-1960s to the end of the 1970s, stereotactic operations for sexual deviations were a means of intervention in Germany (Adler and Saupe 1979). Up to 80 such operations were performed by three different teams, mainly on criminals with often doubtful indications and eventually dubious results (Schorch and Schmidt 1979). This led to vehement public discussions (Fülgraff and Barbey 1978) and to an increasingly critical attitude towards psychosurgery. The number of psychosurgical operations decreased sharply in subsequent years. Another reason was the increasing use of psychotherapy for the treatment of mental disorders such as depression. Although psychoanalysis was only rarely effective for severely ill patients, the introduction of behavioral therapy in the early 1960s justly gave rise to the hope that many patients could be helped who suffered from depression, obsessive-compulsive disorder (OCD), or phobias.

With these developments, psychosurgery as a treatment for mentally ill patients has been greatly reduced. In some countries such as Japan, the United States, and Germany, psychosurgery is performed only in rare cases. However, in certain European countries (e.g. Sweden, the United Kingdom, Spain, Belgium, and the Netherlands), psychosurgery is still performed on a regular basis, in specialized centers, conducted by experienced teams. This is in part due to the new stereotactic methods, which allow a very precise positioning of small lesions, and by the recognition by psychiatrists that, in spite of all the new pharmacological and psychotherapeutic treatments, a group of patients remain who cannot be helped adequately by these conservative means.

3 Ethical Aspects

Psychosurgery is looked upon with some reservation in all countries by both the general public and the medical community. This is especially true in Germany because of its historical background. This concern is understandable considering that psychosurgery has two serious differences compared to other treatments:

1. Operations on anatomically intact brain involve a certain incalculable risk for the patient with regard to possible behavioral consequences in view of the still limited knowledge about the function of individual brain structures. Most significant is the fact that the targeted brain areas are damaged irreversibly.
2. Since psychosurgery is intended to influence the mental and emotional functioning of the patient, it necessarily invades the sphere of what a patient regards as the essence of his or her selfhood. This affects the moral and/or religious beliefs held by some people. It is clear that some patients will have severe difficulties overcoming their ethical reservations to such a procedure.

The ethical problems psychosurgery raises for the physician and the patient include moral philosophy as well as questions regarding acceptable practical means to alleviate human suffering.

In 1988, A.J. Bouckoms from the Harvard Medical School in Boston formulated the following recommendations (Bouckoms 1988):

1. No consideration of ethics in psychosurgery is complete without consideration of both the scientific data and moral conflicts.
2. The liberal advocation of autonomy without responsibility is an amoral, not a liberating, point of view.
3. Politics should be denounced as the most serious ethical problem in medical decision-making. Political intrusion into scientific matters and the doctor–patient relationship has created ethical problems with psychosurgery and continues to do so today.

The probability of the success of scientific research should be weighted against philosophical norms and rules regarding ethic and moral conflicts relating to altruism, autonomy, and human suffering.

Generally, the following principles have to be considered in the context of psychosurgical interventions:

1. All conservative, i.e. reversible, treatments (pharmacological and psychotherapeutic treatment, electroconvulsive therapy, if applicable) must have been applied adequately without yielding satisfactory results. Psychosurgery should be the last resort.
2. Psychosurgical interventions are thus never an emergency treatment; they require intensive preparation and planning with sufficient time for the patient to realize the possible risks and consequences of the operation and to make an appropriate decision. It is desirable that the relatives are involved and support the patient's decision.
3. Fundamentally, psychosurgical interventions must have a therapeutic goal and must not be considered as an experimental trial. This means that, based on previous experience, there must be a realistic chance of alleviating the suffering of a patient with regard to his or her mental illness by psychosurgery.
4. The indication and performance of psychosurgery should only be conducted by experienced psychiatrists and neurosurgeons. The neurosurgeon should be competent and versed in the planning and execution of the appropriate stereotactic intervention. The psychiatrist must be able and willing to provide adequate and intensive pre- and postoperative care.

For this reason some countries such as the Netherlands and Belgium have formed advisory boards comprising independent psychiatrists, neurosurgeons, neurologists, and sometimes ethics specialists and lawyers. The United States, too, have established a board to investigate psychosurgical interventions and to develop new concepts.

4 Indication and Selection of Patients

Considering the controversial nature of the issue, which was outlined above, it is important to point out the psychiatric symptoms for which a neurosurgical intervention may be indicated with some chance of success. While the range of indications was rather diffuse in the euphoric beginnings of psychosurgery, clearly defined criteria have emerged over the past 10–15 years.

Psychosurgical interventions may be indicated in patients with affective disorders (major depressions and bipolar disorders), OCD, and anxiety disorders (generalized anxiety disorders, agoraphobia with panic attacks). In addition, psychosurgery might possibly be indicated in patients showing extremely aggressive behavior combined with mental retardation, although this remains a contentious issue.

Schizophrenia and substance abuse, although

discussed, are not considered an indication by most authors (Balasubramaniam 1997; Meyerson 1996). Cluster A and B personality disorders according to DSM-III-R are generally perceived as a relative contraindication to surgery, but not those of cluster C. Severe axis III symptomatology according to DSM-III-R, e.g. brain tumors or leukodystrophy, are contraindications for psychosurgery.

In contrast to many other areas in the field of psychiatry, there seems to be a wide and interdisciplinary consensus regarding the criteria for the selection of patients. Before psychosurgery should be considered for a patient, the following criteria must be met:

1. The diagnosis of the mental illness should be confirmed by standardized instruments, such as ICD-10 or DSM-III-R.
2. The disorder must be chronic, i.e. the patient must suffer from the illness for a period of at least 3 years (some authors even call for 5 years); during that time, the patient must be severely impaired in his or her private and professional life.
3. The patient must be suffering severely because of his or her mental illness.
4. The mental illness must have led to a significant functional impairment in the prior year, reflected in a global assessment of function score of between 30 and 50 (axis V according to DSM-III-R).
5. The disorder must be refractory to conventional therapies. All pharmacological treatment options must have not resulted in satisfactory improvement in spite of adequate dosage and duration of treatment or must have had to be aborted due to severe side effects. Extensive psychiatric treatment, particularly cognitive psychotherapy and behavioral therapy, must have resulted in no satisfactory change.

Major depression, in particular, is considered refractory when the following therapies have failed:

a) Treatment with various tricyclic antidepressants at dosages of approximately 300–400 mg per day for at least 2 months, as well as equivalent combined treatments. In addition, adequate treatments with monoamine oxidase inhibitors and serotonin reuptake inhibitors should have been tried.
b) At least one therapy with lithium and with lithium in combination with an antidepressant should have been tried.
c) Two cycles of electroconvulsive treatment, with several months between them, should have been tried.

OCD, in particular, are considered refractory when the following therapies have failed:

a) At least three pharmacological treatment cycles with medications that are known to be effective against OCD. Each cycle should last for least 10 weeks, and one of them should have been conducted with clomipramine. Compliance should be checked by testing medication levels in patients' blood samples.
b) Three additional therapy cycles should have been carried out with combined medications, one including clonazepine.
c) Intensive behavioral therapy should have been conducted at the same time as medical treatment, including exposure and response prevention over a period of at least 30 h.
d) Finally, intensive inpatient treatment must have been tried.

6. The prognosis of the mental illness must be considered bad without psychosurgical intervention.
7. The patient should be completely and carefully informed, preferably by a interdisciplinary team of experts, and these discussions should preferably include a relative. The patient should have sufficient time to understand the meaning and the possible consequences of the operation and to form his or her own opinion about it. The patient should not only agree to the operation, but should actively be in favor before it is carried out.
8. Intensive psychiatric pre- and postoperative care, including rehabilitation measures, must be provided.
9. A careful examination and documentation of the pre- and postoperative status is mandatory, using approved psychological tests, e.g., Hamilton depression scale (HAMD) and Yale-Brown obsession-compulsion scale (YBOCS). A similarly precise survey of the postoperative status is desirable in order to critically evaluate the outcome.
10. Exclusion criteria for a psychosurgical intervention are as follows:

    a) Age under 20 or over 65 years.
    b) Patients who are under custody in psychiatric institutions, e.g. prisoners, patients in forensic psychiatry, and involuntarily committed patients.
    c) Incompetent patients who are under legally appointed guardianship.

In the past, there were often controversies over points b) and c), especially with regard to psychosurgical interventions for sexual deviations. On the one hand, the ability of such patients to freely express agreement or disagreement with a psychosurgical intervention possibly entailing nontherapeutic goals such as premature release appears to be restricted. On the other hand, with these patients, a psychosurgical operation may be the only effective therapeutic means. Nowa-

days, it is generally felt that patients in prison and those under custody or guardianship are dependent individuals whose seemingly free decisions are determined by very different factors that cannot be accepted from a medical point of view (Adler and Saupe 1979).

5 Techniques and Target Areas

5.1 General Neurosurgical Technique

In contrast to the beginnings of psychosurgery, almost all neurosurgical interventions in psychiatric disorders are now carried out using stereotactic techniques. This method allows a precise placement of the lesion as well as minimization of the lesion. A stereotactic intervention does not require general anesthesia; most patients are under mild sedation with local anesthesia. Thus the patient can talk and report on the effects of stimulation during the procedure. This way, the targeted location of the lesion can be stimulated intraoperatively in order to evoke symptoms associated with the disorder and, through feedback with the patient, to check and – if necessary – correct the placement of the probe.

The operation involves the fixation of the stereotactic frame with the fiducial plates on the patient's skull followed by stereotactic imaging – formerly ventriculography, now computed tomography (CT) and more frequently magnetic resonance imaging (MRI). The next step is the definition of the target point for the lesion and calculation of the stereotactic coordinates using the images. The settings on the stereotactic frame and the aiming bow are made in accordance with these coordinates, and the frame is fixed to the ring on the patient's head in a predefined position, with the aiming bow serving as instrument guide. Bilateral burr holes are made in front of the coronal suture, 1 inch lateral to the midline, through skin incisions of little more than 1 inch. After incision of the dura mater and coagulation of the cortex, the probe is advanced with the electrode holder on the aiming bow. The accuracy of this procedure is around 1 mm. The lesion itself is produced by heating the electrode and thermocoagulating the tissue within the desired area.

5.2 General Anatomical Considerations

The morphological substrate of psychiatric illnesses and therefore the exact action of lesional therapy are not totally understood. There are indications from neuroimaging investigations – positron emission tomography (PET), single photon emission computed tomography (SPECT), or functional MRI – that for some psychiatric conditions the prefrontal cortex, the cingulate gyrus, and the basal ganglia might play a major role. The target areas for lesioning are closely correlated, and for effective therapy connections with the limbic system and the circuit of Papez are postulated. It is remarkable that lesions at different sites for the same disorder and lesions at identical sites for different disorders seem to be effective. This supports the theory that a dysregulation of control circuits may be involved in the development of psychiatric illness (Meyerson 1996; Cosgrove and Rauch 1995; Cosgrove and Ballantine 1996).

Different target regions will be discussed below; the reader should bear in mind that the lesions are always placed symmetrically and bilaterally.

5.3 Target Regions

5.3.1 Cingulotomy

Anterior cingulotomy, first suggested by Fulton (1951), is the most common psychosurgical intervention, at least in the United States. Initially used as a therapy for chronic pain (Foltz and White 1962), it was later reported to be a safe and successful treatment for a large number of psychiatric patients (Ballantine et al. 1987). This operation is the one that is most noted in current literature (Baer et al. 1995; Jenike et al. 1991; Cosgrove and Ballantine 1996). It involves bilateral lesions in the cingulate gyrus disrupting the limbic pathways of the circuit of Papez contained therein. Target coordinates are calculated for a point within the cingulum 2 cm posterior to the tip of the frontal horns and 7 mm lateral to the midline. The lesions are made by thermocoagulation (Ballantine et al. 1995) and are equivalent to a cylinder of 2 cm in length with a diameter of 1 cm. The adverse effects that can sometimes occur include reversible bladder dysfunction or an elevated temperature. Lasting cognitive or behavioral deficits have not been observed, but some authors report lasting improvements in the performance of memory and intelligence tests, which is attributed to improved concentration (Cosgrove and Ballantine 1996; Ballantine et al. 1995; Tippin and Henn 1982). Thus cingulotomy seems to be the safest of the neurosurgical procedures in psychiatric illness. Psychiatric indications are severe affective disorders, anxiety disorders, and OCD.

5.3.2 Anterior Capsulotomy

First reported by Talairach et al. (1949), anterior capsulotomy was used for a variety of psychiatric conditions on a larger scale mainly by Lars Leksell (Bingley et al. 1977). This procedure seems to be the most frequently used in Europe (Cosgrove and Ballantine 1996). The target area is the anterior limb of the internal capsule, where frontothalamic pathways are located, connecting the limbic system to the frontal lobes. The lesions are produced by thermocoagulation or radiosurgical methods (gamma knife) (Mindus and Meyerson 1995; Leksell and Backlund 1979). Reversible disorientation between postoperative day 2 and 5 has been reported as a complication in 90% of the patients (Mindus and Nyman 1991). Long-lasting adverse effects consisted in weight gain, fatigue, and memory deficits (Mindus and Nyman 1991), whereas no evidence for cognitive dysfunction was found (Bingley et al. 1977; Burzaco 1981; Herner 1961). A tendency for recurrence with time has not been observed (Meyerson 1977), but in 17 of 85 patients a second procedure was necessary due to suboptimal lesion size or positioning, this second procedure being effective in 52% of the patients (Burzaco 1981). Major indications are chronic anxiety disorders, OCD, and major depression.

5.3.3 Subcaudate Tractotomy

Subcaudate tractotomy was introduced by Knight in 1964 (Knight 1964), mainly to reduce the size of the surgical lesions and therefore to reduce the side effects seen with standard prefrontal lobotomies. The lesion can be produced by thermocoagulation or stereotactic placement of radioactive seeds (Cosyns et al. 1994). The target region is the white matter directly below the head of the caudate nuclei with the fiber tracts of the orbital cortex. Complications reported include epileptic seizures in 2.2% and inadvertent personality changes in 6.7%; in addition, transient disinhibition was frequently observed. Main indications are depression, OCD, and anxiety disorders. Patients with personality disorders, schizophrenia, and drug or alcohol abuse had poor outcomes (Goktepe et al. 1975; Balasubramaniam 1997).

5.3.4 Limbic Leukotomy

Limbic leukotomy is a combination of anterior cingulotomy and subcaudate tractotomy and was introduced in 1973 by Kelley (Kelley et al. 1973) in order to improve results in OCD patients. Adverse effects included bladder dysfunction, fatigue, and disorientation during the immediate postoperative period (Cosgrove and Rauch 1995). A total of 12% of the patients complained of a lasting lethargy, while epileptic seizures were not observed and the patients' intelligence quotient (IQ) showed a slight increase. Indications are concurrent with the single procedures and include OCD, affective disorders, and anxiety disorders.

5.3.5 Intralaminar and Dorsomedial Thalamotomy

Introduced by Hassler and Dieckmann (1973), the intralaminar and dorsomedial thalmotomy procedure consists of a lesion within the dorsomedial nucleus of the thalamus encompassing the adjacent medial lamella. Because of its difficult anatomical location close to the important mamillothalamic tract, the target remains controversial. If used at all, it should be employed in disorders with aggressive behavior in combination with mental retardation (Gybels and Cosyns 1995), because there are hardly any alternatives for such patients. Complications include reversible somnolence, development of hydrocephalus, and mild hemiparesis in one patient (Cosyns et al. 1994). Experience with this procedure is only very limited compared to other methods.

5.3.6 Amygdalotomy

According to Narabayashi and Shima (1973), amygdalotomy is based on a lesion of the corticomedial nucleus of the amygdala and, like dorsomedial thalamotomy, it is employed in patients with aggressive disorders with mental retardation. Again, experience is very limited (Cosyns et al. 1994). It should be emphasized that psychosurgery for aggressive behavior is controversial and generally not recommended.

5.3.7 Ventromedial Hypothalamotomy

On the basis of different and controversially assessed results from animal experiments (Orthner 1982; Schorsch and Schmidt 1979), ventromedial hypothalamotomy was applied as therapy for sexual deviations (Müller et al. 1974). The lesion was meant to influence sexual behavior in such a way as to reduce "hypersexuality" and "perversion" as observed in patients

with lesions of both medial temporal lobes, for example (Klüver and Bucy 1939). Experience with this method was gathered practically only in Germany, and it was only carried out to a small extent and was insufficiently documented (Schorsch and Schmidt 1979).

In general, in any kind of neurosurgical intervention in psychiatric illness, there is the risk of complications regardless of the area of lesioning, as with every neurosurgical operation. With varying frequency, general complications such as intracerebral hemorrhage (0%–2%), infections (0%–1%), and development of postoperative epileptic seizures (1%) are reported.

6 Results and Perspectives

An evaluation and comparison of the results of neurosurgical intervention for the treatment of psychiatric symptoms poses some difficulties. One of these difficulties arises from the fact that different techniques, diagnostic criteria, and standards for postoperative examinations/surveys are being used in different centers.

In addition, most of the published papers are based on retrospective surveys, and there are only a few prospective studies, each with only a small number of patients (Baer et al. 1995; Kartsounis et al. 1991). Because of the limited, severely ill population of patients, prospective randomized studies are not possible, and because the treatment involves surgery, double-blind studies are neither feasible nor ethically justifiable. However, with the advent of radiosurgical methods, this may be considered, and consequently a double-blind study is currently being carried out on capsulotomy for the treatment of OCD patients, possibly providing the desired evidence to improve the basis of decision-making for the application of psychosurgery (Mindus et al. 1994).

A total of 18 OCD patients were observed in a prospective study on cingulotomy using strict criteria for therapeutic success, and 30% were felt to be responders (Baer et al. 1995). Applying less strict criteria, Ballantine et al. (1987) reported an improvement of symptoms in 64% of patients with affective disorders, 56% of patients with OCD, and 79% of patients with anxiety disorders in an investigation of 196 patients who underwent cingulotomy for their disorder.

A success rate of between 64% and 70% has been reported for capsulotomy in OCD patients (Mindus and Meyerson 1995; Bingley et al. 1977; Meyerson 1977, 1996), and in patients with anxiety disorders the percentage of markedly improved patients was 71% (Rylander 1979).

Table 1. Probability of a positive outcome in patients with obsessive-compulsive disorder (OCD) and affective disorder using various psychosurgical techniques

	Anterior cingulotomy (%)	Anterior capsulotomy (%)	Subcaudate tractotomy (%)	Limbic leucotomy (%)
OCD	56	67	50	61
Affective disorder	65	55	68	78

Good results have been reported for *subcaudate leukotomy* in 68% of depressive patients, 62.5% of patients with anxiety disorders, and 50% of patients suffering from OCD (Goktepe et al. 1975).

Mitchell-Heggs et al. (1977) reported on an improvement in 89% of OCD patients, 68% of patients suffering from a chronic anxiety disorder, and 78% of depressives after *limbic leukotomy*.

There seems to be a greater likelihood of therapeutic improvement after surgery if significant vegetative symptoms and/or anxiety are concomitant in a patient with an underlying diagnosis of depression or OCD.

A comparison of results across centers is problematic, since imprecise criteria for diagnosis, nonstandardized presurgical evaluation tools, and different outcome assessment scales have been used. However, in nearly all of the published reports, some modification of the Pippard Postoperative Rating Scale (Pippard 1955) may be ascertained; this scale rates outcome in the following five categories: (1) symptom free, (2) much improved, some symptoms remaining but no additional treatment necessary, (3) slightly improved, (4) unchanged, and (5) worse. The probability of a positive outcome in the different combinations of disorders and stereotactic target areas is given in Table 1, with categories 1 and 2 regarded as successful (Cosgrove and Rauch 1995).

In general, applying restrictive criteria and selective indications, an improvement can be expected in more than 50% of patients. However, the published results do not allow a statistically significant evaluation of the effectiveness of the different techniques. Concurrently, it appears that patients with depressive disorders and OCD do not show an improvement until after up to 6 months, which involves intensive postoperative therapy, whereas patients with anxiety disorders often show an immediate improvement.

7 References

Adler M, Saupe R (1979) Psychochirurgie. Enke, Stuttgart

*Baer L, Rauch SL, Ballantine HT Jr, Martuza R, Cosgrove R, Cassem E, Giriunas I, Manzo PA, Dimino C, Jenike MA (1995) Cingulotomy for intractable obsessive-compulsive disorder. Prospective long-term follow-up of 18 patients. Arch Gen Psychiatry 52: 384–392

Balasubramaniam V (1997) Magnetic resonance image-guided stereotactic cingulotomy for intractable psychiatric disease. Neurosurgery 40: 107–108

Ballantine HT, Bouckoms AJ, Thomas EK (1987) Treatment of psychiatric illness by stereotactic cingulotomy. Biol Psychiatry 22: 807–819

Ballantine HT Jr, Cosgrove GR, Giriunas IE (1995) Surgical treatment of intractable psychiatric illness and chronic pain by stereotactic cingulotomy. In: Schmiedek HH, Sweet WH (eds) Operative neurosurgical techniques, 3rd edn. Saunders, Philadelphia, pp 1423–1430

Bingley T, Leksell L, Meyerson BA (1977) Long term results of stereotactic capsulotomy in chronic obsessive compulsive neurosis. In: Sweet WH, Obrador S, Martin-Rodriguez JG (eds) Neurosurgical treatment in psychiatry, pain and epilepsy. University Park Press, Baltimore, pp 287–289

Bouckoms AJ (1988) Ethics of psychosurgery. Acta Neurochir Suppl (Wien) 44: 173–178

Burzaco J (1981) Stereotactic surgery in the treatment of obsessive compulsive neurosis. In: Perris C, Struwe G, Janssen B (eds) Biological psychiatry. Elsevier, Amsterdam, pp 1108–1109

**Cosgrove GR, Ballantine HT (1996) Cingulotomy in psychosurgery. In: Gildenberg PL, Tasker RR (eds) Textbook of stereotactic and functional neurosurgery. McGraw-Hill, New York, pp 1965–1970

Cosgrove GR, Rauch SL (1995) Psychosurgery. Neurosurg Clin North Am 6: 167–176

Cosyns P, Caemaert J, Haaijman W, van Veelen C, Gybels J, van Manen J, Ceha J (1994) Functional stereotactic neurosurgery for psychiatric disorders: an experience in Belgium and the Netherlands. Adv Tech Stand Neurosurg 21: 239–279

Foltz EL, White LE Jr (1962) Pain relief by frontal cingulotomy. J Neurosurg 19: 89–94

Freeman W, Watts JW (1937) Prefrontal lobotomy in the treatment of mental disorders. South Med J 30: 23–31

Freeman W, Watts JW (1950) Psychosurgery in the treatment of mental disorders and intractable pain, 2nd edn. Thomas, Springfield

Fülgraff G, Barbey I (eds) (1978) Stereotaktische Hirnoperationen bei abweichendem Sexualverhalten. Abschlußbericht der Kommission beim Bundesgesundheitsamt. Reimer, Berlin

Fulton JE (1951) Frontal lobotomy and affective behaviour: a neurophysiological analysis. Norton, New York

Goktepe EO, Young LB, Bridges PK (1975) A further review of the results of stereotactic subcaudate tractotomy. Br J Psychiatry 126: 270–280

Gybels JM, Cosyns PR (1995) Cerebral lesions for psychiatric disorders and pain. In: Schmiedek HH, Sweet WH (eds) Operative neurosurgical techniques, 3rd edn. Saunders, Philadelphia, pp 1413–1421

Hassler R, Dieckmann G (1973) Relief of obsessive-compulsive disorders, phobias and tics by stereotactic coagulation of the rostral intralaminar and medial-thalamic nuclei. In: Laitinen LV, Livingston KE (eds) Surgical approaches in psychiatry. Medical and Technical Publishing, Lancaster, pp 206–212

Herner T (1961) Treatment of mental disorders with frontal stereotactic thermo-lesions. A follow-up study of 116 cases. Acta Psychiatr Neurol Scand 36[Suppl 158]: 1–146

Jenike MA, Baer L, Ballantine T, Martuza RL, Tynes S, Giriunas I, Buttolph ML, Cassem NH (1991) Cingulotomy for refractory obsessive-compulsive disorder. A long-term follow-up of 33 patients. Arch Gen Psychiatry 48: 548–555

Kartsounis LD, Poynton A, Bridges PK, Bartlett JR (1991) Neuropsychological correlates of stereotactic subcaudate tractotomy. A prospective study. Brain 114: 2657–2673

Kelley D, Richardson A, Mitchell-Heggs N (1973) Stereotactic limbic leucotomy: neurophysiologic aspects and operative technique. Br J Psychiatry 123: 133–140

Klüver H, Bucy PC (1939) Preliminary analysis of functions of the temporal lobes in monkeys. Arch Neurol Psychiatry 42: 979–1000

Knight GC (1964) The orbital cortex as an objective in the surgical treatment of mental illness. The development of the stereotactic approach. Br J Surg 51: 114–124

Leksell L, Backlund EO (1979) Stereotactic gammacapsulotomy. In: Hitchcock ER, Ballantine HT Jr, Meyerson BA (eds) Modern concepts in psychiatric surgery. Elsevier/North Holland Biomedical, Amsterdam, pp 213–216

Meyerson BA (1977) Stereotactic anterior capsulotomy in the treatment of obsessive-compulsive neurosis. In: Carrea R (ed) Neurological surgery. Excerpta Medica, Amsterdam, pp 307–312 (International congress series no 433, 307th edn)

Meyerson BA (1996) Neurosurgical treatment of mental disorders: introduction and indications. In: Gildenberg PL, Tasker RR (eds) Textbook of stereotactic and functional neurosurgery. McGraw-Hill, New York, pp 1955–1963

Mindus P, Meyerson BA (1995) Anterior capsulotomy for intractable anxiety disorders. In: Schmiedek HH, Sweet WH (eds) Operative neurosurgical techniques, 3rd edn. Saunders, Philadelphia, pp 1443–1455

Mindus P, Nyman H (1991) Normalization of personality characteristics in patients with incapacitating anxiety disorders after capsulotomy. Acta Psychiatr Scand 83: 283–291

Mindus P, Rasmussen SA, Lindquist C (1994) Neurosurgical treatment for refractory obsessive-compulsive disorder: implications for understanding frontal lobe function. J Neuropsychiatry Clin Neurosci 6: 467–477

Mitchell-Heggs N, Kelley D, Richardson A (1977) Stereotactic limbic leucotomy: clinical, psychological and physiological assessment at sixteen months. In: Sweet WH, Obrador S, Martin-Rodriguez JG (eds) Neurosurgical treatment in psychiatry, pain and epilepsy. University Park Press, Baltimore, pp 367–379

Moniz E (1936) Les premières tentatives opératoires dans le traitment de certaines psychoses. Masson, Paris

Moniz E (1937) Prefrontal leukotomy in the treatment of mental disorders. Am J Psychiatry 93: 1379–1385

Müller DH, Orthner H (1973) Further results of stereotaxis in the human hypothalamus in sexual deviations. Neurochiurgia (Stuttg) 16: 113-126

Narabayashi H, Shima F (1973) Which is the better amygdala target, the medial or lateral nuclei (for behaviour problems and paroxysm in epileptics). In: Laitinen LV, Livingston KE (eds) Surgical approaches in psychiatry. Medical and Technical Publishing, Lancaster, pp 129–134

Orthner H (1982) Die theoretischen und tierexperimentellen Grundlagen der vorderen Hypothalamotomie zur Behandlung schwerer Sexualstörungen und die Zielsicherheit der Methode. Fortschr Neurol Psychiatry 50: 316–329

Pippard J (1955) Rostral leucotomy: a report on 240 cases personally followed up after one and one half to five years. J Ment Sci 101: 756–773

Rylander G (1979) Stereotactic radiosurgery in anxiety and obsessive-compulsive states: psychiatric aspects. In: Hitchcock ER, Ballantine HT Jr, Meyerson BA (eds) Modern concepts in psychiatric surgery. Elsevier/North Holland Biomedical, Amsterdam, pp 235–240

Schorsch E, Schmidt G (1979) Hypothalamotomie bei sexuellen Abweichungen. Eine Kritik aus sexualwissenschaftlicher Sicht. Nervenarzt 50: 689–699

Stone EM (ed) (1989) American psychiatric glossary. American Psychiatric Press, Washington, DC

Talairach J, Hecean H, David M (1949) Recherches sur la coagulation thérapeutique des structures sous-corticales chez l'homme. Rev Neurol (Paris) 81: 4–24

Tippin J, Henn F (1982) Modified leukotomy in the treatment of intractable obsessional neurosis. Am J Psychiatry 139: 1601–1608

Part 2
Personality Disorders, Anxiety and Related Disorders, Behavioural and Addictive Disorders

CHAPTER

1

I. Iancu, P.N. Dannon, Y. Sasson, J. Zohar

Obsessive–Compulsive Disorder

1 Introduction

Up to the early 1980s, obsessive-compulsive disorder (OCD) was considered a treatment-refractory, chronic condition of psychological origins. The management of OCD consisted of dynamic psychotherapy, which was of little benefit, and several pharmacological treatments which had been tried without much success (Salzman and Thaler 1981). However, two findings have substantially changed this field: several researchers have reported that OCD might be a prevalent disorder (Robins et al. 1984; Weissman et al. 1994), and numerous studies have reported the efficacy of various specific therapies for this disorder. Today, OCD is believed to be biologically driven and to be highly treatable by serotonin reuptake inhibitors (SRI).

The observation that clomipramine (CMI), a tricyclic antidepressant with a serotonergic profile, is effective in treating symptoms of OCD (Renynghe de Voxrie 1968; Fernandez-Cordoba and Lopez-Ibor 1967) has resulted in intense interest in the relationship between serotonin and this disorder. Substantial evidence now suggests that OCD is almost unique among psychiatric disorders, as only serotonergic medications appear to be effective in this disorder (Dolberg et al. 1996). Researchers agree that non-serotonergic drugs, such as desipramine (DMI), a potent antidepressant and antipanic agent, are ineffective in OCD (Zohar and Insel 1987; Goodman et al. 1990; Leonard et al. 1991). The specific response to serotonergic drugs has paved the way to further research on the pathogenesis of OCD and OCD-related disorders and has improved the prognoses of these patients.

2 Epidemiology

The lifetime prevalence of OCD in the general population is 2%–3% (Karno et al. 1988; Robins et al. 1984), more than 40 times more than previous estimations. OCD is the fourth most common psychiatric disorder after phobias, substance-related disorders, and major depressive disorder(MDD) (Karno et al. 1988), with even higher rates for subclinical OCD (Valleni-Basile et al. 1996). Despite initial estimations that OCD exists mainly in Western cultures, the high prevalence of OCD has been confirmed across cultures (Weissman et al. 1994). The prevalence of OCD among children and adolescents has been reported by several researchers to be as high as among adults (Flament et al. 1988), but most specialists agree that OCD is more frequent in middle-aged adults than in people of other age groups.

Recently, the above-mentioned numbers have been challenged, as Nelson and Rice (1997) and Stein et al. (1997) have suggested that diagnosis of OCD by the Diagnostic Interview Schedule (DIS) and by laymen leads to overdiagnosis. Thus, lower prevalence rates of 1%–2% were suggested.

Men and women are equally likely to be affected, although some reports have suggested a slight female predominance (Weissman et al. 1994). However, during adolescence, boys are more commonly affected than girls. The mean age of onset is about 20 years of age, although OCD can commence earlier. Single people are more commonly affected, representing their difficulty to maintain a relationship.

Patients with OCD are commonly affected by other mental disorders. The lifetime prevalence for MDD in OCD patients is approximately 67% (Weissman et al. 1994; Rasmussen and Eisen 1992). Other common comorbid psychiatric diagnoses include alcohol use disorders, social phobia, specific phobia, panic disorder, and eating disorders. The comorbidity with schizophrenia and with tic disorders raises interesting pathophysiological and therapeutic implications. In juvenile OCD, the rate of tic disorders can be as high as 40% of patients (Geller et al. 1996). Only 15%–35% of OCD patients have had premorbid obsessional traits.

3 Clinical Features and Diagnosis

The growing interest in OCD has drawn attention to the need for accurate diagnostic tools for OCD and the diagnostic criteria of DSM were reevaluated. The DSM-IV diagnosis of OCD is based on the presence of either obsessions or compulsions, which cause marked distress, are time-consuming (more than 1 h per day), or significantly interfere with the person's normal routine and academic and social activities. It is required that at some point during the course of the disorder the person has recognized that the obsessions or compulsions are excessive or unreasonable.

Specific questions to identify OCD include the following:

1. Do you wash or clean a lot?
2. Do you check things a lot?
3. Is there any thought that keeps bothering you that you would like to get rid of but can't?
4. Do your daily activities take a lot of time to finish?
5. Are you concerned about orderliness or symmetry?

If another axis I disorder is present, it is mandatory that the content of the obsessions or compulsions is not restricted to it (e.g. preoccupation with food or weight in eating disorders or guilt ruminations in the

presence of MDD). The disturbance should not be due to the direct effects of a substance (e.g. a drug of abuse or a medication) or a general medical condition.

A subtype of OCD with poor insight is described, if for most of the time during the current episode, the person does not recognize that the obsessions and compulsions are excessive or unreasonable. This subtype should be distinguished from other psychotic disorders (see "Differential Diagnosis").

The obsessions are recurrent, intrusive, and distressing thoughts, images, or impulses, whereas the compulsions are repetitive, seemingly purposeful behaviors or thoughts that a person feels driven to perform (compulsions). Obsessions are usually unpleasant and increase a person's anxiety, whereas carrying out compulsions reduces a person's anxiety. Resisting carrying out a compulsion, however, results in increased anxiety. The patient usually realizes that the obsessions are irrational and experiences both the obsession and the compulsion as ego-dystonic. The patient has a strong desire to resist the obsessions or compulsions. However, about half of the patients hardly resist the compulsion. About 80% of all patients believe that the compulsion is irrational, but sometimes obsessions and compulsions become overvalued to the patients. Some may insist that compulsive cleanliness is morally correct, even though they have lost their job because of the cleaning compulsion.

Patients with both obsessions and compulsions constitute at least 75% of the affected patients. Most patients present with multiple obsessions and compulsions. The symptoms may shift and a patient who had washing rituals during childhood may present with checking rituals as an adult.

Four major symptom patterns exist. The most common pattern is an obsession of contamination by dirt or germs, followed by washing or accompanied by compulsive avoidance of the presumably contaminated object (not touching doorknobs, electrical switches, newspapers, people's hands, phones). The feared object is hard to avoid (e.g. feces, urine, dust, or germs). Patients wash their hands excessively and sometimes avoid leaving home because of fear of germs. Patients feel anxiety, avoid exposure to the feared object, and avoid every contact they believe that may spread the germs.

The second most common pattern is an obsession of doubt, followed by a compulsion of checking. The person checks whether he closed the door of the house or whether he closed the stove (indirect danger of violence), and the checking may involve multiple trips back home to check the stove, for example. Instead of resolving uncertainty, the checking often contributes to even greater doubt, which leads to further checking. The patients exhibit obsessional self-doubt, as they always feel guilty for having forgotten or committed something (i.e. fear of hurting someone while driving, leading to repetitive driving back over the same spot again and again).

The third most common pattern is one with merely intrusive obsessional thoughts without compulsions. Such obsessions are usually repetitious thoughts of some sexual or aggressive act that is reprehensible to the patient. The fourth most common pattern is the need for symmetry or precision, which leads to a compulsion of slowness in an attempt to do things "just right": patients can take hours to eat a meal or shave. Unlike other OCD patients, these patients do not resist the symptoms. Other patterns include hoarding and religious obsessions.

The gap between the knowledge that the symptoms are irrational and the overwhelming urge to perform them (even after resisting and failing) adds more tragic color to the immense suffering associated with this disorder.

About 15% of chronic schizophrenia patients may also suffer from OCD, six times more than in the general population. Although many patients report symptoms that resemble OCD superficially, DSM-IV requires that the obsessive-compulsive symptoms should not be related to the delusions content in order to allow both diagnoses. Despite present attempts to treat various psychopathologies according to symptom clusters, the difference between the "obsessive syndrome" seen in the framework of other mental disorders and OCD as an independent disorder comorbid with other mental disorders (e.g. schizophrenia) should still be borne in mind.

Many schizophrenic patients can distinguish the ego-dystonic obsessive-compulsive symptoms, perceived as coming from within, from the ego-syntonic delusions, perceived as introduced from the outside. Long-term follow-up studies demonstrate diagnostic stability over the years and suggest that the presence of OCD in schizophrenia predicts a poor prognosis. Several studies among schizophrenic patients with OCD reported an improvement in OCD symptomatology after the addition of a specific antiobsessive medication (Zohar et al. 1993). Preliminary data also show some role for the new atypical neuroleptics, such as olanzapine, in treating this subset of patients.

The poor prognosis of schizo-obsessive patients, preliminary data regarding their response to the unique combination of antipsychotic and antiobsessive medications, and the high prevalence of this presentation suggest that a schizo-obsessive category may be of value.

4 Differential Diagnosis

Personal distress and functional impairment, which are required for the diagnosis, differentiate OCD from

ordinary or mildly excessive worries, thoughts, and habits. The medical differential diagnosis includes tic disorders (especially Tourette's syndrome, in which two-thirds of patients meet the diagnostic criteria for OCD), temporal lobe epilepsy, head trauma, and postencephalitic complications.

Psychiatric diagnoses that should be ruled out include schizophrenia, obsessive-compulsive personality disorder (OCPD), phobias, and depressive disorders. OCD can usually be differentiated from schizophrenia by the absence of other schizophrenic symptoms, by the less bizarre nature of symptoms, and by the patients' insight into their disorder. OCPD and the ICD-10, parallel anankastic personality disorder, do not have the degree of functional impairment characteristic of OCD and are ego-syntonic.

Phobias are distinguished by the absence of a relation between the obsessive thoughts and the compulsions. MDD can sometimes be associated with obsessive ideas, but patients with OCD usually fail to meet all the criteria of MDD. Other psychiatric diagnoses that are closely related to OCD are hypochondriasis, body dysmorphic disorder, and trichotillomania. Several researchers have investigated these disorders, their relations to OCD, and the response to various drugs (Hollander 1997).

5 Course and Prognosis

More than half the patients with OCD have a sudden onset of symptoms, usually after a stressful event, such as pregnancy, a loss, or a sexual problem. Due to the secretive nature of the disorder, there is often a 5- to 10-year delay before patients come to psychiatric attention, although the delay might shorten due to the increased public awareness of the disorder through articles, books, and films on the subject. The course is usually long, but variable; some patients experience a fluctuating course, while others experience a chronic course.

About 20%–30% of the patients have significant improvement in their symptoms, and 40%–50% have moderate improvement. The remaining 20%–40% become chronic or have a worsening of their symptoms.

Patients are prone to depression and sometimes even to suicide. A poor prognosis is indicated by yielding to (rather than resisting) compulsions, childhood onset, bizarre compulsions, the need for hospitalization, a coexisting MDD, delusional beliefs, the presence of overvalued ideas (i.e. some acceptance of the obsessions and compulsions), and the presence of personality disorder (especially schizotypal personality disorder). A good prognosis is indicated by a good social and occupational adjustment, the presence of a precipitating event, and an episodic nature of the symptoms. The obsessional content does not seem to be related to the prognosis.

6 Etiology

6.1 Neurotransmitters

The many clinical trials that have been conducted with various serotonergic drugs support the hypothesis that a dysregulation of serotonin is involved in the response to these treatments. However, this does not necessarily reflect on pathogenesis. Abnormality of the serotonergic system and particularly hypersensitivity of postsynaptic 5-HT (hydroxytryptamine) receptors remains the leading hypothesis for the underlying pathophysiology of OCD (Zohar and Insel 1987).

Clinical studies have assayed cerebrospinal fluid (CSF) levels of serotonin metabolites, e.g. 5-hydroxyindoleacetic acid (5-HIAA, a metabolite of 5-HT serving as an index of 5-HT turnover) (Thoren et al. 1980; Insel et al. 1985), and affinities of platelet-binding sites of tritiated imipramine (which binds to serotonin reuptake sites) (Weizman et al. 1986; Marazziti et al. 1992; Vitiello et al. 1991; Kim et al. 1991) and have reported variable findings of those measures in OCD patients. A recent study supporting the relationship between decreased function of the serotonergic system and positive response to selective SRI (SSRI) is that of Marazziti et al. (1997), who showed normalization of the number of platelet 5-HT transporters following treatment with different SSRI.

In a study by Thoren et al. (1980), patients who responded to CMI had higher pretreatment levels of 5-HIAA than the nonresponders, and clinical improvement was positively correlated with the decrease in CSF concentration of 5-HIAA. Flament et al. (1987) examined peripheral measures of serotonergic and noradrenergic function in adolescents with OCD participating in a double-blind, placebo-controlled study with CMI. Clinical improvement during drug therapy closely correlated with pretreatment platelet serotonin concentration and monoamine oxidase (MAO) activity, as well as with the decrease in both measures during CMI administration. Moreover, only the plasma levels of CMI, a potent 5-HT reuptake inhibitor, but not the plasma levels of its primary metabolite, desmethyl CMI, which has noradrenergic properties, correlated significantly with improvement in obsessive-compulsive symptoms. These findings suggested

that the effects of CMI on serotonin function are pertinent to the antiobsessional action observed.

Additional evidence for the derangement in the serotonergic system in OCD was provided by challenge studies which differ from biological markers which examine systems in equilibrium. Challenges with L-tryptophan (Charney et al. 1988), m-chlorophenylpiperazine (mCPP; Zohar and Insel 1987; Hollander et al. 1992), sumatriptan (5-HT1D agonist; Zohar 1996), ipsapirone (a 5-HT1A receptor ligand; Lesch et al. 1991), and MK-212 (affecting 5HT1A and 5HT2C; Bastani et al. 1990), among others, were used to evaluate whether they worsen obsessive-compulsive symptoms or whether they have thermal or neuroendocrine effects. Some of these studies have shown behavioral hypersensitivity and neuroendocrine hyposensitivity to be characteristic of the OCD challenge response. These studies might have the potential to pinpoint the receptor subtype involved in OCD, raising the possibility that 5HT1D and 5HT2C receptors, but not 5HT1A, may be involved in OCD (Sasson and Zohar 1996).

Additional support for the importance of serotonin in the therapeutic response to SRI in OCD was given by Benkelfat et al. (1989). This group of investigators administered the serotonin receptor antagonist metergoline and placebo to ten patients with OCD in a double-blind crossover study. Patients receiving CMI on a long-term basis responded with greater anxiety to a 4-day administration of metergoline when compared with the placebo phase. Obsessive-compulsive symptoms also peaked during the metergoline phase. Metergoline lowered plasma prolactin concentrations, providing further evidence of physiologically significant 5-HT antagonism. In another study, metergoline abolished the above-mentioned increase in obsessive-compulsive symptoms caused by mCPP.

6.2 Dopamine and Serotonin

The most compelling evidence for dopaminergic involvement in OCD comes from the abundance of obsesssive-compulsive symptoms in basal ganglia disorders, such as Tourette's syndrome and postencephalitic Parkinson's disease. The therapeutic benefit obtained with coadministration of dopamine blockers and SRI (McDougle et al. 1990) has suggested a role for dopamine dysfunction. Marazitti et al. (1992) evaluated not only peripheral markers of serotonin, but also levels of platelet sulphotransferase, an enzyme involved in the catabolism of catecholamines, which provides a marker of presynaptic dopamine function. OCD patients had a decreased level of platelet H3-imipramine binding and a parallel increase in the level of sulphotransferase activity compared with controls, a finding consistent with the hypothesis of reduced 5-HT activity and increased dopamine transmission in OCD.

6.3 Immune Factors

The study of autoimmune factors has been prompted by the association between OCD and the autoimmune disease of the basal ganglia, Sydenham's chorea, a complication of rheumatic fever that is accompanied by obsessive-compulsive symptoms in over 70% of patients (Swedo et al. 1994). This group reported that ten out of 11 children had antibodies directed against the caudate. These children had a history of obsessive-compulsive symptoms which started prior to the onset of the chorea, reached a peak in line with the motor symptoms, and declined with their resolution. This led to the postulation that Sydenham's chorea might constitute a medical model of OCD, consistent with the hypothesis of basal ganglia dysfunction in OCD. Antibodies against two peptides of the basal ganglia have also been found (Roy et al. 1994). A strong connection was reported between OCD/Tourette's syndrome and the B cell antibody D8/17, another anti-brain antibody (Swedo et al. 1997). Successful immunomodulatory treatments of Sydenham's chorea with plasmapheresis and I.V. immunoglobulin have provided more evidence for the autoimmune theory of OCD (Allen et al. 1995). Cell-mediated immune function alterations were reported in OCD, but replication studies are needed.

6.4 Brain Imaging Studies

Positron emission tomography (PET) has displayed increased activity (i.e. metabolism and blood flow) in the frontal lobes, the basal ganglia (especially the caudate nucleus), and the cingulum of OCD patients (Rauch 1998). Pharmacological and behavioral treatments reportedly reverse those abnormalities (Baxter et al. 1992). The data from functional imaging studies are consistent with the data from structural brain imaging studies. Both computed tomography and magnetic resonance imaging studies have found decreased sizes of caudates bilaterally. Both functional and structural imaging procedures are also consistent with the observation that neurological procedures involving the cingulum are sometimes effective in the treatment of OCD patients.

Overall, the brain imaging research provides evidence that the underlying dysfunction is likely to be in

the prefrontal cortex–basal ganglia thalamic circuitry rather than in any one single brain region. Dysfunction of these circuits can be explored by neuropsychological testing and evoked potentials. A recent study of OCD patients showed that they are slower in performing tasks involving frontocortical systems, suggesting alterations at this level (Galderisi et al. 1995). An evoked potential study in 17 patients by Towey et al. (1994) showed enhanced processing negativity in the frontal cortex consistent with prefrontal hyperactivity shown in brain imaging studies.

There may be an abnormality of the orbitofrontal cortex in OCD. Patients perform poorly on orbitofrontal cognitive tests (such as the Object Alternation Test), showing a different neuropsychological profile to other disorders such as schizophrenia, where patients classically exhibit deficits on such tests as the Wisconsin Card Sorting Test (Abbruzzese et al. 1995). The data suggest an anatomically specific frontal lobe dysfunction in OCD.

6.5 Genetics

The inheritance of OCD has a significant genetic component, but research is still needed in this area. A significantly higher concordance rate was found for monozygotic twins than for dizygotic twins (Rasmussen and Tsuang 1986). A total of 35% of the first-degree relatives of OCD patients are also afflicted with the disorder (Lenane et al. 1990). Thus, a genetic component of OCD is supported by family and twin studies. Nevertheless, genetic research has yet to find abnormalities at the 5-HT transporter gene level. A recent study exploring the polymorphism of the promoter region of the gene for the 5-HT transporter failed to identify any differences between OCD patients and controls (Billet et al. 1997). A trend towards homozygosity among the OCD group was demonstrated, but there was no correlation with the clinical response to SRI.

6.6 Other Biological Data

Sleep EEG and neuroendocrine studies have found abnormalities similar to those seen in depression, such as decreased rapid eye movement (REM) latency, nonsuppression on the dexamethasone–suppression test (DST), and decreased growth hormone secretion with clonidine infusions (Sasson and Zohar 1996).

6.7 Behavioral Factors

Based on postulations from the learning theory, obsessions are conditioned stimuli. When a relatively neutral stimulus is coupled with an anxiety-provoking one, through conditioning it will produce anxiety even when presented alone. Compulsions are acquired through their anxiety-reducing effect, causing the patient to repeat them in order to avoid anxiety. Avoidance strategies are learned and become fixed.

6.8 Psychosocial Factors

The dynamic aspects of OCD were first described by Sigmund Freud, who referred to OCD as "obsessional neurosis." The disorder was thought to result from a regression from the oedipal phase to the anal phase, with its characteristic ambivalent emotional stage. The ambivalence is connected to the unraveling of the smooth fusion between sexual and aggressive drives characteristic of the oedipal phase. The coexistence of hatred and love toward the same person leaves the patient paralyzed with doubt and indecision. Freud originally suggested that obsessive symptoms result from unconscious impulses of an aggressive or sexual nature. These impulses could cause extreme anxiety, which is avoided by the defense mechanisms of repression and reaction formation. One of the striking features of OCD patients is the degree to which they are preoccupied with aggression or cleanliness (anal phase), either overtly in the content of their symptoms or in the associations that lie behind them. Thus the psychogenesis of OCD may lie in disturbances in normal growth and development related to the anal-sadistic phase.

Freud described three major psychological defense mechanisms that are important in OCD: isolation, undoing, and reaction formation. According to the psychoanalytic formulation, OCD develops when these defenses fail to contain the anxiety. *Isolation* is the separation of the idea and the affect that it arouses, when the patient is only aware of the affectless idea (i.e. a doctor that performs resuscitation and is not overwhelmed by the scene, the patient's condition, and the surroundings). *Undoing* is a secondary defense in order to combat the impulse and quieten the anxiety that its imminent eruption into consciousness arouses. Undoing is a compulsive act that is performed in an attempt to prevent or undo the results that the patient irrationally anticipates from a frightening obsessional thought or impulse. *Reaction formation* is related to the production of character traits rather than symptom formation (characteristic of the above defenses). The

trait seems highly exaggerated and inappropriate. An example can be the switch of anger and hate into exaggerated love and dedication. Ambivalence and magical thinking are additional features of the psychological texture of the OCD patient.

7 Treatment

7.1 Pharmacological Treatment

The outlook for patients with OCD was not very promising in the early 1980s. However, since then, several potent SRI have been studied extensively in OCD. Aggregate statistics for all SRI suggest that 70% of treatment-naive patients will improve at least moderately (Rasmussen et al. 1993).

CMI was the first to be used and the most extensively studied. CMI was reported to be an effective medication for OCD in the late 1960s (Fernandez-Cordoba and Lopez-Ibor 1967; Ronynghe de Voxirie 1968). Since then, numerous placebo-controlled studies have clearly shown CMI's effectiveness. This research direction culminated in the multicenter, controlled trial in the United States ($n = 520$) that conclusively confirmed CMI's effectiveness (Clomipramine Collaborative Study Group 1991). This study showed that, after 10 weeks of treatment, 58% of patients treated with CMI rated themselves much or very much improved versus only 3% of placebo-treated patients.

Beside CMI, other non-tricyclic SRI, such as fluoxetine, fluvoxamine, paroxetine, and sertraline, are gaining acceptance as effective alternatives for the treatment of OCD in controlled studies.

7.1.1 Efficacy of Serotonergic Versus Adrenergic Antidepressants

While anecdotal reports have suggested that clinical benefit can be obtained with a range of antidepressant medications, consistent effectiveness has only been demonstrated for the SRI. Several studies have directly compared CMI to other antidepressants, and a consistent pattern emerges: antidepressant drugs that are less potent SRI than CMI are generally ineffective in OCD (Thoren et al. 1980; Zohar and Insel 1987; Leonard et al. 1991; Goodman et al. 1990).

7.1.2 Other Pharmacological Approaches

The treatment of resistant OCD patients has also contributed to the understanding that serotonergic tricyclics and SSRI are important in the management of these patients. Providing further support for the importance of serotonin neurotransmission in the pathophysiology and treatment of OCD are early observations regarding the effectiveness of isocarboxazid and more recently phenelzine, MAO inhibitors with prominent effects on serotonin metabolism. However, this efficacy was not replicated in recent studies. Further studies on the therapeutic role of MAO inhibitors in OCD, especially in OCD with comorbidity of panic disorder, are warranted. Considered as one of the anxiety disorders according to DSM-IV (but not according to ICD-10), it is not surprising that anxyolitics have been suggested in the treatment of OCD patients. Thus alprazolam and clonazepam have been reported as efficient in several uncontrolled studies and case series, and even in a small double-blind randomized, multiple crossover study. However, since OCD is a chronic disorder, the use of anxyolitics for long periods raises questions of dependency brought about by long-term use of benzodiazepines.

Despite reports in open studies as regards the efficacy of trazodone, buspirone, fenfluramine, and lithium (drugs that have some serotonergic effect), the results in double-blind studies were mainly negative. However, it is important to stress that not all serotonergic drugs are efficacious in this disorder. Thus, the serotonin precursor tryptophan exhibited only mixed efficacy.

Adding drugs affecting dopamine function (pimozide, haloperidol, risperidone) to SRI therapy in treatment-resistant OCD patients may result in improvement in patients with a personal or family history of tics (McDougle 1997).

Neurosurgery has been reported to be effective in some OCD patients, with procedures that disconnect the outflow pathways originating from the orbitofrontal cortex. Cingulotomy can help some intractable patients, but although the immediate results may be striking, the long-term prognosis is more reserved (Jenike et al. 1991) and a second operation or suicide are not infrequent.

7.1.3 Treatment Algorithm

The first line of treatment in an OCD patient is an SSRI. If the patient cannot tolerate the drug or does not respond, a switch to CMI (or to another SSRI) is warranted. The third line of treatment consists of a combination of SSRI and CMI or the addition of small doses of neuroleptics (in comorbidity with Tourette's syndrome or schizophrenia), as well as the addition of one of the following: lithium, trazodone, buspirone, tryptophan. The fourth stage consists of either atypical neuroleptics, thyroid supplementation, clonidine,

MAO inhibitors, I.V. CMI, or clonazepam. In resistant cases, electroconvulsive therapy (ECT) or neurosurgery is indicated.

7.1.4 Onset of Treatment Response

It has been suggested that a relatively long period (as long as 10–12 weeks) is needed for CMI to be significantly effective. Up to 6 months or even longer is needed to achieve maximum response.

7.1.5 Long-Term Treatment

It has been demonstrated that the length of treatment should be considerable and that most patients relapse after premature discontinuation. Pato et al. (1988) reported that 16 out of 18 patients with OCD relapsed within 7 weeks after stopping CMI, although some had been treated for more than 1 year (mean, 10.7 months). All patients regained therapeutic effects when CMI was reintroduced. Leonard et al. (1991) examined the effect of CMI substitution during long-term CMI treatment in 26 children and adolescents with OCD (mean duration of treatment, 17 months; range, 4–32 months). Half of the patients were blindly assigned to 2 months of desipramine (DMI) treatment, and then CMI was reintroduced. Almost 90% relapsed during the 2-month substitution period in comparison with only 18% of those kept on CMI.

OCD patients should be maintained on CMI more than 1 year after the disappearance of symptoms before a very gradual attempt to discontinue the treatment is carried out. The maintenance dose needed in OCD is also unclear. In a recent study that examined this issue, Mundo et al. (1997) investigated the effect of dose reduction in 30 patients previously treated successfully with CMI or fluoxetine. Patients were randomized to receive the same drug dosage, a reduced dose, or a very reduced dose. There was no difference between the groups during the 102 days of the study.

7.1.6 Drug Dosage

Higher doses of antidepressants have been used in the treatment of OCD than of depression, although empirical data supporting this practice is scant. Some fixed-dose studies using fluoxetine have found some advantage with using higher doses, while others have not. The tradition of using high doses has to be supported by further research, although it seems reasonable to use higher doses in nonresponders or when only partial relief is attained.

7.1.7 Comparative Studies of Clomipramine and Selective Serotonin Reuptake Inhibitor

The introduction of SSRI has raised the question regarding the comparative efficacy of CMI versus that of the SSRI. SSRI are important alternatives to CMI, since their range of side effects is clearly different (lacking anticholinergic side effects, sedation, weight gain to name but a few). Although SSRI may provoke nausea, headaches, and sleep disturbances, these side effects are usually less troublesome to most patients.

Fluoxetine was compared with CMI by Pigott et al. (1990) in 11 OCD patients in a 10-week crossover study. Although no significant differences were noted regarding efficacy, the proportion of fluoxetine nonresponders who later responded to CMI tended to be higher in comparison to the CMI nonresponders who were switched to fluoxetine. Patients reported significantly fewer side effects while on fluoxetine.

Freeman et al. (1994) compared the efficacy of fluvoxamine and CMI in a multicenter, randomized, double-blind, parallel-group comparison ($n = 66$). Both drugs were equally effective and well-tolerated, but fluvoxamine produced fewer anticholinergic side effects and caused less sexual dysfunction, and more reports of headache and insomnia. Paroxetine was of comparable efficacy to CMI, and both were significantly more effective than placebo in a multinational, double-blind, placebo-controlled, parallel-group study with 399 OCD patients (Zohar and Judge 1996). Bisserbe et al. (1997) reported that sertraline (50–200 mg/day) was significantly more effective than CMI (50–200 mg/day) in a double-blind study ($n = 160$).

Greist et al. (1995) compared the results from the four large multicenter placebo-controlled trials of the SRI: CMI ($n = 520$), fluoxetine ($n = 355$), fluvoxamine ($n = 320$), and sertraline ($n = 325$). All four agents were significantly more effective than placebo. A significantly greater percentage of patients treated with CMI were rated as much or very much improved in comparison with patients on other treatments. However, the studies included in this meta-analysis had different subject characteristics, and this might have affected the results, as more treatment-naive patients (who are more likely to respond) participated in the CMI study.

7.1.8 Predictors of Treatment Outcome

Several predictors of negative outcome were suggested based on small studies: male gender, longer duration of illness, higher initial Yale-Brown Obsessive-Compulsive Scale (Y-BOCS) score, cleaning rituals, axis II diagnosis, baseline depression, overvalued ideas and

high avoidance, nonsuppression on the DST at baseline, increased prolactin blunting, and behavioural exacerbation on mCPP challenge (Dolberg et al. 1996; Sasson and Zohar 1996; Ravizza et al. 1995). However, based on a large multicenter CMI study, the abovementioned predictors were not confirmed. Additional studies on possible predictors of outcome in OCD are of great importance, with sound methodology and large samples of patients warranted.

7.2 Behavior Therapy

Behavior therapy is as effective as pharmacotherapy in OCD (Marks et al. 1975), and some data indicate that the beneficial effects are longer-lasting with behavior therapy (Greist 1996). About two-thirds of patients with moderate or severe rituals can be expected to improve substantially, but not completely. A combination of behavior therapy and pharmacotherapy might constitute the optimal treatment for OCD. Two recent neuroimaging studies found that patients with OCD who are successfully treated with behavior therapy show changes in cerebral metabolism similar to those produced by successful treatment with SRI (Schwartz et al. 1996; Baxter et al. 1992). Behavior therapy is considered less effective for obsessive thoughts occurring without rituals. Behavior therapy can be conducted in inpatient and outpatients settings. The principal behavioral approaches in OCD are exposure for obsessions and response prevention for rituals. Desensitization, thought stopping, flooding, implosion therapy, and aversion conditioning have also been used in OCD patients. In behavior therapy, the patient must be truly committed to improvement. In a recent study, 18 OCD patients were randomly assigned to receive exposure and response prevention or a general anxiety management intervention (control) for 3 weeks. The first group showed significant improvement, whereas the control group showed no improvement from baseline (Lindsay et al. 1997). Direct comparisons of behavior therapy and pharmacotherapy are few and are limited by methodological issues. Cox et al. (1993) reported equal efficacy in a meta-analysis. However, behavior therapy might have a lower relapse rate, even after interruption of therapy.

7.3 Psychotherapy

Despite the fact that biological interventions are more efficacious in OCD patients, psychodynamic factors might be of considerable benefit in understanding what precipitates exacerbations of the disorder and in treating various forms of resistances to treatment, such as noncompliance to medications or to homework assignments. It is important to remember that the symptoms may have important psychological meanings that make patients reluctant to give them up. A psychodynamic exploration into the patient's resistance to treatment may result in improved compliance.

In the absence of controlled studies of insight-oriented psychotherapy for OCD, the anecdotal reports reporting lasting change do not enable us to generalize about its efficacy. In addition, the efficacy of medications in producing quick improvement has rendered slow and long-term psychotherapy out of favor.

Supportive psychotherapy has a place in the treatment of OCD patients and may help patients to improve their functioning and adjustment. Management should also include attention to family members through the provision of emotional support, reassurance, explanation, and advice on how to manage and respond to the patient. Family therapy may reduce marital discord and build a treatment alliance, also to help in the resistance to compulsions. Group therapy is useful as a support system for some patients.

8 Conclusion

The treatment of OCD was characterized by pessimism until 20 years ago, when effective treatments using behavior therapy and SRI were developed. Although introduced for OCD in 1967, it was only in the 1980s that double-blind studies confirmed the efficacy of CMI, an SRI, for this disorder. This was followed by the further introduction of the SSRI, which also proved effective for OCD patients (Greist et al. 1995). The antiobsessive activity of these drugs was found to be independent of the drugs' antidepressant effect, as established by efficacy both in depressed and nondepressed patients. Overall, serotonergic therapies have enabled a better outlook for these patients and have enlarged our understanding of the pathophysiology of this disorder (Dolberg et al. 1996; Zohar and Insel 1987).

Previously thought to be a rare and untreatable disorder, OCD is now recognized as common, and there is now good reason to expect that OCD patients will benefit substantially from behavior therapy and potent SRI. Unfortunately, some OCD patients do not seek treatment and the disease tends to be chronic. There is a 10-year lag between the onset of symptoms and the seeking of professional help due to feelings of humiliation and embarrassment. A further lag is caused until diagnosis and correct treatment are given (Hollander 1997).

The efficacy of the SRI together with the lack of efficacy on the part of adrenergic antidepressants as medications for OCD have suggested that serotonin is involved in the pathophysiology of OCD. This relationship was validated only later by research on serotonergic markers in OCD and by the challenge paradigm (Dolberg et al. 1996). Which type of serotonergic receptor is involved in the pathogenesis and/or the mechanism of action of antiobsessional drugs is still unclear. The studies by Lesch et al. (1991) with ipsapirone, a 5-HT1A ligand, have shown that ipsapirone had no notable effect on behavioral measures, nor did it produce thermoregulatory or neuroendocrine responses that were significantly different from placebo. Moreover, the lack of therapeutic effectiveness of the 5HT1A agent buspirone adds to the impression that this receptor is not involved in the pathophysiology and management of OCD. Further studies of the involvement of the serotonergic system in OCD are crucial to elucidating the role of this neurotransmitter in the pathophysiology and management of OCD.

U.S. census data suggest that over $8 billion are spent each year on the management of OCD, one fifth of that spent for heart disease (DuPont et al. 1995). Because OCD patients often attempt to conceal their symptoms, it is incumbent on clinicians to screen for OCD in every mental status examination, since appropriate treatment can result in improved quality of life, reduced OCD chronicity, and reduced costs to the individual and society.

9 References

Abbruzzese M, Bellodi L, Ferrari S et al (1995) Frontal lobe dysfunction in schizophrenics and obsessive-compulsive disorder. A neuropsychological study. Brain Cogn 27: 202–212

Allen AJ, Leonard HL, Swedo SE (1995) Case study: a new infection-triggered, autoimmune subtype of pediatric OCD and Tourette's syndrome. J Am Acad Child Adolesc Psychiatry 34: 307–311

Bastani B, Nash JF, Meltzer HY (1990) Prolactin and cortisol responses to MK-212, a serotonin agonist, in obsessive-compulsive disorder. Arch Gen Psychiatry 47: 833–839

Baxter LR Jr, Schwartz JM, Bergman KS et al (1992) Caudate glucose metabolic rate changes with both drug and behaviour therapy for OCD. Arch Gen Psychiatry 49: 681–689

Benkelfat C, Murphy DL, Zohar J et al (1989) Clomipramine in obsessive-compulsive disorder: further evidence for a serotonergic mechanism of action. Arch Gen Psychiatry 46: 23–28

Billet EA, Richter MA, King N et al (1997) Obsessive compulsive disorder, response to serotonin reuptake inhibitors and the serotonin transporter gene. Mol Psychiatry 2: 403–406

Bisserbe JC, Lane RM, Flament MF et al (1997) A double-blind comparison of sertraline and clomipramine in outpatients with obsessive-compulsive disorder. Eur Psychiatry 153: 1450–1454

Charney DS, Goodman WK, Price LH et al (1988) Serotonin function in obsessive-compulsive disorder: a comparison of the effects of tryptophan and m-chlorophenylpyperazine in patients and healthy subjects. Arch Gen Psychiatry 45: 177–185

Clomipramine Collaborative Study Group (1991) Clomipramine in the treatment of patients with obsessive-compulsive disorder. Arch Gen Psychiatry 48: 730–738

Cox BJ, Swinson RP, Morrison B et al (1993) Clomipramine, fluoxetine, and behaviour therapy in the treatment of obsessive-compulsive disorder: a meta-analysis. J Behav Ther Exp Psychiatry 24: 149–153

Dolberg OT, Iancu I, Sasson Y et al (1996) The pathogenesis and treatment of obsessive-compulsive disorder. Clin Neuropharmacol 19: 129–147

DuPont RL, Rice DP, Shiraki S et al (1995) Economic costs of obsessive-compulsive disorder. Med Interface 8(4): 102–109

Fernandez-Cordoba E, Lopez-Ibor AJ (1967) La monoclorimipramina en enfermos psiquiatricos resistenses a otros tratamientos. Actas Luso Esp Neurol Psiquiatr 26: 119–147

Flament MF, Rapoport JL, Murphy DL et al (1987) Biochemical changes during clomipramine treatment of childhood obsessive-compulsive disorder. Arch Gen Psychiatry 44: 219–225

Flament MF, Whitaker A, Rapoport JL et al (1988) Obsessive compulsive disorder in adolescence: an epidemiological study. J Am Acad Child Adolesc Psychiatry 27: 764–771

Freeman CPL, Triable MR, Deakin JFW et al (1994) Fluvoxamine versus clomipramine in the treatment of obsessive-compulsive disorder: a multicenter, randomized, double-blind, parallel group comparison. J Clin Psychiatry 55: 301–305

Galderisi S, Mucci A, Catapano F (1995) Neuropsychological slowness in obsessive-compulsive patients: is it confined to tests involving the fronto-subcortical systems? Br J Psychiatry 167: 394–398

Geller DA, Biederman J, Griffin S et al (1996) Comorbidity of obsessive-compulsive disorder with disruptive behavior disorders. J Am Acad Child Adolesc Psychiatry 35: 1637–1646

*Goodman WK, Price LH, Delgado PL et al (1990) Specificity of serotonin reuptake inhibitors in the treatment of obsessive compulsive disorder: comparison of fluvoxamine and desipramine. Arch Gen Psychiatry 47: 577–585

*Greist JH (1996) New developments in behaviour therapy for obsessive compulsive disorder. Int Clin Psychopharmacol 11[Suppl 5]: 63–73

*Greist JH, Jefferson JW, Kobak KA et al (1995) Efficacy and tolerability of serotonin transporter inhibitors in obsessive-compulsive disorder. Arch Gen Psychiatry 52: 53–60

Hollander E (1997) Obsessive-compulsive disorder: the hidden epidemic. J Clin Psychiatry [Suppl 12]: 3–6

Hollander E, DeCaria CM, Nitescu A et al (1992) Serotonergic function in obsessive-compulsive disorder: behavioral and neuroendocrine responses to oral m-chlorphenylpyperazine and fenfluramine in patients and healthy volunteers. Arch Gen Psychiatry 49: 21–28

*Insel TR, Mueller EA, Alterman I et al (1985) Obsessive-compulsive disorder and serotonin: is there a connection? Biol Psychiatry 20: 1174–1188

Jenike MA, Baer L, Ballantine T et al (1991) Cingulotomy for refractory obsessive-compulsive disorder. Arch Gen Psychiatry 48: 548–555

Karno M, Golding JM, Sorenson SB et al (1988) The epidemiology of obsessive-compulsive disorder in five US communities. Arch Gen Psychiatry 45: 1094–1099

Kim SW, Dysken MW, Pandey GN et al (1991) Platelet 3H-imipramine binding sites in obsessive compulsive behavior. Biol Psychiatry 30: 467–474

Lenane MC, Swedo SE, Leonard H et al (1990) Psychiatric disorders in first degree relatives of children and adolescents with obsessive-compulsive disorder. J Am Acad Child Adolesc Psychiatry 29: 407–412

*Leonard H, Swedo SE, Lenane MC et al (1991) A double-blind desipramine substitution during long-term clomipramine treatment in children and adolescents with obsessive-compulsive disorder. Arch Gen Psychiatry 48: 922–927

Lesch KP, Hoh A, Disselkamp-Tietze J et al (1991) 5-Hydroxytryptamine 1A receptor responsivity in obsessive compulsive disorder. Comparison of patients and controls. Arch Gen Psychiatry 48: 540–547

Lindsay M, Craig R, Andrews G (1997) Controlled trial of exposure and response prevention in obsessive-compulsive disorder. Br J Psychiatry 171: 135–139

Marazziti D, Hollander E, Lensi P et al (1992) Peripheral markers of serotonin and dopamine function in obsessive-compulsive disorder. Psychiatr Res 42: 41–51

Marazziti D, Pfanner C, Palego L et al (1997) Changes in platelet markers of obsessive compulsive patients during a double-blind trial of fluvoxamine versus clomipramine. Pharmacopsychiatry 30: 245–249

Marks IM, Hodgson R, Rachman et al (1975) Treatment of chronic obsessive-compulsive neurosis in vivo exposure: a 2-year follow-up and issues in treatment. Br J Psychiatry 127: 349–364

McDougle CJ (1997) Update on pharmacologic management of OCD: agents and augmentation. J Clin Psychiatry 58[Suppl 12]: 11–17

McDougle J, Goodman WK, Price LH et al (1990) Neuroleptic addition in fluvoxamine-refractory obsessive-compulsive disorder. Am J Psychiatry 147: 652–654

Mundo E, Barregi SR, Pirola R et al (1997) Long-term pharmacotherapy of obsessive-compulsive disorder: a double-blind controlled study. J Clin Psychopharmacol 17: 4–10

Nelson E, Rice J (1997) Stability of diagnosis of obsessive compulsive disorder in the epidemiologic catchment area study. Am J Psychiatry 154: 826–831

*Pato MT, Zohar-Kadouch R, Zohar J et al (1988) Return of symptoms after discontinuation of clomipramine in patients with obsessive compulsive disorder. Am J Psychiatry 145: 1521–1525

Pigott TA, Pato MT, Bernstein SE et al (1990) Controlled comparisons of clomipramine and fluoxetine in the treatment of OCD: behavioral and biological results. Arch Gen Psychiatry 47: 926–932

Rasmussen SA, Tsuang MT (1986) Clinical characteristics and family history in DSM-III obsessive-compulsive disorder. Am J Psychiatry 143: 317–322

Rasmussen SA, Eisen JL (1992) Epidemiology and clinical features of obsessive-compulsive disorder. In: Jenike MA, Baer L, Minichiello WE (eds) Obsessive compulsive disorders. Theory and management. Year Book Medical Publishers, Chicago, pp 10–27

Rasmussen SA, Eisen JL, Pato MT (1993) Current issues in the pharmacological management of obsessive-compulsive disorder. J Clin Psychiatry 54: 4s–9s

Rauch SL (1998) Neuroimaging in OCD: clinical implications. CNS Spectrums 3[Suppl 1]: 26–29

Ravizza L, Barzega G, Bellino S, Bogetto F, Maina G (1995) Predictors of drug treatment response in OCD. J Clin Psychiatry 56: 368–373

Renynghe de Voxrie GV (1968) Anafranil (G34586) in obsessive compulsive neurosis. Arch Neurol Belg 68: 787–792

Robins LN, Helzer JE, Weissman MM et al (1984) Lifetime prevalence of specific psychiatric disorders in three sites. Arch Gen Psychiatry 41: 949–958

Roy BF, Benkelphat C, Hill JL et al (1994) Serum antibody for somatostatin: 14 and prodynorphin 209–240 in patients with obsessive-compulsive disorder, schizophrenia, Alzheimer's disease, multiple sclerosis and advanced HIV infection. Biol Psychiatry 35: 335–344

Salzman L, Thaler FH (1981) Obsessive compulsive disorder: a review of the literature. Am J Psychiatry 138: 286–296

*Sasson Y, Zohar J (1996) New developments in obsessive-compulsive disorder research: implications for clinical management. Int Clin Psychopharmacol 11[Suppl 5]: 3–12

Schwartz JM, Stoessel PW, Baxter LR Jr et al (1996) Systematic changes in cerebral glucose metabolic rate after successful behavior modification treatment of OCD. Arch Gen Psychiatry 53: 109–113

Stein MB, Forde DR, Anderson G, Walker JR (1997) Obsessive compulsive disorder in the community: an epidemiologic survey with clinical reappraisal. Am J Psychiatry 154: 1120–1126

Swedo SE, Leonard HL, Kiessling LS (1994) Speculations on antineuronal antibody-mediated neuropsychiatric disorders of childhood. Pediatrics 93: 323–326

Swedo SE, Leonard HL, Mittelman BB et al (1997) Identification of children with pediatric autoimmune neuropsychiatric disorders associated with streptococcal infections by a marker associated with rheumatic fever. Am J Psychiatr 154: 110–112

Thoren P, Asberg M, Gronholm B et al (1980) Clomipramine treatment of obsessive compulsive disorder. II. Biochemical aspects. Arch Gen Psychiatry 27: 1289–1294

Towey JP, Tenke CE, Bruder GE et al (1994) Brain event-related potential correlates of over focused attention in obsessive-compulsive disorder. Psychophysiology 31: 535–543

Valleni-Basile LA, Garrison CZ, Waller JL et al (1996) Incidence of OCD in a community sample of young adolescents. J Am Acad Child Adolesc Psychiatry 35(7): 898–906

Vitiello B, Shimon H, Behar D et al (1991) Platelet imipramine binding and serotonin uptake in obsessive-compulsive patients. Acta Psychiatr Scand 84: 29–32

*Weissman MM, Bland RC, Canino GJ et al (1994) The cross national epidemiology of OCD. J Clin Psychiatry 55[Suppl 3]: 5–10

Weizman A, Carmi M, Hermesh H et al (1986) High affinity imipramine binding and serotonin uptake in platelets of eight adolescent and ten adult obsessive-compulsive patients. Am J Psychiatry 143: 335–339

Zohar J (1996) Is 5-HT1D involved in obsessive-compulsive disorder? Eur Neuropsychol 6: 54–55

Zohar J, Insel T (1987) OCD: psychobiological approaches to diagnosis, treatment, and pathophysiology. Biol Psychiatry 22: 667–687

Zohar J, Judge R (1996) Paroxetine versus clomipramine in the treatment of obsessive-compulsive disorder. Br J Psychiatry 169: 468–474

Zohar J, Kaplan Z, Benjamin J (1993) Clomipramine treatment of obsessive compulsive symptomatology. J Clin Pscyhiatry 54: 385–388

CHAPTER

2

M. Linden, D. Zubrägel

Anxiety Disorders: Diagnosis and Epidemiology

Translator: E. Taub

1 Introduction

Anxiety is a universal human experience. Since its earliest literary mention in the Epic of Gilgamesh in about 2000 B.C., it has been dealt with in many different ways in literature, philosophy, religion, and, of course, psychology and medicine. In 1826, the Belgian psychiatrist Joseph Guslain first pointed out the important role played by anxiety in the generation of some psychopathological conditions (McReynolds 1985). In 1871, Westphal gave the original description of agoraphobia. In 1895, Sigmund Freud published a paper entitled *Über die Berechtigung, von der Neurasthenie einen bestimmten Symptomenkomplex als 'Angstneurose' abzutrennen* (English title in the Standard Edition: "On the Grounds for Detaching a Particular Syndrome from Neurasthenia Under the Description 'Anxiety Neurosis' "). In 1920, Watson and Rayner first reported the experimental induction of an anxiety reaction by conditioning.

In keeping with this clinical tradition, anxiety disorders were listed under the major category of the neuroses in the diagnostic key and glossary of psychiatric diseases of ICD-8 (Mombour and Kockott 1971), and there subdivided into anxiety neurosis, phobia, and obsessive-compulsive neurosis. Subsequently, DSM-III (APA 1980) provided a major extension and refinement of this classification, which largely remains current today. An entire chapter was devoted to the anxiety disorders, which were subdivided into agoraphobia with panic attacks, agoraphobia without panic attacks, social phobia, simple phobia, panic disorder, generalized anxiety disorder, obsessive-compulsive disorder, post-traumatic stress disorder, and atypical anxiety disorder. DSM-III originally included a hierarchical rule, according to which the diagnosis of an anxiety disorder required the exclusion of other psychiatric disorders, such as depression or schizophrenia. This rule was dropped in the revision of DSM-III, DSM-III-R (APA 1987).

Both the question of the differentiation of anxiety disorders from other mental disorders, which is dealt with in a particular way by the diagnostic hierarchical or stratification rule (Jaspers 1948), and the problem of diagnostic subdivision arise because there is a continuous spectrum of anxiety ranging from normal to pathological conditions (Rapee 1991) and because anxiety is found as a symptom or syndrome in practically every other mental disorder, most typically in depression (Helmchen and Linden 1986). Nevertheless, the anxiety disorders listed in the present systems of classification by their nature unambiguously merit designation as independent illnesses, both because of the degree of subjective suffering and impairment they cause, and because of their phenomenological features. Thus, even when depression is present, it is entirely reasonable to diagnose the simultaneous presence of phobia, obsessive-compulsive manifestations, or a condition in which depressed mood and generalized anxiety are present to an equal degree. If one prefers not to assign double diagnoses in such cases, single diagnostic categories are also available from other classification schemes, such as anxiety-depression (Gurney et al. 1972) or mixed anxiety and depression, as in ICD-10 (F41.2).

The operationalized diagnostic criteria now available make it possible for the first time to collect comparable data from different countries on the frequency of anxiety disorders in the population. Anxiety disorders are among the more common mental illnesses; they are set apart from other mental illnesses by their specific manifestations, course, and treatment. Unless proper treatment is given, the danger exists that they may become chronic and lead to disability. The remainder of this chapter contains a brief overview of the diagnostic criteria for the most important anxiety disorders, followed by a summary of the findings of epidemiologic studies.

2 Diagnosis

While the fourth edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV; APA 1994) contains an independent chapter entitled "Anxiety Disorders," ICD-10 (WHO 1992) lists this group of illnesses in chapter F4 among the neurotic, stress-related, and somatoform disorders. Despite this difference, there is a large degree of agreement between these two classification systems, as seen in Fig. 1. In both systems, for example, phobic disorders are subdivided into three types: specific phobias, social phobias, and agoraphobia.

2.1 Agoraphobia

The two systems differ with respect to the differential diagnosis of panic disorder and agoraphobia. In ICD-10, agoraphobia is listed as an independent clinical syndrome in the group of phobic disorders and can be further specified according to the presence or absence of a (hierarchically subordinated) panic disorder. In DSM-IV, in contrast, panic attacks occupy a higher position in the diagnostic hierarchy than agoraphobia;

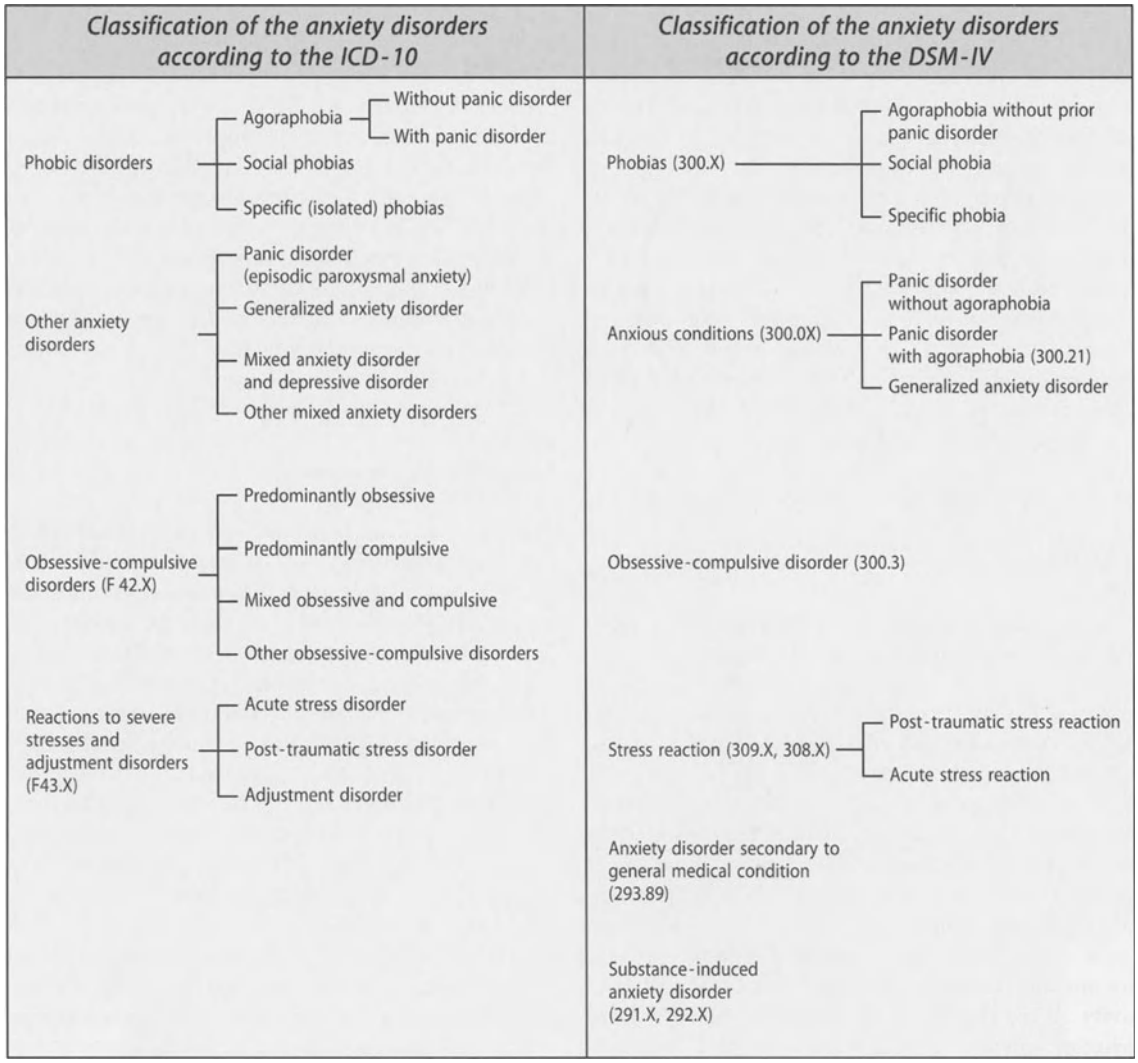


Fig. 1. Classification of the anxiety disorders according to ICD-10 and DSM-IV

this ordering may be traced back to the works of Klein (1964, 1981).

The major feature of agoraphobia is the avoidance of situations or places that are experienced as unpleasant or dangerous and that cause the patient to fear that an anxiety attack or panic-like symptoms will occur. Patients imagine that it is difficult or unpleasant to escape from these situations or places, or that they are helpless. As a rule, anxiety appears in circumscribed situations such as in crowds, in public places, far from home, etc. and these particular situations are avoided. The radius of activity of the affected individuals is thereby greatly constricted.

The term "agoraphobia" ("fear of the marketplace") refers only to the most prominent symptom of this disorder, but patients suffering from it have, as a rule, multiple other anxieties, such as acrophobia, claustrophobia, or thanatophobia. Thus we may speak of agoraphobia as a complex phobia, in contradistinction to the specific phobias. According to DSM-IV, the diagnosis of agoraphobia without panic syndrome may be assigned if the criteria for a panic disorder have never been met, and the diagnosis of a panic disorder with agoraphobia may be assigned if the criteria for both agoraphobia and panic disorder are or have been met at some time.

2.2 Social Phobia

The distinguishing feature of social phobia is an excessive fear of other people and a consequent avoidance of social situations, especially when the

affected individuals think they will be observed or evaluated. They fear that they will fail, appear ridiculous, or be humiliated by their maladroit behavior. Social anxieties may be limited to specific situations (e.g., fear of public speaking) or may be nonspecific, even to the extent of affecting the majority of interpersonal activities. Typically, almost any confrontation with, or anticipation of, a social situation immediately induces fear, which may take the form of a situational panic attack. Social phobia can be understood as a disturbance of patients' positioning of themselves in social groups and rank orderings ("pecking orders"), in which they are always relegated to the omega position - with all of the resulting problems in social adjustment.

2.3 Specific Phobias

The distinguishing feature of the specific (also called isolated or simple) phobias is an inappropriate and excessive fear and avoidance of specific objects or situations. The most common phobias involve animals (spiders, snakes, mice, birds), heights, narrow spaces, airplanes, and the sight of blood or syringes. Almost any confrontation with these situations or objects immediately induces anxiety, which may rise to the severity of a panic attack. The resulting impairment depends on the ease with which the affected person can avoid the phobic situation or object; it may range from a mild impairment to a major limitation of life activities and constriction of the patient's radius of activity. It is typical of the objects and situations provoking specific phobias that they are unconditioned stimuli to which nearly all people spontaneously respond with fear or avoidance, even in the absence of prior learning or experience.

2.4 Panic Disorder

Panic disorder consists of frequently recurring, usually very intense attacks of anxiety, which the affected person does not spontaneously attribute to any particular provoking factor. These suddenly appearing states are characterized by multiple physical and mental symptoms, including palpitations, chest pain, a sensation of asphyxiation, vertigo, and depersonalization or derealization. They generally last only a few minutes, occasionally longer. Patients subsequently develop anxiety about the anxiety or a fear of the consequences of the symptoms connected with anxiety, e.g., fear of dying, losing control, or becoming insane. These symptoms are extremely stressful for the patients and usually produce significant impairment in everyday life.

The fact that the patients cannot state spontaneouslyt why the anxiety attacks occur has given rise to the assumption that they are endogenous. If the diagnostician realizes, however, that thoughts, recollections, or internal images, i.e., cognitions, can induce anxiety reactions just as strongly as, and sometimes even more strongly than, external stimuli, then a correspondingly thorough evaluation usually reveals a provoking factor for the symptoms of anxiety in the pure form of panic disorder, just as in agoraphobia.

2.5 Generalized Anxiety Disorder

The distinguishing feature of generalized anxiety disorder is the presence of excessive, general, and multifarious worries, fears, or anxieties. On most days, over a period of at least 6 months, the patient suffers from excessive worry. These worries, fears, and anxieties, which are not restricted to specific situations or, if accentuated in specific situations, are nonetheless also present at other times, are experienced by the patient as difficult to control and lead to significant impairment in everyday life. In addition, at least three of the following six symptoms must be simultaneously present for the diagnosis of a generalized anxiety disorder to be assigned: muscle tension, restlessness, easy fatigability, difficulty concentrating, sleep disturbance, and irritability. The patients typically think of themselves not as anxious, but as under excessive stress because of the many problematical situations to which they imagine they are exposed. A patient might think that he or she is overburdened by problems with a child, for example, without realizing that the actual source of worry is "self-worry."

2.6 Obsessive–Compulsive Disorder

In obsessive–compulsive disorder, there are either recurrent obsessive thoughts or recurrent compulsive behaviors. Obsessive thoughts (obsessions) are continually recurring ideas, thoughts, or conceptions that patients recognize as their own but nonetheless experience as importunate, tormenting, and burdensome, either because they consider them nonsensical, or because they are of a violent or obscene nature. The patient tries to ignore the obsessive thoughts or to neutralize them through other thoughts or through the performance of rituals.

Compulsive behaviors (compulsions) are recurring, intentional behaviors carried out according to precise

rules or in stereotypic fashion, usually motivated by the desire to avert unpleasant consequences or disasters. The connection between the compulsive behavior and its purpose may be quite far-fetched, e.g., frequent hand-washing to prevent the outbreak of an epidemic. Compulsions, like obsessions, are generally experienced by the affected persons as senseless, ineffectual, and unpleasant. The attempt to suppress them induces severe anxiety or revulsion. Obsessions and compulsions take up a great deal of time. The diagnosis requires the demonstrable presence of obsessions or compulsions that disturb the patient's normal activities on most days for a period of at least 2 weeks. A thorough discussion of obsessive-compulsive disorders is found in Chap. 1 (Vol. 3, Part 2).

2.7 Post-traumatic Stress Disorder

Post-traumatic stress disorder arises as a protracted reaction to an extraordinarily threatening or catastrophic event, such as a natural disaster, wartime experience, serious accident, or violent crime. Typical symptoms of post-traumatic stress disorder include an intense fear and avoidance of activities and situations related to the trauma and, most prominently, the frequent, intense, and unavoidable reliving of the trauma in importunate recollections (nightmares, daydreams, "flashbacks"); emotional insensitivity, blunting, and indifference with respect to the present environment and other persons; and a simultaneous state of vegetative hyperexcitation with increased vigilance, sleeplessness, etc. The symptoms must produce clinically significant stress or impairment. Acute and chronic subtypes are distinguished. A thorough discussion of post-traumatic stress disorder is found in Chap. 4 (Vol. 3, Part 2).

3 Epidemiology

3.1 General Prevalence Rates

Anxiety disorders are among the more common mental illnesses. Table 1 contains a summary of the findings of several epidemiologic studies in different countries. Approximately 5% of the general population suffers from an anxiety disorder. This order of magnitude is roughly comparable to that of the prevalence of depressive illnesses.

The exact figures obtained for prevalence are, however, highly dependent on the detailed criteria for the diagnosis of anxiety disorder applied in such studies. Thus the studies conducted by Bebbington et al. (1981) and Henderson et al. (1979) mentioned in Table 1 were carried out using the Present State Examination (PSE; Wing et al. 1974), while that by Regier et al. (1988) employed the Diagnostic Interview Schedule (DIS; Robins et al. 1981). The latter American study counted simple phobias among the anxiety disorders and thus arrived at a value for prevalence considerably above the mean of those found in other studies (7.3% vs. 3.4%).

As for the prevalence of individual subtypes of the anxiety disorders (Table 2), the different international studies all show that simple phobia is the most common subtype, with a lifetime prevalence of 5.4%–11.3%. Panic disorders are rather uncommon: their lifetime prevalence in the population lies between 1.1% and 3.5%. Obsessive-compulsive disorders occur with a lifetime prevalence of 2%–3%, and generalized anxiety disorders with a 1-year prevalence of 2.3%. There is no reliable estimate to date of the prevalence of post-traumatic stress disorder. Helzer et al. (1987) gave the lifetime prevalence of post-traumatic stress

Table 1. One-month prevalence (%) of anxiety disorders compared to overall psychiatric morbidity and depressive disorders

Study	Sex	Overall psychiatric morbidity	Depressive disorders	Anxiety disorders
USA (ECA, Regier et al. 1988)	Both sexes	11.2[a]	5.1[b]	7.3
	Men	7.6[a]	3.5[b]	4.7
	Women	14.5[a]	6.6[b]	9.7
London (Bebbington et al. 1981)	Both sexes	10.9	7.0	2.9
	Men	6.1	4.8	1.0
	Women	14.9	9.0	4.5
Australia (Henderson et al. 1979)	Both sexes	9.0	–	–
	Men	7.0	2.6	4.1
	Women	11.0	6.7	3.0

ECA, Epidemiologic Catchment Area Study.
[a]Addictions and dementing illnesses excluded.
[b]Overall spectrum of affective disorders.

Table 2. Prevalence (%) of anxiety disorders according to DSM-III and DSM-III-R

Study	Interval for prevalence	Anxiety disorders	Subtypes					
			Panic disorder	Agoraphobia	Simple phobia	Social phobia	Generalized anxiety disorder	Obsessive-compulsive disorder
Edmonton, Canada (Bland et al. 1988a,b)	Point	4.7	0.4	–	3.6[a]	–	–	1.2
	6 months	6.5	0.7	–	5.1[a]	–	–	1.6
	1 year	7.6	0.7	–	6.2[a]	–	–	1.8
	Lifetime	11.2	1.2	2.9	7.2	1.7	–	3.0
Seoul, Korea (Lee et al. 1990)	Point	–	–	–	–	–	–	–
	6 months	–	–	–	–	–	–	–
	1 year	–	–	–	–	–	–	–
	Lifetime	9.2	1.1	2.0	5.4	0.5	3.6	2.3
NCS, USA (Kessler et al. 1994)	Point	–	–	–	–	–	1.6[b]	–
	6 months	–	–	–	–	–	–	–
	1 year	17.2	2.3	2.8	8.8	7.9	3.1	–
	Lifetime	24.9	3.5	5.3	11.3	13.3	5.1	–
ECA, USA (Regier et al. 1988)	Point	7.3	0.6	–	6.2[a]	–	–	1.3
	6 months	8.9	0.8	–	7.7[a]	–	–	1.5
	1 year	–	–	–	–	–	2.3[c]	–
	Lifetime	14.6	1.6	–	12.5[a]	–	–	2.5
Munich, Germany (Wittchen et al. 1992)	Point	–	–	–	–	–	–	–
	6 months	8.1	1.1	3.6	4.1[d]	–	–	1.8
	1 year	–	–	–	–	–	–	–
	Lifetime	13.9	2.4	5.7	8.0[d]	–	–	2.0

ECA, Epidemiologic Catchment Area Study; NCS, National Comorbidity Survey.
[a]Including Social phobias and Agoraphobia.
[b]Wittchen et al. (1994).
[c]Blazer et al. (1987).
[d]Including Social phobias.

disorder as 1% in one of the five study areas of a large-scale American epidemiologic study. The rate was 0.5% in men and 1.3% in women.

As can be seen in Table 2, the point prevalence of the overall category of anxiety disorders, as well as that of its individual subtypes, is generally approximately half of the corresponding lifetime prevalence.

3.2 Comorbidity

It is already clear from a comparison of the prevalences of individual subtypes of the anxiety disorders with that of the overall category of such disorders that many patients suffer from more than one anxiety disorder (see Table 2). In the Munich Follow-Up Study (Wittchen 1991), only 20% of panic disorders and 48% of phobias were pure disorders for life; all other cases met the criteria for at least one other anxiety disorder, and often more than one. According to the findings of the National Comorbidity Survey (NCS; Wittchen and Vossen 1995), approximately 20% of patients with agoraphobia simultaneously meet all of the criteria for a generalized anxiety disorder, and approximately 22% meet all of the criteria for a panic disorder.

A comparable degree of overlap is found with respect to other psychiatric illnesses, including, in particular, depressive illnesses. A total of 21% of patients with anxiety disorders simultaneously meet all of the criteria for a depressive disorder, and 33% of patients with depressive disorders simultaneously have anxiety disorders (Regier et al. 1990).

Clinical findings have long implied a close relationship between anxiety disorders and personality disorders. A study by Sanderson et al. (1993) of 347 patients diagnosed as having an anxiety disorder according to the DSM-III-R criteria revealed that 35% met the diagnostic criteria for at least one personality disorder. A total of 61% of patients with social phobia and 49% of patients with a generalized anxiety disorder had at least one personality disorder, while patients with specific phobias had comorbid personality disorders much less frequently (12% of patients).

These high rates of comorbidity raise the fundamental question as to whether anxiety disorders should indeed be considered independent illnesses. The problem is one of content, as well as of method. In methodological terms, it must be determined to what extent the reported results are an artifact brought about by the admittedly broad symptom profile of the anxiety disorders, many components of which are simultaneously used as criteria for the diagnosis of other disorders. As a matter of content, it must be determined whether distinguishable core symptoms or syndromes occur simultaneously, a state of affairs known in the literature as syndromal association or "co-occurrence," or whether an actual comorbidity is present in the proper sense of the term, i.e., the simultaneous occurrence of different diseases (Wittchen 1993). Finally, syndrome changes over the course of anxiety disorders must also be taken into account (Wittchen et al. 1989).

Empirical research at present offers no more than an attempt at clarifying these issues. Nonetheless, it already seems clear that the use of a highly differentiated diagnostic scheme, and the determination of spectra of comorbidity, are more helpful for an understanding of the course and treatment of these disorders than subsuming their heterogeneity under hierarchically ordered categories (Bronisch 1995).

3.3 Incidence Rates

The initial appearance of anxiety disorders has a characteristic distribution over the human life span. According to the findings of the Munich Follow-up Study (Wittchen 1991), the specific phobias largely arise in childhood and adolescence. Examples include fear of the dark, separation anxiety, and fear of animals. Social phobias are generally said to appear in adolescence (Van Ameringen et al. 1991). Panic disorders, agoraphobia, generalized anxiety disorders, and obsessive–compulsive disorders usually appear in early adulthood; the age-related incidence of generalized anxiety disorder also has a second peak between the ages of 30 and 35 (Scheibe and Albus 1992). In general, these data indicate that early adulthood is a critical time of life for the generation of these disorders.

3.4 Differential Prevalence Rates

As seen in Table 1, anxiety disorders are nearly twice as common in women as in men. This is largely due to differences in the prevalence of phobias and generalized anxiety disorders; differences in the prevalence of the obsessive–compulsive disorders, for example, are much less marked. The prevalence of anxiety disorders is higher in women than in men in all age-groups, as shown in Table 3.

A methodological limitation necessarily applies to the question of the frequency of anxiety disorders in different age-groups: differences in prevalence are not necessarily the result of differences in age and may, instead, reflect cohort effects. Thus older individuals who lived through one or both world wars may be affected by anxiety disorders to a different extent than younger people who were spared such experiences. With these reservations in mind, it may be concluded

Table 3. One-month prevalence of anxiety disorders in different age-groups (after Regier et al. 1988)

Age (years)	Both sexes (%)	Men (%)	Women (%)
18–24	7.7	4.9	10.4
25–44	8.3	4.7	11.7
45–64	6.6	5.1	8.0
6+	5.5	3.6	6.8
All age-groups	7.3	4.7	9.7

from the data shown in Table 3 that anxiety disorders are less common with advancing age and that young adults are affected by them particularly frequently. This conclusion is in accordance with the findings on age-related incidence presented above. Anxiety disorders according to the DSM-III-R criteria are found in only 1.9% of those above age 70 (Helmchen et al. 1996a,b).

A comparison of the frequency of anxiety disorders across social classes yields the interesting finding that they become much more common with lower social status and level of education (Kessler et al. 1994). If we assume that anxiety disorders are at least partly learned disorders, whose appearance and course may be determined by life experiences, by the so-called coping repertoire, or by framing conditions such as alcohol abuse, then this finding opens the way to diverse etiological speculations. Epidemiologic studies alone, however, cannot confirm or refute such hypotheses; to do so, more specific scientific study is required.

With respect to the frequency of anxiety disorders and relationship status, a consistent finding is that they are more prevalent by approximately one third among separated, divorced, and widowed individuals than among married people, people who have never been married, and people living alone (Regier et al. 1990). The question arises as to what extent traumatic experiences in the context of the separation can account for the elevated prevalence rates. Once again, it must be stated that epidemiologic data alone can be used only to generate such hypotheses, not to confirm them.

If we consider the foregoing findings from the perspective of risk estimation, it may be concluded that women between the ages of 15 and 24, of low social and educational status, who are separated, divorced, or widowed, have a significantly elevated risk of developing an anxiety disorder.

3.5 Familial Prevalence

Studies on the frequency of illness in first-degree relatives in comparison to unrelated control subjects suggest that genetic vulnerability is a risk factor for the occurrence of anxiety disorders. When interpreting such data, it must, of course, be borne in mind that family-specific child-rearing styles and life experiences, or the presence of illness in another family member as an environmental stress factor, may favor the familial appearance of certain illnesses.

An overall consideration of the findings of studies on the frequency of panic disorders in first-degree relatives of affected individuals reveals that relatives are approximately four times more likely to be affected than unrelated control subjects (Noyes et al. 1986). First-degree relatives of individuals with anxiety disorders, in comparison with relatives of patients with affective disorders, are approximately seven times as likely to suffer from a panic disorder and twice as likely to suffer from a generalized anxiety disorder (Skre et al. 1994).

Further evidence for a genetic vulnerability factor is derived from studies on mono- and dizygotic twins (Torgersen 1983; Kendler et al. 1992): the concordance rate for monozygotic twins is approximately 25%–30%, twice as high as that for dizygotic twins. These findings imply the existence of a genetically determined constitutional vulnerability for both panic disorders and generalized anxiety disorders.

3.6 Course

Anxiety disorders tend to take a chronic course. According to the findings of the Munich Follow-up Study (Wittchen 1991), a chronic course is found in just over 50% of patients, regardless of the type of the disorder. If we count episodic and less marked cases of illness, then agoraphobia and obsessive-compulsive disorders are found to have an overall chronification rate of approximately 80%. An asymptomatic state is reached, at least temporarily, at some point in the course of 30%–40% of simple phobias and panic disorders. Yonkers et al. (1996) prospectively studied the course of generalized anxiety disorder over an interval of 2 years. The frequency of remission was found to be 15% at 1 year and 25% at 2 years; the presence of a comorbid psychiatric illness reduced the frequency of remission at 1 year by nearly half.

These high rates of chronic disease imply that the anxiety disorders must be considered serious mental illnesses. The same conclusion is reached if we consider the severity of the disorder in the synchronic perspective and the disturbance of social adaptation that it produces. A follow-up study with an interval of 7 years (Wittchen 1991) revealed that a major impairment of social adaptation and a degree of disease-

related disability corresponding to a score of less than 80 on the Global Assessment Scale (GAS; Spitzer et al. 1978) were present in 71% of patients with panic disorders, 50% of those with agoraphobia, 47% of those with simple phobias and social phobia, and 58% of those with obsessive-compulsive disorders.

Data from the American Epidemiologic Catchment Area Study (ECA) reveal the impact of anxiety disorders on patients' ability to function in an occupation, a component of the social cost of these disorders. A total of 60% of affected men and 68% of affected women stated that they were unemployed at the time of the study; 25% of the men and 29% of the women with panic disorders had been unemployed for at least 5 years (Leon et al. 1995). A study by Massion et al. (1993) of the quality of life of patients with generalized anxiety disorder and/or panic disorder revealed that 9% of patients with one or both of these disorders had attempted suicide in the past, 31% had been hospitalized at least once, and 25% of patients with generalized anxiety disorders were receiving disability payments. Chronification of anxiety disorders thus implies a chronic impairment in everyday life, including occupational activity.

3.7 Institutional Prevalence Rates

It is typical of the anxiety disorders that somatic manifestations are often their most prominent feature. We are thus led to ask not just how frequently these disorders come under general medical, as opposed to specifically psychotherapeutic, treatment, but also how frequently the anxiety disorders are in fact recognized as such.

An investigation of the prevalence of mental illness in patients treated in general medical practice, which was carried out as part of an international study by the World Health Organization (Linden et al. 1996), revealed that generalized anxiety disorders are the type of mental illness most likely to be encountered. A total of 9% of general medical patients in Berlin and 7.9% in Mainz suffer from a generalized anxiety disorder. Panic disorders are comparatively uncommon: 0.9% and 1.7% in Berlin and Mainz, respectively. Approximately 60% of all psychiatric illnesses were recognized by the physicians as such; anxiety disorders were less commonly recognized as such in primary care practice than were depressive illnesses (Ormel et al. 1990).

Much more than in general medical practice, anxiety disorders are frequently encountered in psychotherapeutic practice, and particularly in the practice of behavioral therapy. A study of patients treated by behavioral therapists in the out-of-hospital office setting, in the framework of care provided by the German quasi-governmental health insurance scheme (*Krankenkassen*; Linden et al. 1993), revealed that 40% of such patients suffered from an anxiety disorder. Approximately half of these patients had phobias, and approximately 10% obsessive-compulsive disorders. Anxiety disorders are thus second in frequency only to depressive syndromes among patients treated using behavioral therapy. This finding suggests that patients with anxiety disorders are being referred appropriately, as behavioral therapy is currently considered the method of choice for the treatment of these disorders (Michelson and Marchione 1991). It is not clear, however, whether the quantity of treatment currently available and the quantity actually delivered actually satisfy the need for treatment in this patient group.

4 References

APA (1980) Diagnostic and statistical manual of mental disorders, 3rd edn. American Psychiatric Association, Washington DC

APA (1987) Diagnostic and statistical manual of mental disorders, 3rd edn rev. American Psychiatric Association, Washington DC

APA (1994) Diagnostic and statistical manual of mental disorders, 4th edn. American Psychiatric Association, Washington DC

Bebbington P, Hurry J, Tennant C, Sturt E, Wing JK (1981) Epidemiology of mental disorders in Camberwell. Psychol Med 11: 561–579

Bland RC, Newman SC, Orn H (1988a) Period prevalance of psychiatric disorders in Edmonton. Acta Psychiatr Scand 77 [Suppl 338]: 33–42

Bland RC, Orn H, Newman SC (1988b) Lifetime prevalence of psychiatric disorders in Edmonton. Acta Psychiatr Scand 77 [Suppl 338]: 24–32

Blazer D, Hughes D, George LK (1987) Stressful life events and the onset of generalized anxiety syndrome. Am J Psychiatry 144: 1178–1183

Bronisch T (1995) Komorbidität von Angsterkrankungen. In: Kasper S, Möller HJ (eds) Angst- und Panik-Erkrankungen. Fischer, Jena, pp 99–108

Freud S (1895) Über die Berechtigung, von der Neurasthenie einen bestimmten Symptomencomplex als "Angstneurose" abzutrennen. Neurol Zentralblatt 14: 50–66

Gurney CM, Roth M, Garside RF, Kerb TA, Shapira K (1972) Studies in the classification of affective disorders. II. The relationship between anxiety states and depressive illnesses. Br J Psychiatry 121: 162–166

Helmchen H, Linden M (1986) Die Differenzierung von Angst und Depression. Springer, Berlin Heidelberg New York

Helmchen H, Baltes MM, Geiselmann B et al (1996a) Psychische Erkrankungen im Alter. In: Mayer KU, Baltes PB (eds) Die Berliner Altersstudie. Akademie Verlag, Berlin, pp 185–220

Helmchen H, Linden M, Wernicke T (1996b) Psychiatrische Morbidität bei Hochbetagten. Nervenarzt 67: 739–750

Helzer JE, Robins JN, McEvoy L (1987) Posttraumatic stress disorder in the general population. N Engl J Med 317: 1630–1634

Henderson S, Duncan-Jones P, Byrne DG, Scott R, Adcock S (1979) Psychiatric disorders in Canberra: a standardised study of prevalence. Acta Psychiatr Scand 60: 355–374

Jaspers K (1948) Allgemeine Psychopathologie, 5th edn. Springer, Berlin Göttingen Heidelberg

Kendler KS, Neale MC, Kessler RC, Heath AC, Eaves LJ (1992) Generalized anxiety disorder in women. A population-based twin study. Arch Gen Psychiatry 49: 267–272

Kessler RC, McGonagle KA, Zhao S et al (1994) Lifetime and 12-month prevalence of DSM-III-R psychiatric disorders in the United States. Arch Gen Psychiatry 51: 8–19

Klein DF (1964) Delineation of two drug-responsive anxiety syndromes. Psychopharmacologia 5: 397–408

Klein DF (1981) Anxiety reconceptualized. In: Klein DF, Rabkin J (eds) Anxiety: new research and changing concepts. Raven, New York, pp 235–264

Lee CK, Kwak, YS, Yamamoto J et al (1990) Psychiatric Epidemiology in Korea. I. Gender and age differences in Seoul. J Nerv Ment Dis 178: 242–246

Leon AC, Portera L, Weissman MM (1995) The social costs of anxiety disorders. Br J Psychiatry 166: 19–22

Linden M, Maier W, Achberger M, Herr R, Helmchen H, Benkert O (1996). Psychische Erkrankungen und ihre Behandlung in Allgemeinarztpraxen in Deutschland. Ergebnisse aus einer Studie der Weltgesundheitsorganisation (WHO). Nervenarzt 67: 205–215

Linden M, Förster R, Oel M, Schlötelborg R (1993) Verhaltenstherapie in der kassenärztlichen Versorgung. Eine versorgungsepidemiologische Untersuchung. Verhaltenstherapie 3: 101–111

Massion AO, Warshaw MG, Keller MB (1993) Quality of life and psychiatric morbidity in panic disorder and generalized anxiety disorder. Am J Psychiatry 150: 600–607

McReynolds P(1985) Changing conceptions of anxiety: a historical review and a proposed integration. Issues Ment Health Nursing 7:131-158

Michelson LK, Marchione K (1991) Behavioral, cognitive, and pharmacological treatments of panic disorder with agoraphobia: critique and Synthesis. J Consult Clin Psychol 59: 100–114

Mombour W, Kockott G (1971) Diagnosenschlüssel und Glossar psychiatrische Krankheiten. Springer, Berlin Heidelberg New York

Noyes R, Crowe RR, Harris EL, Hamra BJ, McChesney CM, Chaudry DR (1986) Relationship between panic disorder and agoraphobia. A family study. Arch Gen Psychiatry 43: 227–232

Ormel J, Van den Brink W, Koeter MWJ, Giel R, Van der Meer K, Van de Willige G, Wilmink FW (1990) Recognition, management and outcome of psychological disorders in primary care: a naturalistic follow-up. Psychol Med 20: 909–923

Rapee RM (1991) Generalized anxiety disorder: a review of clinical features and theoretical concepts. Clin Psychol Rev 11: 419–440

Regier DA, Boyd JH, Burke JD et al (1988) One-month prevalence of mental disorders in the United States: based on five epidemiologic catchment area sites. Arch Gen Psychiatry 45: 977–986

Regier DA, Narrow WE, Rae DS (1990) The epidemiology of anxiety disorders: the epidemiologic catchment area (ECA) experience. J Psychiatr Res 24: 3–14

Robins LN, Helzer JE, Croughan J, Williams JBW, Spitzer RL (1981) NIMH Diagnostic Interview Schedule: version III (May 1981). National Institute of Mental Health, Rockville MD

Sanderson WS, Wetzler S, Beck AT, Betz F (1993) Prevalence of personality disorders among patients with anxiety disorders. Psychiatry Res 51: 167–174

Scheibe G, Albus M (1992) Age at onset, precipitating events, sex distribution, and co-occurrence of anxiety disorders. Psychopathology 25: 11–18

Skre I, Onstad S, Edvardsen J, Torgersen S, Kringlen E (1994) A family study of anxiety disorders: familial transmission and relationship to mood disorder and psychoactive substance use disorder. Acta Psychiatr Scand 90: 366–374

Spitzer RL, Endicott J, Fleiss L (1978) The global assessment scale: a procedure for measuring overall severity of psychiatric disturbance. Arch Gen Psychiatry 33: 766–771

Torgersen S (1983) Genetic factors in anxiety disorders. Arch Gen Psychiatry 40: 1085–1089

Van Ameringen M, Mancini C, Styan G, Donison D (1991) Relationship of social phobia with other psychiatric illness. J Affect Dis 21: 93–99

Watson JB, Rayner R (1920) Conditioned emotional responses. J Exp Psychol 3: 1–14

Westphal C (1871) Die Agoraphobie. Eine neuropathische Erscheinung. Archiv Psychiatr Nervenheilkd 3: 138–161

Wing JK, Cooper JE, Sartorius N (1974) The measurement and classification of psychiatric symptoms. Cambridge University Press, London

Wittchen HU (1991) Der Langzeitverlauf unbehandelter Angststörungen: Wie häufig sind Spontanremissionen? Verhaltenstherapie 4: 273–282

Wittchen HU (1993) Komorbidität bei Angststörungen. Häufigkeit, ätiologische und klinische Implikationen. In: Kasper S, Möller HJ (eds) Angst- und Panikerkrankungen, Diagnose, Therapie. Socio Medico, Gräfelfing, pp 60–69

Wittchen HU, Vossen A (1995) Implikationen von Komorbidität bei Angststörungen. Ein kritischer Überblick. Verhaltenstherapie 5: 120–133

Wittchen HU, Hand I, Hecht H (1989) Prävalenz, Komorbidität und Schweregrad von Angststörungen. Ergebnisse der Münchner Follow-up Studie (MFS). Z Klin Psychol 18: 117–133

Wittchen HU, Essau CA, Zerssen D von, Krieg JC, Zaudig M (1992) Lifetime and six-month prevalence of mental disorders in the Munich follow-up-study. Eur Arch Psychiatry Clin Neuosci 241: 247–258

Wittchen HU, Zhao S, Kessler RC, Eaton WW (1994) DSM-III-R generalized anxiety disorder in the National Comorbidity Survey. Arch Gen Psychiatry 51: 355–364

WHO (1992) Manual of the international statistical classification of diseases, injuries, and causes of death, 10th edn. World Health Organization, Geneva

Yonkers KA, Warshaw MG, Massion AO, Keller MB (1996) Phenomenology and course of generalised anxiety disorder. Br J Psychiatry 168: 308–313

CHAPTER

3

J.J. López-Ibor, M. Linden, J. Gibert

Pathogenesis and Treatment of Anxiety Disorders

Translator: E. Taub

1 Phenomenology of Anxiety

Anxiety (from the Latin *anxietas,* in turn derived from *angere,* to choke, throttle, trouble, distress) is an emotional state induced by the perception of danger or threat arising from the environment or from within the individual him- or herself. In humans, as in animals, anxiety is necessary for survival, as it protects the individual from repeated contact with threatening situations. It can also be understood as a preparatory reaction for fight or flight. The behavioral inhibition accompanying anxiety also has an important protective function: in both animal and human social conflicts, it brings about the retreat of the aggressor.

Anxiety is seen as one of the so-called primary human emotions (Pekrun 1990). The subjective experience of anxiety comprises the following components (Lòpez-Ibor 1952):

1. An affective component, i.e. an emotional experience
2. A somatosensory component, i.e. perception of physiologic changes
3. A cognitive component, i.e. worry about a threat or impending loss
4. Characteristic behavior that may express itself either as inhibition or as alertness, defense, fight, or flight

Even though certain behavior patterns are characteristic of anxiety (see item 4, above), the subjective experience of anxiety need not be congruent with observable behavior. Thus anxiety research revealed long ago that there may be a considerable discrepancy between observed behavior, physiologic measurements, and subjective reports (Lader and Marks 1971; Fahrenberg 1987; Lang 1993).

The proneness to react in an anxious way is variable among individuals and may thus be thought of as a personality trait. All human beings have a behavioral disposition to experience anxiety at varying intensities for varying lengths of time (Amelang and Bartussek 1985). Anxiousness in this sense (the proneness toward anxiety) is distinct from situation-related fear, which is immediately preceded by a fear-inducing stimulus and which, on the behavioral level, leads either to flight and avoidance or to aggression and fighting, depending on the particular stimulus that induced it (Fröhlich 1983).

For anxiety, unlike other psychopathological phenomena, such as perceptual illusions, a continuous transition exists from "normal" to "pathological" forms (Strian 1983). The criteria for pathological anxiety are as follows:

- Anxiety in the absence of a real threat
- An anxiety reaction that is subjectively or objectively judged to be disproportionate to the inducing stimulus
- Severe anticipatory anxiety
- Fear of anxiety (phobophobia)
- Socially or personally limiting avoidance behavior
- Anxiety that persists even when the threat is no longer present
- Anxiety of such intensity that the individual's ability to cope with life is affected

Anxiety may occur as an disease in itself or as a symptom, or constellation of symptoms, in the context of other mental or organic illnesses.

Some anxiety disorders are listed in ICD-10 (WHO 1992) under chapter headings specifically relating to anxiety (F40, F41, F42), while other anxiety disorders appear under headings such as somatization disorder and hypochondriacal disorders (F45), organic anxiety disorders (F06, F10), and severe personality disorders (F60). Anxiety may also occur as a symptom, or syndrome, accompanying mental disorders of practically any type. Free-floating anxiety is one of the hallmark symptoms of depressive illness; anxiety may also occur in association with delusional mood changes or with delusions of harmful influence or persecution due to schizophrenia.

Physical illness must always be considered in the differential diagnosis of anxious conditions. The following list of differential diagnoses (mental and physical illnesses accompanied by anxiety; after Deister 1995) shows that the initial diagnosis of anxiety disorders always requires a comprehensive medical evaluation:

1. Nonorganic mental disorders:
   - Schizophrenic psychosis
   - Affective psychosis
   - Personality disorders
2. Organic mental disorders:
   - Delirium
   - Organic delusional disorder
   - Organic depressive disorders
   - Organic personality disorders
3. Substance-related disorders:
   - Intoxication with amphetamines, cocaine, hallucinogens, alcohol, nicotine, caffeine
   - Withdrawal of alcohol, opiates, anxiolytics
4. Neurological diseases:
   - Epilepsy
   - Huntington's disease
   - Migraine
   - Multiple sclerosis
   - Elevated intracranial pressure

5. General medical illnesses:
   - Angina pectoris
   - Myocardial infarction
   - Cardiac dysrhythmias
   - Hyperchloremia
   - Hypoxia
   - Pulmonary embolism
   - Hyperthyroidism
   - Carcinoid
   - Pheochromocytoma
   - Anaphylaxis

2 Neurobiology of Anxiety Disorders

2.1 Genetic Aspects

Both animal experiments and human twin studies have yielded evidence that genetic factors play a role in anxiety and anxiety disorders.

A number of rat strains have been bred selectively according to their rate of defecation when at large, and these rats may be classified as either reactive or nonreactive (Weyers and Fritze 1995). It has been shown that reactive rats tolerate stress less well in conflict tests, that chlordiazepoxide has a lesser anxiolytic effect in such rats, and that young reactive rats, when separated from their mothers, vocalize more than nonreactive rats.

There are also significant differences in anxiety reactions among different breeds of dog. Thus two lines of pointers were bred (McKinney 1988). The "nervous" line is characterized by extreme fearfulness, avoidance of human beings, reduced exploratory behavior, a more extreme fright reaction, and an elevated incidence of atrioventricular heart block. The administration of chlordiazepoxide to such dogs enables them to cope with an anxiety-producing lever-pressing task, while the administration of amphetamines or cocaine leads to a markedly more severe behavioral disorder. Nervous pointers were found to have a higher concentration of norepinephrine in the reticular formation, and a lower concentration of serotonin in the septal nuclei, than normal pointers (Gurgius et al. 1990).

Studies of human families have yielded evidence that genetic load affects the risk of developing an anxiety disorder (Linden 1993). Studies have revealed that the first-degree relatives of patients with panic disorders have such disorders themselves with a prevalence of 15%–20%, significantly higher than the rates of 2%–4% found among controls (Crowe et al. 1983; Harris et al. 1983; Noyes et al. 1986). Twin studies (Kendler et al. 1992; Torgersen 1983) have shown a concordance rate of approximately 20%–30% for generalized anxiety disorder among monozygotic twins, approximately three times as high as the rate among dizygotic twins. These findings imply that there is a genetically transmitted, constitutional vulnerability to panic disorders and to generalized anxiety disorder.

2.2 Neural Structures

Anxiety, as discussed above, is a reaction that prepares the organism to respond to a threat in any of several different ways. It is thus not surprising that anxiety activates a number of different areas of the brain simultaneously. The limbic system plays an important role in this process; occupying an anatomical position between the neocortex, the hypothalamic-pituitary axis, the reticular formation, and the sensory system, the limbic system is strategically placed to influence a large number of different cerebral functions and modes of behavior. It has important connections both to lower brain centers, such as the locus ceruleus, and to higher brain centers, such as the frontal cortex. Brain stimulation experiments have revealed that stimulation of the amygdala, in particular, evokes fear (Kuhar 1986). In animal experiments, electrical stimulation of the amygdala leads to intense anxiety reactions involving increased fearfulness, a change in pulse and respiratory rate, and the interruption of behavior currently being performed. If lateral, basal, or central amygdaloid nuclei are destroyed, the symptoms of fear and anxiety disappear (Davis 1992).

It is likely that the septohippocampal system plays a role in anxiety, because destruction of these areas has been found to have a sedative and anxiolytic effect which is similar to that of the administration of benzodiazepines or barbiturates.

Studies employing positron emission tomography (PET) in patients with panic disorders before and during an infusion of lactate have revealed an abnormal asymmetry of perfusion in the parahippocampal gyrus; during panic attacks, these patients have a significant increase of cerebral blood flow, particularly in the temporal lobes bilaterally (Reiman et al. 1984). This activation of the temporal lobe was also observed in normal subjects trained to expect a painful electric shock (Reiman et al. 1989). The limbic and septohippocampal systems may be activated by lower brain areas, such as the reticular formation and especially the locus ceruleus (Gray 1982).

2.3 Transmitter Systems

The benzodiazepines are, the substance class with the greatest anxiolytic effect. Their mechanism of action involves the so-called benzodiazepine γ-aminobutyric acid (GABA) receptor (see Chap. 6, Vol. 1). It is now known that there are several different types of $GABA_A$ receptor that are differentially distributed in different brain areas; some types of $GABA_A$ receptor may be more closely related to anxiety than others (Sieghart 1995). Changes in $GABA_A$ receptor activity can both induce fear and counteract it. Any increase of GABAergic transmission, e.g. after the administration of benzodiazepines, barbiturates, anxiolytic steroids, or alcohol, has an anxiolytic effect, while any decrease of GABAergic transmission, e.g. after receptor blockade with substances such as pentylenetetrazol and picrotoxin, is anxiogenic. These effects occur immediately, without delay, after the administration of the corresponding substances. It therefore has to be concluded that $GABA_A$ receptors are directly involved in the modulation of the anxiety reaction.

It is conceivable that differences among individuals in the proneness to anxiety might be explained by differences in $GABA_A$ receptor affinity. It is also possible that anxiety disorders lead to an overproduction of endogenous reverse agonists, or to a change of receptor density, and thereby to prolonged states of anxiety and, finally, to neural structural changes (Sieghart 1995). A number of endogenous ligands for the benzodiazepine binding site of the $GABA_A$ receptor have been discovered to date, and endogenous agonists are also known.

A number of experimental findings reveal that fear and anxiety can be induced by electrical stimulation of the primate brain by way of a mechanism involving overfunction of the noradrenergic system, in particular of the locus ceruleus, whose predominantly adrenergic neurons make synaptic connections throughout the brain (Gray 1982). If the locus ceruleus is destroyed, such reactions become less pronounced. It has also been shown that benzodiazepines reduce the activity of locus ceruleus neurons. Furthermore, pharmacologic studies have shown that the administration of α_2-adrenergic antagonists such as yohimbine can induce anxiety, both in animals and in humans (Charney et al. 1984). These α_2-adrenergic receptors are typically localized presynaptically on noradrenergic cell bodies and serve to regulate the release of norepinephrine.

It is known that benzodiazepines also inhibit the electrical activity of serotonergic neurons in the dorsal raphe nuclei and reduce the functioning of the serotonin system in general. Conversely, serotonin agonists can have an anxiogenic effect in patients with panic disorders (Sieghart 1995). It seems, however, that long-term stimulation of the serotonin system has different physiologic effects than acute serotonergic activation, which causes acute excitement or anxiety. Long-term pharmacologic serotonergic stimulation, either with serotonin reuptake inhibitors or with $5\text{-}HT_{1a}$ agonists such as buspirone, has a clearly anxiolytic clinical effect (Rickels et al. 1982b). This may be explained as the result of an activation of presynaptic autoreceptors, which would then, in turn, lead to a reduction of serotonin release.

2.4 Neuroendocrinology

It is generally recognized that anxiety is associated with a large number of vegetative manifestations, such as alterations of skin conductivity, of blood pressure, and of heart rate. These effects may be explained as the result of increased release of epinephrine, norepinephrine, adrenocorticotropic hormone (ACTH), cortisol, prolactin, and growth hormone (Hemmeter and Holsboer-Trachsler 1995; see also Chap. 8, Vol. 1, Part 1).

These vegetative phenomena are mainly short-term, nonspecific adaptive changes, which may be understood as reactions preparing the organism for fight or flight. Long-term adaptive biological changes, as seen in the anxiety disorders, take place mainly along the hypothalamic-pituitary axis. It has been known since the studies carried out by Selye (1956) that anxiety and stress are associated with cortisol secretion. Many experimental studies have documented a positive relationship between anxiety reactions and cortisol release in normal individuals (Nesse et al. 1985; Schedlowski et al. 1992); furthermore, in similar studies, subjects with an elevated proneness to anxiety as a personality trait respond with a greater rise of the serum cortisol level than others (Hemmeter et al. 1991). Those who have lived in anxiety-producing circumstances for prolonged periods have elevated salivary cortisol levels, even after the acute threat is removed (Rahe et al. 1990).

Counterregulatory adaptation also appears to occur in chronic anxiety disorders. Thus some patients with post-traumatic stress disorder (PTSD), when given dexamethasone, actually have lower cortisol levels than normal control subjects (Halbreich et al. 1989; Yehuda et al. 1993). Holsboer (1993) hypothesizes that an increased endogenous secretion of corticotropin-releasing hormone (CRH) during the anxious state leads

to secondary hyposensitivity of CRH receptors in the pituitary gland.

Not only adrenal cortical function, but also the regulation of growth hormone secretion depends on the function of the hypothalamic-pituitary axis. In individuals exposed to anxiety-producing situations, an increased secretion of growth hormone can be observed, though the threshold for growth hormone secretion is apparently higher (Greenwood and Landon 1966; Greene et al. 1970). Patients with panic disorders, and those with social phobia, have been found in several studies to have a diminished sensitivity of this system, as shown, e.g. by measurements after clonidine infusion. This is also thought to be a result of hyposensitivity of postsynaptic α-adrenergic receptors (Charney and Heninger 1986; Tancer et al. 1993; Holsboer 1993).

These central and peripheral neurobiological changes associated with anxiety have not yet been shown to be specific enough to allow a distinction between different anxiety, or even between pathological and normal states of anxiety. They have been useful to date mainly as aids in the development of explanatory hypotheses for the phenomenon we call "anxiety."

3 Psychology of Anxiety Disorders

In psychoanalytic theories, it has been proposed that excessive separation anxiety in childhood predisposes to panic attacks in adulthood (Klein 1981), although the former may well be an early manifestation of the latter, or an actual etiologic factor for it. Other authors were unable to provide support for this association (Deltito et al. 1986, Margraf et al. 1986).

With respect to parental child-raising styles, it has been claimed of patients with social phobia, in particular, that they had been held back by their parents in social situations and had been led to overvalue the opinions of others and to undervalue the attitudes of family members (Bruch et al. 1989). Agoraphobia has been said to result from the reactivation of early bonding behavior in patients with excessively solicitous mothers (Solyom et al. 1976; Parker 1979; Faravelly et al. 1991; Willi 1972).

The learning-theoretic view of anxiety disorders describes a number of types of stimulus that unconditionally induce anxiety by way of an innate mechanism. Examples include physically painful stimuli, social threat, heights, and narrow spaces. It can be shown in animal experiments that the coupling of an initially nonaversive stimulus (i.e. an conditioned stimulus, such as a light signal) with a definitely aversive stimulus (i.e. and unconditioned stimulus) leads to a learned anxiety response to the first stimulus, according to classical conditioning (Brady and Hunt 1955). Such conditioning processes are potentiated in the presence of an already high level of anxiety, e.g. after previous aversive stimulation (Davis 1988).

Such conditioning is especially powerful when certain forms of aversive stimuli are used for which there is a genetic predisposition, e.g. with olfactory or gustatory stimuli (Weyers and Fritze 1995; Seligman 1971). In accordance with this learning principle, agoraphobia is thought to start from an initially unconditioned aversive stimulus, which is followed by an unconditioned anxiety reaction (Mowrer 1960). Patients, when questioned, can often state the exact time and place at which the illness began. Because of association with a conditioned stimulus, which thereby obtains signal character for the actual threat (Reinecker 1993), the affected individual learns to experience anxiety even in the absence of the primary stimulus. Just as in animal experiments, such associations are formed more or less rapidly depending on the particular stimulus involved.

In the further course of the disorder, operant conditioning leads to generalization, as the anxiety-inducing signal stimuli lead to avoidance behavior, and the anxiety reactions thus become generalized to other situations rather than simply forgotten. Avoidance, which may be understood as negative reinforcement, accounts for the generalization and further course of phobic disorders.

In addition to these conditioning processes, model learning also plays a role in the generation of anxiety (Bandura and Menlove 1968; Reinecker 1993). With respect to model learning, it should be borne in mind that observation alone does not induce an anxiety reaction; rather, the model must be perceived as self-relevant by additional evaluative processes of the observer.

Social skills are also important in the development of anxiety. It has been shown that social phobics, in particular, suffer from a lack of social skills (Dow et al. 1985). In such cases, however, the affected person's cognitive assessment of his or her own competence, and of the degree of competence demanded in particular situations, plays a larger role than factual competence (Edelmann 1985).

In summary, anxiety reactions are the final common pathway of a multidimensional array of contingencies. These include individual excitability, fearfulness, and conditionability, the nature and intensity of the anxiety-inducing stimulus, its duration, rate of repetition, and repetition interval, the conditioning processes involved, cognitive processes for evaluating the situation and the individual him- or herself, the

individual's own available resources, and the selective direction of attention toward potentially anxiety-inducing stimuli.

These mechanisms are involved in all clinically observed anxiety disorders, though to different degrees in different disorders. As mentioned above, agoraphobia can be explained as a product of classical and operant conditioning processes (Reinecker 1993; Margraf and Schneider 1989, 2000). Anticipatory cognitions are of central pathogenetic importance in the further development of the disorder. In panic disorder, unlike agoraphobia, anxiety states are induced not by external, but rather by internal stimuli, such as memories or associations; as a rule, they can be triggered at any time by the corresponding stimuli (contrary to the original definition) (Margraf and Schneider 2000).

Traumatic, unconditioned anxiety-inducing stimuli play an important etiologic role in PTSD (Foa et al. 2000; cf. also Chap. 4 of this volume). The diagnostic criteria for PTSD require an initial event that led to panic-like anxiety reactions: examples of such events include rape, wartime experiences, and traffic accidents. Later, this unconditioned panic reaction can be reactivated by signal stimuli, at progressively decreasing threshold levels. A major pathogenetic role is played by memories that bear such a strong emotional charge that they come into consciousness over and over again, calling forth a stronger emotional reaction each time.

In generalized anxiety disorder, the most important etiologic and pathogenetic factors are primary, anticipatory, catastrophizing cognitions and, above all, a direction of attention toward potential anxiety-inducing stimuli in the environment (Linden and Zubrägel 2000; Turowsky and Barlow 2000). Patients suffering from social phobia (Juster et al. 2000) have been found to have a lack of social competence as well as a cognitive perception of their own inadequacy and lack of social skills. Social phobia may possibly also result from a genetically based, unconditionally anxiety-inducing stimulus related to social hierarchy or "pecking order."

Inborn unconditioned stimuli of this type are, indeed, the central pathogenetic factor of the monophobias (Öst 2000), such as those relating to spiders, snakes, mice, or birds; what all of these phobias have in common is that rapid rustling movements are followed by an unconditioned fright reaction.

4 Treatment of Anxiety Disorders

4.1 Pharmacotherapy

Nearly all psychotropic drugs have anxiolytic effects (Linden et al. 1988). This is true not only of benzodiazepines and 5-HT_{1a} antagonists, which are classified as anxiolytics, but also of antidepressants, whether sedating (e.g. trimipramine) or nonsedating (serotonin reuptake inhibitors, such as paroxetine, and monoamine oxidase (MAO) inhibitors, such as moclobemide). Sedating neuroleptics of low potency (e.g. thioridazine), as well as highly potent neuroleptics in low doses (e.g. fluspinilene), also have anxiolytic effects. There are also a large number of herbal drugs with anxiolytic effects, both sedating (e.g. baldrian) and nonsedating (e.g. kavain) (Kapfhammer 1995; Wurthmann 1995; Laux 1995; Volz 1995; Kasper 1995). Both the existence of so many different substances that may be used to treat the anxiety disorders and the diversity of the disorders themselves are reflections of the multiplicity of biological and psychological mechanisms participating in anxiety and further indicate that their mechanisms of action are highly nonspecific.

Benzodiazepines, acting through the benzodiazepine GABA receptor, open a chloride channel in the cell membrane and thereby bring about a hyperpolarization of the membrane potential, which makes the cell less responsive to excitatory synaptic input. Benzodiazepines act immediately and are thus highly effective in the acute treatment of anxiety. They are less suitable for the long-term treatment of anxiety disorders, though some authors do regard long-term benzodiazepine administration as an option in special cases (Rickels et al. 1982a,b, 1985, 1993b; Ballenger et al. 1988; Burrows et al. 1993; Curtis et al. 1993). The high efficacy of the benzodiazepines and the immediate onset of their effect give rise to the danger that patients may become psychologically dependent. There are patients with agoraphobia, for example, who will not leave the house unless they have the medication in their pocket.

The antidepressants are currently regarded as the agents of choice for prophylactic long-term treatment of the anxiety disorders. Either tricyclic antidepressants, such as imipramine (Kahn et al. 1986; Rickels et al. 1993a), or serotonin reuptake inhibitors (Sternbach 1990; Schneier et al. 1990; Black et al. 1993) may be used. Data derived mainly from studies of 6 weeks' to 6 months' duration reveal that syndrome scores on anxiety scales decline significantly over the course of treatment. Tricyclic antidepressants, serotonin reuptake inhibitors, and MAO inhibitors are all effective

against panic disorders, agoraphobia, generalized anxiety disorders, and social phobia.

4.2 Psychotherapy

Anxiety disorders have played a major role in psychotherapy ever since Freud (Freud 1926). The basic psychoanalytical model of anxiety postulates that certain situational conditions induce anxiety because their unconscious processing recalls previously repressed traumatic situations to memory, usually events that took place in childhood (Brenner 1982; Rocah 1991). Thus phobic symptoms, or other expressions of anxiety, are to be understood as expressions of latent content, whose analysis can then lead to a resolution of the anxiety neurosis (Lewin 1935). Though quite elaborate theories of the treatment of anxiety have been developed, no controlled studies on the psychoanalytical and psychodynamic treatment of the anxiety disorders are available to date, and the effectiveness of such methods thus remains a matter of speculation.

The current state of the empirical literature suggests that anxiety disorders should be regarded as a primary indication for cognitive behavioral therapy (Margraf 2000). The underlying principle of the therapeutic intervention is the so-called vicious circle of anxiety. In this model, external stimuli, perceptions, cognitive judgments, physiologic and psychological reactions, and behavior are all regarded as components of a single feedback loop. The explanatory power of this approach is relevant not only in the theoretical sense, but also for the patient. The patient, bearing this model in mind, can observe how certain stimuli lead to the evocation of certain thoughts and associations, which, in turn, lead to excitatory reactions, to the experience of anxiety, and, finally, to the corresponding behavioral consequences. As the latter often consist of flight and avoidance behavior, the model further explains why anxiety becomes generalized, rather than subsiding, and grows ever more intense through a process of continued self-reinforcement.

After individual behavioral analysis with reference to the vicious circle of anxiety, the most important component of the next phase of therapy is the so-called reframing process, through which subjective experience is given a new meaning. Physical reactions that may themselves induce anxiety, such as cardiac palpitations, are discussed in detail with the patient, their a priori interpretations are analyzed, and alternative interpretations are suggested. This process results in a reattribution of the patients' personal reactions. Then, in a further step, the patient tries to control anxiety reactions by internal dialogue and other coping strategies. An important part of this form of treatment is the reorientation of the patient away from anxiety-producing stimuli toward his or her own reactions, so that so-called reaction management can take place.

Once this phase of treatment is completed, the next, decisive step is a confrontation with anxiety-inducing situations in the framework of exposure exercises, the purpose of which is to enable the application of the learned coping mechanisms in real-life situations and thus give the patient confidence in his or her own ability to deal with anxiety and to overcome the situations that induce it. It is a common mistake to assume that confrontation alone reduces anxiety; in fact, it is more likely to lead to a reinforced learning of anxiety. Many empirical studies have now shown that behavioral therapy, when performed as recommended, can yield lasting therapeutic success in approximately 80% of agoraphobic patients so treated (Clum 1989; Margraf and Schneider 1989; Chambless and Gillis 1993; Clum et al. 1993; Clark 1994; Emmelkamp et al. 1994; Hollon and Beck 1994).

The treatment of generalized anxiety disorder (Brown et al. 1993; Turowsky and Barlow 2000; Linden and Zubrägel 2000) is primarily focused on the typical "sensitizer orientation" of these patients and the catastrophizing cognitions attached to it. These patients do not realize at first that they have an anxiety disorder; rather, they believe that their perception of multiple dangers in the environment is correct. Thus the first step of treatment is to make the patient understand that he or she suffers from a perceptual stereotype that is not limited to a particular situation, i.e. that any situation is seen only in terms of the disasters it may produce. The patients learn to recognize that many other people deal with everyday situations in a far more relaxed way, and they often come to wish that they, too, could do so. Building on this wish, the therapist works with the patient to change both the perceptual stereotype and the associated expectations of disaster. This is done by means of internal dialogue and exercises to promote the alternative direction of attention or the self-observation of catastrophizing conditions. Finally, exposure exercises are carried out, both in sensu and in real life, in which the patients learn to control thoughts of disaster and the accompanying emotions (e.g. "My daughter is on vacation and hasn't phoned yet today; she might have had an accident").

A number of well-controlled studies are now available in which different forms of intervention are compared with one another with respect to clinical efficacy (Fisher and Durham 1999). Remission rates of approximately 70% were obtained with either of two types of cognitive behavioral therapy specifically focusing on coping with anxiety and on changing

catastrophizing cognitions. Interestingly, behavioral-therapy strategies directed toward external problems, rather than the central pathogenetic process of worrying behavior, as well as psychodynamic interventions aimed at the uncovering of problems from a historical perspective turned out to yield even lower rates of remission than placebo therapy. The experience to date with group therapy has also failed to show an adequate therapeutic effect.

The treatment of social phobia is multidimensional, in accordance with the theoretical models of its underlying pathogenesis (Hoppe and Heimberg 1993; Juster et al. 2000). Its initial emphasis is the analysis of automatic thoughts in social situations. The patient's fantasies of failure, self-reproach, and inadequacy are identified. In the next step, typical cognitive distortions are dealt with, such as all-or-nothing thinking, magnification (i.e. the exaggerated perception of one's own minor mistakes as major disasters), and arbitrary inferences (i.e. the unfounded perception of reference to oneself in any situation, e.g. the patient connects the boss's bad mood to him- or herself, without examining the issue any more closely). In a further step, these dysfunctional cognitions are replaced with alternative ones. Thus, in social phobia, too, confrontation exercises play an important role.

Moreover, possible deficiencies of social competence must be worked on, though the basic problem in most cases is not an actual lack of social skills, but an inhibition in the application of skills that are, in fact, present. Effective treatment, therefore, must go beyond exposure exercises to include exercises aimed at a general improvement of "experienced self-effectiveness." Self-confidence is understood, in this sense, as a self-attribution. The goal of the treatment of social phobia is a change in the patient's self-image.

5 Concluding Remarks

Anxiety disorders are among the common mental illnesses (Achberger and Linden 1998; Linden 1993). The danger is, however, that their effect on patients and on society may be underestimated. One reason is that these disorders are often not particularly conspicuous; in social phobia, for example, patients are even characterized by their high degree of unobtrusiveness. A further reason is that the patients' anxiety may merely evoke astonishment or amusement, because the observer would feel no anxiety himself in such situations (e.g. a subway train) and thus lacks any understanding of the patients' reaction. Yet anxiety disorders can destroy these patients' lives.

Patients with agoraphobia can no longer fulfill their occupational, social, and family responsibilities. They may become completely disabled and bound to their own home. Patients with generalized anxiety disorders are subject to constant stress and worry and are thus in danger of creating becoming a burden to their social partners, which, in turn, results in further negative reactions. Patients with social phobia are unable to pursue a career because they always come out behind in competitive situations.

In view of the high prevalence of these disorders, their severity, and their significance in society as a whole, the development of pharmacologic and psychotherapeutic treatments with comparatively high rates of success must be considered one of the major therapeutic advances of recent years.

6 References

Achberger M, Linden M (1998) Generalized anxiety disorder (GAD) in primary health care. Frequencies, recognition and interventions. Psychiatric Danubia 10: 376–378

Amelang M, Bartussek D (1985) Differentielle Psychologie und Persönlichkeitsforschung. Kohlhammer, Stuttgart

Ballenger JC, Burrows GD, DuPont RL, et al (1988) Alprazolam in panic disorder and agoraphobia: results from a multicenter trial. I. Efficacy in short-term treatment. Arch Gen Psychiatry 45: 413–422

Bandura A, Menlove FL (1968) Factors determining vicarous extinction of a avoidance behavior through symbolic modeling. J Pers Soc Psychol 6: 99

Black DW, Wesner R, Bowers W, Gabel J (1993) A comparison of fluvoxamine, cognitive therapy and placebo in the treatment of panic disorder. Arch Gen Psychiatry 50: 44–50

Brady JV, Hunt HF (1955) An experimental approach to the analysis of emotional behavior. J Psychol 40: 313–324

Brenner C (1982) Elemente des seelischen Konflikts. Fischer, Frankfurt

Brown TA, O'Leary TA, Barlow DH (1993) Generalized anxiety disorder. In: Barlow DH (ed) Clinical handbook of psychological disorders. Guilford, New York, pp 137–188

Bruch MA, Heimberg RG, Berger P, Collins TM (1989) Social phobia and perceptions of early parental and personal characteristics. Anx Res 2: 57–65

Burrows GD, Judd FK, Norman TR (1993) Long-term treatment of panic disorder. J Psychiatr Res 27: 111–125

Chambless DL, Gillis MM (1993) Cognitive therapy of anxiety disorders. J Consult Clin Psychol 61: 248–260

Charney DS, Heninger GR (1986) Abnormal regulation noradrenergic function in panic disorder. Arch Gen Psychiatry 43: 1042–1054

Charney DS, Heninger GR, Breier A (1984) Noradrenergic function and panic anxiety effects of yohimbine in healthy subjects and patients with agoraphobia and panic disorder. Arch Gen Psychiatry 41: 751–763

Clark DM (1994) Cognitive therapy for panic disorder. In: Wolfe BE, Maser JD (eds) Treatment of panic disorder: a consensus

development conference. American Psychiatric Press, Washington
Clum GA (1989) Psychological interventions vs. drugs in the treatment of panic. Behav Ther 20: 429–457
Clum GA, Clum GA, Surls R (1993) A meta-analysis of treatments for panic disorder. J Consult Clin Psychol 61: 317–326
Crowe RR, Noyes R, Pauls DL, Slymen D (1983) A family study of panic disorder. Arch Gen Psychiatr 40: 1065–1069
Curtis GC, Massana J, Udina C et al (1993) Maintenance drug therapy of panic disorder. J Psychiatr Res 27: 127–142
Davis M (1988) The potentiated startle response as a measure of conditioned fear and its relevance to the neurobiology of anxiety. In Simon P, Soubrie P, Wildlocher D (eds) Animal models of psychiatric disorders, vol 1. Selected models of anxiety, depression and psychosis. Karger, Basel, pp 61–89
Davis M (1992) The role of the amygdala in fear-potentiated startle: implications for animal models of anxiety. Trends Pharmacol Sci 13: 35–41
Deister A (1995) Diagnostik und Psychopathalogie der Angsterkrankungen. In: Kasper S, Möller HJ (eds) Angst- und Panikerkrankungen. Fischer, Jena, pp 39–51
Deltito JA, Perugi G, Maremmani J, Mignani V, Cassano GB (1986) The importance of separation anxiety in the differentiation of panic disorder from agoraphobia. Psychiatr Dev 3: 227–236
Dow MG, Bigland A, Glaser SR (1985) Multimethod assessment of socially anxious and socially unanxious women. Behav Assess 7: 27–28
Edelmann RJ (1985) Dealing with embarrassing events: socially anxious and non-socially anxious groups compared. Br J Clin Psychol 24: 281–288
Emmelkamp PMG (1994) Behavior therapy with adults. In: Bergin AE, Garfield SL (eds) Handbook of psychotherapy and behavior change. Wiley, New York
Faravelli G, Panichi C, Pallanti S, Paterniti S, Grecu LM, Rivelli S (1991) Perception of early parenting in panic and agoraphobia. Acta Psychiatr Scand 84: 6–8
Fahrenberg J (1987) Zur psychophysiologischen Methodik: Konvergenz, Fraktionierung oder Synergismen? Diagnostica 33: 272–287
Fisher PL, Durham RC (1999) Recovery rates in generalized anxiety disorder following psychological therapy: an analysis of clinically significant change in the STAI-T across outcome studies since 1990. Psychol Med 29: 1425–1434
Foa EB, Rothbaum BO, Maerker A (2000) Posttraumatische Belastungsstörungen. In: Margraf J (ed) Lehrbuch der Verhaltenstherapie. Springer, Berlin Heidelberg New York
Freud S (1926) Hemmung, Symptom und Angst. Fischer, Frankfurt
Fröhlich WD (1983) Perspektiven der Angstforschung. In: Enzyklopädie der Psychologie, Themenbereich C – Theorie und Forschung. Series IV: Motivation und Emotion. Vol 2: Psychologie der Motive. Hogrefe, Göttingen, pp 110–320
Gray JA (1982) The neuropsychology of anxiety. Oxford University Press, Oxford
Greene WA, Conron G, Schalch DS, Schreiner BF (1970) Psychologic correlation of growth hormone and adrenal secretory response of patients undergoing cardiac catherization. Psychosom Med 32: 599–614
Greenwood FC, Landon J (1966) Growth hormone secretion in response to stress in man. Nature 210: 540–541
Gurgius GNM, Klein E, Mefford IN, Uhde TW (1990) Biogenic amines distribution in the brain of nervous and normal pointer dogs. A genetic animal model of anxiety. Neuropsychopharmacology 3: 297–303
Halbreich U, Olympia J, Carson S, Glogowski J, Yeh C, Axelrod S, Desu MM (1989) Hypothalamo-pituitary-adrenal activity in endogenously depressed post-traumatic stress disorder patients. Psychoneuroendocrinology 14: 365–370
Harris EL, Noyes RJ, Crowe RR, Chandhry DR (1983) Family study of agoraphobia. Arch Gen Psychiatr 40: 1061–1064
Hemmeter U, Holsboer-Trachsler E (1995) Neuroendokrinologie und Biochemie bei Angsterkrankungen. In: Kasper S, Möller HU (eds) Angst- und Panikerkrankungen. Fischer, Jena
Hemmeter UM, Burkhardt H, Netter P (1991) Der Einfluß depressionsassoziierter Persönlichkeitsmerkmale auf die Cortisolwerte unter experimentellem Streß und Fastenbedingung. Z Klin Psychol 20/2: 166–176
Hollon S, Beck AT (1994) Cognitive and behavioral therapies. In: Bergin AE, Garfield SL (eds) Handbook of psychotherapy and behavior change. Wiley, New York
Holsboer F (1993) Neurobiologie. In: Holsboer F, Philipp M (eds) Angststörungen. SMV, Gräfelfing
Hope DA, Heimberg RG (1993) Social phobia and social anxiety. In: Barlow DH (ed) Clinical handbook of psychological disorders. A step-by-step treatment manual, 2nd edn. Guilford, New York, pp 99–136
Juster HR, Brown EJ, Heimberg RG (2000) Sozialphobie. In: Margraf J (ed) Lehrbuch der Verhaltenstherapie. Springer, Berlin Heidelberg New York
Kahn RJ, McNair DM, Lipman RS et al (1986) Imipramine and chlordiazepoxide in depressive and anxiety disorders. II. Efficacy in anxious outpatients. Arch Gen Psychiatry 43: 79–85
Kapfhammer HP (1995) Psychopharmakologische Behandlung von Angsterkrankungen. In: Kasper S, Möller HU (eds) Angst- und Panikerkrankungen. Fischer, Jena
Kasper S (1995) Neue psychopharmakologische Strategien bei der Behandlung von Angsterkrankungen. In: Kasper S, Möller HU (eds) Angst- und Panikerkrankungen. Fischer, Jena
Kendler KS, Neale MC, Kessler RC, Heath AC, Eaves LJ (1992) Generalized anxiety disorder in women. A population-based twin study. Arch Gen Psychiatry 49: 267–272
Klein DF, Rabkin J (eds) (1981) Anxiety. New research and changing concepts. Raven, New York
Kuhar MJ (1986) Neuroanatomical substrates of anxiety: a brief survey. Trends Neurosci 9: 307–311
Lader MH, Marks IM (1971) Clinical anxiety. Heinemann, London
Lang PJ (1993) The three-system approach to emotion. In: Birbaumer N, Öhmann A (eds) The structure of emotion. Hogrefe and Huber, Toronto
Laux G (1995) MAO-Hemmer bei Angsterkrankungen. In: Kasper S, Möller HU (eds) Angst- und Panikerkrankungen. Fischer, Jena
Lewin BD (1925) Claustrophobia. Psychoanal Q 4: 227–233
Linden M (1993) Epidemiologie von Angst- und Panikerkrankungen. In: Kasper S, Möller HJ (eds) Angst- und Panikerkrankungen. SMV, Gräfelfing
Linden M, Zubrägel D (2000) Generalisierte Angsterkrankungen. In: Butra A, Wassmann R, Buchkremer G (eds) Verhaltenstherapie. Thieme, Stuttgart
Linden M, Geiselmann B, Helmchen H (1988) Axiolytika und Sedativa. MMW 130: 571–574
López-Ibor JJ (1952) La angustia vital. Paz Montalvo, Madrid

Margraf J (ed) (2000) Lehrbuch der Verhaltenstherapie. Springer, Berlin Heidelberg New York

Margraf J, Schneider S (1989) Panik, Angstanfälle und ihre Behandlung. Springer, Berlin Heidelberg New York

Margraf J, Schneider S (2000) Paniksyndrom und Agoraphobie. In: Margraf J (ed) Lehrbuch der Verhaltenstherapie. Springer, Berlin Heidelberg New York

Margraf J, Ehlers A, Roth WT (1986) Sodium lactate infusions and panic attacks. A review and critique. Psychosom Med 48: 23–26

McKinney WT (1988) Models of mental disorders. Plenum, New York

Mowrer OH (1960) Learning theory and behavior. Wiley, New York

Nesse RM, Curtis GC, Thyer BA, McCann DS, Huber-Smith MJ, Knopf RF (1985) Endocrine and cardiovascular responses during phobic anxiety. Psychosom Med 47: 320–332

Noyes RJ, Crowe RR, Harris EL, McChessney CM, Chaudhry DR (1986) Relationship between panic disorder and agoraphobia. A family study. Arch Gen Psychiatry 43: 227–232

Öst LG (2000) Spezifische Phobien. In: Margraf J (ed) Lehrbuch der Verhaltenstherapie. Springer, Berlin Heidelberg New York

Parker G (1979) Reported parental characteristics of agoraphobics and social phobics. Br J Psychiatry 135: 555–560

Pekrun R (1990) Emotion: Klassifikation und Diagnostik. In: Baumann U, Perrez M (eds) Klinische Psychologie, vol I. Huber, Bern

Rahe RH, Karson S, Howard NS Jr, Rubin RT, Poland RE (1990) Psychological physiological assessments on American hostages freed from captivity in Iran. Psychosom Med 52: 1–16

Reiman EM, Raichle ME, Butler FK, Herscovitch P, Robins E (1984) A focal brain abnormality in panic disorder, a severe form of anxiety. Nature 310: 683–685

Reiman EM, Fusselman MJ, Fox PT, Raichle ME (1989) Neuroanatomical correlates of anxiety. Science 243: 1071–1074

Reinecker H (1993) Phobien. Hogrefe, Göttingen

Rickels K, Case WG, Downing RW (1982a) Issues in longterm treatment with diazepam therapy. Psychopharmacol Bull 18: 38–41

Rickels K, Case WG, Downing RW et al (1982b) Buspirone and diazepam in anxiety: a controlled study. J Clin Psychiatry 43: 81–86

Rickels K, Case WG, Downing RW et al (1985) Indications and contraindications for chronic anxiolytic treatment: is there tolerance to the anxiolytic effect? In: Kemali D, Racagni G (eds) Chronic treatment in neuropsychiatry. Raven, New York

Rickels K, Downing R, Schweizer E, Hassman H (1993a) Antidepressants for the treatment of generalized anxiety disorder. A placebo-controlled comparison of imipramine, trazodone and diazepam. Arch Gen Psychiatry 50: 884–895

Rickels K, Schweizer E, Weiss S, Zavodnick S (1993b) Maintenance drug treatment for panic disorder. II. Short- and long-term outcome after drug taper. Arch Gen Psychiatry 50: 61–68

Rocah B (1991) A clinical study of a phobic illness: the effects of traumatic scars on symptom formation and treatment. In: Moraitis G (ed) Recent developments in the concept und treatment of phobias and panic states. Psychoanal Inquiry 11: 351–375

Schedlowsky M, Wiechert D, Wagner TO, Tewes U (1992) Acute psychological stress increases plasma levels of cortisol, prolactin and TSH. Life Sci 50/17: 1201–1205

Schneier RR, Liebowitz MR, Davies SO, Fairbanks J, Hollander E, Campeas R, Klein DF (1990) Fluoxetine in panic disorder. J Clin Psychopharmacol 10: 199–121

Selye H (1956) The stress of life. McGraw-Hill, New York

Sieghart W (1995) Transmittersysteme und Rezeptoren in der Neurobiologie der Angst. In: Kasper S, Möller HU (eds) Angst- und Panikerkrankungen. Fischer, Jena

Seligman M (1971) Phobias and preparedness. Behav Ther 2: 307–320

Solyom L, Silberfeld M, Solyom C (1976) Maternal overprotection in the etiology of agoraphobia. Can Psychiatr Assoc J 21: 109–113

Sternbach H (1990) Fluoxetine treatment of social phobia. J Clin Psychopharmacol 10: 230

Strian F (1983) Klinik der Angst. In: Strian F (ed) Angst. Grundlagen und Klinik. Springer, Berlin Heidelberg New York

Tancer ME, Stein MB, Uhde TW (1993) Growth hormone response to intravenous clonidine in social phobia: comparison to patients with panic disorder and healthy volunteers. Biol Psychiatry 34: 591–595

Torgersen S (1983) Genetic factors in anxiety disorders. Arch Gen Psychiatry 40: 1085–1089

Turowsky J, Barlow DH (2000) Generalisiertes Angstsyndrom. In: Margraf J (ed) Lehrbuch der Verhaltenstherapie. Springer, Berlin Heidelberg New York

Volz HP (1995) Serotonin-Wiederaufnahmehemmer (SSRI) bei Angsterkrankungen. In: Kasper S, Möller HU (eds) Angst- und Panikerkrankungen. Fischer, Jena

Weyers P, Fritze J (1995) Modelle für depressive Störungen: Möglichkeiten der Weiterentwicklung. In: Debus G, Erdmann G, Kallus G, Kallus KW (eds) Biopsychologie von Streß und emotionalen Reaktionen. Hogrefe, Göttingen, pp 209–231

WHO (1992) International statistical classification of diseases and related health problems, 10th revision. World Health Organization, Geneva

Willi J (1972) Die angstneurotische Ehe. Nervenarzt 43: 399–408

Wurthmann C (1995) Trizyklika bei Angsterkrankungen. In: Kasper S, Möller HU (eds) Angst- und Panikerkrankungen. Fischer, Jena

Yehuda R, Southwick SM, Krystal JH, Bremner D, Charney DS, Mason JW (1993) Enhanced suppression of cortisol following dexamethasone administration in posttraumatic stress disorder. Am J Psychiatry 150: 83–86

CHAPTER

4

E. Vermetten, D.S. Charney, J.D. Bremner

Post-traumatic Stress Disorder

1 Introduction

Society is increasingly affected by all types of human violence, including rape, robberies, assault, natural disaster, and accidents. This can leave the individual with intense terror, fear, and paralyzing helplessness. About 60% of men and 50% of women have experienced psychological trauma (defined as threat to life of self or significant other) at some time in their lives. As many as 39% of these individuals exposed to a traumatic event will develop post-traumatic stress disorder (PTSD). The lifetime prevalence of PTSD in the United States is estimated at 7.8% and is twice as common in women as in men. This is more than twice the prevalence of bipolar disorder or schizophrenia. PTSD is strongly comorbid with other lifetime psychiatric disorders. More than one third of people with an index episode of PTSD fail to recover even after many years (Kessler et al. 1995). PTSD has long been recognized as being related to wartime experiences, and only more recently has been recognized as being associated with other types of trauma (Weisaeth and Eitinger 1993; Wilson 1994; Saigh and Bremner 1999). PTSD is increasingly recognized as being present in diverse cultures (Kleber et al. 1995; I.M. Allen 1996). The diagnosis of PTSD was first included in the third edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-III) in 1980 (APA 1980), where it was categorized as an anxiety disorder. Its first appearance in the *International Classification of Diseases* (ICD) is of more recent date (WHO 1992).

Recently, there has been a rapid expansion of empirical studies that have led to a significant increase in our knowledge of the phenomenology and biology of PTSD (Ursano et al. 1994). Over 4000 papers have been published on PTSD in the last 35 years, as well as numerous books and dissertations (see Fig. 1). The *American Journal of Psychiatry* has shown the greatest interest in PTSD overall, publishing 221 papers on this disorder in the last 10 years alone, by far more than any other journal in the field of general psychiatry. *Biological Psychiatry* has showed a marked increase in publication in PTSD in recent years. The *Journal of Traumatic Stress* is a specialty journal that started in 1987 and deals with all PTSD-related issues. This chapter will highlight these findings in PTSD and give a comprehensive overview on the history, diagnosis, epidemiology, and treatment of the disorder.

2 History and Conceptual Development of the Disorder

2.1 From "Soldier's Heart" to Post-traumatic Stress Disorder

Although the diagnosis of PTSD was only recently established, there is a long history of observations of the effects of stress on the individual (for a review, see Saigh and Bremner 1999). Much of the history of stress and psychiatric diagnosis has been determined by ebb and flow of attention paid to this area with each new major military conflict. During the American Civil War, DaCosta (1871) first described a syndrome involving symptoms of exhaustion and increased physiological responsivity ("soldier's heart" or DaCosta's syndrome) seen in soldiers exposed to the stress of war. DaCosta felt that this syndrome was a physical disorder involving the cardiovascular system that was caused by the extreme stress of war. DaCosta's approach was similar to theories of the time advanced by Kraepelin (1919), who also believed that schizo-

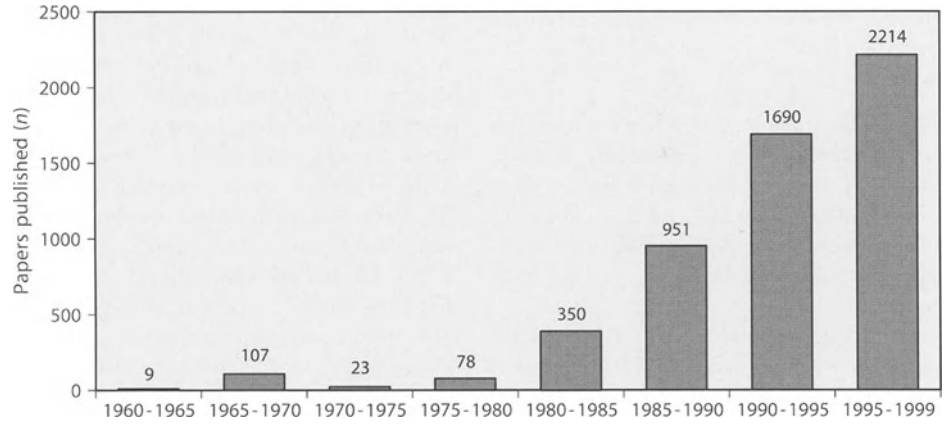


Fig. 1. Increase in the number of publications in the field of PTSD over time (1960–1999)

phrenia had its basis in the constitution, leading to abnormalities in the brain and physiology. Brain-based explanations of psychiatric disorders left the scene at the turn of the century with Sigmund Freud. Freud originally believed that his famous case of Anna O. was a victim of exposure to traumatic sexual experiences in childhood and only later changed his views into the theory that fantasies, and not the reality, of childhood sexuality led to mental illness (Nemiah 1998). His final formulation of psychodynamic theory did not incorporate environmental events such as traumatic stress in the development of mental disorders.

During the First World War, the large number of psychiatric casualties of combat forced attention on the effects of the stress of war and led to the description of "combat fatigue," or "physioneurosis." Experimental studies revealed intolerance in these patients to carbon dioxide and exaggerated psychological and physiological responses to epinephrine. Psychiatrists described phenomena such as amnesia on the battlefield, where soldiers forgot their name or who they were. After the war, however, the effects of combat stress on the mind were soon forgotten. With World War II, interest in the mental health effects of the stress of war was revived. Again, psychiatrists described amnesia and other dissociative responses to trauma (Sargent and Slater 1941; Torrie 1944). Studies in Danish police interned in German concentration camps noted symptoms in the survivors including recurrent memories of the camps, feelings of detachment and estrangement from others, sleep disturbance and hyperarousal, as well as problems with memory and concentration (Thygesen et al. 1970). Abraham Kardiner, who treated many casualties in World War II, which he described in *Traumatic Neuroses of War* in 1941, recognized a set of symptoms that underlies the current conceptualization of PTSD, a condition he referred to as "physioneurosis":

1. Preoccupation with the trauma
2. Constriction of personality
3. Atypical dream life
4. Startle response
5. Irritability

Survivors of death camps in this war were found to have symptoms of anxiety, motor restlessness, hyperarousal, sleeping difficulties, nightmares, fatigue, feeling worse with traumatic reminders, and a constant preoccupation with recollections of persecutory experiences. These symptoms became known as the "concentration camp syndrome."

The experience of the Second World War was still fresh in the minds of psychiatrists when the first edition of the *Diagnostic and Statistical Manual* (DSM-I), nicknamed the "Grey Manual," was formulated in 1952 (APA 1952). This led to the addition of gross stress reaction in DSM-I. Gross stress reaction described a series of stress-related symptoms in response to an extreme stressor that would be traumatic for almost anyone. This may have stemmed from the experience of military psychiatrists in the Second World War, who observed during the war that many normal men were having mental breakdowns in the face of combat. However, gross stress reaction specified that the individual must have a normal prestressor personality and that the symptoms should naturally resolve with time. This disorder did not take into account the fact that individuals with preexisting psychiatric disorders may develop a new disorder that is specifically related to the stressor, or that acute responses to stress can translate into long-term pathology. It is as if gross stress reaction was a response to the reality that extreme stressors such as war can lead to psychiatric outcomes that are not secondary to "bad personalities" (military psychiatrists in the Second World War had tried in vain to find premilitary personality traits that would help them predict who was most vulnerable to the stress of combat). Embodied in gross stress reaction was the ambivalence that has pervaded psychiatry until the current time about whether stress has merely transient effects or whether it can lead to permanent psychopathology.

It was perhaps the forgetting of the horrors of the Second World War that resulted in gross stress reaction being dropped from DSM-II in 1968 (APA 1968). Oddly enough, this took place just in the same year as the major outburst of the Vietnam War. The gross stress reaction was replaced in DSM-II by "transient adjustment reactions." These included acute reactions to overwhelming environmental stress, but did not name any specific catastrophes, such as war, death camps, or Hiroshima. Combat cropped up in one line, as an example of the adult type of adjustment reaction: "fear associated with military combat and manifested by trembling, running and hiding." The important work of Kardiner (1941), *Traumatic Neuroses of War*, of Grinker and Spiegel (1945), *Men Under Stress*, and the impressive follow-up studies conducted by Archibald and Tuddenham (1965) were overlooked by many. These studies suggested that the argument of preexisting disorders as an explanation for traumatic stress disorders was largely irrelevant. It was only later in the aftermath of the Vietnam War that the *lasting* effects of traumatic stress on the mind were recognized. Researchers and clinicians such as Robert Jay Lifton, Chaim Shattan, and Charles Figley argued at that time that the stress of war itself led to psychopathology as opposed to factors such as "bad character" (preexistent and preceding the war) (see Figley 1978a). This was the background leading to the inclusion of PTSD (with both acute and chronic types) as a disorder in DSM-III in 1980. In the DSM-III-based criteria for

PTSD, there was fortunately resistance by Nancy Andreasen (Andreasen et al. 1971; Andreasen 1985) and others to those who argued against the uniqueness of individual traumas. For example, at the time, some advocated a "Vietnam syndrome" which would describe a constellation of symptoms unique to Vietnam veterans, while others have argued for "post-rape syndrome" or specific childhood sexual abuse syndromes (ideas for which there is little empirical evidence). This resistance was indispensable to the scientific advancement of the study of traumatic stress.

DSM-III represented a departure from the approaches to stress represented in earlier versions of DSM. DSM-I and DSM-II followed the view that, if the patient had good adaptive capacity, his or her symptoms usually receded as the stress diminished. Both DSM-I and DSM-II added that, if the symptoms persisted "after the stress was removed, the diagnosis of another mental disorder was indicated." A neurotic or psychotic label replaced the diagnosis of stress disorder or adjustment reaction. Stress disorders were held to be transient and reversible, with no rubric for the continuance of stress-induced symptoms as such. With DSM-III-based PTSD, there was finally a diagnosis that recognized the lasting pathological effects of traumatic stress. DSM-III also found a place for code 308.33 delayed catastrophic stress disorder following an asymptomatic interval ("incubation period"). This phenomenon was seen by clinicians, resembling ongoing, chronic, low-grade, or latent subclinical disturbance that years later can be triggered into acute disorder by events which symbolize or recapture the original trauma (helicopters, smells, war in Kosovo), probably best understood by a model of stress sensitization.

Some authors still noted gaps in the classification of PTSD in DSM-III. These were related to the etiology of the disorder, its natural history, and diagnostic specificity. A demand was made for research in theoretical issues and to collect empirical data in order to be more precise in later diagnostic descriptions and understanding (Green et al. 1985). DSM-IIIR (APA 1987) included specification of generic characteristics of traumatic stressors, clearer organization of symptoms around three dimensions of stress response (reexperiencing, avoidance and numbing, and physiological arousal), inclusion of symptoms specific to children, and specification of onset and duration of the disorder (Brett et al. 1988).

2.2 Definition of Psychological Trauma: The A Criterion

One of the most vexing issues in the field of traumatic stress is the Criterion A problem, or the definition of psychological trauma. In earlier versions of DSM, this was centered on questions of how to define a traumatic event and whether experiencing a traumatic event should be a requirement for a PTSD diagnosis. The definitions in DSM-III and DSM-IIIR defined psychological trauma as an event "outside the range of human experience" and required the presence of psychological trauma for the diagnosis. This definition seems unsatisfactory, however, because traumatic events are actually fairly common. For example, a large epidemiological study found that 60% of men and 51% percent of women experienced at least one traumatic event in their lives (Kessler et al. 1995). Other epidemiological studies showed similar prevalence figures (for a review, see Acierno et al. 1999). In response to these concerns, a two-part definition for DSM-IV was developed, where the first part was "the person experienced, witnessed or was confronted with an event that involved actual or threatened death or serious injury, or a threat to the physical integrity of self or others," which distinguishes traumatic events from ordinary stressors. The second part required that the person's response involved intense fear, helplessness, or horror, giving recognition to the fact that an individual's subjective response is an important dimension to consider in defining a traumatic event. Although the definition of psychological trauma continues to be a matter of debate, there appears to be a growing consensus that traumatic events can be distinguished from ordinary stressors and that the use of specific criteria for psychological trauma serves a useful gatekeeping function that preserves the meaningfulness of the disorder as a distinct diagnostic entity and prevents trivialization of the suffering of survivors of overwhelming stressors.

2.3 Classification Systems: DSM Versus ICD

The *International Statistical Classification of Diseases, Injuries, and Causes of Death* (ICD) developed in parallel with DSM over the past half-century. This classification system was originally formalized as the *1892 Bertillon Classification List of Causes of Death.* ICD is meant to cover the worldwide classification of diseases and has been used more in Europe than DSM. In 1948, the World Health Organization (WHO) decided to include mental disorders in the sixth revision of the ICD (ICD-6; WHO 1948), since many psychiatric hospitals were dealing with a nomenclature that did not appropriately describe the majority of the cases handled. DSM, which was developed in the United States, has only focused on disorders of mental health.

The formulation of stress-related psychopathology in the later ICD systems and in the DSM classification is

based upon attempts to describe post-traumatic psychopathology. The perspective that traumatic reactions were short-lived responses in essentially normal individuals with no premorbid psychopathology can be originally found in both the international and American classifications. Prototypes of PTSD were called "acute situational maladjustment" in ICD-6 and "transient situational personality disturbance" in DSM-I. There were no changes in the mental disorders section in ICD-7. In ICD-8 (WHO 1968), there was a rephrasing to "transient situational disturbance." DSM-II was mostly based on the mental disorder section of this eighth edition of ICD. DSM-II focused on the diagnosis "adjustment reaction" and specified the phase in life to which this was related. In DSM-III and ICD-9, there were significant revisions of the conceptualization of post-traumatic stress reactions, with the most important change being the notion that stress disorders were no longer acute responses in healthy individuals. Traumatic stress was considered to cause chronic reactions, and responses to stress were considered to occur with previous and simultaneous conditions. Another significant change was the categorization of the diagnosis in the anxiety disorders section of DSM-III. ICD went along the same line; however, it also incorporated a variation of PTSD in its classification. In ICD, under Sect. F62.0, there is also a diagnosis named "enduring personality changes after catastrophic experience" alluding to some variation of a personality disorder. The American approach was to consider anxiety a core phenomenon of both PTSD and obsessive-compulsive disorder (OCD), whereas the Europeans regarded anxiety as a common and nonspecific feature of many disorders and located OCD and PTSD in their classification system in ICD on the basis of other features.

The most recent version of ICD is ICD-10 (1992). This version now consists of a three-volume work, with one section on mental health. ICD-10 was field-tested in over 100 institutions in 40 countries, making this book the product of a very large research effort. PTSD is coded as F43.1 and subsectioned under "neurotic stress-related and somatoform disorders" and further under "other anxiety disorders and reactions to severe stress," where it is grouped with "acute stress reaction," "adjustment disorders," "other reactions to severe stress," and "reaction to severe stress unspecified." PTSD is defined as follows:

A mental disorder characterized by a preoccupation with traumatic events beyond normal human experience; events such as rape or personal assault, combat, violence against civilians, natural disasters, accidents or torture precipitate this mental disorder; patients suffer from recurring flashbacks of the trauma and often feel emotionally numb, are overly alert, have difficulty remembering, sleeping or concentrating, and feel guilty for surviving.

ICD-10 defines the A criterion "beyond normal human experience," whereas DSM-IV is more stringent. Although an essential criterion, the A criterion does not predict PTSD or the course of PTSD. As stated earlier, the latest revision of ICD-10 has a wider variety of diagnoses for traumatic reactions, with the inclusion of enduring personality changes after catastrophic experience (F62.0) (WHO 1992). In their latest releases, both classification systems seem to be more in line in their description of PTSD (see Table 1). However, in a comparison of 12-month prevalence of PTSD using ICD-10 and DSM-IV criteria , Peters et al. (1999) found

Table 1. Comparison of ICD and DSM diagnoses in traumatic stress over time

Year	ICD	DSM
1948	Acute situational maladjustment (ICD-6; WHO 1948)	
1952		Transient situational personality disturbance; gross stress reaction; adult situational reaction; adjustment reaction of infancy, childhood, adolescence, or late-life (DSM-I; APA 1952)
1968	Transient situational disturbance (ICD-8; WHO 1968)	Adjustment reaction of infancy, childhood, adolescence, or late-life (DSM-II; APA 1968)
1977	Acute reaction to stress with predominant disturbance of emotions, consciousness, or psychomotor disturbance or mixed (ICD-9; WHO 1977)	
1980		Post-traumatic stress disorder (DSM-III; APA 1980)
1987		Post-traumatic stress disorder (DSM-III-R; APA 1987)
1992	Acute stress reaction; post-traumatic stress disorder; enduring personality changes after catastrophic experience (ICD-10; WHO 1992)	
1994		Acute stress disorder; post-traumatic stress disorder (DSM-IV; APA 1994)

that, using the ICD criteria, the prevalence was more than double (3% using DSM-IV and 7% using ICD-10 criteria). Almost half of the discrepancies between the classification systems could be explained by the F criterion in DSM-IV requiring clinically significant distress or impairment in social, occupational or other important areas of functioning. Another factor that accounted for 18% of the discrepancies was the C criterion in DSM-IV (see the Appendix for an overview of the diagnostic criteria for PTSD; see also Lundin and Lotfi 1996).

3 Diagnostic Features

3.1 Traumatic Stress: The A Criterion

PTSD is the only psychiatric condition whose definition demands that a particular stressor precede its appearance. PTSD is characterized by specific symptoms which develop following exposure to psychological trauma, defined as a situation in which a person experienced, witnessed, or was confronted with an event or events that involved actual or threatened death or serious injury, or a threat to the physical integrity of self or others, and where the person's response involved intense fear, helplessness, or horror. According to DSM-IV (APA 1994), PTSD can result from combat, violent personal assault (sexual assault, physical assault, robbery, mugging), being kidnapped, being taken hostage, terrorist attack, torture, incarceration as a prisoner of war or in a concentration camp, or natural or manmade disasters (severe automobile accidents, being diagnosed with a life-threatening disease, or other threat to life) (e.g. Tucker et al. 1997; Foa 1997; Schnurr and Spiro 1999; Walker et al. 1999; Alonzo 1999; Conlon et al. 1999). Sexually traumatic events may include developmentally inappropriate sexual experiences without threatened or actual violence (e.g. Davidson et al. 1998a). PTSD may also be induced by observing events such as the serious injury or unnatural death of another person due to violent assault, accident, war, or disaster or unexpectedly witnessing a dead body or body parts (e.g. Zisook et al. 1998; Ursano et al 1999a). Traumatization can also occur where the individual receives information about the stressful experiences of others such as personal assault, serious accident or serious injury experienced by a family member or a close friend, or learning that one's child has a life-threatening disease, if this is associated with intense fear, helplessness, or terror (e.g. Manne et al. 1998). Psychological trauma may occur as a single episode or repetitive ongoing trauma (as sexual or physical abuse); however; studies demonstrate that cumulative effects of trauma are an important determinant of risk for PTSD (Bremner et al. 1993a; Bremner et al. 1995a).

3.2 Diagnostic Categories

Symptoms of PTSD are divided into three categories:

1. Reexperiencing of the event
2. Avoidance of stimuli
3. Persistent symptoms of increased arousal

At least one symptom of reexperiencing is required for the diagnosis, including recurrent and intrusive distressing recollections of the event, recurrent distressing dreams of the event, acting or feeling as if the traumatic event were recurring, intense psychological distress at exposure to internal or external cues that resemble an aspect of the traumatic event, or physiological reactivity upon exposure to internal or external cues that symbolize an aspect of the traumatic event. Three avoidance symptoms are required, including avoidance of stimuli associated with the trauma and numbing of general responsiveness by efforts to avoid thoughts, feelings, or conversations associated with the trauma, efforts to avoid thoughts, feelings, activities, places, or people that arouse recollections of the trauma, inability to recall an important aspect of the trauma, markedly diminished interest or participation in significant activities, feeling of detachment or estrangement from others, restricted range of affect (e.g. unable to have loving feelings), a sense of a foreshortened future (e.g. does not expect to have a career, marriage, children, or a normal life span). The third cluster consists of persistent symptoms of increased arousal that were not present before the trauma. At least two of these symptoms are required, including difficulty falling or staying asleep, irritability or outbursts or anger, difficulty concentrating, hypervigilance, or an exaggerated startle response. The duration of the disturbance needs to be more than 1 month and the symptoms must be associated with significant distress or impairment in social, occupational, or other areas of functioning. DSM categorized three subtypes of the disorder: acute, chronic, and a subtype with delayed onset. PTSD is described as delayed onset if the symptoms appear more than 6 months after the event has passed (an overview of diagnostic criteria is given in the Appendix).

Two typical cases of PTSD are given below.

- *Case example I:* A 27-year-old female reported freezing responses, anxiety, and uncontrollable flashbacks. She stated she sometimes lashes out at

her children for no reason. She sleeps 4 h a night and weighs 245 lb. She is living with her boyfriend and her two children aged 13 and 8. Her eldest child was born when she was 13 years old. She states that her PTSD symptoms gradually became worse as she became older. She hesitated volunteering for the clinical research we were conducting, but did so because she felt she was in despair. She disclosed that her current feeling of despair is the result of experiencing and witnessing traumatic events during childhood (e.g. she was raped at age 6; she witnessed several shootings in the neighborhood where she was living and saw her brothers being molested by strangers; her youngest brother was once kidnapped and was found days later bruised and naked). She is the eldest of four children and has three brothers. All her brothers are currently into drugs and deal on the street. She stated that her brothers may have fathered several children, but she is uncertain of this for sure. She disclosed that her father "never cared" for her. She reported that both her father and mother are "crackheads." She begins crying as she says "my life has holes in it." She says that she has no memory for things others tell her she has done as a teenager. No matter how hard she tries, the memories do not come back. She was told she was dating men and that she often went out with them and slept with them.

She attempted suicide when she was 16 years, but was never hospitalized at the time of the attempt. She has been in counseling in the past. She is currently employed as a nurse assistant and works 80–100 h per week. She stated, "I have to keep going, for if I don't I'll fall apart." She reports she is afraid of being alone. Recently, she smelled the scent of grease and fried eggs, which overwhelmed her and placed her in a flashback. As a result, she "lost control," dissociated, and became extremely frightened. This apparently is related to one of the traumatic scenes from her childhood when she was raped near a kitchen by a neighbor. He had been baking eggs. It happened 3 days before her birthday. She now "hates" her birthday and spends it alone away from home. She reports she has always had a clear memory of this incident, stating, "that is what made it all start," referring to her current symptoms.

She reported later that after disclosing information about her childhood trauma to the phone screener, she showered for 1 h. In addition, she wanted her boyfriend to be with her all the time, did not want to be alone at night, and was thinking of binding a bell around his legs so that she would hear him if he were to leave.

- *Case example II:* A 50-year-old Vietnam veteran seeks help on the encouragement of his wife. He had always tried to avoid clinical treatment, stating he felt fine periodically attending a Veterans Administration (VA) support group. His wife urged him to participate in our clinical research as his irritability has been gravely impacting his family. She felt that the news about the Kosovo war had triggered his irritability and his own memories of war. Upon our meeting, he presented as a friendly man, cooperative and quiet. He complained of irritability and of often lashing out. He reported concentration problems at work, memory problems, crying spells, and often wanting to be left alone. He felt easily startled, jumpy, and on edge. His stated his sleep has always been severely disruptive as he sleeps only 3–4 h a night, frequently wakes up sweating, derealized and extremely afraid. This had been going on for the last 30 years, unchanged. He disclosed that after his return from Vietnam, the relationship with his fiancée had ended. Soon after the breakup, he married his current wife, who, at the time, was actively protesting against the war. With her he had two children. They never discussed the war in any way until years later. After returning from Vietnam, he began working for a telephone company and continues to this day, stating he feels he cannot perform on another level. He complains about his increasing forgetfulness. He also reports suicidal thoughts, but has never acted upon them. He cannot picture himself years from now, feeling his future has been taken away by Vietnam. He disclosed that he was on the frontline during the war, seeing several of his fellow soldiers die or get wounded. He reported that, during that time, he often imagined himself dead during active combat. He then pictured himself saying goodbye to his family, which resulted in him feeling detached from his body. He still wears some representation of Vietnam to show he is a veteran, but does not want people to comment on it. He avoids going to any annual ceremony. He reported that he avoids talking about Vietnam because it is "too intense," that he "cannot cope" as it has left him feeling irritable for the remainder of the day when speaking about it in the past. He is convinced he will die a premature death.

The longitudinal course of PTSD was retrospectively assessed in a sample of Vietnam combat veterans with PTSD (Bremner et al. 1996a). The onset of symptoms in this group typically occurred at the time of exposure to combat trauma in Vietnam and increased rapidly during the first few years after the war. Symptoms plateaued within a few years after the war, following which the disorder became chronic and unremitting. Hyperarousal symptoms such as feeling on guard and feeling easily startled developed first, followed by avoidant symptoms and finally by symptoms from the

intrusive cluster. Onset of alcohol and substance abuse typically was associated with the onset of symptoms of PTSD, and the increase in use paralleled the increase in symptoms. Patients reported a tendency for alcohol, marijuana, heroin, and benzodiazepines to make PTSD symptoms better, while cocaine made symptoms in the hyperarousal category worse. No relationship was found between treatment interventions and the natural course of PTSD. Figure 2 shows the longitudinal course of PTSD in this patient sample.

Symptoms of PTSD generally become evident within the first 3 months following the trauma; sometimes acute stress disorder (ASD) develops into PTSD. ASD is a rather similar disorder compared with PTSD that may occur immediately after traumatic stress exposure and may last from 2 days to 4 weeks. ASD also includes symptoms of dissociation, such as derealization and depersonalization, as well as one symptom from each of the PTSD symptom clusters. Both the occurrence of dissociative symptoms at the time of trauma and their duration afterward predict later PTSD, as well as dissociative reactions to subsequent trauma (Bremner and Brett 1997; Classen et al. 1998). Brewin et al. (1999) found that ASD diagnosis predicts 83% of cases of PTSD at 1 year. This study also found that reexperiencing and hyperarousal (but not avoidance) were equally adept at predicting development of chronic PTSD. Most individuals who, shortly after trauma, express symptoms of PTSD recover within 1 year of their traumatic experiences (Kulka et al. 1990). In contrast, those who remain ill for 1 year rarely recover completely (Freedman et al. 1999). In many individuals, PTSD can be a chronic disorder that they take with them to their graves, putting a burden on physical and mental health and also on health providers (Blank 1993; Friedman and Schnurr 1995; Solomon and Davidson 1997).

3.3 Conditions Associated with Post-traumatic Stress Disorder

Psychological trauma can have can have far more pervasive consequences than the current PTSD criteria would suggest. A close relationship exists between PTSD, dissociation, somatization, and a variety of other problems. Some authors have argued that chronic interpersonal trauma, especially with a childhood onset such as incest, physical abuse, torture, or neglect, leads to a much broader range of symptoms, described as complex PTSD (Herman 1992; Van der Kolk et al. 1996). Individuals with PTSD may also experience profound feelings of guilt and may blame themselves for surviving when others did not, keeping the guilt inside. This conflict, in its most acute presentation, typically resembles an agitated depression and is described as being associated with frequent dreams of friends dying (e.g. in battle) and by avoidance of interpersonal intimacy because they fear the other party may abandon them or die (e.g. Glover 1984; Henning and Frueh 1997). This can also cause phobic avoidance that interferes with daily activities and social interactions. Interpersonal stressors, such as childhood sexual abuse, domestic violence, or being held a hostage or a prisoner of war, may be associated with a constellation of impaired affect modulation, self-destructive and impulsive behavior, dissociative symptoms, feelings of ineffectiveness, difficulties in sexual involvement, shame, despair, or hopelessness. It may also lead to feeling permanently damaged, a loss of previously sustained beliefs, hostility, social withdrawal, feeling constantly threatened, and impaired relationships with others or may induce a change away from the individuals' previous personality characteristics (DSM-IV; APA 1994).

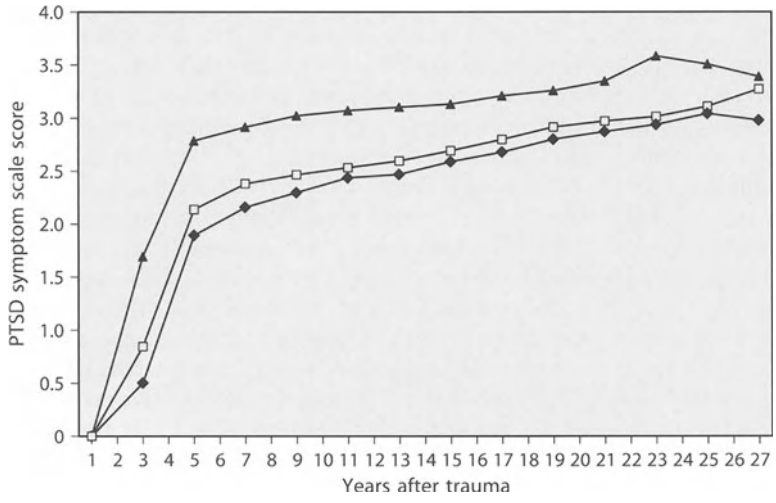


Fig. 2. Longitudinal course in PTSD in a population of Vietnam war veterans. *Triangles*, hyperarousal; *squares*, avoidance; *diamonds*, intrusions

PTSD is also frequently comorbid with other psychiatric disorders, such as depression, substance abuse, and anxiety disorders. These comorbid disorders may predate PTSD, may be related to traumatic stress, or may also be a reflection of inadequacies of our diagnostic schema. For example, data from the National Vietnam Veterans Readjustment Study (Kulka et al. 1990) show that 98% of patients with Vietnam combat-related PTSD have a comorbid lifetime psychiatric disorder. Studies to date show an increased risk for comorbid anxiety disorders (agoraphobia, panic disorder, OCD, social phobia) as well as major depressive disorder, somatization disorder, and substance-related disorder in PTSD (for a review, see Keane and Kaloupek 1997; Weathers and Keane 1999). Clinicians assessing victims of chronic interpersonal trauma need to be particularly aware that the presentation may very well include many other problems than the core symptoms of PTSD. Comorbidity may also reflect a more general vulnerability to psychopathology that renders some individuals more susceptible to developing a variety of disorders, including PTSD (Weathers and Keane 1999).

Depression is one disorder that often occurs in conjunction with PTSD. Workers in the PTSD field have long been forced to explain the "comorbidity" of depression in patients with PTSD. However, Koren and colleagues have used a prospective design to show that depression and non-PTSD anxiety disorders develop in conjunction with PTSD in stress survivors (Koren et al. 1999). This is consistent with other data, e.g. from the National Vietnam Veterans Readjustment Study, which showed that rates of depression and anxiety disorders (as well as alcohol and substance abuse) are increased in combat veterans with PTSD (Kulka et al. 1990). Zimmerman and Mattia (1999) investigated whether an association exists between psychotic subtyping of major depressive disorder and PTSD. They interviewed 500 psychiatric outpatients using the Structured Clinical Interview for DSM-IV. Almost half of these patients had a major depressive disorder ($n = 235$). Nineteen percent of this sample met criteria for PTSD. These results indicate that the presence of psychosis in psychiatric outpatients with major depressive disorder is associated with concurrent PTSD. Clearly, trauma results in a range of outcomes, and although PTSD patients may meet symptom criteria for other psychiatric disorders (e.g. major depression), they may not be equivalent to nontraumatized patients with these disorders.

From a practical assessment perspective, comorbid disorders complicate differential diagnosis and treatment planning. For example, PTSD and depression share anhedonia, sleep disturbance, and impaired concentration as core diagnostic criteria, and there is an arguable overlap in terms of restricted range of affect, impairment in personal functioning, and guilt. Interestingly enough, there are no explicit exclusion criteria for PTSD as there are for other axis I disorders. For an overview of the range of symptomatology among psychiatric disorders linked with trauma, see Fig. 3.

Bereavement or traumatic grief (involving a pathological reaction to loss of a loved one) could be argued to be subsumable under PTSD using current criteria, since loss of a loved one can be included under "threat to life of self or a loved one," although it would not be under previous criteria of an event "beyond the range of human experience" (Prigerson et al. 1999). Murphy et al. (1999) studied the prevalence of PTSD among parents bereaved by the violent deaths of their 12- to 28-year-old children. A community-based sample of 171 bereaved mothers and 90 fathers was followed for 2 years. Both parents' gender and children's causes of death significantly affected the prevalence of PTSD symptoms; twice as many mothers and fathers whose children were murdered met PTSD caseness (full diagnostic criteria) compared with accident and suicide bereavement; symptoms in the reexperiencing domain were the most commonly reported. PTSD symptoms persisted over time, with 21% of the mothers and 14% of the fathers who provided longitudinal data still meeting caseness criteria 2 years after the deaths.

3.4 Dissociative Disorders and Post-traumatic Stress Disorder

The requirement of exposure to a traumatic stressor for PTSD has led to an odd dichotomy over the years between PTSD and other psychiatric disorders. However, now that a strong and well-documented relationship between trauma (especially in early childhood) and development of the dissociative disorders has been established (e.g. Putnam et al 1986; Spiegel and Cardena 1991), the dissociative disorders still need to be theoretically linked with traumatic exposure and somehow integrated with PTSD and ASD. There is ample evidence to support a close integration between these disorders. For instance, we found that 86% of a PTSD sample met criteria for a "comorbid" dissociative disorder (Bremner et al. 1998), while essentially 100% of patients with dissociative identity disorder (DID) meet criteria for PTSD (R. Loewenstein, personal communication). We found that Vietnam combat veterans with PTSD had increased dissociative symptom levels compared to combat veterans without PTSD, and that individuals with dissociative responses to trauma are at increased risk for PTSD (Bremner et al. 1992) and continue to have dissociative responses to subsequent stressors (Bremner et al. 1998).

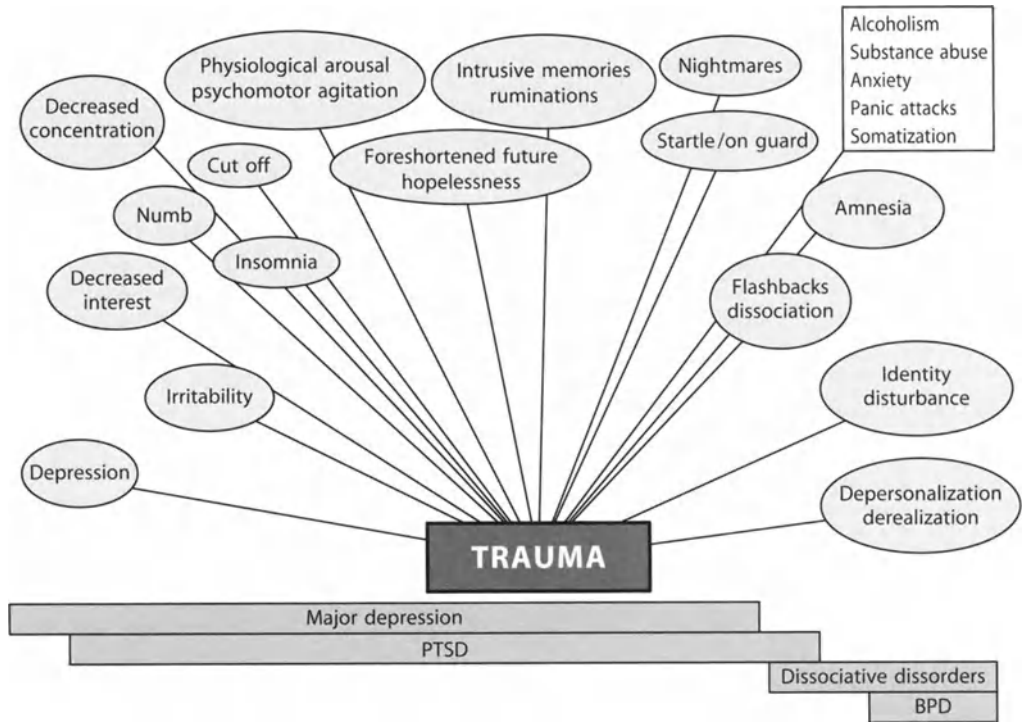


Fig. 3. Range of symptomatology among psychiatric disorders linked with trauma; BPD = borderline personality disorders with the underlying disorder that represents them

Similarly, Marmar et al. (1996) found that dissociative responses to trauma predict long-term PTSD in emergency personnel, while more recent prospective studies have documented the association between dissociative states at the time of trauma and the development of chronic PTSD (Koopman et al. 1994; Shalev et al. 1996a).

The fact that the dissociative disorders are rarely used in routine diagnosis highlights the limitations of our current diagnostic schema for clinical practice. For example, we recently searched the medical records of Psychiatric Hospital Maastricht in The Netherlands for medical chart diagnoses of dissociative amnesia and were only able to find five reported cases in well over 17,000 discharge diagnoses in the period 1980–1999. However, when structured interviews were performed on a ward at the West Haven VA, Connecticut, United States, of patients with a history of exposure to the stress of Vietnam combat and the diagnosis of PTSD, we found that 29% met DSM-IV criteria for dissociative amnesia (Bremner et al. 1998), excluding the possibility that this is an extremely rare disorder. Currently, PTSD is a rule-out for dissociative amnesia. We found that symptom levels from different dissociative symptom areas (e.g. amnesia, depersonalization) are highly correlated with one another (Bremner et al. 1998), challenging the uniqueness of individual dissociative disorders. The artificial separation of dissociative disorders such as dissociative amnesia, dissociative fugue, and depersonalization disorder, however, contributes to the fact that they are not used in clinical practice (Vermetten and Bremner 2000).

3.5 Proposal for Trauma Spectrum Disorders

The validity of diagnostic criteria of PTSD has recently been criticized (Maes et al. 1998a,b). There may be two subtypes of acute trauma response, one primarily dissociative and the other intrusive/hyperarousal, that both can lead to chronic PTSD. The weak predictive value of avoidance raises questions about this symptom cluster. Inspection of the symptoms of avoidance shows that several are in fact dissociative (e.g. amnesia and emotional numbing), others represent the behavioral response to trauma (being cut off from others or avoiding reminders), and some simply do not make sense (i.e. avoiding thinking about the trauma). Sense of foreshortened future is related to the central existential dilemma of the trauma survivor, and it might be questioned whether it should be included as a symptom of a disorder. The current intrusion and avoidance clusters were originally based on the influential theory of alternating waves of intrusions and avoidance proposed in 1976 by Mardi Horowitz in

Stress Response Syndromes. This theory has survived as dogma in the PTSD field for many years, although there has not been supportive empirical data. In a retrospective longitudinal study of the course of symptoms in PTSD patients (Bremner et al. 1996a), we found no evidence for the theory of alternating intrusions and avoidance or for another popular but unsupported theory, that of delayed onset. Rather, symptoms increased soon after the trauma, following which they plateaued and remained largely unchanged over time. In the earlier cited prospective study, Koren et al. (1999) found a similar pattern of increase in symptoms soon after the trauma, followed by a plateau, with no evidence of delayed onset. These findings may finally allow us to lay to rest the theories of alternating intrusion/avoidance and of delayed onset.

An accurate description of psychiatric responses to trauma might be served by the development of a new category of trauma spectrum disorders. This would include both acute PTSD (the current ASD) and chronic PTSD (using revised criteria to be in line with ASD), dissociative identity disorder, conversion disorder, adjustment disorders, and acute and chronic dissociative disorder not otherwise specified (NOS). A diagnosis to capture pathological responses to bereavement (traumatic grief) has been proposed as a separate diagnosis (Prigerson et al. 1999). If traumatic grief is introduced as a diagnosis in subsequent editions of DSM, it should be included in the trauma spectrum disorders. PTSD and ASD would be dropped from the anxiety disorders, since they share little relationship with several anxiety disorders (i.e. OCD).

4 Epidemiology

4.1 Prevalence

When PTSD was first defined in DSM-III (APA 1980), the original stressor criterion characterized traumatic experiences as being outside the range of human experience. However, when the prevalence of such events was systematically examined, it became apparent that trauma is surprisingly commonplace. Several studies have investigated the overall prevalence of traumatic events in the general population, looking both at community-based populations and populations of individuals at high risk for trauma or exposed to events such as natural disasters (Acierno et al. 1999). In a study by Resnick et al. (1993), 4008 adult women randomly selected from a U.S. nationwide sampling were interviewed over the phone. Prevalence of crime and non-crime civilian traumatic events, lifetime PTSD, and PTSD in the past 6 months were assessed by telephone interview. The authors found a lifetime exposure to any type of traumatic event of 69%, whereas exposure to crimes that included sexual or aggravated assault or homicide of a close relative or friend occurred among 36%. Overall sample prevalence of PTSD was 12.3% lifetime and 4.6% within the past 6 months. The rate of PTSD was significantly higher among crime versus non-crime victims (25.8% vs. 9.4%). Norris (1992) assessed the frequency and impact of ten potentially traumatic events in a sample of 1000 men and women drawn from four southeastern U.S. cities. Over their lifetimes, 69% of the sample experienced at least one traumatic event, while 21% experienced a traumatic event in the past year alone. The ten events varied in importance, with tragic death occurring most often, sexual assault yielding the highest rate of PTSD, and motor vehicle accident presenting the most adverse combination of frequency and impact. Numerous differences were observed in the prevalence of these events across demographic groups. Lifetime exposure was higher among whites and men than among blacks and women; past-year exposure was highest among younger adults. When impact was analyzed as a continuous variable (perceived stress), black men appeared to be most vulnerable to the effects of events, and young people showed the highest rates of PTSD.

Another study by Breslau et al. (1999a) involved 2181 individuals in the United States, aged 18–45 years, who were interviewed by telephone to assess the lifetime history of traumatic events and PTSD. PTSD was assessed using a modified version of the DSM-IV and ICD-10 criteria. The risk of PTSD following exposure to trauma was 9.2%. The highest risk of PTSD was associated with assaultive violence (20.9%). The trauma most often reported as the precipitating event among individuals with PTSD (31% of all PTSD cases) was sudden unexpected death of a loved one, an event experienced by 60% of the sample, and with a moderate risk of PTSD (14.3%), indicating the importance of traumatic grief. Women were at higher risk for PTSD than men, controlling for type of trauma. Although recent research had focused on combat, rape, and other assaultive violence as causes of PTSD, sudden unexpected death of a loved one was considered a far more important cause of PTSD in the community, accounting for nearly one third of PTSD cases.

Kessler et al. (1995) reported a study in a large national sample of 5877 individuals aged 15–54 years. This study obtained information on estimated life-time prevalence of trauma and PTSD, the kinds of traumas that were most often associated with PTSD, sociodemographic correlates, the comorbidity of PTSD with

other lifetime psychiatric disorders, and the duration of an index episode. The estimated lifetime prevalence of PTSD was found to be 7.8%. Prevalence was elevated among women and the previously married. The traumas most commonly associated with PTSD were combat exposure and witnessing of violence among men and rape and sexual molestation among women. PTSD was strongly comorbid with other lifetime DSM-III-R disorders. Survival analysis showed that more than one third of people with an index episode of PTSD failed to recover even after many years. These studies led to the conclusion that PTSD is more prevalent than previously believed and that PTSD often persists throughout the lifespan.

4.2 Childhood Abuse

Childhood abuse is a phenomenon that occurs on a wide scale and has quite dramatic proportions. Childhood physical or sexual abuse is a strong predictor of PTSD later in life. One million children are documented to have been exposed to abuse each year, and the prevalence of self-reported abuse is much higher. To investigate the prevalence of type of trauma and relate it with adult health problems, including physical symptoms, psychological problems, and substance abuse, McCauley et al. (1997) examined 1931 women of varied age and looked at prevalence of each symptoms cluster. Of this sample, 424 (22.0%) reported childhood or adolescent physical or sexual abuse. Compared with women who reported never having experienced abuse ($n = 1257$), women who reported abuse as children ($n = 204$) had more physical symptoms and had higher scores for depression, anxiety, somatization, and interpersonal sensitivity; they were more likely to be abusing drugs or to have a history of alcohol abuse; were more likely to have attempted suicide; and were more likely to have had a psychiatric admission.

Violence is a major public health problem that increasingly involves children and adolescents as both victims and witnesses (e.g. for reviews, see Lyons 1987; Terr 1991; Young 1993; Pfefferbaum 1997). In one study, victims of substantiated child abuse and neglect were assessed and compared with a group of matched nonabused and non-neglected children and followed into adulthood. Victims of child abuse (sexual and physical) and neglect were found to be at increased risk for developing PTSD. Slightly more than a third of the childhood victims of sexual abuse (37.5%), 32.7% of those physically abused, and 30.6% of victims of childhood neglect met DSM-III-R criteria for lifetime PTSD. Family, individual, and lifestyle variables also placed individuals at risk and contributed to the symptoms of PTSD (Widom 1999).

Interventions related to childhood abuse tend to include efforts within families, schools, and/or communities. A variety of factors influence response to trauma and affect recovery. They include characteristics of the stressor and exposure to it; individual factors such as gender, age and developmental level; and psychiatric history; family characteristics; and cultural factors (Pfefferbaum 1997). Since there is growing recognition that violence deeply affects the lives of children, there is great concern over the effect these terrible experiences will have on present and future generations. It has not only awakened the consciousness of mental health workers, but also awakened the collective consciousness of our society that the impact of violence extends well beyond the child who is physically victimized to other larger groups of children such as those who witness these events (Knapp 1998).

4.3 Disaster Situations

The Buffalo Creek disaster of 1972 involving the collapse of a slag dam and subsequent flood in West Virginia was one of the most extensively studied natural disasters of recent history. The rate of disaster-related PTSD was 7%, down from a post-flood rate at follow-up of 32%. There were no differences by age-group in their current psychological status; however, women evidenced more PTSD-related symptoms than did men (Green et al. 1990, 1994). Initially, the intrusive symptoms and nightmares and sleep disturbances were the most frequent difficulties. An exaggerated startle response was somewhat more frequent than the avoidance and numbing symptoms, such as loss of interest and caring (Green 1993). Survivor guilt was a relatively uncommon phenomenon in disaster-affected populations and was possibly one of the reasons why it was left out of the DSM-IV criteria for PTSD.

In other disaster, North et al. (1999) report about a follow-up on the terrorist Oklahoma City bombing in 1995, where a bomb blast killed 168 people. In their sample of 182 participating adults, 45% had a postdisaster psychiatric disorder and 34% had PTSD. The onset of PTSD was swift, with 76% reporting same-day onset. The relatively uncommon avoidance and numbing symptoms virtually dictated the diagnosis of PTSD (94% meeting avoidance and numbing criteria had full PTSD diagnosis) and were further associated with psychiatric comorbidity, functional impairment, and treatment received. Intrusive reexperiencing and hyperarousal symptoms were nearly universal, but by themselves were generally not associated with other psychopathology or impairment in functioning (North et al. 1999).

Disasters have a dramatic impact pending on the region of occurrence. Focusing solely on disasters as a traumatic stress experience, McFarlane reports about a Red Cross report issued in 1993 that highlighted the differential impact of disasters in Third World countries. In the period from 1967 to 1991, an average of 17 million people living in developing countries were affected by disasters each year, as compared to about 700,000 in developed countries, which is a difference of 166:1 (McFarlane and Potts 1999, p. 94).

Table 2 presents an overview of recent epidemiological data (1996–1999) on PTSD. This list is not exhaustive, but gives an impression of the wide range of target populations as well as differences in prevalence (for a comprehensive review of epidemiological data on child/adolescent PTSD, see Saigh et al. 1999a; on adults in criminal victimization, see Acierno et al. 1999; on combat-related PTSD, see Schlenger et al 1999; for epidemiological data on PTSD in disasters, see McFarlane and Potts 1999). In this table, the

Table 2. Prevalence of PTSD in epidemiologic studies 1996–1999

Author(s)	Trauma	Population	Sample	Method of measurement	Prevalence (%)
Widom(1999)	Child abuse and neglect	Adults	676	Diagnostic Interview Schedule (DIS) for PTSD	Current 30.9, lifetime 17.8
Harvey (1999)	Motor vehicle accidents	Adults	62	Structured Clinical Interview DSM-IV	16.1
Thabet and Vostanis (1999)	War	Children (6–11 years)	239	Rutter A2, B2, Gaza Traumatic Event Checklist, Child Post-Traumatic Stress Reaction Index	Mild 72.8, moderate 41
Dansky et al. (1999)	Intimate physical assault	Cocaine users	78	National Womens Study PTSD module based on DSM-III	Current 24.4, lifetime 46.2
Ursano et al. (1999b)	Motor vehicle accident	Adults	122	SCID for DSM-III and DSM-IV	1-month 34.4, 6-month 18.2, 1-year 14.0
Murphy et al. (1999)	Violent deaths of child	Parents	171 mothers, 90 fathers	Traumatic Experiences Questionnaire	Mothers 21, fathers 41
Thulesius and Hakansson (1999)	War exposure	Bosnian refugees	206	Modified version of self report instrument PTSD-10	Females 17.2, males 17.6
Amir and Sol (1999)	Wide range of civilian and combat-related traumatic events	Israeli students	983, of which 661 were exposed to at least one traumatic event	Traumatic Event Questionnaire, Impact of Event Scale, PTSD Scale	4 (6 in those exposed to traumatic events)
Stallard et al. (1998)	Road traffic accidents	Children (5–18 years)	119	CAPS – child version	34.5 (current)
	Sports injuries		66		3 (current)
Zisook et al. (1998)	Chronic illness	Adults	350	Structured Clinical Interview based on DSM-IV	10
	Unexpected death				9
	Unnatural death				36
Wagner et al. (1998)	Witnessing suffering of others	Male professional firefighters	402	PTSD symptoms scale	24.5 (current)
Manne et al. (1998)	Pediatric cancer survivors	Mothers	65	Self report checklist PTSD based on SCID	6.2

Table 2 (Continued)

Author(s)	Trauma	Population	Sample	Method of measurement	Prevalence (%)
Landolt et al. (1998)	Serious accidents, cancer	Children	23 high-risk group (HRG = accidents) vs. 11 low-risk group (LRG = simple surgery)	Structured Clinical Interview based on DSM-IV, Kinder-DIPS	52 vs. 9
Widom(1999)	Child abuse and neglect	Adults	676	Diagnostic Interview Schedule (DIS) for PTSD	Current 30.9, lifetime 17.8
		Parents	HRG parents vs. LRG parents		68 vs. 20
Stretch et al. (1998)	Premilitary and military trauma	Soldiers	573 women	National Womens Study PTSD module based on DSM-III	Current 8.6, lifetime 37.5
			555 men		Current 5, lifetime 27.6
Taal and Faber (1998)	Burns	Hospitalized adults	174	Symptoms checklist for PTSD	31.3
Thompson and Kingree (1998)	Sexual and physical assault	Pregnant substance abusers	96	Traumatic Stress Schedule, Civilian Mississippi Scale for combat-related PTSD	62 (lifetime)
Green et al. (1998)	Early-stage breast cancer	Women	160	Trauma History Questionnaire, SCID-PTSD module	Cancer related: current 2.5, lifetime 5
Farley and Barkan (1998)	Childhood abuse, violence in prostitution	Prostitutes	130	Questionnaire, PTSD checklist based on DSM-III-R	68 (current and lifetime)
Wijma et al. (1997)	Childbirth	Women	1640	Interview based on DSM-IV criteria	1.7
Robin et al. (1997)	Civilian trauma (physical assault, combat, accidents, death)	Adults in Indian community	247	SCID-PTSD module based on SM-III-R	21.9 (17.9 lifetime vs. 5.7 current in men; 25.4 lifetime vs. 6.2 current in women)
Engdahl et al. (1997)	War	Former POW	262	Structured Clinical Interview based on DSM-III and DSM-IV criteria	Lifetime 53, current 29
Bisson et al. (1997)	Facial trauma (assaults, accidents)	Adults	47	PSS-I	27 (after 7 weeks)
Deykin and Buka (1997)	Rape, assault	Drug-dependent children (15–19 years)	297	DIS based on DSM-III criteria	Lifetime: 45.3 in men vs. 24.3 in women
Steiner et al. (1997)	Violent offenses	Children (16 years)	85	SCID-PTSD module	Current 31.7
Shaw et al. (1996)	Hurricane	School-age children (7–13 years)	30	Pynoos PTSD Reaction Index	70
Schnurr et al. (1996)	Mustard gas test WWII	Adult men	24	SCID-PTSD module	Current 17, lifetime 17

PTSD, post-traumatic stress disorder; SCID, Structured Clinical Interview for DSM; DIS, Diagnostic Interview Schedule; CAPS, Clinician Administered PTSD Scale; PSS, PTSD Symptom Scale-Interview.

prevalence of PTSD varies (from 1.7%–86%) according to the population studied, the nature of the traumatic stress, length of follow-up (not included in the table), and the instrument used. Most of these recent studies do report variables that are necessary to make an estimate of prevalence and to interpret the relevance in relation to the course of PTSD: nature or type of the trauma, sample size, recruitment of study population (help seeking, mailing, or advertisement), method of assessment, interval between stress exposure and assessment, lifetime and current PTSD. However, better standardization is needed to make comparisons between studies.

5 Etiology

There are multiple variables that play a role in the development of trauma-related psychopathology. Models of PTSD take into account genetic constitution, prestressor factors (defined as vulnerability), peritraumatic factors (characteristics of stress, of length of dissociative response), and poststressor factors in the development of PTSD (for a review, see Bremner et al. 1995a).

5.1 Genetic Contribution

Genetic contributions to PTSD have been studied using monozygotic twins. Goldberg et al. (1990) studied the impact of military service during the Vietnam era (1965–1975) on PTSD using a sample of 2092 male–male, monozygotic, veteran twin pairs. In 715 monozygotic twin pairs who were discordant for military service in Southeast Asia (SEA), PTSD was found to be strongly associated with military service in SEA. The prevalence of PTSD was 16.8% in twins who served in SEA compared with 5.0% in co-twins who did not serve in SEA. There was a ninefold increase in the prevalence of PTSD comparing twins who experienced high levels of combat with their co-twin who did not serve in SEA (Goldberg et al. 1990). True et al. (1993) followed up on this and looked at effects of heredity, shared environment, and unique environment on the liability for 15 self-reported PTSD symptoms. He studied 4042 Vietnam era veteran monozygotic and dizygotic male twin pairs. Quantitative genetic analysis revealed that inheritance had a substantial influence on liability for all symptoms. This study showed (a) that there are significant genetic influences on symptom liability, even after adjusting for differences in combat exposure, and (b) that genetic factors account for 13%–30% of the variance in liability for symptoms in the reexperiencing cluster, 30%–34% for symptoms in the avoidance cluster, and 28%–32% for symptoms in the arousal cluster (True et al. 1993). These findings were supported by a study conducted by Skre et al. (1993), who found a genetic contribution in the etiology of panic disorder, generalized anxiety disorder, and PTSD. Simple and social phobia were found to be mainly caused by environmental experiences.

5.2 Early Trauma

As stated earlier, early trauma in the form of childhood physical or sexual abuse has been associated with adult psychopathology in the stress-sensitization model: prior stressors increase risk for PTSD with reexposure to stress (for reviews, see Bremner et al. 1999a; Bremner 1999b). Consistent with this, Vietnam veterans with PTSD had higher rates of childhood physical abuse than Vietnam veterans without PTSD (26% vs. 7%). The association between childhood abuse and PTSD persisted after controlling for the difference in level of combat exposure between the two groups. Patients with PTSD also had a significantly higher rate of total traumatic events before joining the military than patients without PTSD. This supported the notion that childhood physical abuse may be an antecedent to the development of combat-related PTSD in Vietnam combat veterans (Bremner et al. 1993a).

In a large epidemiological study, Breslau et al. (1999a) found that subjects who experienced multiple events involving assaultive violence in childhood were more likely to experience PTSD from trauma in adulthood. Furthermore, previous events involving assaultive violence – single or multiple, in childhood or later on – were also associated with a higher risk of PTSD in adulthood (Breslau et al 1999a).

5.3 Predictive Factors

Predicting who would develop PTSD in response to a stressor can have many implications for the development of screening instruments, treatment strategies, and policy. The influence of Freud's original and somewhat misleading "war neurosis" led the military to attempt to predict vulnerability to combat stress based on premilitary developmental variables assuming they could be found in conflicts in early development. In different epidemiological studies, prestressor factors in the development of combat-related PTSD are found to be age of traumatization, years of education, minority status, pretrauma and psychiatric disorder, substance abuse, history of childhood abuse, antisocial

behavior, academic difficulty, family history of psychiatric disorders, family environment, and lack of social support. In civilian PTSD, a previous history of stress and previous psychiatric disorders may increase both the risk of exposure to a traumatic stressor as well as the development of PTSD. Factors in PTSD associated with combat veterans are severity and length of exposure, seeing others killed or wounded, participating in atrocities, and dissociation at time of trauma. Research has not found a direct association between drug and alcohol abuse in Vietnam combat veterans (Fontana and Rosenheck 1994). A postmilitary risk factor is lack of social support. Since most victims of civilian trauma, such as rape or childhood sexual abuse, feel shame and guilt, which discourages them from talking openly about their feelings and seeking social support, this may contribute to a decrease in social support analogous to that of combat veterans with PTSD (for a review, see Bremner et al. 1995a).

6 Diagnosis and Assessment

Trauma measures vary widely in scope and format, ranging from self-report checklists assessing the presence or absence of a limited range of potentially traumatic events to comprehensive protocols assessing a wide range of stressors through both self-report and interview (for an overview of psychological assessment in PTSD, the reader is referred to Wilson and Keane 1997; for an overview of assessment of PTSD in children, see March 1999; on assessment in traumatized adults, see Weathers and Keane 1999; on forensic assessment, see Sparr and Pitman 1999). Recommended clinician-administered instruments with good reliability and validity include the following:

1. *Structured Clinician Interview for DSM-IIIR and DSM-IV* (SCID; Spitzer et al. 1990). The SCID is a comprehensive structured interview that assesses all of the major axis I disorders. A revised module of the original DSM-III-R PTSD module is now introduced in the DSM-IV section of the SCID. This module appears to have adequate reliability and validity (Kulka et al. 1991).
2. *Diagnostic Interview Schedule* (DIS; Robins et al. 1981). The DIS is a highly structured, comprehensive interview designed for use by lay interviewers in the context of epidemiological research. Like the SCID, the DIS provides a standard prompt question for each of the 17 PTSD symptoms. Each symptom is scored dichotomously on its presence or its absence. The DIS was found to be one of the best predictors of a clinical diagnosis of PTSD (Kulka et al. 1991).
3. *Clinician-Administered PTSD Scale* (CAPS; Blake et al. 1995). The CAPS is a comprehensive scale developed by the U.S. National Center for PTSD (Blake et al. 1990 1995). It is intended to be used by clinicians and addresses some of the limitations of the other instruments. It assesses the 17 core symptoms for PTSD, as well as associated symptoms, response validity, overall symptom severity, and the impact of symptoms on social and occupational functioning. The CAPS assesses the frequency and intensity of each symptom on a 5-point scale, yielding continuous and dichotomous scores for each symptom and across the 17 symptoms. The CAPS also contains behaviorally anchored prompt questions and rating scales to help increase the reliability of symptom inquiry and severity ratings. These are specific guidelines for assessing lifetime diagnostic status. With its recent revision for DSM-IV, the CAPS is divided into CAPS-DX, for diagnostic assessment of current and lifetime PTSD, and CAPS-SX, assessing a 1-week symptom status. The CAPS takes significantly more time to administer than the other interviews.
4. *PTSD Symptom Scale-Interview* (PSS-I; Foa et al. 1993). The PSS-I is a structured interview specifically designed to assess DSM-III-R PTSD symptoms (Foa et al. 1993). It contains 17 items; interviewers rate the severity of each symptom over the past 2 weeks. The severity of each symptom is rated on a scale from 0 to 3, where 0 is "not at all," and 3 "very much." A PTSD diagnosis is obtained by considering symptom ratings of 1 or higher as present and then following the DSM-III algorithm. The PSS-I has excellent psychometric properties. The advantages of this scale are that it yields continuous and dichotomous scores, is easy to administer, and has good reliability and validity. The disadvantages are that it only uses one single prompt for each item, its rating anchors are not explicitly defined, and it assesses symptoms over a 2-week rather than a 1-month period. The instrument lacks lifetime diagnostic status.
5. *Structured Interview for PTSD* (SI-PTSD; Davidson et al. 1989). The SI-PTSD was originally designed to assess both DSM-III and DSM-III-R criteria for PTSD (Davidson et al. 1989). Items consist of initial prompt questions and follow-up questions that clarify the initial question with concrete behavioral samples. The severity is rated on a 5-point scale, both for the past month and for the worst period since the trauma. Descriptors are given for scale anchors. A total severity score is obtained by summing ratings of all 17 symptoms, and symptoms are counted if they are scored as 2 or higher.
6. *PTSD Interview* (PTSD-I; Watson et al. 1991). The PTSD-I is another structured interview for assessing

the DSM-III-R criteria for PTSD (Watson et al. 1991). The scale follows the same course as the other scales except its recommended format for administration. Interviewers are instructed to give respondents a copy of the rating scale, read the questions aloud, and ask respondents to rate themselves.

In order to assess traumatic events that may account for later-life PTSD, we developed a semistructured interview (Early Trauma Inventory, ETI). This assesses traumatic life events in the first 18 years of life. The questionnaire gives a composite score of four domains of early trauma: natural disasters, emotional trauma, physical trauma, and sexual trauma. The interview combines narrative and structured approaches to inquiring about childhood emotional, physical, and sexual abuse as well as nonabusive traumas. Each subsection begins with an open-ended format in which subjects are asked in a general way about their experiences related to that domain and are allowed to tell their story in their own words. Following this initial introduction, subjects are assessed using a series of structured questions within that particular domain. The ETI assesses a wide range of abuse experiences, including physical abuse items, such as a more common "were you ever spanked with a hand?" as well as less common events, such as "were you ever locked in a closet?" The ETI has a range of sexual abuse items, including "were you ever exposed to someone flashing" and "were you ever forced to have anal sex against your will?" Emotional abuse items range from "were you often shouted at?" to "did your parents or caretakers fail to understand your needs?" The general trauma component assesses events ranging from parental loss and natural disaster to criminal victimization. Data show excellent interrater reliability. Test–retest studies performed in a patient and nonpatient population show excellent reliability (Bremner et al. in press).

There is now considerable controversy about the validity of recall of childhood abuse. There have been studies which reported long-term correlates of abuse events which were documented at the time which they occurred (Chu et al. 1999). The relationship between reported abuse and the historical accuracy of abuse events is not known. Our findings of a strong level of agreement between test and retest administration of ETI, however, supports the stability of recalled memories of abuse and other traumas over time. Regardless of the controversy surrounding the accuracy of reported abuse, the development of standardized instruments for the assessment of reported abuse and other traumas, with demonstrated reliability and validity, will be beneficial in furthering careful research in this area (Bremner et al. 1999a).

Breslau et al. (1999b) report a short screening scale consisting of seven screening symptoms for use in computer-assisted telephone interviews. They chose five of the symptoms from the avoidance and numbing group, and two from the hyperarousal group. A score of 4 or greater on this scale defined positive cases of PTSD with a sensitivity of 80%, a specificity of 97%, a positive predictive value of 71%, and a negative predictive value of 98%, showing that this short screening scale can be an efficient method to screen for PTSD in epidemiologic and clinical studies with limited burden on respondents (Breslau et al. 1999b).

Other instruments are summarized in Stamm's *Measurement of Stress, Trauma,* and Adaptation (Stamm 1996). This book covers a rich variety of structured interviews, paper and pencil tests, and psychobiological tests. Instruments are classified according to many characteristics, including their psychometric maturity, applicability across race, sex, ethnicity, language, suitability for age-groups, types of data yielded, diagnosis, and theoretical perspectives.

7 Brain, Neurohormonal, and Transmitter Regulations

Long-standing alterations in biological stress response systems may underlie symptoms of PTSD (for an overview, see Friedman et al. 1995; Charney and Bremner 2000). Research in animals, in a paradigm of maternal deprivation of exposure to inescapable stress, has been highly relevant to our understanding of PTSD, e.g. in the model of inescapable stress, animals are exposed to repeated stressors, such as electric foot shock or being forced to swim in cold water. These stressors result in an acute release of stress-related neuropeptides, hormones, and transmitters, including corticotropin-releasing factor (CRF) – which activates the hypothalamic–pituitary–adrenal (HPA) axis to cause release of adrenocorticotropic hormone (ACTH) and cortisol – norepinephrine, benzodiazepines, serotonin (5-HT), dopamine, and opiates. Animals exposed to chronic stressors have shown to develop long-term dysregulation of these systems and typically potentiated release of transmitters with exposure to subsequent stressors. These neurochemical changes are also accompanied by behavioral changes (sometimes referred to as "learned helplessness") that are similar to human anxiety, including increased defecation and avoidance of novel stimuli such as an open field. Thus chronic stress results in long-term abnormalities in the neurochem-

ical systems that are necessary for appropriately coping with stressful situations. Specific brain areas that play an important role in a variety of types of memory are preferentially affected by stress (and mediate the stress response), including amygdala, hippocampus, hypothalamus, thalamus, medial prefrontal and parietal cortex, visual association cortex, and cingulate (Bremner et al. 1999b).

7.1 The Hypothalamic–Pituitary–Adrenal Axis in the Stress Response

The HPA axis is an important component of the stress response system. CRF is released from the paraventricular nucleus (PVN) of the hypothalamus, causing release of ACTH from the pituitary, which stimulates release of cortisol (the major stress hormone) from the adrenals. This axis is involved in a negative feedback loop that regulates cortisol release (as well as regulatory feedback with the noradrenergic system, which is discussed in more detail below).

7.1.1 Adrenocorticotropic Hormone

Acute stress of many types results in release of CRF as well as ACTH and cortisol. The mechanism responsible for transient stress-induced hyperadrenocorticism and feedback resistance may involve a downregulation of glucocorticoid receptors (J.P. Herman et al. 1989; Sapolsky and Plotsky 1990). High glucocorticoid levels (such as those elicited by acute stress) decrease the number of hippocampal glucocorticoid receptors, resulting in increased corticosterone secretion and feedback resistance. Following stress termination, when glucocorticoid levels decrease, receptor numbers are increased and feedback sensitivity normalizes (Sapolsky et al. 1984a/b). The effects of chronic stress on ACTH and corticosterone secretion vary depending on the experimental paradigm. It has been reported that an adaptation to chronic stress may occur, resulting in decreased plasma ACTH and corticosterone levels compared with levels following a single stressor (Kant et al. 1987). However, other investigations have revealed enhanced corticosterone secretion after chronic stressor regimens (Irwin et al. 1986). There is also evidence that the experience of prior stress may result in augmented corticosterone responses to a subsequent stress exposure (Dallman and Jones 1973; Caggiula et al. 1989). It is not known which factors determine whether adaptation or sensitization of glucocorticoid activity will occur following chronic stress (Yehuda et al. 1991a). The HPA axis has important functional interactions with the norepinephrine system that facilitate a sophisticated range of responses to stress. Glucocorticoids inhibit stress-induced activation of catecholamine synthesis in the PVN (Pacak et al. 1993; 1995; Vetrugno et al. 1993). CRF increases activity of the locus ceruleus (Valentino and Foote 1988), and CRF injected into the locus ceruleus intensifies anxiety-related responses (Butler et al. 1990). These findings support the notion that CRF serves as an excitatory neurotransmitter in the locus ceruleus, which may represent the pathway for the behavioral effects of CRF.

7.1.2 Norepinephrine

Norepinephrine release in the brain represents an important part of the stress response (for reviews, see Bremner et al. 1996b,c). The majority of noradrenergic cell bodies are located in the brainstem, in the locus ceruleus region of the pons, with axons that extend throughout the cerebral cortex and to multiple subcortical areas. Neurons in the locus ceruleus are activated in association with fear and anxiety states (Redmond 1987; Abercrombie and Jacobs 1987), and the limbic and cortical regions innervated by the locus ceruleus are those thought to be involved in the elaboration of adaptive responses to stress. Stressors such as a cat seeing a dog result in and increase in firing of neurons in the locus ceruleus (Levine et al. 1990) and enhanced norepinephrine release in the hippocampus (Nisenbaum et al. 1991; Petty et al. 1993) and medial prefrontal cortex (Finlay et al. 1995).

Stress sensitization refers to a stressor-induced increase in behavioral, physiologic, and biochemical responding to subsequent stressors of the same or lesser magnitude. For example, chronically stressed animals had increased norepinephrine release in hippocampus with subsequent stressors (Abercrombie et al. 1989). Significantly greater behavioral, cardiovascular, and biochemical responses to equivalent doses of yohimbine in combat veterans with PTSD compared with healthy controls are an example of evidence supporting this model (Southwick et al. 1993, 1999). It has been proposed that multiple neurobiological systems, including catecholamine systems and the HPA axis, can become sensitized over time by traumatic stress and as a result contribute to PTSD symptoms such as hypervigilance, poor concentration, insomnia, exaggerated startle response, and intrusive memories. It has been proposed that this facilitation of memory is caused by endogenous neuromodulators, such as epinephrine and norepinephrine, which are released during arousing and stressful circumstances. Overstimulation of these stress-related neuromodulators during traumatic events can cause overconsolidation and deeply engraved memories for the event.

7.1.3 Dopamine, Serotonin, Benzodiazepines, and Neuropeptides

The role of dopamine, serotonin (5-HT), benzodiazepines and neuropeptides will not be reviewed here in detail. Readers are referred to recent findings in the dopaminergic innervation of the medial prefrontal cortex, which appears to be particularly vulnerable to stress (for a review, see Thierry et al. 1998). The effects of stress on serotonin systems have been studied less thoroughly than noradrenergic systems. Although there are only a limited number of studies of serotonergic function in PTSD (L.L. Davis et al. 1997; Southwick et al. 1997; Maes et al. 1999a), there is a large body of indirect evidence suggesting that this neurotransmitter may be important in the pathophysiology of trauma-related symptomatology. In humans, low 5-HT functioning has been associated with aggression, impulsivity, and suicidal behavior. Patients with PTSD are frequently described as aggressive or impulsive and often suffer from depression, suicidal tendencies, and intrusive thoughts that have been likened to obsessions. Endogenous benzodiazepines also play an important role in the stress response and anxiety (for a review, see Guidotti et al. 1990). Benzodiazepine receptors are present throughout the brain with the highest concentration in cortical gray matter. Several neuropeptides also mediate the response to stress. Cholecystokinin (CCK) is an anxiogenic neuropeptide present in the gastrointestinal tract as well as the brain and has recently been suggested as a neural substrate for human anxiety. Stress is associated with an increase in endogenous opiate release (Maier et al. 1981; Madden et al. 1977) with decreased density of μ-opiate receptors (Stuckey et al. 1989), which may mediate the analgesia associated with stress. Other neuropeptides under investigation that appear to play role in the stress response are neuropeptide Y, somatostatin, and thyrotropin. Stress also has important effects on the immune system that are not reviewed here in detail.

7.2 Linkage with Other Neurobiological Systems

Coordinated functional interactions between the HPA axis and noradrenergic neuronal systems may be critical in promoting adaptive responses to stress, anxiety, or fear. CRF increases locus ceruleus firing, resulting in enhanced norepinephrine release in cortical and subcortical areas throughout the brain. The PVN of the hypothalamus, the site of the majority of CRF-containing neurons in the hypothalamus, is an important site in effecting cardiovascular and neuroendocrine responses to stress. Norepinephrine increases CRF in the PVN of the hypothalamus. In chronically stressed animals, the locus ceruleus (as opposed to other norepinephrine neurons in the medulla) may be preferentially responsible for norepinephrine release in the PVN. Increased CRF release from the PVN results in stimulation of ACTH secretion from the pituitary and consequently cortisol release from the adrenal gland. High levels of circulating cortisol act through a negative feedback pathway to decrease both CRF and norepinephrine synthesis at the level of the PVN. Glucocorticoid inhibition of norepinephrine-induced CRF stimulation may be evident primarily during stressor-induced cortisol release and not under resting conditions. High levels of cortisol likely inhibit the effects of norepinephrine on CRF release from the PVN, serving to restrain the stress-induced neuroendocrine and cardiovascular effects mediated by the PVN. Norepinephrine, cortisol, and CRF thus appear to be tightly linked in a functional system that offers a broad homeostatic mechanism for coping with stress.

7.3 Brain Circuitry

Based on animal research and emerging work in clinical neuroscience of PTSD, a model for a neural circuitry of anxiety and fear that is also applicable to PTSD has been described (Charney and Bremner, 2000). The model explains how information related to a threatening stimulus (e.g. someone approaches you with a gun in a dark alley) enters the primary senses (smell, sight, touch, hearing), is integrated into a coherent image that is grounded in space and time, activates memory traces of prior similar experiences with the appropriate emotional valence (necessary in order to evaluate the true threat potential of the stimulus), and subsequently triggers an appropriate motor response. Specific brain circuits that mediate these responses make up the neural circuitry of anxiety and fear and can be adapted to the field of traumatic stress (see Fig. 4). Critical processes in these circuits include failure of inhibition, stress sensitization, and fear condition. Critical brain structures are hippocampus, amygdala, and orbitifrontal cortex.

In the development of human fear or anxiety, afferent sensory input enters through the eyes, ears, nose, sense of touch, the body's own visceral information, or any combination of these. These sensory inputs are relayed through the dorsal thalamus to cortical brain areas, such as primary visual (occipital), auditory (temporal), or tactile (postcentral gyrus) cortical areas. Olfactory sensory input, however, has

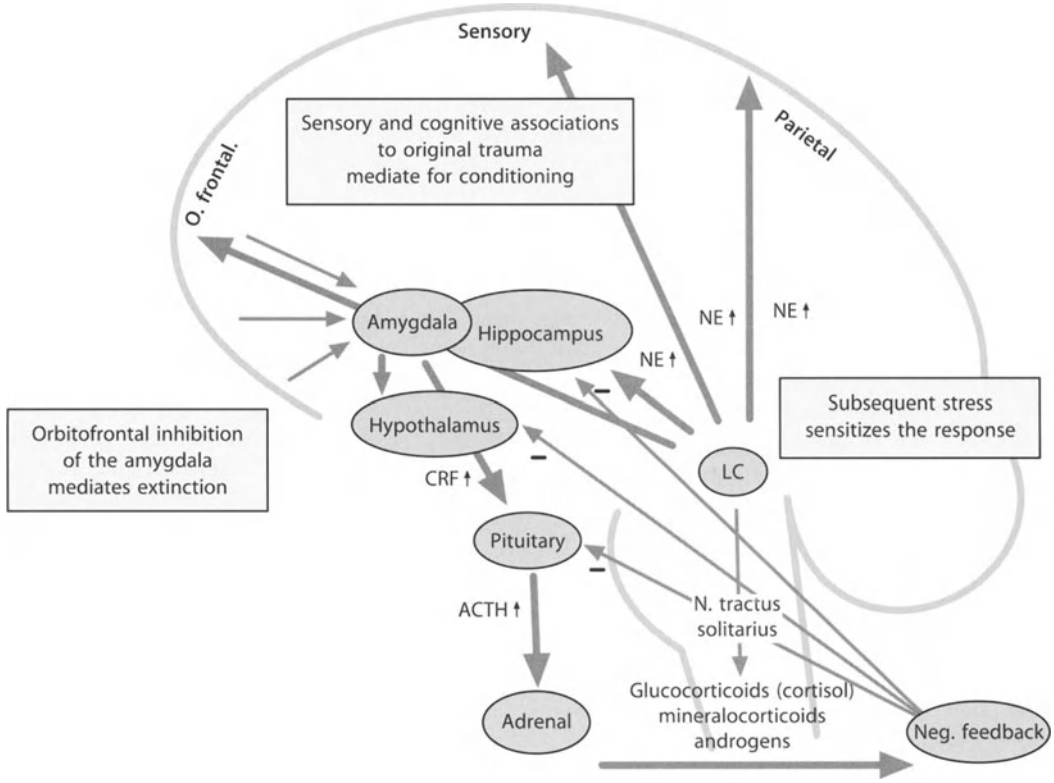


Fig. 4. Psychobiological mechanisms and their site of action, including the hypothalamic–pituitary–adrenal (HPA) axis, in PTSD. CRF, corticotropin-releasing factor; *ACTH*, adenocorticotropic hormone; *NE*, norepinephrine; *LC*, locus ceruleus. (Based on Charney et al. 1993)

direct inputs to the amygdala and entorhinal cortex (Turner et al. 1978). Input from peripheral visceral organs is relayed in the brainstem to the locus ceruleus, the site of the majority of the brain's noradrenergic neurons, and from here to central brain areas.

These brain areas have projections to multiple areas, including amygdala, hippocampus, entorhinal cortex, orbitofrontal cortex, and cingulate, that are involved in mediating memory and emotion (Vogt and Miller 1983). Cognitive appraisal of potential threat is also an important aspect of the stress response. The cognitive response to threat involves placing the threatening object in space and time. Specific brain areas are involved in these functions (such as localizing an object in space, visuospatial processing, memory, cognition, action, or planning). The anterior cingulate gyrus (Brodmann area 32) is involved in selection of responses for action as well as emotion (Devinsky et al. 1995). This area and other medial portions of the prefrontal cortex, including area 25 and orbitofrontal cortex, modulate emotional and physiological responses to stress and are discussed in more detail below.

Another important aspect of the stress response is incorporation of a person's prior experience (memory) into the cognitive appraisal of stimuli. For example, if you are approached in a potentially threatening situation, it will be important to determine whether the face of the person is someone known to you or is a stranger who may be more threatening. In addition, it is important to place the situation in time and place. Entering a dark alleyway may trigger prior memories of being robbed, with associated negative emotions and physiological arousal. These memories may have survival value, in that the individual will avoid the situation where the previous negative event took place. Finally, it is critical to effectively lay down memory traces related to a potential threat in order to avoid this type of threat in the future.

7.3.1 Hippocampus

The hippocampus, which is particularly vulnerable to stress, plays an important role in memory. The

hippocampus and adjacent cortex mediate declarative memory function (e.g. recall of facts and lists) and has been hypothesized to play an important role in integration of memory elements at the time of retrieval and in assigning significance for events within space and time (Squire and Zola-Morgan 1991). The hippocampus also plays an important role in mediating emotional responses to the context of a stressor, e.g. in animal studies, lesions of the hippocampus disrupted the formation of emotional memories of the context (i.e. the box) where the stressor (i.e. electric footshock) took place (Ross and Randich 1984; Phillips and LeDoux 1992; Kim and Fanselow 1992). High levels of glucocorticoids released during stress were also associated with damage to the CA3 region of the hippocampus (Uno et al. 1989; Sapolsky et al. 1990) as well as related deficits in memory (Arbel et al. 1994; Luine et al. 1994; McEwen et al. 1992; reviewed in more detail below).

With long-term storage, memories are felt to be shifted from hippocampus to the neocortical areas, where the initial sensory impressions take place (Squire and Zola-Morgan 1991). The shift in memory storage to the cortex may represent a shift from conscious representational memory to unconscious memory processes that indirectly affect behavior. "Traumatic cues" such as a particular sight or sound reminiscent of the original traumatic event will trigger a cascade of anxiety and fear-related symptoms will ensue, often without conscious recall of the original traumatic event. In patients with PTSD, however, the traumatic stimulus is always potentially identifiable. Symptoms of anxiety in panic or phobic disorder patients, however, may be related to fear responses to a traumatic cue (in individuals who are vulnerable to increased fear responsiveness, either through constitution or previous experience), where there is no possibility that the original fear-inducing stimulus will ever be identified.

7.3.2 Amygdala

The amygdala is involved in memory for the emotional valence of events. The paradigm of conditioned fear has been utilized as an animal model for stress-induced abnormalities of emotional memory (M. Davis 1992). Conditioned fear, in which pairing of a neutral ("conditioned") stimulus to a fear-inducing ("unconditioned") stimulus results in fear responses to the neutral (conditioned) stimulus alone, has been used as a probe of amygdala function (LeDoux 1993). Lesions of the central nucleus of the amygdala have been shown to completely block fear conditioning (Hitchcock and Davis 1986; Hitchcock et al. 1989), while electrical stimulation of the central nucleus increases acoustic startle (Rosen and Davis 1988). The central nucleus of the amygdala projects to a variety of brain structures via the stria terminalis and the ventral amygdalofugal pathway. One pathway is from the central nucleus to the brainstem startle reflex circuit (nucleus reticularis pontis caudalis). Pathways from the amygdala to the lateral hypothalamus effect peripheral sympathetic responses to stress (Iwata et al. 1986). Electrical stimulation of the amygdala in cats resulted in peripheral signs of autonomic hyperactivity and fear-related behaviors seen in the wild when the animal is being attacked or is attacking, including alerting, chewing, salivation, piloerection, turning, facial twitching, arching of the back, hissing, and snarling, associated with an increase in catecholamine turnover (Hilton and Zbrozna 1963). Electrical stimulation of the amygdala in human subjects resulted in signs and symptoms of fear and anxiety, including an increase in heart rate and blood pressure, increased muscle tension, subjective sensations of fear or anxiety, and increases in peripheral catecholamines (Chapman et al. 1954; Gunne and Reis 1963). These findings demonstrated that the amygdala plays an important role in conditioned fear and emotional responding, as well as modulating peripheral stress responses. There are also important connections between cortical association areas, thalamus, and amygdala that are important in shaping the emotional valence of the cognitive response to stressful stimuli (LeDoux et al. 1988). In addition to thalamo-cortico-amygdala connections, there are direct pathways from thalamus to amygdala, which could account for fear responding below the level of conscious awareness (Romanski and LeDoux 1992).

7.3.3 Frontal Cortex

Frontal cortical areas modulate emotional responsiveness through inhibition of amygdala function, and we have hypothesized that dysfunction in these regions may underlie pathological emotional responses in patients with PTSD and possibly other anxiety disorders. Medial prefrontal cortex (area 25) (subcallosal gyrus) has projections to the amygdala which are involved in the suppression of amygdala responsiveness to fearful cues. Dysfunction of this area may be responsible for the failure of extinction to fearful cues, which is an important part of the anxiety response (Morgan and LeDoux 1995). This area is involved in regulation of peripheral responses to stress, including heart rate, blood pressure, and cortisol response (Roth et al. 1988). Finally, case studies of humans with brain lesions have implicated medial prefrontal cortex (including orbitofrontal cortex, area 25, and anterior cingulate area 32) in "emotion" and socially appro-

priate interactions (Damasio et al. 1994). Auditory association areas (temporal lobe) have also been implicated in animal studies as mediating extinction to fear responding (Jarrell et al. 1987; Romanski and LeDoux 1993). As reviewed later, we found dysfunction of medial prefrontal cortex and auditory cortex with traumatic reminders in PTSD (Bremner et al. 1997b, 1999c,d).

7.3.4 Other Brain Structures

A final component of the stress response involves preparation for a response to potential threat. Preparation for responding to threat requires integration between brain areas involved in assessing and interpreting the potentially threatening stimulus, and brain areas involved in response. For instance, prefrontal cortex and anterior cingulate play an important role in the planning of action and in holding multiple pieces of information in "working memory" during the execution of a response (Goldman-Rakic 1988). Parietal cortex and posterior cingulate are involved in visuospatial processing, which is an important component of the stress response. Motor cortex may represent the neural substrate of planning for action. The cerebellum has a well-known role in motor movement, which would suggest that this region is involved in planning for action; however, recent imaging studies are consistent with a role in cognition as well (Ashkoomoff and Courchesne 1992). Connections between parietal and prefrontal cortex are required in order to permit the organism to rapidly and efficiently execute motor responses to threat. It is therefore not surprising that these areas have important innervations to precentral (motor) cortex, which is responsible for skeletal motor responses to threat, which facilitate survival. The striatum (caudate and putamen) modulates motor responses to stress. The dense innervation of the striatum and prefrontal cortex by the amygdala indicates that the amygdala can regulate both of these systems. These interactions between the amygdala and the extrapyramidal motor system may be very important for generating motor responses to threatening stimuli, especially those related to prior adverse experiences (McDonald 1991a,b).

7.4
Developmental and Hereditary Factors

As reported earlier, it has been estimated that between 13% and 34% of the variance for specific PTSD symptom clusters is genetically transmitted. It is possible that inherited variations in the capacity for fear conditioning, sensitization, and memory consolidation help to determine who is likely to develop PTSD. From a developmental standpoint, it has been proposed that psychological trauma in childhood may differentially affect maturation of various brain regions, particularly those regions involved in fear and alarm (De Bellis et al. 1999a,b).

8
Long-Term Alterations in Neurobiological Systems

8.1
Hypothalamic–Pituitary–Adrenal Axis

Alterations in HPA axis function have been demonstrated in PTSD (for a review, see Yehuda et al. 1991b). A decrease in urinary cortisol levels has been found in Vietnam veterans with chronic PTSD in comparison to controls and patients with other psychiatric disorders in some studies (Mason et al. 1986; Yehuda et al. 1991a; Yehuda et al. 1993) but not others (Pitman and Orr 1990; Mason et al. 1988), while decreased plasma cortisol was found in 24-h sampling in patients with combat-related PTSD relative to healthy controls and patients with depression (Yehuda et al. 1994). On the other hand, women with a history of childhood sexual abuse-related PTSD (Lemieux and Coe 1995) and patients with PTSD related to a natural disaster (Baum et al. 1993) had elevated levels of urinary cortisol relative to controls. Male patients with combat-related PTSD (Kosten et al. 1990; Kudler et al. 1987; Olivera and Fero 1990) and female patients with sexual assault-related PTSD (Dinan et al. 1990) have been shown to suppress cortisol normally with the standard 1 mg dexamethasone suppression test (DST). Studies utilizing lower doses of dexamethasone (0.5 mg) suggest that PTSD may be associated with a super-suppression of the cortisol response in comparison to normal controls (Yehuda et al. 1993; Stein et al. 1997), which appears to be the opposite of patients with major depression who are nonsuppressers with the standard 1 mg DST test. PTSD patients have also been found to have a significantly lower ("blunted") ACTH response to CRF than controls, suggesting an increased release of neuronal CRF. Consistent with this are findings of elevated levels of CRF in the cerebrospinal fluid of Vietnam combat veterans with PTSD relative to healthy subjects (Bremner et al. 1997a; Baker et al. 1999). Other studies showed that patients with combat-related PTSD had an increase in lymphocyte glucocorticoid receptors in comparison to healthy subjects, non-PTSD combat veterans, and patients with other psychiatric disorders (Yehuda et al. 1991a, 1993). These studies demonstrate that alterations in cortisol

and HPA axis function are associated with PTSD. One possible explanation of clinical findings to date is an increase in neuronal CRF release, with resultant blunting of ACTH response to CRF, increased central glucocorticoid receptor responsiveness, and resultant low levels of peripheral cortisol due to enhanced negative feedback. Interestingly, nonhuman primates with variable foraging mothers (a model for early-life stress) had elevated cerebrospinal fluid CRF and decreased CSF cortisol levels in adulthood, a picture that is closer to PTSD than depression (Coplan et al. 1996).

Studies showing decreased cortisol in chronic PTSD raise the question of how elevated cortisol can represent the etiology of hippocampal atrophy in PTSD. We have hypothesized that high levels of cortisol *at the time of the stressor* result in damage to hippocampal neurons which can persist for many years after the original trauma, leading to reductions in hippocampal volume as measured by magnetic resonance imaging (MRI) (Bremner et al. 1995c, 1997c). In this scenario, decreased cortisol characterizes the chronic stages of the disorder due to adaptation and long-term changes in cortisol regulation. Longitudinal studies of cortisol in sexually abused girls supports an elevation in cortisol around the time of the stressor, with decreased cortisol developing later in development in patients who develop chronic symptoms of PTSD (F. Putnam, personal communication). However, an alternative hypothesis for hippocampal atrophy is that small hippocampal volume, which is present from birth, is a risk factor for the development of PTSD – in this scenario, high levels of cortisol associated with stress would have nothing to do with hippocampal atrophy.

Findings of no increase in cortisol levels in the aftermath of rape in women who subsequently develop PTSD (Yehuda et al. 1998) have also been used to argue against the glucocorticoid hypothesis of stress-induced hippocampal damage. In an initial report, cortisol samples obtained days to weeks after exposure to the trauma of rape found a relationship between low cortisol in the aftermath of rape and prior history of trauma (Resnick et al. 1995). In a subsequent analysis of samples obtained 8–48 h after the rape, the authors found a relationship between low cortisol levels in the aftermath of rape and prior history of trauma. However, there was no relationship between cortisol and risk for subsequent development of PTSD (Yehuda et al. 1998). Given the finding that history of prior trauma increases the risk for PTSD with subsequent victimization, this raises the question of whether these patients had prior PTSD or were physiologically distinct from the non-stress-exposed subjects.

In animal studies, it has been demonstrated that a variety of early stressors resulted in chronic increased levels of CRF and increased responsiveness of the HPA axis to subsequent stressors in adult life. As stated earlier, several observations suggest that early-life adverse experience permanently affects the HPA axis (Plotsky and Meaney 1993; van Oers et al. 1998). In human studies, women with a history of childhood sexual abuse-related PTSD were found to have elevated levels of the stress hormone cortisol in 24-h urine (Lemieux and Coe 1995). Sexually abused girls had a blunted ACTH response to CRF, consistent with hypersecretion of CRF (De Bellis et al 1994). Consistent with this, we found increased levels of CRF in the cerebrospinal fluid in PTSD (Bremner et al. 1997a). Children with PTSD had increased levels of cortisol (De Bellis et al. 1999a/b). Our own preliminary data in women with PTSD related to early childhood sexual abuse show decreased baseline cortisol based on 24-h diurnal assessments of plasma; cortisol levels were lower at all time points of CRF and ACTH challenge. There was a blunted ACTH response to CRF and a normal cortisol response to ACTH challenge. Studies using low doses (0.5 mg) of the DST suggest that adult women with a history of childhood abuse may show a super-suppression of the cortisol response in comparison to normal controls (Stein et al. 1997), which appears to be the opposite of patients with major depression. Children with abuse-related PTSD also had smaller intracranial and cerebral volume as well as smaller hippocampal volume (De Bellis et al. 1999a,b). These studies demonstrate that early-life trauma has lasting effects on neurochemical systems that mediate the stress response (Bremner et al. 1995b; Bremner 1999a,b) and that alterations in cortisol and HPA axis function are associated with abuse-related PTSD.

8.2 Norepinephrine

Several studies have shown long-term alterations in norepinephrine systems in PTSD. PTSD is characterized by tonic autonomic hyperarousal and increases in autonomic system activity in response to trauma-relevant stimuli. The most frequently used measures are electrodermal activity as presented by skin conductance levels; skin temperature; responses such as heart rate and systolic and diastolic blood pressure; and electromyography (EMG) activity of various facial muscles. These variables reflect in part activity of the peripheral sympathetic system. Exposure to traumatic reminders and neutral scenes utilized in the psychophysiology paradigm includes slides, sounds, or scenes similar to the original trauma or reading scripts that describe what actually happened during the original trauma. Comparisons are either made between expo-

sure to trauma-related material and both the baseline and/or the neutral exposures.

Over the past 20 years, a large number of psychophysiology studies have reported heightened sympathetic nervous system activity in veterans with PTSD. Although most studies have found no difference in resting baseline heart rate and blood pressure, the majority of studies have reported exaggerated increases in cardiovascular reactivity among subjects with PTSD compared with normal control subjects when exposed to trauma-specific stimuli such as laboratory-stimulated sights and sounds of combat or tape recordings of personally experienced traumas. Such exaggerated increases have not been found in combat veterans without PTSD, in combat veterans with anxiety disorders other than PTSD, or in response to generic stressors (such as the film of an automobile accident) that have never been experienced by the trauma survivor.

Studies on autonomic arousal show some variability in their reports. Most of the studies assessing physiological characteristics in PTSD have been conducted with veteran subjects and women with histories of childhood sexual abuse (Prins et al. 1995; Carlson et al. 1997; Metzger et al. 1999; Liberzon et al. 1999). A general finding in these studies was that PTSD patients showed heightened responsivity to trauma-related cues, consistent with increased norephinephrine responsivity. Patients with combat-related PTSD were found to have elevated norepinephrine and epinephrine in 24-h urine in comparison with normal controls and patients with other psychiatric disorders (Mason et al. 1988; Spivak et al. 1999). Relative elevations of the norepinephrine metabolite 3-methoxy-4-hydroxyphenylglycol (MHPG) were found in nighttime samples in PTSD (Mellman et al. 1995). No differences were found, however, in urinary norepinephrine between patients with combat-related PTSD and combat-exposed non-PTSD subjects (Pitman and Orr 1990) or in baseline levels of plasma norepinephrine in combat-related PTSD versus healthy subjects (for a review, see Murberg 1994). Women with PTSD secondary to childhood sexual abuse had significantly elevated levels of catecholamines (norepinephrine, epinephrine, dopamine) and cortisol in 24-h urine samples (Lemieux and Coe 1995). Sexually abused girls excreted significantly greater amounts of catecholamine metabolites, metanephrine, vanillymandelic acid, and homovanillic acid (HVA) than non-sexually abused girls (DeBellis et al. 1994) and increased in abuse-related PTSD (De Bellis et al. 1999a,b). Exposure to traumatic reminders in the form of combat films resulted in increased epinephrine and norepinephrine release, and increased MHPG with physical exercise (Hamner and Hitri 1992) has been found in Vietnam veterans with PTSD in comparison to healthy subjects. Children with PTSD were found to have increased orthostatic heart rate response, suggesting noradrenergic dysregulation (Perry 1994).

Studies of peripheral norepinephrine receptor function have also shown alterations in α_2-receptor and cyclic adenosine 3′,5′-monophosphate (cAMP) function in patients with PTSD, which are similar to those in panic disorder. A decrease in platelet adrenergic α_2-receptor number as measured by total binding sites for the α_2-antagonist [3H]rauwolscine (Perry and U'Prichard 1981), and a significantly greater reduction in number of platelet a_2-receptors after exposure to agonist (epinephrine), has been observed in PTSD patients in comparison to healthy controls (Perry et al. 1987). A decrease in platelet basal adenosine, isoproterenol, and forskolin-stimulated cAMP signal transduction and basal platelet monoamine oxidase (MAO) activity (Davidson et al. 1985) was found in PTSD patients in comparison to controls. These findings may reflect chronic high levels of norepinephrine release, which lead to compensatory receptor downregulation and decreased responsiveness.

Patients with combat-related PTSD compared to healthy controls had enhanced behavioral, biochemical (norepinephrine metabolite MHPG), and cardiovascular (heart rate and blood pressure) responses to the α_2-antagonist yohimbine, which stimulates central norepinephrine release (Southwick et al. 1993). Moreover, as noted previously, a recent positron emission tomography (PET) study demonstrated that PTSD patients have a cerebral metabolic response to yohimbine consistent with increased norepinephrine release, showing failure of activation in medial prefrontal cortex and decreased metabolism in hippocampus (Bremner et al. 1997b). In summary, although there is inconsistent evidence for elevations in norepinephrine at baseline in PTSD, there is evidence for increased noradrenergic responsivity in this disorder.

8.3 Memory Function

There are a number of reasons why the investigation of memory function is clinically relevant to PTSD. As reported earlier, patients with combat-related PTSD (as well as PTSD related to other causes) demonstrate alterations in memory, including nightmares, flashbacks, intrusive memories, and amnesia for the trauma (Figley 1978b; Egendorf et al. 1981). In addition, descriptions from all wars of this century document alterations in memory occurring in combat veterans during or after the stress of battle. These include the forgetting of one's name or identity and the forgetting of events which had just taken place during the previous battle. Vietnam veterans with PTSD have been found to have an increase in current amnestic

symptomatology in comparison to Vietnam combat veterans without PTSD as measured with the SCID for dissociative disorders (Bremner et al. 1993b). Amnestic memory disturbances should not be confused with deficits in short-term memory; we have reviewed the distinction between these two phenomena in greater detail elsewhere (Bremner et al. 1995a). Explicit memory is expressed on tests that require conscious recollection of previous experiences (e.g. free recall). Implicit memory is revealed when these experiences affect performance on a test that does not require conscious recollection (e.g. perceptual identification). PTSD patients have been found to have alterations in both explicit and implicit memory (McNally 1997).

Studies have found deficits in explicit declarative memory in PTSD. Danish survivors of the concentration camps in the Second World War were described to have persistent self-reported difficulties in memory after release from internment (Thygesen et al. 1970). Korean prisoners of war have been found to have an impairment of short-term verbal memory measured by the logical memory component of the Wechsler Memory Scale in comparison to Korean veterans without a history of containment (Sutker et al. 1991). The same was found in Persian Gulf War veterans (Vasterling et al. 1998). We have found deficits in short-term memory in Vietnam combat veterans with combat-related PTSD as measured by the logical memory component of the Wechsler Memory Scale and the visual and verbal components of the Selective Reminding Test. PTSD patients in that study did not have lower scores on the Wechsler Adult Intelligence Scale-Revised (WAIS-R) than controls (Bremner et al. 1995c,d). Similar results were found in patients with abuse-related PTSD (Bremner et al. 1997c). Studies have also found deficits in explicit short-term memory as assessed with the Auditory Verbal Learning Test (AVLT) in Vietnam combat veterans with PTSD in comparison to National Guard veterans without PTSD (Uddo et al. 1993) and the California Verbal New Learning Test in Vietnam combat veterans with PTSD in comparison to controls (Yehuda et al. 1995). A decrease in IQ in combat veterans with PTSD relative to controls may be due to an increased risk for the development of PTSD with lower IQ or may be a secondary effect of exposure to trauma (McNally et al. 1995). Other studies in patients with PTSD have shown enhanced recall on explicit memory tasks for trauma-related words relative to neutral words in comparison to controls (Zeitlin and McNally 1991; McNally et al. 1998). These findings are consistent with both deficits in encoding on explicit memory tasks, deficits in retrieval, and enhanced encoding or retrieval for specific trauma-related material. These findings could be related to stress-induced hippocampal damage.

8.4 Glucocorticoids and the Hippocampus

Current research strongly advocates that stress is associated with damage to hippocampal neurons. The work of Sapolsky and McEwen is seminal in this respect. The hippocampus, a major target organ for glucocorticoids in the brain (McEwen et al. 1986), modulates the pituitary–adrenocortical response to stress (Sapolsky and McEwen 1988). Monkeys who died spontaneously following exposure to severe stress were found on autopsy to have multiple gastric ulcers, consistent with exposure to chronic stress, and hyperplastic adrenal cortices, consistent with sustained glucocorticoid release. These monkeys also had damage to the CA3 subfield of the hippocampus (Uno et al. 1989).

Follow-up studies suggested that hippocampal damage was associated with direct exposure of glucocorticoids to the hippocampus (Sapolsky et al. 1990). Early studies in a variety of animal species (Aus der Muhlen and Ockenfels 1969) suggest that direct glucocorticoid exposure results in decreased dendritic branching (Wooley et al. 1990) and a loss of neurons (Uno et al. 1990) that are steroid and tissue specific (Packan and Sapolsky 1990). Prenatal exposure to elevated levels of glucocorticoids also results in hippocampal damage (Uno et al. 1990). Glucocorticoids appear to exert their effect through disruption of cellular metabolism (Lawrence and Sapolsky 1994) and by increasing the vulnerability of hippocampal neurons to a variety of insults, including endogenously released excitatory amino acids (Sapolsky 1986; Armanini et al. 1990). Glucocorticoids have also been shown to augment extracellular glutamate accumulation (Stein-Behrens et al. 1994). Furthermore, reduction of glucocorticoid exposure prevents the hippocampal cell loss associated with chronic stress (Landfield et al 1981; Meaney et al. 1988). Differing strains of rats may have varying glucocorticoid responses to stress, suggesting the possibility that constitutional factors may influence the glucocorticoid-mediated effects of stress on hippocampal neurons (Dhabhar et al. 1993). There are also substantial gender differences in the concentrations of glucocorticoid receptors at baseline and in response to stress, suggesting that studies related to this area which are performed exclusively in males will not be applicable to females (McCormick et al. 1994). In summary, findings to date are consistent with the idea that stress results in damage to neurons of the hippocampus, possibly through the effects of increased levels of glucocorticoids. More recently, emerging studies have found evidence for the idea that other factors besides glucocorticoids, such as brain-de-

rived neurotrophic factor (BDNF), trkB mRNA, and nerve growth factor (NGF), which have a regulatory effect on neuronal morphology and proliferation, may mediate stress-induced alterations in hippocampal morphology (Nibuya et al. 1995 1996; Smith et al. 1995).

Glucocorticoids also have other actions besides a damaging effect on hippocampal neurons. For instance, low levels of glucocorticoids following adrenalectomy result in damage to neurons of the dentate gyrus of the hippocampus (Vaher et al. 1994). Glucocorticoids also have effects on brain function through modulation of gene expression and have a variety of effects on immunity, reproduction, bone formation, and other physiological functions. These effects may have a protective effect on the organism during certain situations of stress, but in other situations the effects of glucocorticoids may be damaging (McEwen et al. 1992). Hippocampal damage may also play a role in other aspects of the long-term dysregulation of brain function associated with stress. The hippocampus is in general felt to have an inhibitory effect on the HPA axis. Stress-induced damage has been shown to be associated with an increase in levels of CRF mRNA in the PVN of the hypothalamus (J.L. Herman et al. 1989) as well as a decrease in the sensitivity of rats to dexamethasone suppression of HPA function (Feldman and Conforti 1980; Magarinos et al. 1987). Consistent with this, we have found increased levels of CRF in the cerebrospinal fluid of patients with combat-related PTSD in comparison to controls (Bremner et al. 1997a).

One of the mysteries of the neurobiology of traumatic stress is why physiological systems, which are designed to protect the organism during stress, may also have detrimental effects in some situations.

8.5
Hyperarousal and Sleep Dysfunction

Alterations in sleep function may be secondary to altered pontine function and noradrenergic dysregulation in PTSD. Sleep dysfunction has been documented following acute stress and appears to be related to development of chronic PTSD (Mellman et al. 1995). PTSD patients have been found to have an increase in phasic rapid eye movement (REM) activity (Ross et al. 1994a,b), decreased total sleep time, and increased "micro-awakenings" relative to controls. These abnormalities may play a role in nightmares and sleep disturbance in PTSD patients, which according to some authors can be considered to be the hallmark symptoms of PTSD (for a review, see Ross et al. 1989).

8.6
Thyroid Function

Thyroid function tests are frequently abnormal, hyperthyroid, in PTSD and these abnormalities tend to be in an opposite direction from the abnormalities found in depressive disorder. Thyroid-stimulating hormone (TSH) has a range of actions which include energy utilization within the cell (important in stress), and stress results in long-lived elevations in thyroid hormone (Mason et al. 1986). Although few studies have looked at thyroid function in anxiety disorders, Mason et al. (1994) found elevated levels of thyroxine (T3) in patients with combat-related PTSD. While patients with major depression exhibit a blunted TSH response to thyreotropin-releasing hormone (TRH), some patients with PTSD have exhibited an augmented response to TRH (Kosten et al. 1990). Recent research includes the involvement of prolyl endopeptidase, a cytosolic endopeptidase which degrades neuropeptides, such as TRH, and which can play a role in stress responsivity PTSD (Maes et al. 1999b). More research in this area is needed.

9
Neuroimaging Studies

Preclinical and clinical investigations provide strong evidence for linking several brain structures to the signs and symptoms of anxiety and fear associated with trauma. Chief among these are the amygdala, hippocampus, thalamus, periaquaductal gray, and orbitofrontal cortex. Although there were no published studies on PTSD as recently as 5 years ago, since that time there has been a rapid growth of studies in this area. This is in part due to a growing appreciation of the neurobiologic contributions to PTSD. Another factor that explains this is that initial studies well represented direct extensions of basic science research that amplified and broadly extended the knowledge about the effects of stress on neurobiological systems.

9.1
Structural Neuroimaging with Magnetic Resonance Imaging

Studies using MRI in PTSD have measured the volume of the hippocampus. This line of research was prompted by studies in animals showing that high levels of cortisol seen during times of stress are associated with damage to the hippocampus (reviewed above). We used MRI to measure hippocampal volume in 26 Vietnam combat veterans with PTSD and 22

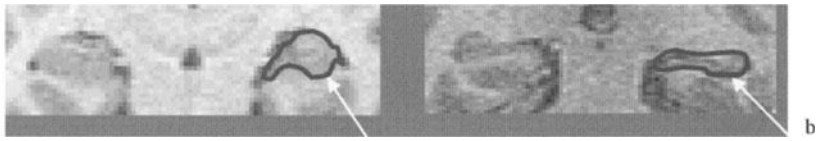


Fig. 5a,b. Magnetic resonance imaging (MRI) scan of the hippocampus in **a** a normal control and **b** a patient with PTSD secondary to childhood abuse. The hippocampus (*outlined*) is visibly smaller in PTSD. Overall, there was a 12% reduction in volume in the PTSD patient population in comparison to trauma controls without PTSD

controls matched for age, sex, race, handedness, years of education, body size, and years of alcohol abuse. We found an 8% reduction in right hippocampal volume in PTSD patients in comparison to controls, without differences in volume of comparison regions. The magnitude of reduction in hippocampal volume was associated with magnitude of deficits in short-term verbal memory (Bremner et al. 1995d). We have replicated these findings in 17 patients with PTSD related to severe childhood physical and/or sexual abuse in comparison to 17 controls matched on a case-by-case basis for several factors and found a 12% reduction in left hippocampal volume in the patients in comparison to controls in this study (Bremner et al. 1997c). Our findings of reduction in hippocampal volume have been replicated in patients with PTSD related to combat and civilian stressors (Gurvits et al. 1996; Stein et al. 1997). Figure 5 illustrates the reduction in volume in the hippocampus.

9.2 Functional Neuroimaging with Positron Emission Tomography

Sophisticated techniques have been developed for the measurement of cerebral blood flow using PET. $H_2[^{15}O]$ provides a good measure of cerebral blood flow. Cerebral blood flow has been shown to be highly correlated with local cerebral glucose metabolism. Since neurons almost exclusively utilize glucose for cell processes, glucose utilization provides a measure of local neuronal activity. Studies in PTSD have begun to use PET during pharmacologic and cognitive provocation of symptom states in order to identify neural correlates of PTSD symptomatology and of traumatic remembrance.

We also used PET and fluorodeoxyglucose (FDG) in the measurement of cerebral glucose metabolic rate following administration of yohimbine and placebo in Vietnam combat veterans with PTSD and healthy controls. Increased noradrenergic function has been hypothesized to underlie many of the symptoms of PTSD. Administration of the α_2-antagonist yohimbine, which stimulates brain norepinephrine release, resulted in increased PTSD symptoms and anxiety in the PTSD group. Norepinephrine has a U-shaped curve type of effect on brain function, with lower levels of release causing an increase in metabolism, while very high levels of release actually cause a decrease in metabolism. We hypothesized that yohimbine would cause a relative decrease in metabolism in patients with PTSD in cortical brain areas which receive noradrenergic innervation. Indeed, yohimbine resulted in differences in metabolism in orbitofrontal, temporal, parietal, and prefrontal cortex in PTSD patients relative to controls, with PTSD patients showing a pattern of decreased and normal subjects a pattern of increased metabolism in these areas. PTSD patients (but not normal subjects) had decreased hippocampal metabolism with yohimbine (Bremner et al. 1997b).

Several studies have now used PET $H_2[^{15}O]$ in a challenge paradigm. In a study of combat-related PTSD using PET and $H_2[^{15}O]$ measurement of cerebral blood flow, ten Vietnam veterans with PTSD and ten Vietnam veterans without PTSD were studied during exposure to combat-related and neutral slides and sounds. Vietnam veterans with combat-related PTSD (but not non-PTSD veterans) demonstrated a decrease in blood flow in the medial prefrontal cortex (Brodman's area 25, or subcallosal gyrus) and middle temporal cortex (auditory cortex) during exposure to combat-related slides and sounds. A failure of activation was found in anterior cingulate (area 32 and 24), and increased activation in posterior cingulate, motor cortex, and lingual gyrus in PTSD veterans (Bremner et al. 1999c). In another study, cerebral blood flow correlates of exposure to personalized scripts of childhood sexual abuse were looked at in women with histories of childhood abuse with ($n = 10$) and without ($n = 12$) PTSD. PTSD women showed decreased blood flow in medial prefrontal cortex (area 25) and failure of activation in anterior cingulate, with increased blood flow in posterior cingulate and motor cortex (replicating findings in combat-related PTSD) and anterolateral prefrontal cortex. PTSD women also had decreased blood flow in right hippocampus and parietal and visual association cortex (Bremner et al. 1999d). Other studies of traumatic imagery in combat-related PTSD found alterations in orbitofrontal and temporal cortex in PTSD (Rauch et al. 1996; Shin et al. 1997, 1999).

Another field is receptor imaging in PTSD. Based on findings of decreased benzodiazepine binding in fron-

tal cortex in animal models of stress, we measured benzodiazepine binding with single photon emission computed tomography (SPECT) [^{123}I]iomazenil in 13 combat-related PTSD patients and 13 healthy controls. We found a decrease in benzodiazepine receptor binding in prefrontal cortex (Brodmann's area 9) in patients with combat-related PTSD compared to matched healthy controls (Bremner et al. 1999e).

Human subjects with lesions of medial prefrontal cortical areas have deficits in interpretation of emotional situations that are accompanied by impairments in social relatedness. Anterior cingulate (area 32) activation might well represent a "normal" brain response to traumatic stimuli that serves to inhibit feelings of fearfulness when there is no true threat. From our studies, we learn that dysfunction in this area and/or decreased blood flow in adjacent medial prefrontal cortex (subcallosal gyrus) in PTSD may lead to increased fearfulness that is not appropriate for the context, a behavioral response that is highly characteristic of patients with PTSD (Hamner et al. 1999).

10 Pharmacological Treatment

10.1 General Considerations

Compared with other psychiatric disorders, few studies of psychopharmacology have been performed in PTSD (see Friedman 1996; Sutherland and Davidson 1999). However, with the introduction of new potential targets such as beta blockers and CFR antagonists, this field is rapidly expanding.

10.2 Pharmacotherapy

Antidepressant medications are the mainstay of treatment and are the best studied in controlled clinical trials, but there are also reports in the literature of alleviation of specific symptoms of PTSD with the use of other medications such as sympatholytic agents, mood stabilizers such as lithium and anticonvulsants, benzodiazepines, and drugs which affect the dopamine, opioid, and serotonergic system. Given the different clusters of PTSD symptoms, the clinician may find only partial response in individual patients with a single medication and will find it necessary to consider addressing the multiple symptoms with a combination of medications. Such medications will be found in the following groups: antidepressants, anxiolytics, adrenergic inhibitors, mood stabilizers, and anticonvulsants (see Table 3). The individual therapeutic dose range is listed for each drug.

No medication is approved yet for the indication of PTSD. Following expert guidelines, it is recommended that the selective serotonin reuptake inhibitors (SSRI) and nefazodone should be used as first-line drugs for PTSD (Foa et al. 1999a). In adults, it is recommended

Table 3. Classes of medication in PTSD (adapted from Sutherland and Davidson 1999)

Class	Group	Generic name	Dose range (mg)
Adrenergic inhibitors		Propanolol	20–160
		Clonidine	0.1–0.4
Anxiolytics	Benzodiazepines	Clonazepam	0.5–6
		Lorazepam	0.5–4
		Clordiazepoxide	5–40
		Diazepam	2–40
	Azapirones	Buspirone	5–60
Antidepressants	MAOI	Phenelzine	15–90
		Tranylcypromine	20–60
	SSRI	Fluvoxamine	50–300
		Fluoxetine	10–60
		Sertraline	50–200
		Paroxetine	10–60
	Phenylpiperazine	Nefazodone	200–600
	Triazolopyradine	Trazodone	150–300
	TCA	Amitryptyline	50–300
		Imipramine	50–300
Mood stabilizers	Lithium	Lithium carbonate	300–1200
Anticonvulsants		Carbamazepine	200–1500
		Valproic acid	125–2000
		Dilantin	300–600

MAOI, monoamine oxidase inhibitors; SSRI, selective serotonin reuptake inhibitors; TCA, tricyclic antidepressants.

that buspar or benzodiazepines should be added in situations of persistent fear reactivity. There are eight double-blind placebo-controlled medication trials to date in the last 10 years showing no or on average only modest positive outcomes. Shestatzky et al. (1988) found no difference in response between placebo and 45–75 mg phenelzine in a population of ten PTSD patients with varied trauma after treating them for 4 weeks. Frank et al. (1988) gave 46 male veterans 50–300 mg imipramine or 15–75 mg phenelzine for an average duration of 6 weeks. They found that intrusive scores on the Impact of Events Scale (IES; Horowitz et al. 1979) decreased both with phenelzine and imipramine. There was some decrease in the avoidance subscale with phenelzine. However, there was no difference in measures with the placebo group. Another study in imipramine was performed by Robert et al. (1999) looking at the effects of imipramine in pediatric burn patients with symptoms of acute stress disorder in 25 children. They received either imipramine or chloral hydrate for 7 days, suggesting a place for cautious initial use of imipramine to reduce ASD symptoms in burned children. Reist et al. (1989) studied 100–200 mg desipramine in 18 male veterans with concurrent diagnoses in a 4-week cross-over design. They found significant improvement on the IES intrusion subscale and depressive symptoms only in one subgroup with concurrent major depression. There was no change overall in anxiety and PTSD symptoms. Davidson et al. (1990, 1993) studied the effects of amitriptylline in 62 male veterans, treating them for 8 weeks. The results were that amitriptylline was superior to placebo on anxiety and depression, with marginally effects on IES avoidance and intrusive scales. Braun et al. (1990) found aprazolam superior to placebo on anxiety in 16 patients with PTSD due to varied causes. They treated the patients for 5 weeks with 2.5–6 mg alprazolam. There was no difference on IES or depression. Van der Kolk et al. (1994) looked at the effects of 20–40 mg fluoxetine in a trial consisting of 5 weeks double-blind and 5 weeks of open medication in 31 veterans and 33 civilian trauma victims. They reported significant reduction in symptoms including intrusive numbing. One half of the population no longer met criteria for PTSD. Baker et al. (1995) assessed the efficacy of brofaromine in 118 PTSD outpatients, giving 56 brofaromine and 58 placebo. Their results showed that both groups showed significant reductions in the symptoms, without significant differences in the between-groups analysis.

The effect of these medications may well begin as early as 2 weeks after treatment in drug-responsive patients. Whereas earlier findings suggested that SSRI may be less effective against reexperiencing and arousal than they are against avoidant/numbing symptoms, it is still not yet clear to what extent the medication shows its efficacy in the different clusters (Davidson et al. 1997). Reports of the efficacy of different treatments are being published (sertraline: Brady et al. 1995; mirtazapine: Connor et al. 1999; Falkai 1999; fluvoxamine: Davidson et al. 1998b; nefazadone: Davidson et al. 1998c; Hidalgo et al. 1999). More double-blind, placebo-controlled clinical trials on PTSD are in progress to assess the utility of different SSRI and selective norephinephrine reuptake inhibitors (SNRI) as a treatment in this disorder.

Interestingly, antiadrenergic drugs (such as α_2-agonists or beta blockers) have so far received little systematic attention in clinical trials, despite evidence for adrenergic dysregulation in PTSD. Some authors recommended starting pharmacotherapy in new PTSD patients with an antiadrenergic agent such as clonidine or propanolol (Friedman 1998). Clonidine can be very effective in this respect, where the reduced adrenergic activity is often accompanied by a reduction in dissociative symptoms. The advantage of clonidine or a beta-adrenergic agonist is that this medication can be titrated over the course of 1–2 weeks. We suggest switching to a drug with longer half-life if a clonidine responder appears to develop tolerance to the drug. If the symptoms persist, as they often do, after optimal titration, the next drug to add is an SSRI. Using another medication from the same class is not recommended if the first drug is not effective. If patients develop insomnia or agitation, the best choice is nefazadone or trazodone at bedtime. If patients are still nonresponders after 8–10 weeks, a different class of medication (e.g. adrenergic inhibitors, anxiolytics, mood stabilizers, anticonvulsants) should be considered. The importance of pharmacoeducation has never been well studied, but every clinician knows or should know the importance of this part of the treatment.

10.3 New Developments in the Treatment of Stress-Related Brain Changes

Studies in animals have demonstrated several agents with potentially beneficial effects on the reversibility of the glucocorticoid-mediated hippocampal toxicity. It has been found that phenytoin (Dilantin) reverses stress-induced hippocampal atrophy, probably through excitatory amino acid-induced neurotoxicity. Agents such as tianeptine and dihydroepiandosterone (DHEA) have been described to produce similar effects (Watanabe et al. 1993). Neurons within the hippocampus were found to be unique within the brain in showing the capacity to regenerate themselves (Gould et al. 1999). Since these are in vitro studies that report

on neurogenesis in the hippocampus through a regulation of BDNF and cAMP by SSRI, we hypothesize that the effects of SSRI may play a role in the reversibility of the accelerated aging process in humans (Duman et al. 1997; Nibuya et al. 1999). There are findings indicating that a common action of antidepressant treatments is through upregulation of cAMP response element binding protein (CREB) and that this may lead to regulation of specific target genes. Such treatment effects on BDNF and trkB mRNA can have long-term effects on brain function. In these studies, however, it is evident that chronic, but not acute, administration of several different classes of antidepressants, including SSRI and SNRI, increase the expression of CREB mRNA. Interestingly, in contrast, chronic administration of several non-antidepressant psychotropic drugs did not influence expression of CREB mRNA, demonstrating the pharmacological specificity of this effect (Nibuya et al. 1996). If through hippocampal neurogenesis the medication has an effect on the glucocorticoid-induced impairment in declarative memory, this may have promising results if found in humans (Newcomer et al. 1994).

11 Psychological Treatment

According to expert guidelines (Foa et al. 1999a) concerning the treatment of PTSD, current psychological treatments of PTSD include the following: (a) behavioral therapy (flooding; systematic desensitization [SD] eye movement desensitization reprocessing [EMDR]), (b) cognitive-behavioral therapy, and (c) anxiety management training (stress inoculation training [SIT] biofeedback). These treatments are described in detail in Meadows and Foa (1999). Each intervention has its own merits, and all three can be very useful regardless of the stage of PTSD. It is recommended that patients should be seen in the acute situation for 3 months, with boost sessions every 2–4 weeks; in a chronic situation, the therapy can last up to 6 months, with booster sessions every 2–4 weeks. Studies have been performed to assess efficacy of psychological interventions in PTSD (see Foa et al. 1999b; Spiegel 1999).

11.1 Behavioral Treatments

Exposure techniques involve confronting one's fears. Exposure-based therapeutic regimens have had a long history in the treatment of anxiety-related disorders.

11.1.1 Flooding

Flooding involves two essential steps: (1) the administration of multiple component assessment packages and (2) the provision of imaginal flooding regimens. In the assessment phase, the goal is to identify the pathological behaviors, determine the etiological variables that maintain distress, and establish baseline data to determine treatment efficacy over time. Different sources of information should be used. The intervention phase is presented in five basic steps, which should be presented in sequential order: (1) education, (2) imagery training, (3) relaxation training, (4) presentation of traumatic scenes, and (5) debriefing. Flooding is an aversive process in that PTSD symptoms may increase during the initial phases of therapeutic exposure. (For an overview of the method, see Saigh et al. 1999b.)

11.1.2 Systematic Desensitization

Systematic desensitization falls at the other end of the dimensions of exposure methods, using brief, imaginal, and minimally arousing exercises. Pioneered by Wolpe (1958), SD was among the earliest behavioral treatments studies for PTSD. It involves pairing imaginal exposure with relaxation, so that the anxiety elicited by the confrontation with the feared stimuli is inhibited by relaxation. First, the patient is instructed in muscle relaxation exercises. When a state of relaxation is achieved, the feared stimuli are introduced, via imagined scenarios, in a graded hierarchical manner, with the least anxiety-provoking scenarios presented first. When the patient begins to feel anxious, the instruction is given to erase the screen, focus on relaxation, and begin again. The scenario is repeated until it no longer elicits anxiety, at which point the next scenario is introduced. This process continues until the stimuli on the hierarchy no longer elicit anxiety. For example, a therapist may teach a patient to head off panic attacks by taking slow deep breaths. The therapists may gradually expose the patient to images or sensations that remind him or her of the trauma (battle photos, loud noises, smells) and then help him or her deal with the fears that come up.

11.1.3 Eye-Movement Desensitization Reprocessing

EMDR is a form of imaginal exposure accompanied by saccadic eye movements. It is conducted by having the patient focus on a disturbing image or memory while the therapist moves a finger across the patient's visual field (Shapiro 1989). The saccadic eye movements

result from the patient's tracking the therapist's finger. (For a variety of reviews, see Lohr et al. 1995; Devilly and Spence 1999; Cusack and Spates 1999; Lazrove et al. 1998.) The most recent review by Shapiro describes four recent, independent, rigorously controlled studies of EMDR that reported that 84%–100% of single-trauma victims no longer maintain the PTSD diagnosis after the equivalent of three 90-min sessions (Shapiro 1999). Since its inception, EMDR has been the focus of much controversy.

11.2 Cognitive Therapy

Cognitive therapy was pioneered by Beck (1972) and further developed by others to help patients modify dysfunctional cognitions. The basic assumption is that dysfunctional thoughts drive negative emotional states such as fear or anger. In a given situation, this may lead to different emotions depending upon the interpretation of the situation. Pathological emotions are generated by distorted, dysfunctional thoughts. Typically, cognitive restructuring (CR) aims to teach patients to identify dysfunctional thoughts, to evaluate their validity, to challenge their erroneous or unhelpful thoughts, and to replace them with more beneficial ones. CR would focus on identifying the thoughts that precede strong emotions, such as "I am going to be assaulted." Then the validity of this belief would be evaluated. Erroneous beliefs are replaced by rational ones, suggested by the evidence reviewed in one of the previous steps.

11.3 Anxiety Management Therapies

The assumption that seems to underlie the rationale for anxiety management programs centers on the notion that pathological anxiety stems from skill deficits. This implies that providing patients with appropriate skills would enable them to manage their anxiety. Skills such as relaxation training, positive self-statements, breathing retraining, biofeedback, social skill training, and distraction methods aim at managing the anxiety when it occurs, rather than preventing pathological anxiety from occurring by correcting the supposedly underlying mechanisms.

11.3.1 Stress Inoculation Training

SIT was developed by Meichenbaum (1974) as a treatment for anxious patients. The method was adopted by Veronen et al. (1979) for treatment of rape-related disturbances. This modified SIT includes psychoeducation, muscle relaxation training, breathing training, role playing, covert modeling, guided self-dialogue, and thought-stopping. The three-component model of SIT includes patients beginning to see their responses in the physiological, behavioral, and cognitive domains. SIT is one of the most commonly used anxiety management treatments for PTSD.

11.3.2 Biofeedback

Biofeedback is a procedure in which patients learn to gain control over their physiological processes. This control is achieved by having patients observe displays or listen to tones of their physiologic activity (EMG activity) and then try to change the display or change the tone.

11.4 Other Therapeutic Interventions and Modalities

Several other therapeutic strategies are reported in PTSD. Shay and Munroe report on a treatment setting in which the encounter is the focus of attention in therapy (Shay and Munroe 1999); other reports describe psychodrama therapy (Carbonell and Parteleno-Barehmi 1999), psychodynamic approaches (Huller and Barash-Kishon 1998), or use of hypnosis (Spiegel 1992; Kluft 1992; M. Phillips 1993).

Most clinicians who treat PTSD have an eclectic approach to treatment, and systematic treatment outcome studies have yet to be performed (Brom et al. 1989; Spiegel 1999). Combining biological, psychological, and psychosocial treatment may well yield best results. Rehabilitative goals should replace curative techniques in those patients with chronic PTSD (Shalev et al. 1996b). Similar to the importance of pharmacoeducation, the importance of psychoeducation should not be underestimated (J.G. Allen et al. 1997; Lubin et al. 1998).

12 Concluding Remarks

The field of psychiatry appears to be finally coming to terms with the fact that traumatic stress can result in chronic psychiatric disorders, including PTSD. We may be ready to go beyond the thinking inherent in the gross stress reaction of DSM-I, which described a reversible mental response to stress that can affect anyone and does not constitute a true psychiatric disorder. These changes have been partially driven by

the finding, surprising to many, that events which enter the eye and the ear (trauma) can have lasting effects on the brain and physiology (Andreasen 1995; Bremner 1999b). In PTSD, we are at the crossroads between mind and brain, and our field is moving toward a breakdown between the artificial distinctions between psychology and biology. This will bring us back to the viewpoint originally encompassed by DaCosta and Kraepelin that mental disorders have their basis in the brain (DaCosta 1871; Kraepelin 1919).

The more traumatic stress is examined, the more it becomes apparent that stress may have also far-reaching influences on all of the major psychiatric disorders. This has led to a tendency for PTSD and the stress response to become like a fast-moving train that threatens to take along everything in its path. In fact, workers in the field of PTSD are now in danger of being victimized by their own success. Once again, psychiatrists may be asked to bear the responsibility for the unfairness and discrepancies of our society. The challenge in the future will be to determine the most appropriate way to use our psychiatric nomenclature to describe the relationship between environmental events such as extreme stress and psychopathology.

Future studies will need to focus on estimating age-at-onset distributions, cohort effects, and the conditional probabilities of PTSD from different types of trauma. These future epidemiologic studies will also need to assess PTSD for all lifetime traumas rather than for only a small number of retrospectively reported "most serious" traumas. As for the neurobiology and treatment of the disorder, the wealth of laboratory research and animal models has helped us understand the pathophysiology of the disorder. It is for future studies to translate these findings into optimal pharmacotherapeutic interventions that can be combined with psychological treatments for alleviation of symptoms, reduction of medical consumption, and better quality of life for patients currently suffering from PTSD – since it is not likely that traumatic stress will be eliminated from our society.

Appendix. Diagnostic Criteria for Post-traumatic Stress Disorder (from DSM-IV; APA 1994)

A. The person has been exposed to a traumatic event in which both of the following were present:
(1) the person experienced, witnessed, or was confronted with an event or events that involved actual or threatened death or serious injury, or a threat to the physical integrity of self or others
(2) the person's response involved intense fear, helplessness, or horror

Note: In children, this may be expressed instead by disorganized or agitated behavior

B. The traumatic event is persistently reexperienced in at least one of the following ways:
(1) recurrent and intrusive distressing recollections of the event, including images, thoughts, or perceptions

Note: In young children, repetitive play may occur in which themes or aspects of the trauma are expressed.

(2) recurrent distressing dreams of the event

Note: In children there may be frightening dreams without recognizable content

(3) acting or feeling as if the traumatic event were recurring (includes a sense of reliving the experience, illusions, hallucinations, and dissociative flashback episodes, including those that occur on awakening or when intoxicated)

Note: In young children, trauma-specific reenactment may occur

(4) intense psychological distress at exposure to internal or external cues that symbolize or resemble and aspect of the traumatic event
(5) physiological reactivity on exposure to internal or external cues that resemble an aspect of the traumatic event.

C. Persistent avoidance of stimuli associated with the trauma and numbing of general responsiveness (not present before the trauma), as indicated by three (or more) of the following:
(1) efforts to avoid thoughts, feelings, or conversations associated with the trauma
(2) efforts to avoid activities, places, or people that arouse recollections of the trauma
(3) inability to recall an important aspect of the trauma
(4) markedly diminished interest or participation in significant activities
(5) feeling of detachment or estrangement from others
(6) restricted range of affect (e.g. unable to have loving feelings)
(7) sense of a foreshortened future (e.g. does not expect to have a career, marriage, children, or a normal life span).

D. Persistent symptoms of increased arousal (not present before the trauma), as indicated by two (or more) of the following:
(1) difficulty falling or staying asleep
(2) irritability or outbursts or anger
(3) difficulty concentrating
(4) hypervigilance
(5) exaggerated startle response.

E. Duration of the disturbance (symptoms in Criteria B, C and D) is more than 1 month.
F. The disturbance causes clinically significant distress or impairment in social, occupational, or other important areas of functioning

Specify if: Acute: if duration of symptoms is less than 3 months
Chronic: if duration of symptoms is 3 months or more
Specify if: With delayed onset: if onset of symptoms is at least 6 months after the stressor

13 References

Abercrombie ED, Jacobs BL (1987) Single-unit response of noradrenergic neurons in the locu coeruleus of freely moving cats. I. Acutely presented stressful and nonstressful stimuli. J Neurosci 7(9): 2837–2843

Abercrombie ED, Keefe KA, DiFrischia DS, Zigmond MJ (1989) Differential effect of stress on in vivo dopamine release in striatum, nucleus accumbens, and medial frontal cortex. J Neurochem 52: 1655–1658

*Acierno R, Kilpatrick DG, Resnick HS (1999) Child-adolescent PTSD, prevalence, risk factors and comorbidity. In: Saigh PA, Bremner JD (eds) Posttraumatic stress disorder, a comprehensive text. Ally and Bacon, Boston, pp 18–44

Allen IM (1996) PTSD among African Americans. In: Marsella AJ, Friedman MJ (eds) Ethnocultural aspects of posttraumatic stress disorder: issues, research and clinical applications. American Psychological Association, Washington, pp 209–238

Allen JG, Kelly KA, Glodich A (1997) A psychoeducational program for patients with trauma-related disorders. Bull Menninger Clin 61(2): 222–239

Alonzo AA (1999) Acute myocardial infarction and posttraumatic stress disorder: the consequences of cumulative adversity. J Cardiovasc Nurs 13(3): 33–45

Andreasen NJC (1985) Posttraumatic stress disorder. In: Kaplan HI, Sadock BJ (eds) Comprehensive textbook of psychiatry, vol IV. Williams and Wilkins, Baltimore, pp 918–924

Andreasen NC (1995) Posttraumatic stress disorder: psychology, biology, and the Manichean warfare between false dichotomies. Am J Psychiatry 152: 963–965

Andreasen NJC, Norris AS, Hartford CE (1971) Incidence of long-term psychiatric complications in severely burned adults. Ann Surg 174: 785

APA (1952) Diagnostic and statistical manual of mental disorders, 1st edn. American Psychiatric Association, Washington, DC

APA (1968) Diagnostic and statistical manual of mental disorders, 2nd edn. American Psychiatric Association, Washington, DC

APA (1980) Diagnostic and statistical manual of mental disorders, 3rd edn. American Psychiatric Association, Washington, DC

APA (1987) Diagnostic and statistical manual of mental disorders, 3rd edn, revised. American Psychiatric Association, Washington, DC

APA (1994) Diagnostic and statistical manual of mental disorders, 4th edn. American Psychiatric Association, Washington, DC

Archibald HC, Tuddenham RD (1965) Persistent stress reaction after combat. Arch Gen Psychiatry 12: 475–481

Amir M, Sol O (1999) Psychological impact and prevalence of traumatic events in a student sample in Israel: the effect of multiple traumatic events and physical injury. J Trauma Stress 12(1): 139–154

Arbel I, Kadar T, Silberman M, Levy A (1994) The effects of long-term corticosterone administration on hippocampal morphology and cognitive performance of middle-aged rats. Brain Res 657: 227–235

Armanini MP, Hutchins C, Stein BA, Sapolsky RM (1990) Glucocorticoid endangerment of hippocampal neurons is NMDA-receptor dependent. Brain Res 532: 7–11

Ashkoomoff NA, Courchesne E (1992) A new role for the cerebellum in cognitive operations. Behav Neurosci 106: 731–738

Aus der Muhlen K, Ockenfels H (1969) Morphologische Veränderungen im Diencephalon und Telencephalon nach Störungen des Regelkreises Adenohypophyse-Nebennierenrinde. III. Ergebnisse beim Meerschweinchen nach Verabreichung von Cortison und Hydrocortison. Z Zellforsch 93: 126–138

Baker DG, Diamond BI, Gillette G, Hamner M, Katzelnick D, Keller T, Mellman TA, Pontius E, Rosenthal M, Tucker P et al (1995) A double-blind, randomized, placebo-controlled, multi-center study of brofaromine in the treatment of posttraumatic stress disorder. Psychopharmacology 122(4): 386–389

Baker DG, West SA, Nicholson WE, Ekhator NN, Kasckow JW, Hill KK, Bruce AB, Orth DN, Geracioti TD Jr (1999) Serial CSF corticotropin-releasing hormone levels and adrenocortical activity in combat veterans with posttraumatic stress disorder. Am J Psychiatry 156(4): 585–588

Baum A, Cohen L, Hall M (1993) Control and intrusive memories as possible determinant of chronic stress. Psychosom Med 55(3): 274–286

Beck AT (1972) Depression: causes and treatment. University of Philadelphia Press, Philadelphia

Bisson JI, Shepherd JP, Dhutia M (1997) Psychological sequelae of facial trauma. J Trauma Injury Infect Crit Care 43(3): 496–500

Blake DD, Keane TM, Wine PR, Mora C (1990) Prevalence of PTSD symptoms in combat veterans seeking medical treatment. J Trauma Stress 3(1): 15–27

*Blake DD, Weathers FW, Nagy LM, Kaloupek DG, Gusman FD, Charney DS, Keane TM (1995) The development of a clinician-administered PTSD scale. J Trauma Stress 8(1): 75–90

**Blank AS (1993) The longitudinal course of PTSD. In: Davidson JR, Foa EB (eds) Posttraumatic stress disorder: DSM-IV and beyond. American Psychiatric Press, Washington, DC, pp 3–22

Brady KT, Sonne SC, Roberts JM (1995) Sertraline treatment of comorbid posttraumatic stress disorder and alcohol dependence. J Clin Psychiatry 56(11): 502–505

Braun P, Greenberg D, Dasberg H, Lere B (1990) Core symptoms of posttraumatic stress disorder unimproved by alprazolam treatment. J Clin Psychiatry 51(6): 236–238

Bremner JD (1999a) Acute and chronic responses to psychological trauma: where do we go from here? Am J Psychiatry 156(3): 349–351

**Bremner JD (1999b) Does stress damage the brain? Biol Psychiatry 45(7): 797–805

Bremner JD, Brett E (1997) Trauma-related dissociative states and long-term psychopathology in posttraumatic stress disorder. J Trauma Stress 10: 37–50

*Bremner JD, Southwick SM, Rosenheck R, Brett E, Fontana A, Charney DS (1992) Dissociation and posttraumatic stress disorder in Vietnam combat veterans. Am J Psychiatry 149(3): 328–332

Bremner JD, Southwick SM, Johnson DR, Yehuda R, Charney DS (1993a) Childhood physical abuse and combat-related posttraumatic stress disorder in Vietnam veterans. Am J Psychiatry 150(2): 235–239

Bremner JD, Steinberg M, Southwick SM, Johnson DR, Charney DS (1993b) Use of the Structured Clinical Interview for DSMIV-dissociative disorders for systematic assessment of dissociative symptoms in posttraumatic stress disorder. Am J Psychiatry 150: 1011–1014

Bremner JD, Southwick SM, Charney DS (1995a) Etiologic factors in the development of posttraumatic stress disorder. In: Mazure CM (ed) Does stress cause psychiatric disease? American Psychiatric Press, Washington, DC, pp 43–64

Bremner JD, Krystal JH, Southwick SM, Charney DS (1995b) Functional neuroanatomical correlates of the effects of stress on memory. J Trauma Stress 8(4): 527–554

*Bremner JD, Randall PR, Scott TM, Bronen RA, Delaney RC, Seibyl JP, Southwick SM, McCarthy G, Charney DS, Innis RB (1995c) MRI-based measurement of hippocampal volume in posttraumatic stress disorder. Am J Psychiatry 152: 973–981

Bremner JD, Randall P, Scott TM, Capelli S, Delaney R, McCarthy G, Charney DS (1995d) Deficits in short-term memory in adult survivors of childhood abuse. Psychiatry Res 59(1–2): 97–107

Bremner JD, Southwick SM, Darnell A, Charney DS (1996a) Chronic PTSD in Vietnam combat veterans: course of illness and substance abuse. Am J Psychiatry 153(3): 369–375

Bremner JD, Krystal JH, Southwick SM, Charney DS (1996b) Noradrenergic mechanisms in stress and anxiety. I. Preclinical studies. Synapse 23: 28–38

Bremner JD, Krystal JH, Southwick SM, Charney DS (1996c) Noradrenergic mechanisms in stress and anxiety. II. Clinical studies. Synapse 23: 39–51

Bremner JD, Licinio J, Darnell A, Krystal JH, Owens M, Southwick SM, Nemeroff CB, Charney DS (1997a) Elevated CSF corticotropin-releasing factor concentrations in posttraumatic stress disorder. Am J Psychiatry 154: 624–629

Bremner JD, Innis RB, Ng CK, Staib L, Duncan J, Bronen R, Zubal G, Rich D, Krystal JH, Dey H, Soufer R, Charney DS (1997b) PET measurement of central metabolic correlates of yohimbine administration in posttraumatic stress disorder. Arch Gen Psychiatry 54: 246–256

*Bremner JD, Randall P, Vermetten E, Staib L, Bronen RA, Mazure C, Capelli S, McCarthy G, Innis RB, Charney DS (1997c) Magnetic resonance imaging-based measurement of hippocampal volume in posttraumatic stress disorder related to childhood physical and sexual abuse – a preliminary report. Biol Psychiatry 41(1): 23–32

Bremner JD, Krystal JH, Putnam F, Marmar C, Southwick SM, Lubin H, Charney DS, Mazure CM (1998) Measurement of dissociative states with the Clinician Administered Dissociative States Scale (CADSS). J Trauma Stress 11: 125–136

*Bremner JD, Narayan M, Staib LH, Southwick SM, McGlashan T, Charney DS (1999a) Neural correlates of memories of childhood sexual abuse in women with and without posttraumatic stress disorder. Am J Psychiatry 156: 1787–1795

**Bremner JD, Southwick SM, Charney DS (1999b) The neurobiology of posttraumatic stress disorder: an integration of animal and human research. In: Saigh PA, Bremner JD (eds) Posttraumatic stress disorder; a comprehensive text. Allyn and Bacon, Boston, pp 103–144

*Bremner JD, Staib LH, Kaloupek D, Southwick SM, Soufer R, Charney DS (1999c) Positron emission tomographic (PET)-based measurement of cerebral blood flow correlates of traumatic reminders in Vietnam combat veterans with and without posttraumatic stress disorder. Biol Psychiatry 45: 806–816

*Bremner JD, Narayan M, Staib LH, Southwick SM, McGlashan T, Charney DS (1999d) Neural correlates of memories of childhood sexual abuse in women with and without posttraumatic stress disorder. Am J Psychiatry 156(11): 1787–1795

Bremner JD, Baldwin R, Horti A, Staib LH, Ng CK, Tan P-Z, Zea-Ponce Y, Zoghbi S, Seibyl JP, Soufer R, Charney DS, Innis RB (1999e) Quantitation of benzodiazepine receptor binding with PET [^{11}C]iomazenil and SPECT [^{123}I]iomazenil: preliminary results of a direct comparison in healthy human subjects. Psych Res Neuroimaging 91: 79–91

Bremner JD, Vermetten E, Mazure CM. Early Trauma Inventory: development, reliability, and validity. Anxiety Depression (in press)

Breslau N, Peterson EL, Kessler RC, Schultz LR (1999a) Short screening scale for DSM-IV posttraumatic stress disorder. Am J Psychiatry 156(6): 908–911

Breslau N, Chilcoat HD, Kessler RC, Davis GC (1999b) Previous exposure to trauma and PTSD effects of subsequent trauma: results from the Detroit Area Survey of Trauma. Am J Psychiatry 156(6): 902–907

Brett EA, Spitzer RL, Williams JB (1988) DSM-III-R criteria for posttraumatic stress disorder. Am J Psychiatry 145(10): 1232–1236

Brewin CR, Andrews B, Rose S, Kirk M (1999) Acute stress disorder and posttraumatic stress disorder in victims of violent crime. Am J Psychiatry 156(3): 349–351

Brom D, Kleber RJ, Defares PB (1989) Brief psychotherapy for posttraumatic stress disorders. J Consult Clin Psychol 57(5): 607–612

Butler P, Weiss J, Stout J, Nemeroff C (1990) Corticotropin-releasing factor produces fear-enhancing and behavioral activating effects following infusion into the locus coeruleus. J Neurosci 10: 176–183

Caggiula AR, Antelman SM, Aul E, Knopf S, Edwards DJ (1989) Prior stress attenuates the analgesic response but sensitizes the corticosterone and cortical dopamine responses to stress 10 days later. Psychopharmacology 99(2): 233–237

Carbonell DM, Parteleno-Barehmi C (1999) Psychodrama groups for girls coping with trauma. Int J Group Psychother 49(3): 285–306

Carlson JG, Singelis TM, Chemtob CM (1997) Facial EMG responses to combat-related visual stimuli in veterans with and without posttraumatic stress disorder. Appl Psychophysiol Biofeedback 22(4): 247–259

Chapman WP, Schroeder HR, Guyer G, Brazier MAB, Fager C, Poppen JL, Solomon HC, Yakolev PI (1954) Physiological evidence concerning the importance of the amygdaloid nuclear region in the integration of circulating functions and emotion in man. Science 129: 949–950

**Charney DS, Bremner JD (2000) Psychobiology of PTSD. In: Charney DS, Nestler E, Bunny B (eds) Neurobiology of mental illness. Oxford University Press, Oxford, pp 494–517

*Charney DS, Deutch AY, Krystal JH, Southwick SM (1993) Psychobiologic mechanisms of posttraumatic stress disorder. Arch Gen Psychiatry 50(4): 294–305

Chu JA, Frey LM, Ganzel BL, Matthews JA (1999) Memories of childhood abuse: dissociation, amnesia, and corroboration. Am J Psychiatry 156(5): 749–755

Classen C, Koopman C, Hales R, Spiegel D (1998) Acute stress disorder as a predictor of posttraumatic stress symptoms. Am J Psychiatry 155: 620–624

Conlon L, Fahy TJ, Conroy R (1999) PTSD in ambulant RTA victims: a randomized controlled trial of debriefing. J Psychosom Res 46(1): 37–44

Connor KM, Davidson JR, Weisler RH, Ahearn E (1999) A pilot study of mirtazapine in post-traumatic stress disorder. Int Clin Psychopharmacol 14(1): 29–31

Coplan JD, Andrew MW, Rosenblum LA, Owens MJ, Friedman S, Gorman JM, Nemeroff CB (1996) Persistent elevations of cerebrospinal fluid concentrations of corticotropin-releasing factor in adult nonhuman primates exposed to early-life stressors: implications for the pathophysiology of modd and anxiety disorders. Proc Natl Acad Sci USA 93(4): 1619–1623

Cusack K, Spates CR (1999) The cognitive dismantling of eye movement desensitization and reprocessing (EMDR) treatment of posttraumatic stress disorder (PTSD). J Anxiety Dis 13(1–2): 87–99

DaCosta JM (1871) On irritable heart: a clinical study of a form of functional cardiac disorder and its consequences. Am J Med Sci 161: 17–52

Dallman MF, Jones MT (1973) Corticosteroid feedback control of ACTH secretion: effect of stress-induced corticosterone secretion on subsequent stress responses in the rat. Endocrinology 92: 1367–1375

Damasio H, Grabowski T, Frank R, Galaburda AM, Damasio AR (1994) The return of Phineas Gage: clues about the brain from the skull of a famous patient. Science 264: 1102–1105

Dansky BS, Byrne CA, Brady KT (1999) Intimate violence and post-traumatic stress disorder among individuals with cocaine dependence. Am J Drug Alcohol Abuse 25(2): 257–268

Davidson J, Lipper S, Kilts CD, Mahorney S, Hammett E (1985) Platelet MAO activity in posttraumatic stress disorder. Am J Psychiatry 142(11): 1341–1343

Davidson JR, Smith RD, Kudler HS (1989) Validity and reliability of the DSM-III criteria for posttraumatic stress disorder: experience with a structured interview. J Nerv Mental Dis 177: 336–341

Davidson J, Kudler H, Smith R, Mahorney SL, Lipper S, Hammett E, Saunders WB, Cavenar JO Jr (1990) Treatment of posttraumatic stress disorder with amitriptyline and placebo. Arch Gen Psychiatry 47(3): 259–266

Davidson JR, Kudler HS, Saunders WB, Erickson L, Smith RD, Stein RM, Lipper S, Hammett EB, Mahorney SL, Cavenar JO Jr (1993) Predicting response to amitriptyline in posttraumatic stress disorder. Am J Psychiatry 150(7): 1024–1029

*Davidson JR, Malik ML, Sutherland SN (1997) Response characteristics to antidepressants and placebo in post-traumatic stress disorder. Int Clin Psychopharmacol 12(6): 291–296

Davidson JR, Tupler LA, Wilson WH, Connor KM (1998a) A family study of chronic post-traumatic stress disorder following rape trauma. J Psychiatr Res 32(5): 301–309

Davidson JR, Weisler RH, Malik M, Tupler LA (1998b) Fluvoxamine in civilians with posttraumatic stress disorder. J Clin Psychopharmacol 18(1): 93–95

Davidson JR, Weisler RH, Malik ML, Connor KM (1998c) Treatment of posttraumatic stress disorder with nefazodone. Int Clin Psychopharmacol 13(3): 111–113

Davis LL, Suris A, Lambert MT, Heimberg C, Petty F (1997) Posttraumatic stress disorder and serotonin: new directions for research and treatment. J Psychiatry Neurosci 22(5): 318–326

**Davis M (1992) The role of the amygdala in fear and anxiety. Annu Rev Neurosci 15: 353–375

De Bellis MD, Lefter L, Trickett PK, Putnam FW Jr (1994) Urinary catecholamine excretion in sexually abused girls. J Am Acad Child Adolesc Psychiatry 33(3): 320–327

De Bellis MD, Baum AS, Birmaher B, Keshavan MS, Eccard CH, Boring AM, Jenkins FJ, Ryan ND (1999a) A.E. Bennett Research Award. Developmental traumatology. I. Biological stress systems. Biol Psychiatry 45(10): 1259–1270

De Bellis MD, Keshavan MS, Clark DB, Casey BJ, Giedd JN, Boring AM, Frustaci K, Ryan ND (1999b) A.E. Bennett Research Award. Developmental traumatology. II. Brain development. Biol Psychiatry 45(10): 1271–1284

Devilly GJ, Spence SH (1999) The relative efficacy and treatment distress of EMDR and a cognitive-behavior trauma treatment protocol in the amelioration of posttraumatic stress disorder. J Anxiety Disord 13(1–2): 131–157

Devinsky O, Morrell MJ, Vogt BA (1995) Contributions of anterior cingulate to behaviour. Brain 118: 279–306

Deykin EY, Buka SL (1997) Prevalence and risk factors for posttraumatic stress disorder among chemically dependent adolescents. Am J Psychiatry 154(6): 752–757

Dhabhar FS, McEwen BS, Spencer RL (1993) Stress response, adrenal steroid receptor levels and corticosteroid-binding globulin levels – a comparison between Sprague-Dawley, Fischer 344 and Lewis rats. Brain Res 616: 89–98

Dinan TG, Barry S, Yatham LN, Mobayed M, Brown I (1990) A pilot study of a neuroendocrine test battery in posttraumatic stress disorder. Bio Psychiatry 28(8): 665–672

Duman RS, Heninger GR, Nestler EJ (1997) A molecular and cellular theory of depression. Arch Gen Psychiatry 54(7): 597–606

Egendorf A, Kaduschin C, Laufer RS, Rothbart G, Sloan L (1981) Legacies of Vietnam: comparative adjustment of veterans and their peers. Government Printing Office, Washington, DC

Engdahl B, Dikel TN, Eberly R, Blank A Jr (1997) Posttraumatic stress disorder in a community group of former prisoners of war: a normative response to severe trauma. Am J Psychiatry 154(11): 1576–1581

Falkai P (1999). Mirtazapine: other indications. J Clin Psychiatry 60[Suppl 17]: 36–40, 46–48

Farley M, Barkan H (1998) Prostitution, violence, and posttraumatic stress disorder. Women Health 27(3): 37–49

Feldman S, Conforti N (1980) Participation of the dorsal hippocampus in the glucocorticoid feedback effect on adrenocortical activity. Neuroendocrinology 30: 52–55

Figley C (ed) (1978a) Stress disorders among Vietman veterans. Brunner/Mazel, New York

Figley CR (1978b) Psychological adjustment among Vietnam veterans: an overview of the research. In: Figley CR (ed) Stress disorders among Vietnam veterans. Brunner/Mazel, New York

Finlay JM, Zigmond MJ, Abercrombie ED (1995) Increased dopamine and norepinephrine release in medial prefrontal cortex induced by acute and chronic stress: effects of diazepam. Neuroscience 64: 619–628

Frank JB, Kosten TR, Giller EL Jr, Dan E (1988) A randomized clinical trial of phenelzine and imipramine for posttraumatic stress disorder. Am J Psychiatry 145(10): 1289–1291

Friedman MJ (1996) Biological alterations in PTSD: implications for pharmacotherapy. Baillieres Clin Psychiatry 2: 245–262

*Friedman MJ (1998) Current and future drug treatment for posttraumatic stress disorder patients. Psychiatr Ann 8: 461–468

Friedman MJ, Schnurr PP (1995) Relationship-trauma, PST and physical health. In: Friedman MJ, Charney DS, Deutch AY (eds) Neurobiological and clinical consequences of stress: from normal adaptation to post-traumatic stress disorder. Lippincott-Raven, Philadelphia, pp 507–524

**Friedman MJ, Charney DS, Deutch AY (1995) Neurobiological and clinical consequences of stress: from normal adaptation to PTSD. Lippincott-Raven, Philadelphia

Foa EB (1997) Trauma and women: course, predictors, and treatment. J Clin Psychiatry 58[Suppl 9]: 25–28

Foa EB, Riggs DS, Dancu CV, Rothbaum BO (1993) Reliability and validity of a brief instrument for assessing post-traumatic stress disorder. J Trauma Stress 6: 459–473

**Foa EB, Davidson JRT, Frances A, Culpepper L, Ross R, Ross D (eds) (1999a) The expert consensus guideline series: treatment of posttraumatic stress disorder. J Clin Psychiatry 60 [Suppl 16]: 4–76

Foa EB, Dancu CV, Hembree EA, Jaycox LH, Meadows EA, Street GP (1999b) A comparison of exposure therapy, stress inoculation training, and their combination for reducing posttraumatic stress disorder in female assault victims. J Consult Clin Psychol 67(2): 194–200

Fontana A, Rosenheck R (1994) Posttraumatic stress disorder among Vietnam theater veterans. A causal model of etiology in a community sample. J Nerv Ment Dis 182(12): 677–684

Freedman SA, Brandes D, Peri T, Shalev A (1999) Predictors of chronic post-traumatic stress disorder: a prospective study. Br J Psychiatry 174: 353–359

Glover H (1984) Survival guilt and the Vietnam veteran. J Nerv Ment Dis 172(7): 393–397

Goldberg J, True WR, Eisen SA, Henderson WG (1990) A twin study of the effects of the Vietnam War on posttraumatic stress disorder. JAMA 263(9): 1227–1232

Goldman-Rakic PS (1988) Topography of cognition: parallel distributed networks in primate association cortex. Annu Rev Neurosci 11: 137–156

Gould E, Beylin A, Tanapat P, Reeves A, Shors TJ (1999) Learning enhances adult neurogenesis in the hippocampal formation. Nat Neurosci 2(3): 260–265

Green BL (1993) Disasters and posttraumatic stress disorder. In: Davidson JRT, Foa EB (eds) Posttraumatic stress disorder: DSM-IV and beyond. American Psychiatric Press, Washington, DC

Green BL, Lindy JD, Grace MC (1985) Posttraumatic stress disorder. Toward DSM-IV. J Nerv Ment Dis 173(7):406–411

Green BL, Lindy JD, Grace MC et al (1990) Buffalo Creek survivors in the second decade: stability of stress symptoms. Am J Orthopsychiatry 60(1): 43–54

Green BL, Grace MC, Vary MG, Kramer TL, Gleser GC, Leonard AC (1994) Children of disaster in the second decade: a 17-year follow-up of Buffalo Creek survivors. J Am Acad Child Adolesc Psychiatry 33(1): 71–79

Green BL, Rowland JH, Krupnick JL et al (1998) Prevalence of posttraumatic stress disorder in women with breast cancer. Psychosomatics 39(2): 102–111

Grinker RR, Spiegel JP (1945) Men under stress. Blakiston, Philadelphia

Guidotti A, Baraldi M, Leon A, Costa E (1990) Benzodiazepines: a tool to explore the biochemical and neuro-physiological basis of anxiety. Fed Proc 39: 1039–1042

Gunne LM, Reis DJ (1963) Changes in brain catecholamines associated with electrical stimulation of amygdaloid nucleus. Life Sci 11: 804–809

Gurvits TV, Shenton ME, Hokama H, Ohta H, Lasko NB, Gilbertson MW, Orr SP, Kikinis R, Jolesz FA, McCarley RW, Pitman RK (1996) Magnetic resonance imaging study of hippocampal volume in chronic, combat-related posttraumatic stress disorder. Biol Psychiatry 40(11): 1091–1099

Harvey AG, Bryant RA (1999) Predictors of acute stress following motor vehicle accidents. J Trauma Stress 12(3): 519–525

Hamner MB, Hitri A (1992) Plasma beta-endorphin levels in post-traumatic stress disorder: a preliminary report on response to exercise-induced stress. J Neuropsychiatry Clin Neurosci 4(1): 59–63

*Hamner MB, Lorberbaum JP, George MS (1999) Potential role of the anterior cingulate cortex in PTSD: review and hypothesis. Depression Anxiety 9(1): 1–14

Henning KR, Frueh BC (1997) Combat guilt and its relationship to PTSD symptoms. J Clin Psychol 53(8): 801–808

*Herman JL (1992) Trauma and recovery. Basic Books, New York

Herman JL, Perry JC, van der Kolk BA (1989) Childhood trauma in borderline personality disorder. Am J Psychiatry 146: 490–495

Herman JP, Schafer MK, Young EA, Thompson R, Douglass J, Akil H, Watson SJ (1989) Evidence for hippocampal regulation of neuroendocrine neurons of the hypothalamus-pituitary-adrenocortical axis. Neuroscience 9(9): 3072–3082

Hidalgo R, Hertzberg MA, Mellman T et al (1999) Nefazodone in post-traumatic stress disorder: results from six open-label trials. Int Clin Psychopharmacol 14(2): 61–68

Hilton SM, Zbrozyna AW (1963) Amygdaloid region for defense reactions and it efferent pathway to the brain stem. J Physiol 165: 160–173

Hitchcock J, Davis M (1986) Lesions of the amygdala, but not of the cerebellum or red nucleus, block conditioned fear as measured with the potentiated startle paradigm. Behav Neurosci 100(1): 11–22

Hitchcock JM, Sananes CB, Davis M (1989) Sensitization of the startle reflex by footshock: blockade by lesions of the central nucleus of the amygdala or its efferent pathway to the brainstem. Behav Neurosci 103(3): 509–518

Horowitz M (1976) Stress response syndromes. Aronson, New York

Horowitz M, Wilner N, Alvarez W (1979) Impact of Event Scale: a measure of subjective stress. Psychosom Med 41(3): 209–218

Huller U, Barash-Kishon R (1998) Psychodynamic-supportive group therapy model for elderly Holocaust survivors. Int J Group Psychother 48(4): 461–475

Irwin J, Ahluwalia P, Zacharko RM, Anisman H (1986) Central norepinephrine and plasma corticosterone following acute and chronic stressors: influence of social isolation and handling. Pharmacol Biochem Behav 24: 1151–1154

Iwata J, LeDoux JE, Meeley MP, Arneric S, Reis DJ (1986) Intrinsic neurons in the amygdaloid field projected to by the medial geniculate body mediate emotional responses conditioned to acoustic stimuli. Brain Res 383: 195–214

Jarrell TW, Gentile CG, Romanski LM, McCabe PM, Schneiderman N (1987) Involvement of cortical and thalamic auditory regions in retention of differential bradycardiac conditioning to acoustic conditioned stimuli in rabbits. Brain Res 412: 285–294

Kant GJ, Leu JR, Anderson SM, Mougey EH (1987) Effects of chronic stress on plasma corticosterone, ACTH and prolactin. Physiol Behav 40: 775–779

Kardiner A (1941) The traumatic neuroses of war. Hober, New York

*Keane TM, Kaloupek DG (1997) Comorbid psychiatric disorders in PTSD. Implications for research. Ann NY Acad Sci 821: 24–34

**Kessler RC, Sonnega A, Bromet E, Hughes M, Nelson CB (1995) Posttraumatic stress disorder in the National Comorbidity Survey. Arch Gen Psychiatry 52(12): 1048–1060

Kim JJ, Fanselow MS (1992) Modality-specific retrograde amnesia of fear. Science 256: 675–677

*Kleber RJ, Figley CR, Gersons BPR (eds) (1995) Beyond trauma: cultural and societal dynamics. Plenum, New York

Kluft RP (1992) The use of hypnosis with dissociative disorders. Psychiatr Med 10(4): 31–46

Knapp JF (1998) The impact of children witnessing violence. Pediatr Clin North Am 45(2): 355–364

*Koopman C, Classen C, Spiegel D (1994) Predictors of posttraumatic stress symptoms among survivors of the Oakland/Berkeley, Calif, firestorm. Am J Psychiatry 151: 888–894

Koren D, Arnon I, Klein E (1999) Acute stress response and posttraumatic stress disorder in traffic accident victims: a 1 year prospective follow-up study. Am J Psychiatry 156(3): 367–373

Kosten TR, Wahby V, Giller E Jr, Mason J (1990) The dexamethasone suppression test and thyrotropin-releasing hormone stimulation test in posttraumatic stress disorder. Bio Psychiatry 28(8): 657–664

Kraepelin E (1919) Dementia praecox and paraphrenia. Reprinted in 1971 by Krieger, Huntington

Kudler H, Davidson J, Meador K, Lipper S, Ely T (1987) The DST and posttraumatic stress disorder. Am J Psychiatry 144(8): 1068–1071

Kulka RA, Schlenger WE, Fairbank JA, Hough RL, Jordan BK, Marmar CR, Weiss DS (1990) Trauma and the Vietnam War generation: report of findings from the National Vietnam Veterans Readjustment Study. Brunner/Mazel, New York

**Kulka RA, Schlenger WE, Fairbank JA, Hough RL, Jordan BK, Marmar CR, Weiss DS (1991) Assessment of posttraumatic stress disorder in the community: prospects and pitfalls from recent studies of Vietnam veterans. Psychol Assess 3: 547–560

Landfield P, Baskin R, Pitler T (1981) Brain aging correlates: retardation by hormonal-pharmacological treatments. Science 214: 581–584

Landolt MA, Boehler U, Schwager C, Schallberger U, Nuessli R (1998) Post-traumatic stress disorder in paediatric patients and their parents: an exploratory study. J Paediatr Child Health 34(6): 539–543

Lawrence MS, Sapolsky RM (1994) Glucocorticoids accelerate ATP loss following metabolic insults in cultured hippocampal neurons. Brain Res 646: 303–306

Lazrove S, Triffleman E, Kite L, McGlashan T, Rounsaville B (1998) An open trial of EMDR as treatment for chronic PTSD. Am J Orthopsychiatry 68(4): 601–608

*LeDoux JE (1993) Emotional memory systems in the brain. Behav Brain Res 58: 69–79

LeDoux JE, Iwata J, Cicchetti P, Reis DJ (1988) Different projections of the central amygdaloid nucleus mediate autonomic and behavioral correlates of conditioned fear. J Neurosci 8(7): 2517–2529

Lemieux AM, Coe CL (1995) Abuse-related posttraumatic stress disorder: evidence for chronic neuroendocrine activation in women. Psychosom Med 57(2): 105–115

Levine ES, Litto WJ, Jacobs BL (1990) Activity of cat locus coeruleus noradrenergic neurons during the defense reaction. Brain Res 531: 189–195

Liberzon I, Abelson JL, Flagel SB, Raz J, Young EA (1999) Neuroendocrine and psychophysiologic responses in PTSD: a symptom provocation study. Neuropsychopharmacology 21(1): 40–50

Lohr JM, Kleinknecht RA, Tolin DF, Barrett RH (1995) The empirical status of the clinical application of eye movement desensitization and reprocessing. J Behav Ther Exp Psychiatry 26(4): 285–302

Lubin H, Loris M, Burt J, Johnson DR (1998) Efficacy of psychoeducational group therapy in reducing symptoms of posttraumatic stress disorder among multiply traumatized women. Am J Psychiatry 155(9): 1172–1177

Luine V, Villages M, Martinex C, McEwen BS (1994) Repeated stress causes reversible impairments of spatial memory performance. Brain Res 639: 167–170

Lundin T, Lotfi M (1996) Posttraumatic stress disorder in DSM-III-R, DSM-IV, and ICD-10: a comparison and evaluation of the significance of the respective diagnostic criteria. SO Nord J Psychiatry 50(1): 11–15

Lyons JA (1987) Posttraumatic stress disorder in children and adolescents: a review of the literature. J Dev Behav Pediatr 8(6): 349–356

Madden J, Akil H, Patrick RL, Barchas JD (1977) Stress induced parallel changes in central opioid levels and pain responsiveness in the rat. Nature 265: 358–360

Maes M, Delmeire L, Schotte C et al (1998a) Epidemiologic and phenomenological aspects of post-traumatic stress disorder: DSM-III-R diagnosis and diagnostic criteria not validated. Psychiatry Res 81(2): 179–193

Maes M, Delmeire L, Schotte C et al (1998b) The two-factorial symptom structure of post-traumatic stress disorder: depression-avoidance and arousal-anxiety. Psychiatry Res 81(2): 195–210

Maes M, Lin AH, Verkerk R, Delmeire L, Van Gastel A, Van der Planken M, Scharpe S (1999a) Serotonergic and noradrenergic markers of post-traumatic stress disorder with and without major depression. Neuropsychopharmacology 20(2): 188–197

Maes M, Lin AH, Bonaccorso S, Goossens F, Van Gastel A, Pioli R, Delmeire L, Scharpe S (1999b) Higher serum prolyl endopeptidase activity in patients with post-traumatic stress disorder. J Affect Disord 53(1): 27–34

Magarinos A, Somoza G, DeNicola A (1987) Glucocorticoid negative feedback and glucocorticoid receptors after hippocampectomy in rats. Horm Metab Res 19: 105–109

Maier SF, Davies S, Grau JW, Jackson RL, Morrison DH, Moye T, Madden J, Barchas JD (1981) Opiate antagonists and long-term analgesic reaction induced by inescapable shock in rats. J Compar Physiol Psychol 94: 1172–1183

Manne SL, Du Hamel K, Gallelli K, Sorgen K, Redd WH (1998) Posttraumatic stress disorder among mothers of pediatric cancer survivors: diagnosis, comorbidity, and utility of the PTSD checklist as a screening instrument. J Pediatr Psychol 23(6): 357–366

March J (1999) Assessment of pediatric posttraumatic stress disorder. In: Saigh PA, Bremner JD (eds) Posttraumatic stress disorder: a comprehensive text. Allyn and Bacon, New York, pp 199–219

Marmar CR, Weiss DS, Metzler TJ, Delucchi K (1996) Characteristics of emergency services personnel related to peritraumatic dissociation during critical incident exposure. Am J Psychiatry 153[Suppl 7]: 94–102

Mason JW, Giller EL, Kosten TR, Ostroff RB, Podd L (1986) Urinary free-cortisol levels in posttraumatic stress disorder patients. J Nerv Ment Dis 174(3): 145–149

Mason JW, Giller EL, Kosten TR, Harkness L (1988) Elevation of urinary norepinephrine/cortisol ratio in posttraumatic stress disorder. J Nerv Ment Dis 176(8): 498–502

Mason JW, Southwick S, Yehua R, Wang S, Riney S, Bremner D, Johnson D, Luin H, Blake D, Zhous G et al (1994) Elevation of serum free triiodothyronine, total triiodothyronine, thyroxine-binding globulin, and total thyroxine levels in combat-related posttraumatic stress disorder. Arch Gen Psychiatry 51(8): 629–641

*McCauley J, Kern DE, Kolodner K et al (1997) Clinical characteristics of women with a history of childhood abuse: unhealed wounds. JAMA 277(17): 1362–1368

McCormick CM, Smythe JW, Beers D (1994) Sex differences in type I corticosteroid receptor binding in selective brain areas of rats. Ann NY Acad Sci 746: 431–433

McDonald AJ (1991a) Organization of amygdaloid projections to prefrontal cortex and associated striatum in the rat. Neuroscience 44: 1–14

McDonald AJ (1991b) Topographical organization of amygdaloid projections to the caudatoputamen nucleus accumbens, and related striatal-like areas of the rat brain. Neuroscience 44: 15–33

McEwen BS, DeKloet ER, Rostene W (1986) Adrenal steroid receptors and actions in the nervous system. Physiol Rev 66: 1121–1188

McEwen BS, Angulo J, Cameron H, Chao HM, Daniels D, Gannon MN, Gould E, Mendelson S, Sakai R, Spencer R, Woolley C (1992) Paradoxical effects of adrenal steroids on the brain: protection versus degeneration. Biol Psychiatry 31: 177–199

McFarlane AC, Potts N (1999) Posttraumatic stress disorder: prevalence and risk factors relative to disasters. In: Saigh PA, Bremner JD (eds) Posttraumatic stress disorder: a comprehensive text. Allyn and Bacon, Boston, pp 92–102

McNally RJ (1997) Implicit and explicit memory for trauma-related information in PTSD. Ann NY Acad Sci 821: 219–224

McNally RJ, Shin LM (1995) Association of intelligence with severity of posttraumatic stress disorder symptoms in Vietnam combat veterans. Am J Psychiatry 152(6): 936–938

McNally RJ, Metzger LJ, Lasko NB, Clancy SA, Pitman RK (1998) Directed forgetting of trauma cues in adult survivors of childhood sexual abuse with and without posttraumatic stress disorder. J Abnorm Psychol 107(4): 596–601

Meadows EA, Foa EB (1999) Cognitive-behavioral treatment of traumatized adults. In: Saigh PA, Bremner JD (eds) Posttraumatic stress disorder: a comprehensive text. Allyn and Bacon, Boston, pp 376–390

Meaney MJ, Aitken DH, van Berkel C, Bhatnagar S, Sapolsky RM (1988) Effect of neonatal handling on age-related impairments associated with the hippocampus. Science 239: 766–768

Meichenbaum D (1974) Self-instructional methods. In: Kanfer FH, Goldstein AP (eds) Helping people change. Pergamon, New York, Pergamon, pp 357–391

Mellman TA, Kulick-Bell R, Ashlock LE, Nolan B (1995) Sleep events among veterans with combat-related posttraumatic stress disorder. American Journal of Psychiatry 152(1): 110–115

Metzger LJ, Orr SP, Berry NJ, Ahern CE, Lasko NB, Pitman RK (1999) Physiologic reactivity to startling tones in women with posttraumatic stress disorder. J Abnorm Psychol 108(2): 347–352

Morgan MA, LeDoux JE (1995) Differential contribution of dorsal and ventral medial prefrontal cortex to the acquisition and extinction of conditioned fear in rats. Behav Neurosci 109: 681–688

Murberg MM (ed) (1994) Catecholamine function in posttraumatic stress disorder: emerging concepts. American Psychiatric Press, Washington DC

Murphy SA, Braun T, Tillery L, Cain KC, Johnson LC, Beaton RD (1999) PTSD among bereaved parents following the violent deaths of their 12- to 28-year-old children: a longitudinal prospective analysis. J Trauma Stress 12(2): 273–291

*Nemiah JC (1998) Early concepts of trauma, dissociation, and the unconscious: their history and current implications. Progr Psychiatry 54: 1–26

Newcomer JW, Craft S, Hershey T, Askins K, Bardgett ME (1994): Glucocorticoid-induced impairment in declarative memory performance in adult humans. J Neurosci 14: 2047–2053

Nibuya M, Morinobu S, Duman RS (1995) Regulation of BDNF and trkB mRNA in rat brain by chronic electroconvulsive seizure and antidepressant drug treatments. J Neurosci 15: 7539–7547

Nibuya M, Nestler EJ, Duman RS (1996) Chronic antidepressant administration increases the expression of cAMP response element binding protein (CREB) in rat hippocampus. J Neurosci 16(7): 2365–2372

Nibuya M, Takahashi M, Russell DS, Duman RS (1999) Repeated stress increases catalytic TrkB mRNA in rat hippocampus. Neurosci Lett 267(2): 81–84

Nisenbaum LK, Zigmund MJ, Sved AF, Abercrombie ED (1991) Prior exposure to chronic stress results in enhanced synthesis and release of hippocampal norepinephrine in response to a novel stressor. J Neurosci 11: 1473–1484

Norris FH (1992) Epidemiology of trauma: frequency and impact of different potentially traumatic events on different demographic groups. J Consult Clin Psychol 60(3): 409–418

North CS, Nixon SJ, Shariat S, Mallonee S, McMillen JC, Spitznagel EL, Smith EM (1999) Psychiatric disorders among survivors of the Oklahoma City bombing. JAMA 282(8): 755–762

Olivera AA, Fero D (1990) Affective disorders, DST, and treatment in PTSD patients: clinical observations. J Trauma Stress 3: 407–414

Packan DR, Sapolsky RM (1990) Glucocorticoid endangerment of the hippocampus: tissue, steroid and receptor specificity. Neuroendocrinology 51: 613–618

Pacak K, Kvetnansky R, Palkovits M, Fukuhara K, Vadid G, Kopin IJ, Goldstein DJ (1993) Adrenalectory augments in vivo release of NE in the PVN during stress. Endocrinology 133: 1404–1410

Pacak K, Palkovits M, Kopin IJ, Goldstein DJ (1995) Stress induced NE release in the hypothalamic PVN and pituitary-adrenal and sympathodrenal activity: in vivo microdialysis studies. Front Neuroendocrinol 16: 89–150

Perry BD (1994) Neurobiological sequelae of childhood trauma: PTSD in children. In: Murberg MM (ed) Catecholamine function in posttraumatic stress disorder: emerging concepts. American Psychiatric Press, Washington DC, pp 233–256

Perry BD, U'Prichard DC (1981) [3H]rauwolscine (alpha-yohimbine): a specific antagonist radioligand for brain alpha 2-adrenergic receptors. Eur J Pharmacol 76(4): 461–464

Perry BD, Giller EL Jr, Southwick SM (1987) Altered platelet alpha 2-adrenergic binding sites in posttraumatic stress disorder. Am J Psychiatry 144(11): 1511–1512

*Peters L, Slade T, Andrews G (1999) A comparison of ICD10 and DSM-IV criteria for posttraumatic stress disorder. J Trauma Stress 12(2): 335–343

Petty F, Kramer G, Wilson L, Chae Y-L (1993) Learned helplessness and in vivo hippocampal norepinephrine release. Pharmacol Biochem Behav 46: 231–235

Pfefferbaum B (1997) Posttraumatic stress disorder in children: a review of the past 10 years. J Am Acad Child Adolesc Psychiatry 36(11): 1503–1511

Phillips M (1993) Turning symptoms into allies: utilization approaches with posttraumatic symptoms. Am J Clin Hypn 35(3): 179–189

**Phillips RG, LeDoux JE (1992) Differential contribution of amygdala and hippocampus to cued and contextual fear conditioning. Behav Neurosci 106: 274–285

*Pitman R, Orr S (1990) Twenty-four hour urinary cortisol and catecholamine excretion in combat-related posttraumatic stress disorder. Biol Psychiatry 27: 245–247

Plotsky PM, Meaney MJ (1993) Early, postnatal experience alters hypothalamic corticotropin-releasing factor (CRF) mRNA, median eminence CRF content and stress-induced release in adult rats. Brain Res Mol Brain Res 18(3): 195–200

*Prigerson HG, Shear MK, Jacobs SC, Reynolds CF 3rd, Maciejewski PK, Davidson JR, Rosenheck R, Pilkonis PA, Wortman CB, Williams JB, Widiger TA, Frank E, Kupfer DJ, Zisook S (1999) Consensus criteria for traumatic grief. A preliminary empirical test. Br J Psychiatry 174: 67–73

Prins A, Kaloupek DG, Keane TM (1995) Psychophysiological evidence for autonomic arousal and startle in tramatized adult populations. In: Friedman MJ, Charney DS, Deutch A (eds) Neurobiological and clinical consequences of stress: from normal adaptation to post-traumatic stress disorder. Lippincott-Raven, Philadelphia, pp 291–314

Putnam FW, Guroff JJ, Silberman EK, Barban L, Post RM (1986) The clinical phenomenology of multiple personality disorder: a review of 100 recent cases. J Clin Psychiatry 47: 285–293

Rauch SL, van der Kolk BA, Fisler RE, Alpert NM, Orr SP, Savage CR, Fischman AJ, Jenika MA, Pitman RK (1996) A symptom provocation study of posttraumatic stress disorder using positron emission tomography and script-driven imagery. Arch Gen Psychiatry 53(5): 380–387

Redmond DE Jr (1987) Studies of the nucleus locus coeruleus in monkeys and hypotheses for neuropsychopharmacology. In: Meltzer HY (ed) Psychopharmacology: the third generation of progress. Raven, New York, pp 967–975

Reist C, Kauffmann CD, Haier RJ, Sangdahl C, DeMet EM, Chicz-DeMet A, Nelson JN (1989) A controlled trial of desipramine in 18 men with posttraumatic stress disorder. Am J Psychiatry 146(4): 513–516

Resnick HS, Kilpatrick DG, Dansky BS, Saunders BE, Best CL (1993) Prevalence of civilian trauma and posttraumatic stress disorder in a representative national sample of women. J Consult Clin Psychol 61(6): 984–991

Resnick HS, Yehuda R, Pitman RK, Foy DW (1995) Effect of previous trauma on acute plasma cortisol level following rape. Am J Psychiatry 152(11): 1675–1677

Robert R, Blakeney PE, Villarreal C, Rosenberg L, Meyer,WJ 3rd (1999) Imipramine treatment in pediatric burn patients with symptoms of acute stress disorder: a pilot study. J Am Acad Child Adolesc Psychiatry 38(7): 873–882

Robin RW, Chester B, Rasmussen JK, Jaranson JM, Goldman D (1997) Prevalence and characteristics of trauma and post-traumatic stress disorder in a southwestern American Indian community. Am J Psychiatry 154(11): 1582–1588.

Robins L, Helzer J, Croughan J, Ratcliff K (1981) National Institute of Mental Health Diagnositc Interview Schedule: its history, characteristics, and validity. Arch Gen Psychiatry 38: 381–389

Romanski LM, LeDoux JE (1992) Equipotentiality of thalamo-amygdala and thalamo-cortico-amygdala circuits in auditory fear conditioning. J Neurosci 12: 4501–4509

Romanski LM, LeDoux JE (1993) Information cascade from primary auditory cortex to the amygdala: corticocortical and corticoamygdaloid projections of temporal cortex in the rat. Cerebr Cortex 3: 515–532

Rosen JB, Davis M (1988) Enhancement of acoustic startle by electrical stimulation of the amygdala. Behav Neurosci 102(2): 195–202, 324

Ross RT, Randich A (1984) Unconditioned stress-induced analgesia following exposure to brief footshock. J Exp Psychol Anim Behav Process 10(2): 123–137

Ross RJ, Ball WA, Sullivan KA, Caroff SN (1989) Sleep disturbance as the hallmark of posttraumatic stress disorder. Am J Psychiatry 146(6): 697–707

Ross RJ, Ball WA, Dinges DF, Kribbs NB, Morrison AR, Silver SM, Mulvaney FD (1994a) Rapid eye movement sleep disturbance in posttraumatic stress disorder. Biol Psychiatry 35: 195–202

Ross RJ, Ball WA, Dinges DF, Kribbs NB, Morrison AR, Silver SM, Mulvaney FD (1994b) Motor dysfunction during sleep in posttraumatic stress disorder. Sleep 17(8): 723–732

Roth RH, Tam SY, Ida Y et al (1988) Stress and the mesocorticolimbic dopamine systems. Annu Rev NY Acad Sci 537: 138–147

**Saigh PA, Bremner JD (eds) (1999) Posttraumatic stress disorder: a comprehensive text. Allyn and Bacon, New York

Saigh PA, Bremner JD (1999) The history of posttraumatic stress disorder. In: Saigh PA, Bremner JD (eds) Posttraumatic stress disorder: a comprehensive text. Allyn and Bacon, New York, pp 1–17

Saigh PA, Yasik AE, Sack WH, Koplewicz HS (1999a) Child-adolescent posttraumatic stress disorder: prevalence, risk factors and comorbidity. In: Saigh PA, Bremner JD (eds) (1999a) Posttraumatic stress disorder: a comprehensive text. Allyn and Bacon, Boston, pp 18–43

Saigh PA, Yasik AE, Oberfield RA, Inamdar SC (1999b) Behavioral treatment of child-adolescent posttraumatic stress disorder. In: Saigh PA, Bremner JD (eds) Posttraumatic stress disorder: a comprehensive text. Allyn and Bacon, Boston, pp 355–375

Sapolsky R (1986) Glucocorticoid toxicity in the hippocampus: synergy with an excitotoxin. Neuroendocrinology 43: 440–446

*Sapolsky RM (1996) Why stress is bad for your brain. Science 273: 749–750

Sapolsky RM, McEwen BS (1988) Why dexamethasone resistance? Two possible neuroendocrine mechanisms. In: Schatzberg AF, Nemeroff CB (eds) The hypothalamic-pituitary-adrenal axis: physiology, pathophysiology, and psychiatric implications. Raven, New York

Sapolsky RM, Plotsky PM (1990) Hypercortisolism and its possible neural bases. Biol Psychiatry 27: 937–952

Sapolsky R, Krey L, McEwen BS (1984a) Glucocorticoid-sensitive hippocampal neurons are involved in terminating the adrenocortical stress response. Proc Natl Acad Sci USA 81: 6174–6178

Sapolsky R, Krey L, McEwen BS (1984b) Stress downregulates corticosterone receptors in a site specific manner in the brain. Endocrinology 114: 287–293

Sapolsky RM, Uno H, Rebert CS, Finch CE (1990) Hippocampal damage associated with prolonged glucocorticoid exposure in primates. J Neurosci 10(9): 2897–2902

Sargent W, Slater E (1941) Amnesic syndromes in war. Proc R Soc Med 34: 757–674

Schlenger WE, Fairbank JA, Jordan BK, Caddell JM (1999) Combat-related PTSD: prevalence, risk factors an comorbidity. In: Saigh PA, Bremner JD (1999) Posttraumatic stress disorder: a comprehensive text. Allyn and Bacon, Boston, pp 69–91

Schnurr PP, Spiro A 3rd (1999) Combat exposure, posttraumatic stress disorder symptoms, and health behaviors as predictors of self-reported physical health in older veterans. J Nerv Ment Dis 187(6): 353–359

Schnurr PP, Friedman MJ, Green BL (1996) Post-traumatic stress disorder among World War II mustard gas test participants. Mil Med 161(3): 131–136

Shalev AY, Peri T, Canetti L, Schreiber S (1996a) Predictors of PTSD in injured trauma survivors: a prospective study. Am J Psychiatry 153: 219–225

Shalev AY, Bonne O, Eth S (1996b) Treatment of posttraumatic stress disorder: a review. Psychosom Med 58(2): 165–182

Shapiro F (1989) Eye movement desensitization: a new treatment for post-traumatic stress disorder. J Behav Ther Exp Psychiatry 20(3): 211–217

Shapiro F (1999) Eye movement desensitization and reprocessing (EMDR) and the anxiety disorders: clinical and research implications of an integrated psychotherapy treatment. J Anxiety Disord 13(1–2): 35–67

Shaw JA, Applegate B, Schorr C (1996) Twenty-one-month follow-up study of school-age children exposed to Hurricane Andrew. J Am Acad Child Adolesc Psychiatry 35(3): 359–364

Shay J, Munroe J (1999) Group and milieu therapy for veterans with complex posttrauamtic stress disorder. In: Saigh PA, Bremner JD (eds) Posttraumatic stress disorder: a comprehensive text. Allyn and Bacon, Boston, pp 391–413

Shestatzky M, Greenberg D, Lerer BA (1988) A controlled trial of phenelzine in posttraumatic stress disorder. Psychiatry Res 24(2): 149–155

Shin LM, McNally RJ, Kosslyn SM, Thompson WL, Rauch SL, Alpert NM, Metzger LJ, Lasko NB, Orr SP, Pitman RK (1997) A positron emission tomographic study of symptom provocation in PTSD. Ann NY Acad Sci 821: 521–523

Shin LM, McNally RJ, Kosslyn SM, Thompson WL, Rauch SL, Alpert NM, Metzger LJ, Lasko NB, Orr SP, Pitman RK (1999) Regional cerebral blood flow during script-driven imagery in childhood sexual abuse-related PTSD: a PET investigation. Am J Psychiatry 156(4): 575–584

Skre I, Onstad S, Torgersen S, Lygren S, Kringlen E (1993) A twin study of DSM-III-R anxiety disorders. Acta Psychiatr Scand 88(2): 85–92

Smith MA, Makino S, Kvetnansky R, Post RM (1995) Stress and glucocorticoids affect the expression of brain-derived neurotrophic factor and neurotrophin-3 mRNA in the hippocampus. J Neurosci 15: 1768–1777

Solomon SD, Davidson JR (1997) Trauma: prevalence, impairment, service use, and cost. J Clin Psychiatry 58[Suppl 9]: 5–11

Southwick SM, Krystal JH, Morgan CA, Johnson D, Nagy LM, Nicolaou A, Heninger GR, Charney DS (1993) Abnormal noradrenergic function in PTSD. Arch Gen Psychiatry 50(4): 266–274

*Southwick SM, Krystal JH, Bremner JD, Morgan CA 3rd, Nicolaou AL, Nagy LM, Johnson DR, Heninger GR, Charney DS (1997) Noradrenergic and serotonergic function in posttraumatic stress disorder. Arch Gen Psychiatry 54(8): 749–758

Southwick SM, Morgan CA, Charney DS, High JR (1999) Yohimbine use in a natural setting: effect on PTSD. Biol Psychiatry 46(3): 442–444

Sparr LF, Pitman RK (1999) Forensic assessment of traumatized adults. In: Saigh PA, Bremner JD (eds) Posttraumatic stress disorder: a comprehensive text. Allyn and Bacon, Boston, pp 284–309

Spiegel D (1992) The use of hypnosis in the treatment of PTSD. Psychiatr Med 10(4): 21–30

Spiegel D (ed) (1999) Efficacy and cost-effectiveness of psychotherapy. Clinical practice. American Psychiatric Association, Washington, DC

*Spiegel D, Cardena E (1991) Disintegrated experience: the dissociative disorders revisited. J Abnorm Psychol 100: 366–378

Spitzer RL, Williams JBW, Gibbon M, First MB (1990) Structured Clinical Interview for DSM-IIIR. American Psychiatric Press, Washington, DC

Spivak B, Vered Y, Graff E, Blum I, Mester R, Weizman A (1999) Low platelet-poor plasma concentrations of serotonin in patients with combat-related posttraumatic stress disorder. Biol Psychiatry 45(7): 840–845

Squire LR, Zola-Morgan S (1991) The medial temporal lobe memory system. Science 253: 1380–1386

Stallard P, Velleman R, Baldwin S (1998) Prospective study of post-traumatic stress disorder in children involved in road traffic accidents. Br Med J 317(7173): 1619–1623

*Stamm BH (ed) (1996) Measurement of stress, trauma, and adaptation. Sidran, Lutherville

Stein MB, Koverola C, Hanna C, Torchia MG, McClarty B (1997) Hippocampal volume in women victimized by childhood sexual abuse. Psychol Med 27(4): 951–959

Stein-Behrens BA, Lin WJ, Sapolsky RM (1994) Physiological elevations of glucocorticoids potentiate glutamate accumulation in the hippocampus. J Neurochem 63: 596–602

Steiner H, Garcia IG, Matthews Z (1997) Posttraumatic stress disorder in incarcerated juvenile delinquents. J Am Acad Child Adolesc Psychiatry 36(3): 357–365

Stretch RH, Knudson KH, Durand D (1998) Effects of premilitary and military trauma on the development of post-traumatic stress disorder symptoms in female and male active duty soldiers. Mil Med 163(7): 466–470

Stuckey J, Marra S, Minor T, Insel TR (1989) Changes in mu opiate receptors following inescapable shock. Brain Res 476: 167–169

Sutker PB, Winstead DK, Galina ZH, Allain AN (1991) Cognitive deficits and psychopathology among former prisoners of war and combat veterans of the Korean conflict. Am J Psychiatry 148(1): 67–72

**Sutherland SM, Davidson JRT (1999) Pharmacological treatment of posttraumatic stress disorder. In: Saigh PA, Bremner JD (1999) (eds) Posttraumatic stress disorder: a comprehensive text. Allyn and Bacon, Boston, pp 327–354

Taal L, Faber AW (1998) Posttraumatic stress and maladjustment among adult burn survivors 1 to 2 years postburn. II. The interview data. Burns 24(5): 399–405

*Terr LC (1991) Childhood traumas: an outline and overview. Am J Psychiatry 148(1): 10–20

Thabet AA, Vostanis P (1999) Post-traumatic stress reactions in children of war. J Child Psychol Psychiatry 40(3): 385–391

Thierry AM, Pirot S, Gioanni Y, Glowinski J (1998) Dopamine function in the prefrontal cortex. Adv Pharmacol 42: 717–720

Thompson MP, Kingree JB (1998) The frequency and impact of violent trauma among pregnant substance abusers. Addict Behav 23(2): 257–262

Thulesius H, Hakansson A (1999) Screening for posttraumatic stress disorder symptoms among Bosnian refugees. J Trauma Stress 12(1): 167–174

Thygesen P, Hermann K, Willanger R (1970) Concentration camp survivors in Denmark: persecution, disease, compensation. Dan Med Bull 17: 65–108

Torrie A (1944) Psychosomatic casualties in the Middle East. Lancet 29: 139–143

True WR, Rice J, Eisen SA, Heath AC, Goldberg J, Lyons MJ, Nowak J (1993) A twin study of genetic and environmental contributions to liability for posttraumatic stress symptoms. Arch Gen Psychiatry 50(4): 257–264

Tucker P, Dickson W, Pfefferbaum B, McDonald NB, Allen G (1997) Traumatic reactions as predictors of posttraumatic stress six months after the Oklahoma City bombing. Psychiatr Serv 48(9): 1191–1194

Turner B, Gupta KC, Mishkin M (1978) The locus and cytoarchitecture of the projection areas of the olfactory bulb in Macaca mulatta. J Compar Neurol 177: 381–396

Uddo M, Vasterling JJ, Brailey K, Sutker PB (1993) Memory and attention in combat-related post-traumatic stress disorder (PTSD). J Psychopathol Behav Assess 15(1): 43–52

*Uno H, Tarara R, Else JG, Suleman MA, Sapolsky RM (1989) Hippocampal damage associated with prolonged and fatal stress in primates. J Neurosci 9: 1705–1711

Uno H, Lohmiller L, Thieme C, Kemnitz JW, Engle MJ, Roecker EB (1990) Brain damage induced by prenatal exposure to dexamethasone in fetal rhesus monkeys. I. Hippocampus. Dev Brain Res 53: 157–167

**Ursano RJ, McCaughey BG, Fullerton CS (eds) (1994) Individual and community responses to trauma and disaster: the structure of human chaos. Cambridge University Press, Cambridge

Ursano RJ, Fullerton CS, Vance K, Kao TC (1999a) Posttraumatic stress disorder and identification in disaster workers. Am J Psychiatry 156(3): 353–359

Ursano RJ, Fullerton CS, Epstein RS et al (1999b) Acute and chronic posttraumatic stress disorder in motor vehicle accident victims. Am J Psychiatry 156(4): 589–595

Vaher PR, Luine VN, Gould E, McEwen BS (1994) Effects of adrenalectomy on spatial memory performance and dentate gyrus morphology. Brain Res 1004: 656: 71–78

Valentino R, Foote S (1988) Corticotropin-releasing hormone increases tonic but not sensory-evoked activity of noradrenergic locus coeruleus neurons in unanesthetized rats. J Neurosci 8: 1016–1025

**Van der Kolk, Bessel A, McFarlane, Alexander C (eds) (1996) Traumatic stress: the effects of overwhelming experience on mind, body, and society. Guildford, New York

Van der Kolk BA, Dryfuss D, Michaels M, Berkowitz R, Saxe G, Goldenberg I (1994) Fluoxetine in posttraumatic stress disorder. J Clin Psychiatry 55: 517–522

Vasterling JJ, Brailey K, Constans JI, Sutker PB (1998) Attention and memory dysfunction in posttraumatic stress disorder. Neuropsychology 12(1): 125–133

van Oers HJ, de Kloet ER, Levine S (1998) Early vs. later maternal deprivation differentially alters the endocrine and hypothalamic responses to stress. Brain Res Dev Brain Res 111(2): 245–252

Vermetten E, Bremner JD (2000) Dissociative amnesia: re-remembering traumatic memories. Cambridge University Press, Cambridge, pp 400–432

Veronen LJ, Kilpatrick DG, Resick PA (1979) Treating fear and anxiety in rape victims: implications for the criminal justice system. In: Parsonage WH (ed) Perspectives on victimology. Sage, Beverly Hills, pp 148–159

Vetrugno GC, Lachuer J, Perego C, Miranda E, DeSimons MG, Tappaz M (1993) Lack of glucocorticoids sustains the stress induced release of noradrenaline in the anterior hypothalamus. Neuroendocrinology 57: 835–842

Vogt BA, Miller MW (1983) Cortical connections between rat cingulate cortex and visual, motor, and postsubicular cortices. J Compar Neurol 216: 192–210

Wagner D, Heinrichs M, Ehlert U (1998) Prevalence of symptoms of posttraumatic stress disorder in German professional firefighters. Am J Psychiatry 155(12): 1727–1732

Walker AM, Harris G, Baker A, Kelly D, Houghton J (1999) Post-traumatic stress responses following liver transplantation in older children. J Child Psychol Psychiatry 40(3): 363–374

Watanabe Y, Sakai RR, McEwen BS, Mendelson S (1993) Stress and antidepressant effects on hippocampal and cortical 5HT1A and 5HT2 receptors and transport sites for serotonin. Brain Res 615: 87–94

Watson CG, Juba MP, Manifold V, Kucala T, Anderson PE (1991) The PTSD interview: rationale, description, reliability, and concurrent validity of a DSM-III-based technique. J Clin Psychol 47(2): 179–188

Weathers FW, Keane TM (1999) Psychological assessment of traumatized adults. In: Saigh PA, Bremner JD (eds) Posttraumatic stress disorder: a comprehensive text. Allyn and Bacon, Boston, pp. 219–247

Weisaeth L, Eitinger L (1993) Posttraumatic stress phenomena: common themes across wars, disasters, and traumatic events. In: Wilson JP, Raphael B (eds) International handbook of traumatic stress syndromes. Plenum, New York, pp 69–77 (Plenum series on stress and coping)

WHO (1948) ICD-6 classification of mental and behavioral disorders: clinical descriptions and diagnostic guidelines. World Health Organization, Geneva

WHO (1968) ICD-8 classification of mental and behavioral disorders: clinical descriptions and diagnostic guidelines. World Health Organization, Geneva

WHO (1977) ICD-9 classification of mental and behavioral disorders: clinical descriptions and diagnostic guidelines. World Health Organization, Geneva

WHO (1992) ICD-10 classification of mental and behavioral disorders: clinical descriptions and diagnostic guidelines. World Health Organization, Geneva

Widom CS (1999) Posttraumatic stress disorder in abused and neglected children grown up. Am J Psychiatry 156(8): 1223–1229

Wijma K, Soderquist J, Wijma B (1997) Posttraumatic stress disorder after childbirth: a cross sectional study. J Anxiety Disord 11(6): 587–597

*Wilson JP (1994) The historical evolution of PTSD diagnostic criteria: from Freud to DSM-IV. J Trauma Stress 7(4): 681–698

Wilson JP, Keane TM (1997) Assessing psychological trauma and PTSD. Guilford, New York

Wolpe J (1958) Psychotherapy by reciprocal inhibition. Stanford University Press, Stanford

Wooley CS, Gould E, McEwen BS (1990) Exposure to excess glucocorticoids alters dendritic morphology of adult hippocampal pyramidal neurons. Brain Res 531: 225–231

Yehuda R, Lowy MT, Southwick SM, Shaffer S, Giller EL (1991a) Increased number of glucocorticoid receptor number in posttraumatic stress disorder. Am J Psychiatry 149: 499–504

Yehuda R, Giller EL, Southwick SM, Lowy MT, Mason JW (1991b) Hypothalamic-pituitary-adrenal dysfunction in posttraumatic stress disorder. Biol Psychiatry 30(10): 1031–1048

**Yehuda R, Southwick SM, Krystal JH, Bremner D, Charney DS, Mason JW (1993) Enhanced suppression of cortisol following dexamethasone administration in posttraumatic stress disorder. Am J Psychiatry 150(1): 83–86

Yehuda R, Teicher MH, Levengood RA, Trestman RL, Siever LJ (1994) Circadian regulation of basal cortisol levels in posttraumatic stress disorder. Ann NY Acad Sci 746: 378–380

Yehuda R, Keefe RS, Harvey PD, Levengood RA, Gerber DK, Geni J, Siever LJ (1995) Learning and memory in combat veterans with posttraumatic stress disorder. Am J Psychiatry 152(1): 137–139

Yehuda R, Resnick HS, Schmeidler J, Yang RK, Pitman RK (1998) Predictors of cortisol and 3-methoxy-4-hydroxyphenylglycol responses in the acute aftermath of rape. Biol Psychiatry 43(11): 855–859

Young WC (1993) Sadistic ritual abuse. An overview in detection and management. Primary Care Clin Office Pract 20(2): 447–458

Zeitlin SB, McNally RJ (1991) Implicit and explicit memory bias for threat in post-traumatic stress disorder. Behav Res Ther 29(5): 451–457

Zimmerman M, Mattia JI (1999) Psychotic subtyping of major depressive disorder and posttraumatic stress disorder. J Clin Psychiatry 60(5): 311–314

Zisook S, Chentsova-Dutton Y, Shuchter SR (1998) PTSD following bereavement. Ann Clin Psychiatry 10(4): 157–163

CHAPTER

5

U. Schweiger, F. Hohagen

Adjustment Disorders

Translator: E. Taub

1 Introduction

The diagnostic category of the adjustment disorders consists of temporally limited, event-related disorders that do not undergo a transition to one of the major psychiatric disorders, such as post-traumatic stress disorder (PTSD), severe depressive episode, or schizophrenia, but instead take a "benign," transient course. The concept of the adjustment disorders as event-related psychogenic disorders is in the tradition of the "abnormal reaction to experience" (*abnorme Erlebnisreaktion*; Schneider 1980). Adjustment disorder is a diagnostic category often applied in psychiatric practice. The development of mental symptoms as a consequence of life events or changes in life is a practically universal phenomenon (see Rahe 1995). The diagnosis of an adjustment disorder has the advantage of being plausible to laymen while not being laden with the "stigma" of a severe mental illness. It thus meets the need of the clinician for a widely accepted label for the common, transient mental disorders that are presumed to have a benign prognosis.

These same characteristics, however, make the adjustment disorders one of the more problematic categories of DSM-IV and ICD-10. DSM-IV is intended to be atheoretical; however, unlike the disorders in most other categories, the adjustment disorders have to be linked to an "etiological" event. The diagnosis of an adjustment disorder is thus exceptional in requiring more than just a clinical description. There is an implicit requirement for an assessment of the relative contributions of "exogenous" and "endogenous" factors, with a qualitative judgment of an event as a decisive change in life, as opposed to a stress of the more "usual" kind, which might coincidentally be present at the same time as the abnormal mental manifestations. Furthermore, an assessment of prognosis regarding the severity of the disorder in its future course, and its projected duration, is also required. Unfortunately, these implicitly required assessments rest on an insecure empirical foundation and are of only limited reliability. A further, potentially serious problem is that the wide acceptance of the diagnosis of adjustment disorders among both clinicians and patients leads to an excessive use of this category.

The concept of the adjustment disorders presupposes several seemingly plausible assumptions regarding the nature of life events and the manner in which individuals cope with them. One of these is the assumption that life events strike people essentially at random. This assumption, however, is not supported by the available literature. Rather, prospective studies indicate that many individuals consistently experience life events at a high frequency, while others consistently experience them at a low frequency. The frequency of life events has a moderately high, statistically significant correlation within twin pairs, and the correlation coefficient is higher for monozygotic than for dizygotic twins. Possible explanations for this that have been proposed include similar social status, level of social competence, extent of general psychopathology, lifestyle, and consumption of psychotropic substances (Kendler et al. 1993). There are as yet no research findings available concerning the relative contributions of high-risk behavior, lack of self-protective behavior, and social standing to the variation among individuals in their predisposition to life events.

A further assumption underlying the concept of the adjustment disorders is that normative coping processes exist and that the ability to cope with stressors is normally distributed around this norm. This assumption, however, conflicts with the known extreme heterogeneity of human reactions.

2 Diagnosis and Clinical Features

The diagnosis of an adjustment disorder can be made on an operationalized basis in accordance with either DSM-IV (Appendix A) or ICD-10 (Appendix B). An adjustment disorder must be linked to a specific external stressor, but its manifestations need not appear immediately. DSM-IV allows an interval of up to 3 months before the appearance of manifestations, while ICD-10 allows no more than 1 month. Furthermore, the manifestations need not remit immediately upon removal of the stressor, but they must do so within 6 months. If the stressor remains present, a chronic adjustment disorder may result; DSM-IV gives no upper limit for the duration of such a disorder, while ICD-10 allows a maximum of 2 years.

The manifestations that may be associated with the diagnosis of an adjustment disorder are extremely diverse. Characteristic manifestations include transient emotional symptoms, such as feelings of dejection, worry, grief, hopelessness, discouragement, anxiety, fear, discontentment, irritation, guilt, shame, jealousy, suspicion, weariness, humiliation, anger, rage, and aggressiveness, but also helplessness, a feeling of having no feelings, emotional depletion, or a transient affective instability, with sudden mood fluctuations lasting minutes or hours. There may be disturbances of sleep, appetite, and sexual function.

Behavioral changes are common, including neglect of occupational, scholastic, academic, social, and interpersonal responsibilities and avoidance behavior, with

transient anxious avoidance of social situations or of situations that might aggravate manifestations of anxiety or make them visible in public (e.g. use of public transport, elevators, cable cars, tunnels). Transient increased consumption of substances (alcohol, nicotine) is also common. Transient compulsive behaviors or obsessive thoughts are occasionally observed.

Somatoform symptoms, i.e. pain and other physical complaints for which there is no medical explanation, are also common, as are transient, relatively mild dissociative symptoms. Particularly in children and young adults, there are transient abnormalities of conduct with oppositional or rebellious behavior, destruction of one's own or others' property, theft, or violence.

There may be autoaggressive behavior such as physical injury to oneself or suicidal actions. Adjustment disorders are among the more common diagnoses in groups of patients studied after a first attempt at suicide (Johnsson et al. 1996). Currently available studies, however, allow no precise estimate of the risk of suicide associated with the diagnosis of adjustment disorder.

The intensity and duration of all manifestations may vary considerably. The diagnosis of an adjustment disorder, as opposed to patterns of coping lying within the behavioral "norm," requires that they be intense enough to cause marked distress and/or impair social or occupational performance ability. Furthermore, the manifestations should not merely represent a simple grief reaction. Nonetheless, the differentiation of the adjustment disorders from other, major mental disorders requires that the intensity, duration, and multiplicity of manifestations lie below the threshold values set by whichever diagnostic manual is used (DSM-IV or ICD-10) and, in particular, that they not merely represent an exacerbation of the manifestations of a chronic, or chronically recurring, anxiety disorder, depressive disorder, somatoform disorder, or personality disorder. This is especially true with respect to personality disorders with a high degree of reactivity. In such cases, the additional diagnosis of an adjustment disorder should be assigned only if the manifestations arising in response to the stressor are atypical for the personality disorder in question.

The adjustment disorders thus constitute a "transitional category" between severe mental disorders and states considered to lie within the range of normal coping patterns (Fabrega et al. 1987). The presence of perceptual disturbances such as hallucinations, formal thought disorders beyond mere brooding, disorders of thought content with delusions, or impairment of cognitive abilities is incompatible with the diagnosis of an adjustment disorder. Both DSM-IV and ICD-10 allow the coding of various subtypes of adjustment disorder according to their predominant manifestations.

3 Differential Diagnosis

The differentiation of the adjustment disorders from depressive episodes and recurrent depressive disorders (DSM-IV 296.xx and ICD-10 F32, F33), anxiety disorders, post-traumatic stress disorders, somatoform disorders, dissociative disorders (DSM-IV 300.xx and ICD-10 F40–45), disorders of conduct (DSM-IV 312.xx and ICD-10 F91, F92), and personality disorders (DSM-IV 301.xx and ICD-10 F60, F61) is of major clinical importance.

Several factors may complicate the differential diagnosis: preceding episodes of depression or anxiety, or behavioral abnormalities, may be forgotten or trivialized by the patient during history-taking ("recall bias"). A feeling that it is necessary to establish a cause for the patient's ailment, or an inappropriate attachment to an external model of disease, may lead the diagnostician to exaggerate the importance of a current external event. On the other hand, the intensity and significance of preceding mental manifestations, and of persistent dysfunctional patterns of inner experience and behavior, may occasionally be underestimated because of lack of insight into the illness and the egosyntonic character of the symptoms. Patients with cluster B personality disorders (borderline, histrionic, and narcissistic personality disorders), in particular, have a marked reactivity to external events and attribute their mood fluctuations to them. When the clinician is inadequately informed about the patient's previous history, this may lead to the erroneous diagnosis of an adjustment disorder.

The symptoms of post-traumatic disorders, such as flashbacks and nightmares, are sometimes suppressed by patients in talking with psychiatrists and psychotherapists because of associated feelings of anxiety, shame, or guilt; post-traumatic disorders are therefore underdiagnosed. The time at which the diagnosis is assigned also influences the result. If the patient is seen by the clinician early in the course of the illness, the manifestations may be less than fully developed, and the criteria for a specific disorder not yet fulfilled.

4 Etiology and Pathogenesis

By definition, adjustment disorders are induced by one or more stressors. The severity of the stressor is typically not linearly related to the nature and extent of the resulting psychopathology in the affected individuals. The level of stress experienced by the patient is a complex function not only of the intensity, duration,

controllability, and reversibility of the stressor, but also of individual factors such as the resources and coping strategies at the patient's disposal, social support, personal context, and personal value systems.

DSM-IV has no upper or lower limit for the severity of the stressor, but ICD-10 excludes catastrophic or unusual stressors while also requiring a minimal level of severity (decisive change in life). Thus, in DSM-IV, but not in ICD-10, the etiological factors of the adjustment disorders overlap with those of post-traumatic stress disorder. DSM-IV distinguishes these entities from each other solely on the basis on the severity of their manifestations. On the other hand, life events also play a major role in severe depressive episodes, particularly in the induction of the initial depressive phase, and there is therefore some overlap here as well.

Stressors may appear in the form of single events such as the loss of a partner, family member, or significant other through moving away, separation, divorce, or death; rejection or severe interpersonal conflict with significant others or with the social environment; severe physical illnesses in the affected individual, family members, or significant others; loss of employment; scholastic or academic events, such as failing an examination or having to repeat a school year; judicial events such as arrest, conviction, or losing a lawsuit; political events such as the loss of office or public disputes; mobbing; political, religious, or ethnic discrimination or persecution; crime, accidents, and natural disasters. Even "positive" life events such as marriage, pregnancy, birth of children, promotions, and financial gains can become stressors, particularly when these place demands on the affected individual that he or she cannot meet.

Stressors can also be chronic or recurrent. Examples include poverty, lack of access to health care, chronic illness in the affected individual or significant others, chronic intrafamilial conflicts, and seasonal financial or health problems (e.g. seasonal employment, extreme climatic conditions). Stressors may affect individuals, families, or entire social groups or communities. Thus, in summary, the assessment of stressors requires attention to their degree of severity and to the differentiation of adjustment disorder from post-traumatic stress disorder. DSM-IV and ICD-10 draw the distinction between these two entities in somewhat different ways.

Although the events listed above can lead, quite understandably, to the appearance of emotional symptoms and behavioral abnormalities, major negative life events do not, as a rule, produce mental functional disturbances in the clinical sense. It is assumed that the mental consequences of such events arise through an interaction of stressors, coping mechanisms, and a predisposition to mental disturbances. This interaction has indeed been the subject of scientific investigation, yet little attention has been paid to the category of the adjustment disorders: the available studies on the consequences of highly negative or catastrophic life events deal with general mental manifestations and/or the presence or absence of post-traumatic stress disorder, but hardly any data exist regarding the frequency of adjustment disorders as defined by DSM-III-R or DSM-IV.

Thus, in a study of 38 children from stable families, the death of a parent was associated with significant grief, but with no further limitation of function. The affected children were similar in most dimensions to age-matched normal control subjects, and they were in significantly better condition than a depressive control group (Fristad et al. 1993). According to another study, in the 2 years following an isolated, catastrophic accident, the prevalence of post-traumatic stress disorder in individuals providing assistance to the victims is 7%–12% (Epstein et al. 1998). Among professional firefighters, who are constantly exposed to potentially traumatizing events, the prevalence of post-traumatic stress disorder is approximately 18%, while 49% have individual manifestations of post-traumatic stress disorder (Wagner et al. 1998). In general, significant negative life events raise the risk of psychiatric disorders in the following year by a factor of 3–6 (Goodyer et al. 1987).

There may be predisposing psychological risk factors, including unstable interpersonal relationships, lack of social support, impaired social competence, poor ability to communicate, and a tendency toward overdramatic or dichotomous assessment of external events, or toward inward attribution of external problems. One of the few prospective studies to be performed did indeed reveal that a preexisting tendency to anxiety, behavioral problems, and poor school performance influenced the extent of mental symptoms in children who experienced a natural disaster (La Greca et al. 1998).

Psychobiological studies on adjustment disorder are largely absent. There is evidence, however, that the functioning of the serotonergic system may be deficient in individuals suffering from adjustment disorders associated with suicidality (Mann et al. 1999).

5 Epidemiology

Data on the lifetime prevalence of the adjustment disorders in the general population are lacking. The prevalence of adjustment disorder has been studied only in various patient populations and has been found to depend both on the care structure and on the therapeutic setting. Even though the diagnosis of the

adjustment disorders has been operationalized, traditional diagnostic habits still exert a major influence; thus biologically oriented psychiatrists assign this diagnosis less frequently than psychodynamically oriented psychiatrists. All of these factors together account for the marked variation in the reported values for the prevalence of the adjustment disorders.

An adjustment disorder was diagnosed in 5% of all patients seen in the University of Iowa psychiatry department over 4 years. At the level of manifestations, behavioral abnormalities were more common in adolescent patients than in adults (running away, rule infractions, poor school performance, substance use). Adults more commonly had depressive manifestations and disturbances of appetite and sleep (Andreasen and Wasek 1980). A multicenter study carried out by the psychiatric liaison services of seven American, Canadian, and Australian teaching hospitals revealed the presence of an adjustment disorder in 12% of patients as a single diagnosis, and in a further 4% as a second diagnosis in addition to another disorder on axis I or II (Strain et al. 1998).

In the Monroe County Psychiatric Case Register, the adjustment disorders accounted for 28% of all diagnoses in individuals aged 12–18. The frequency of the adjustment disorders was higher in outpatient care and lower in academic psychiatry departments. There was no difference between the sexes. The number of psychiatric contacts following diagnosis was similar to that of patients with neurotic disorders or personality disorders, but much lower than that of patients with schizophrenia. A study carried out in several outpatient care facilities in the United States revealed that adjustment disorder accounted for 24%–65% of all psychiatric diagnoses in patients under the age of 18 (Jacobson et al. 1980). At least one further mental disorder was diagnosed in more than half of the patients in long-term follow-up (Weiner and Del Gaudio 1976; Cantwell and Baker 1989).

6 Course and Prognosis

Adjustment disorders have a relatively benign short-term prognosis. Most patients return to their original functional level. Thus a study at the University of Pittsburgh showed that 29 of 30 study participants with an adjustment disorder went into remission, although the average duration of an episode in adolescents was longer than expected (8.9 months; Kovacs et al. 1994). Despite this favorable short-term prognosis, however, the same study revealed a relatively unfavorable long-term prognosis. Only 59% of these subjects were predominantly in good health over a follow-up period of 5 years. The frequency of new mental disturbances was considerable: 45% received the diagnosis of an adjustment disorder for a second time, while the diagnosis of an anxiety disorder was made in 10%, substance abuse in 10%, and a severe depressive episode in 17%. Thus, after correction for the extent of mental manifestations, the long-term prognosis of adjustment disorder was no more favorable than that of the other psychiatric diagnostic groups.

Adolescents often have a less favorable course than adults. In the University of Iowa study, 71% of adults with an adjustment disorder were in good mental condition 5 years after the index treatment, as compared to only 44% of adolescent study participants. A total of 19% of the adolescents and 13% of the adults had gone on to develop a severe depressive episode. Schizophrenia had been diagnosed in 6% of the adolescents. Alcohol dependency or alcohol abuse was found in 17% of the adolescents and 13% of the adults, while 2% of the adolescents and 4% of the adults had committed suicide (Andreasen and Hoenk 1982).

A follow-up study of patients treated by a German crisis intervention unit revealed a somewhat more favorable prognosis. A total of 70% of the patients were functioning at a good or very good psychosocial level, and had received no further psychiatric diagnoses after an interval of 5 years, while 30% had been treated, mostly in an inpatient setting, for personality disorders, depressive disorders, or (new) adjustment disorders (Bronisch 1991). The time to remission in patients with adjustment disorders was not significantly shorter than that in patients with severe depressive disorders. The cumulative number of patients in remission after 18 months of follow-up was no different in these two patient groups, although there was a significant difference in this respect between the adjustment disorders and the dysthymic disorders (Kovacs et al. 1984).

These studies of disease course indicate that a favorable prognosis, though implicit in the concept of the adjustment disorders, cannot be assumed for all patients. A large subgroup went on to receive the diagnoses of severe depressive disorders, anxiety disorders, post-traumatic stress disorder, somatoform disorders, psychotic disorders, and personality disorders.

7 Treatment

Psychotherapy, in the form of psychotherapeutic counseling or short-term cognitive-behavioral or depth-psychological therapy, is generally regarded as the treatment of choice. The goals of therapy include

relief from acute distress and help and support in the process of restructuring and reassessing the patient's living situation. A further goal is the prevention of further mental disturbances. Treatment may be individual or in a group; it may be offered in the outpatient setting or in an inpatient crisis intervention center. In general, treatment is given according to principles similar to those governing the treatment of post-traumatic stress disorder, albeit in modified form (Horowitz 1997). To date, there has been no systematic, indication-based evaluation of these treatment processes.

Among the pharmacotherapeutic treatment methods available, the short-term administration of benzodiazepines, or of antipsychotic substances in low doses, may be beneficial. Patients with predominantly anxious or depressive manifestations, severe mood fluctuations, or suicidality may benefit from treatment with antidepressants (Verkes et al. 1998). However, psychopharmacological treatment is only rarely recommended for the adjustment disorders. In the multicenter American–Canadian–Mexican study of liaison psychiatric services mentioned above, the administration of an anxiolytic or sedative substance was recommended in 3% of the patients treated, an antipsychotic substance in 1%, and an antidepressant in 7% (Strain et al. 1998). The possible effect of psychoactive medications on the prognosis of the adjustment disorders has not been studied to date.

8 Overview

In summary, much more research is needed in the field of the adjustment disorders. As for the definition of operationalized diagnostic criteria, the literature on the course of these disorders indicates that the temporal criteria, at least, require further empirical testing. Furthermore, longitudinal studies have shown that the long-term prognosis of these disorders is by no means as favorable as originally assumed, particularly in adolescents; rather, the diagnosis of an adjustment disorder is associated with an elevated risk of developing another mental disorder at a future time.

Studies of disease course must also be performed to examine the question of whether the phenomena currently conceptualized under the heading of the adjustment disorders might better be considered minor or subthreshold versions of other psychiatric disorders, and whether the present concept of an event-associated disorder ought to be retained. Finally, there is a lack of controlled studies on the pharmacological and psychotherapeutic treatment of adjustment disorders.

Appendix A. Diagnostic Criteria for Adjustment Disorders According to DSM-IV

A. The development of emotional or behavioral symptoms in response to an identifiable stressor(s) occurring within three months of onset of the stressor(s).
B. These symptoms or behaviors are clinically significant as evidenced by either of the following:
   1. marked distress that is in excess of what would be expected from exposure to the stressor
   2. significant impairment in social or occupational (academic) functioning
C. The stress related disturbance does not meet the criteria for another specific Axis I disorder and is not merely an exacerbation of a preexisting Axis I or Axis II disorder.
D. The symptoms do not represent bereavement.
E. Once the stressor (or its consequences) has terminated, the symptoms do not persist for more than an additional six months.

Adjustment disorder is specified as acute or chronic, where

- acute applies if the disturbance lasts less than six months; or
- chronic applies if the disturbance lasts for six months or longer.

Adjustment disorders are coded according to the subtype corresponding to the predominant symptoms. The specific stressors can be coded on Axis IV:

309.0 with depressed mood
309.24 with anxiety
309.28 with mixed anxiety and depressed mood
309.3 with disturbance of conduct
309.4 with mixed disturbance of emotions and conduct
309.5 not otherwise specified

Appendix B. Diagnostic Criteria for Adjustment Disorders According to ICD-10

A. Onset of symptoms must occur within one month of exposure to an identifiable psychosocial stressor, not of an unusual or catastrophic type.
B. The individual manifests symptoms or behaviour disturbance of the types found in any of the affective disorders (F30–39) (except for delusions and hallucinations), any disorders in F40–F48 (neurotic, stress-related and somatoform disorders) and conduct disorders (F91.–), but the criteria for an individual disorder are not fulfilled. Symptoms may be variable in both form and severity.

The predominant feature of the symptoms may be further specified by the use of a fifth character:

F43.20 Brief depressive reaction

A transient mild depressive state of a duration not exceeding 1 month.

F43.21 Prolonged depressive reaction

A mild depressive state occurring in response to a prolonged exposure to a stressful situation but of a duration not exceeding 2 years.

F43.22 Mixed anxiety and depressive reaction

Both anxiety and depressive symptoms are prominent, but at levels no greater than those specified for mixed anxiety and depressive disorder (F41.2) or other mixed anxiety disorders (F41.3).

F43.23 With predominant disturbance of other emotions

The symptoms are usually of several types of emotion, such as anxiety, depression, worry, tensions, and anger. Symptoms of anxiety and depression may meet the criteria for mixed anxiety and depressive disorder (F41.2) or for other mixed anxiety disorders (F41.3), but they are not so predominant that other more specific depressive or anxiety disorders can be diagnosed. This category should also be used for reactions in children in whom regressive behaviour such as bedwetting or thumb-sucking is also present.

F43.24 With predominant disturbance of conduct

The main disturbance is one involving conduct, e.g. an adolescent grief reaction resulting in aggressive or dissocial behaviour.

F43.25 With mixed disturbance of emotions and conduct

Both emotional symptoms and disturbances of conduct are prominent features.

F43.28 With other specified predominant symptoms

C. Except in prolonged depressive reaction (F43.21), the symptoms do not persist for more than 6 months after the cessation of the stress or its consequences. However, this should not prevent a provisional diagnosis being made if this criterion is not yet fulfilled.

9 References

*Andreasen NC, Hoenk PR (1982) The predictive value of adjustment disorders: a follow-up study. Am J Psychiatry 139: 584–590

Andreasen NC, Wasek P (1980) Adjustment disorders in adolescents and adults. Arch Gen Psychiatry 37: 1166–1170

Bronisch T (1991) Adjustment reactions: a long-term prospective and retrospective follow-up of former patients in a crisis intervention ward. Acta Psychiatr Scand 84: 86–93

Cantwell DP, Baker L (1989) Stability and natural history of DSM-III childhood diagnoses. J Am Acad Child Adolesc Psychiatry 28: 691–700

Epstein RS, Fullerton CS, Ursano RJ (1998) Posttraumatic stress disorder following an air disaster: a prospective study. Am J Psychiatry 155: 934–938

Fabrega H, Mezzich JE, Mezzich AC (1987) Adjustment disorder as a marginal or transitional illness category in DSM-III. Arch Gen Psychiatry 44: 567–572

Fristad MA, Jedel R, Weller RA, Weller EB (1993) Psychosocial functioning in children after the death of a parent. Am J Psychiatry 150: 511–513

Goodyer IM, Kolvin I, Gatzanis S (1987) The impact of recent undesirable life events on psychiatric disorders in childhood and adolescence. Br J Psychiatry 151: 179–184

*Horowitz MJ (1997) Stress response syndromes. PTSD, grief and adjustment disorders. Aronso, Northvale

Jacobson AM, Goldberg IG, Burns BJ, Hoeper EW, Hankin JR, Hewitt K (1980) Diagnosed mental disorder in children and use of health services in four organized health care settings. Am J Psychiatry 137: 559–565

Johnsson FE, Öjehagen A, Träskman-Bendz L (1996) A 5-year follow-up study of suicide attempts. Acta Psychiatr Scand 93: 151–157

*Kendler KS, Neale M, Kessler R, Heath A, Eaves L (1993) A twin study of recent life events and difficulties. Arch Gen Psychiatry 50: 789–796

Kovacs M, Feinberg TL, Crouse-Novak MA, Paulauskas SL, Finkelstein R (1984) Depressive disorders in childhood. I. A longitudinal prospective study of characteristics and recovery. Arch Gen Psychiatry 41: 229–237

*Kovacs M, Gatsonis C, Pollock M, Parrone PL (1994) A controlled prospective study of DSM-III adjustment disorder in childhood. Arch Gen Psychiatry 51: 535–541

La Greca AM, Silverman WK, Wasserstein SB (1998) Children's predisaster functioning as a predictor of posttraumatic stress following hurricane Andrew. J Consult Clin Psychology 66: 883–892

Mann JJ, Oquendo M, Underwood MD, Arango V (1999) The neurobiology of suicide risk: a review for the clinician. J Clin Psychiatry 60[Suppl 2]: 7–11

Schneider K (1980) Klinische Psychopathologie. Stuttgart,Thieme

Strain JJ, Smith GC, Hammer JS et al (1998) Adjustment disorder: a multisite study of its utilization and interventions in the consultation-liaison psychiatry setting. Gen Hosp Psychiatry 20: 139–149

Verkes RJ, van der Mast RC, Hengeveld MW, Tuyl JP, Zwinderman AH, van Kempen GMJ (1998) Reduction by paroxetine of suicidal behavior in patients with repeated suicide attempts but not major depression. Am J Psychiatry 155: 543–547

Wagner D, Heinrichs M, Ehlert U (1998) Prevalence of symptoms of posttraumatic stress disorder in German professional firefighters. Am J Psychiatry 155: 1727–1732

Weiner IB, Del Gaudio AC (1976) Psychopathology in adolescence. An epidemiological study. Arch Gen Psychiatry 33: 187–193

CHAPTER

6

H.P. Kapfhammer

Dissociative Disorders and Conversion Disorders

Translator: E. Taub
This chapter is an abbreviated and revised version of Kapfhammer (1999a,b).

1 Introduction

The terms "dissociation" and "désaggrégation" (disaggregation) were coined by Pierre Janet (1889). Janet viewed the mental life of the individual as an aggregate of mental elements, which he designated as "psychological automatisms." Each automatism, according to his view, is a complex behavioral tendency, comprising both a performance and an emotion, which is released by a specific provoking situation. Under normal circumstances, psychological automatisms are unified in phenomenal consciousness and are accessible to voluntary control. Under traumatic, stressful conditions, however, single automatisms may become split off – dissociated – from the rest of consciousness and exert their effects in an autonomous fashion. Genetically transmitted temperamental traits as well as earlier experiences and the individual's present physical condition together define the capacity of each individual for the mental integration of new information in general, and of traumatic experiences in particular.

Janet's model of dissociation, though oriented toward trauma psychology, is nonetheless essentially based on the idea of a constitutional predisposition to dissociation ("dégénérescence"). The individual's premorbid vulnerability to dissociative disorders thus plays a decisive role. A dissociation (which is to be understood as passive) by no means requires an intense external trauma for it to be released. Rather, a reactivity to exaggerated emotions, inherent in the individual's personality, often exerts a traumatogenic effect and leads to the generation of a psychopathological disorder.

The original position of Freud and Breuer (1895) was still very close to that of Janet. They, too, viewed a traumatic exposure during a sensitive developmental phase and a predisposition to dissociative psychopathology as decisive. Like Janet, they also recognized the importance of autosuggestion, used as a defense against traumatic experiences, in the pathogenesis of dissociative syndromes, as well as the special role of hypnotic techniques in the treatment of this unique class of disorders.

As psychoanalytic theory came to be further developed, however, a major shift of emphasis took place in the conception of the dissociative disorders. The dominant idea at first was that of external events leading to traumatic excitation beyond the individual's subjective ability to cope and thus to a feeling of psychophysical helplessness. Later, Freud increasingly identified a strong influence of unconscious fantasies in the attribution of meaning to "traumatic situations." What was originally thought of as an external, traumatic situation came to be viewed as an intrapsychic situation of danger, for which the ego can prepare itself in advance with controlled amounts of signal anxiety and to which it can react with specific defense mechanisms, such as repression (Freud 1926). While developing these ideas, Freud in no way denied the existence and clinical relevance of real trauma. Yet his concept of repression and of the defense neuroses became the leading paradigm of dynamic psychiatry in the years that followed and consigned Janet's concept of dissociation to obscurity for many decades thereafter (Nemiah 1998).

Hilgard (1986), making use of the newly developed concepts of cognitive psychology, proposed a neodissociative theory that had many points of contact with the psychology of Janet. In his model of the "divided consciousness," the psyche is viewed as an organized system of mental structures controlling perceptual, cognitive, and behavioral processes in different areas. These mental subsystems bear a certain resemblance to Janet's psychological automatisms, but may also be thought of as comparable to the modules and cognitive units postulated in cognitive theories of parallel information processing. Each of these subsystems can, in principle, regulate its own input and output independently of the others, though the subsystems do communicate with one another under normal circumstances. The central element of the overall system is a structure exercising the executive functions of monitoring and control. This structure also provides the mental basis for the experience of phenomenal consciousness and for voluntary control.

According to Hilgard, this central executive system may become dysfunctional and cease to integrate and organize the individual control structures, resulting in a state of divided consciousness. Hilgard's conception is based primarily on the findings of psychological experiments involving hypnosis in mentally healthy subjects, but it is also relevant to the modern understanding of dissociative processes in psychopathological states. While Hilgard was developing these theoretical ideas, central elements of Janet's trauma psychology experienced a revival in the clinical field as well (van der Hart and Horst 1989).

2 The Concept of Dissociation

The concept of dissociation is multilayered. On one level, broadly understood, it includes a number of normal psychological phenomena, such as automatic modes of behavior, subconscious perception, implicit versus explicit memory, meditation, focused concen-

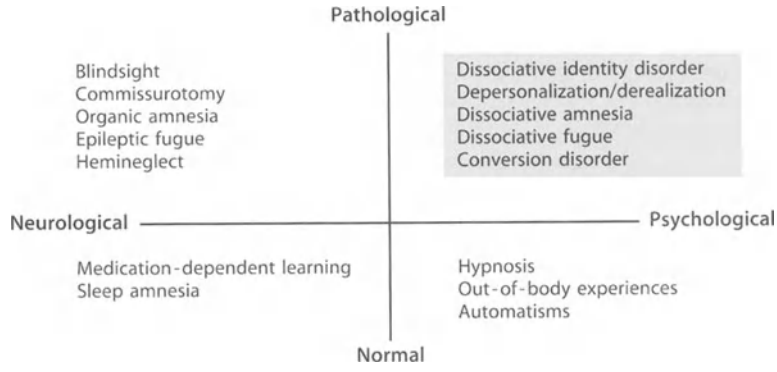


Fig. 1. Dissociative phenomena. (After Cardeña 1994)

tration, and hypnosis. On another level, it is restricted to pathological phenomena, such as disturbances of memory, identity, consciousness, and perception of the self and of the environment. It may also include certain pseudo-neurological and special neurological syndromes (Fig. 1). The term is used in both a descriptive and an explanatory sense.

Cardeña (1994) emphasized the relevant aspects of dissociation for clinical psychopathology. He opposed the extension of the term to cover stimulus-processing phenomena that do not reach consciousness because of neurophysiological or structural limitations or because of the finite capacity of attention – phenomena that can only be made perceptible to the individual, if at all, by special techniques such as biofeedback. Rather, dissociation, in his view, reflects a failure of the normally expected integration of central personality functions, leading to typical disturbances of memory, consciousness, and personal identity. Use of the term in this sense implies the inclusion of pseudoneurological conversion syndromes as well. Furthermore, dissociation is related to certain alterations of consciousness in which the central feature of phenomenal perception is an alienation from oneself or from the environment.

In the foregoing discussion, "dissociation" has been used as a descriptive term for certain clinically relevant phenomena. The term may also be used in the etiological sense to describe the defense mechanisms that are presumed to underlie these psychopathological states. Yet a conceptual differentiation between dissociative and other defense mechanisms, such as repression, may prove difficult. They may be distinguished in a psychodynamic sense as follows: Repression involves the disguise and concealment of information and, consequently, the need to transform the implicit meaning of dreams by reinterpretation, for example; it usually also involves a lack of temporal organization of repressed mental content, and it is decisively provoked by intrapsychic conflicts. In dissociation, on the other hand, information is not transformed. The information in question can often be brought to consciousness through direct psychotherapeutic techniques, such as hypnosis. Dissociation usually occurs in relation to a circumscribed period of time in the individual's biography and represents a basic mechanism of withdrawal from a mentally intolerable reality. Dissociation as a defense mechanism is thus explicitly related to an external trauma (Singer 1990).

3 Establishment of Terminology

In the modern psychiatric classification systems (DSM-IV and ICD-10), the hallmark of the diagnostic category of the dissociative disorders is a partial or total loss of the integrative functions of consciousness, memory, personal identity, and perception of the self and of the environment. These disorders were referred to in earlier diagnostic manuals as "hysterical neurosis of dissociative type." ICD-10 subsumes the complementary type of hysterical neurosis, i.e. the conversion disorders, under the same category, classifying them as dissociative motor disorders, anesthesia and sensory loss, or convulsions. In DSM-IV, in contrast, the conversion disorders appear under the separate diagnostic category of somatoform disorders. The difference between the two classification systems is based on a difference in underlying assumptions. In ICD-10, dissociation, i.e. a failure of integration of mental functioning, is viewed broadly as a classifying principle. In DSM-IV, the presence of physical symptoms with no organic medical cause in a certain group of mental disorders is taken as the defining feature of a separate diagnostic category.

It is conceivable that future editions of DSM will bring the conversion disorders and the dissociative disorders together under a single category. The two classes of disorders share a number of important features. Both are characterized by pseudo-neurological manifestations, with dysfunction in the area of conscious experience. Particular present and past

experiences become temporarily inaccessible to conscious attention, but nonetheless continue to influence the experiences, thoughts, and actions of the individual in the form of implicit perceptions and recollections. Thus certain consciously planned, goal-oriented acts cannot be performed, while other unconsciously planned acts are experienced as involuntary (Kihlstrom 1994).

ICD-10, unlike DSM-IV, has removed the depersonalization and derealization disorders from the category of the dissociative disorders and assigned them to that of "other neurotic disorders," even though they may be counted among the core syndromes in the spectrum of dissociative psychopathology.

The discussion in this chapter is organized essentially along the lines of DSM-IV. We will treat the spectrum of the dissociative disorders as composed of five clusters of manifestations that may appear in varying degrees of severity and in varying combinations (Steinberg 1994a):

- *Amnesia*: the forgetting of personal data, more severe than could be attributed to ordinary forgetfulness
- *Depersonalization*: an altered or distorted experience and perception of oneself and one's own body, with feelings of alienation and emotional removal from oneself
- *Derealization*: a feeling of alienation and emotional removal from the environment, with a possible connotation of unreality
- *Identity confusion*: the subjective feeling of uncertainty, confusion, or conflict regarding personal identity
- *Identity change*: a change of identity, possibly with the use of different names and biographies and the demonstration of otherwise unavailable knowledge and personal capabilities

This chapter also includes a discussion of conversion disorders, which involve pseudo-neurological abnormalities of voluntary motor activity, sensation, or the regulation of consciousness.

4 Dissociative Disorders

4.1 Definition and Clinical Features

4.1.1 *Dissociative Amnesia*

Dissociative or psychogenic amnesia is the forgetting of important personal data that cannot be attributed to "ordinary" forgetfulness, a neurological or general medical illness, or the effects of psychotropic substances. The amnesia is typically retrograde, i.e. it is related to a particular episode or to a stretch of time before a defined event that is usually of an unpleasant, stress-inducing, or traumatic nature. It is usually reversible. Janet (1894) provided a classification of dissociative amnesia into several distinct types, and this system remains valid in the modern clinical view:

- *Circumscribed amnesia*: the inability to remember events that took place within a defined period of time.
- *Selective amnesia*: the ability to remember some, but not all relevant events in a defined period of time.
- *Generalized amnesia*: failure to remember anything about one's own life, including one's own name.
- *Continuous amnesia*: failure to remember successively occurring events.
- *Systematic amnesia*: amnesia for highly specific categories of memory, such as recollections of one's own family or of a specific person.

In the subjective experience of the sufferer, dissociative amnesia manifests itself as a vague loss of time over the course of the preceding days or weeks. Patients find that other people behave in ways that are not familiar from the past and thus appear strange. Unknown objects suddenly appear among their possessions. Confusing changes occur in their interpersonal relationships for which they have no explanation. They may demonstrate unusual fluctuations in their capabilities, habits, preferences, or knowledge. They sometimes experience brief, trance-like amnestic episodes (microamnesia). Not uncommonly, a period of confusion occurs before the patients cognitively realize their loss of memory and then try to reorganize their life situations around this concept (Loewenstein 1996).

Dissociative amnesia has a wide-ranging differential diagnosis. It may be a manifestation of another dissociative disorder, such as dissociative fugue or dissociative identity disorder; it may appear in the context of an acute or post-traumatic stress disorder (PTSD); it may be associated with a somatoform disorder (which may be a conversion disorder); or it may be present in combination with a borderline personality disorder (Davidson and Foa 1993; Spitzer et al. 1994). In terms of clinical psychopathology, it may prove to be a simulation or an artificial disorder (Kapfhammer et al. 1998a,b; Schacter 1986). Amnestic episodes may occur early in the course of a dementing illness or during an episode of delirium, as a component of epilepsy or migraine, or in the context of transient global amnesia. Aside from these neurological disorders, other causes include toxic influences such as massive alcohol consumption or the use of

barbiturates, benzodiazepines, phencyclidine, LSD, and steroids, as well as numerous metabolic disorders, including uremia, hypoglycemia, and porphyria (Akhtar and Brenner 1979; Cummings 1985). A difficult problem in differential diagnosis may arise in the aftermath of mild head trauma in cases of post-contusional amnesia (Kopelman 1987; Lishman 1998).

4.1.2 Dissociative Fugue

A dissociative fugue (from the Latin *fuga,* flight or escape) is characterized by a sudden departure from home or the workplace. The individual does not know where he or she is going or why, and, when discovered in a strange or far-off place, can give no information about his or her own history. The dissociative amnestic component of a fugue state is broader than an isolated dissociative amnesia. There is confusion regarding personal identity, and, occasionally, a new identity is partially or completely assumed. The diagnosis of a dissociative fugue cannot be assigned when similar phenomena appear in the framework of a more comprehensive dissociative disorder, such as multiple personality disorder or dissociative identity disorder, or when they can be explained as the consequence of physical illness or other medical conditions. The behavior of patients in fugue states seems thoroughly well-organized and goal-oriented. The loss of memory and the uncertainty or disturbance of the consciousness of identity contrast markedly with the otherwise intact higher cortical functions. A new identity is assumed in only a minority of cases.

Fisher (1945) classified dissociative fugues into three clinical types:

1. Amnesia as to personal biography, change of identity, and settlement in a strange living environment (classical type)
2. Only amnesia as to personal identity, without change of identity status
3. Return to an earlier period of one's own life, with amnesia for the intervening time up to the present, and without change of identity status

When a new identity is assumed, its basic personality traits are often more expansive and less inhibited than those of the previous identity. After the change of identity status, the individual generally becomes fully integrated in and adapted to the new surroundings, in accordance with sociocultural expectations (Hilgard 1986). Considerable emotional distress accompanies the gradual return of such an individual to his or her original identity and the coming to terms with the amnestic interval (Kihlstrom 1990).

The most important item in the differential diagnosis of dissociative fugue is a post-ictal state of aimless wandering, accompanied by retrograde amnesia and disorientation, in patients with complex partial epilepsy. Further diagnoses to be excluded include poriomania, other non-epileptic organic brain disorders such as migraine and brain tumors, schizophrenia, and alcohol- and drug-induced fugue states (Akhtar and Brenner 1979; Good 1993).

4.1.3 Depersonalization and Derealization

Depersonalization is a feeling of alienation from one's own body and from oneself. Derealization is an analogous feeling with respect to the environment. In both cases, reality judgment remains intact. The subjectively experienced perceptual changes are unpleasant and often difficult to express verbally; descriptions are often of the nature of an "as if." Depersonalization and derealization constitute a continuum and may occur together or separately in the individual's experience. In the literature, these two states are often subsumed together under the term "depersonalization." Depersonalization and derealization may occur as manifestations of a large number of psychiatric and neurological disorders, but they may also constitute independent clinical syndromes. Depersonalization and derealization frequently occur as accompanying manifestations of other dissociative disorders.

In depersonalization, the individual feels entirely changed in comparison to his or her previous state. This change is felt both in the self and in the outside world. The individual perceives his or her own actions and behavior as automatic and observes them from the vantage point of an external observer. The outside world seems strange and new, and not as real as before (Schilder 1935). This altered perception of the self and the environment bears an intensively unpleasant and, at times, tormenting affect quality. It is often associated with a feeling of reduced self-esteem, humiliation, and social isolation and with the fear of losing bodily control or mental health, particularly after a traumatic experience. Diffuse symptoms of somatization, a distressing dazedness, an elementary sensation of vertigo, and a loss of clear perception of time may accompany depersonalization. Autoscopic phenomena, metamorphopsia, and disorders of the body scheme may also appear. Contact with the environment seems to be lost, without the usual affective connotation, but rather with a lack of emotional participation, a dreamlike trance state, a flatness of affect, etc. (Sims 1995).

In the differential diagnosis of depersonalization, it must be remembered that depersonalization is the third most common symptom of an anxiety disorder or of a depressive disorder (Cattell and Cattell 1974). It

appears to a particularly marked degree in cases of the phobic anxiety-depersonalization syndrome (Roth 1959) and of nonparticipatory depression (*teilnahmsarme Depression*) (Leonhard 1995).

4.1.4 Dissociative Identity Disorder

A dissociative identity disorder or multiple personality disorder is a complex, chronic, dissociative psychopathological state characterized by disturbances of memory and identity. It manifests itself in the long-term coexistence of relatively consistent, subjectively separate identities that appear in alternation, and in recurring episodes of distorted memory or overt amnesia.

Patients suffering from a dissociative identity disorder typically do not seek psychiatric or psychotherapeutic help for their primary symptoms. As a rule, they receive wide-ranging diagnoses over the course of several years of illness, usually with reference to secondary, complicating psychiatric disturbances. Psychiatric comorbidity at the time of diagnosis does, in fact, occur very prominently and may include extremely varied psychopathological syndromes.

Kluft (1996) lists the comorbid disturbances accompanying dissociative identity disorders on the basis of the available empirical studies:

- Anxious manifestations (psychophysiological, 100%; phobic, 60%; panic attacks, 55%; obsessive-compulsive manifestations, 35%)
- Affective manifestations (depressive, 90%; hypomanic, 15%–73%)
- Associated dissociative manifestations (amnesia, 57%–100%; fugue, 48%–60%; depersonalization, 38%)
- Somatoform manifestations (general, 90%; conversion, 60%; sexual dysfunction, 60%–84%)
- Suicide attempts (60%–68%)
- Self-injury (34%)
- Substance abuse (40%–45%)
- Eating disorders (16%–40%)
- Sleep disorders (65%)
- Manifestations of schizophrenia (depending on particular manifestations: 35%–73%)
- Manifestations of PTSD (70%–85%)
- Characteristics of borderline personality disorder (70%)

A great many patients display manifestations of dissociative identity disorder only temporarily, often after a prolonged course of treatment. The features of the disorder usually appear in a very mild form (Franklin 1990). The gamut of alternately appearing dissociative and post-traumatic manifestations includes amnesia, depersonalization and derealization, experiences of passive influence, sudden changes in behavior, intrusive visual images (flashbacks), and a hearing of voices in the head. This fluctuating psychopathology may prepare the way for the appearance of the disorder proper. Bizarre temporal gaps, fugue episodes, contradictory behaviors without any subjective explanation, and a confusing forgetfulness have all been reported.

The emergence of pathognomonic multiple personalities – "alter egos" – in the therapeutic conversation may take place spontaneously or only after therapeutic focusing on the amnestic symptoms, in which the patient is prompted to explore their context more closely. The use of hypnotic techniques and pharmacological interventions (e.g. amobarbital), a prolonged interview lasting 2–3 h, or the meticulous keeping of a daily diary may promote the emergence of these alter egos. The coexisting, subjectively separate identities in a single person may be conceptualized as highly distinct states of consciousness organized around a special affect or self-perception (including body image), each with its own limited repertory of behavioral modes and its own set of state-dependent memories (Putnam 1989).

Braun (1993) summarized the phenomenology of a typical dissociative identity disorder, which is as fascinating as it is confusing, as follows:

- The patient possesses two or more personalities at a given point in time (occasionally one personality with executive control and another personality in the position of an observer or in the role of an advisor).
- Each of the personalities has a complete, or nearly complete, set of personality traits; the traits of the different personalities are different and often opposite to one another.
- Transitions between the personalities may occur suddenly or gradually. Some patients may undergo this transition in the presence of the therapist.
- Amnestic barriers between the personalities may exist. One personality may know of the existence of the other(s) (but not vice versa), and two or more personalities may sometimes share a limited amount of information.
- Each personality typically manifests "temporal gaps" in the periods in which other personalities exercise executive control. The sum of the recollections of the various alter egos generally spans a longer interval than the actual elapsed time.

The differential diagnosis of dissociative identity disorder must include a consideration of its extensive psychiatric comorbidity, in particular the possible coexistence of a borderline personality disorder. Considerable difficulties in differential diagnosis may arise when patients claim to be under the influence of

multiple beings coexisting within themselves or when they report other first-rank symptoms, such as hearing commenting voices, generally localized inside the head. Even if other, determinative diagnostic criteria for schizophrenic psychosis are absent, this diagnosis is often considered probable for want of a better alternative explanation; yet modes of psychopharmacological therapy that are otherwise effective for schizophrenia, such as the administration of neuroleptics, are generally ineffective in such situations. Furthermore, an artificial disorder or simulation must also be ruled out.

4.2 Epidemiology

A number of older studies yielded unreliable estimates of the frequency of dissociative disorders, as they were exclusively symptom-oriented and could not make use of the currently accepted diagnostic criteria of DSM-III/R, DSM-IV, and ICD-10. To date, there have been only a few studies using standardized screening (e.g. the Dissociative Experiences Scale, DES; Bernstein and Putnam 1986) followed by diagnostic validation with a structured clinical interview (e.g. the Dissociative Disorders Interview Schedule, DDIS; Steinberg 1994b). The study of Ross (1991) and Ross et al. (1990) is a notable exception: in this study, the DES was administered to 1055 randomly stratified, selected inhabitants of Winnipeg, Manitoba (Canada), of whom 502 were further tested with the DDIS (Table 1).

A total of 6% of the subjects interviewed met the diagnostic criteria for dissociative amnesia. The frequency of psychogenic amnesia is probably even higher in the context of massive trauma, e.g. among survivors of the Nazi Holocaust (Kuch and Cox 1992) or war refugees (Carlson and Rosser-Hogan 1991). Among 53 Swedish survivors of the Estonia ferry disaster, 29% displayed manifestations of dissociative amnesia (Eriksson and Landin 1996). It is hardly surprising that the prevalence figures are much higher among patients presenting to special outpatient clinics for the treatment of dissociative disorders (Coons and Milstein 1992). On the other hand, considerable skepticism has been expressed regarding a linear relationship between dissociative amnesia and trauma when there has been no detailed exploration of the remembered portions of the underlying traumatic events (Pope et al. 1998).

Table 1. Prevalence of dissociative disorders in the general population ($n = 502$) according to DSM-III-R criteria (after Ross 1991)

Diagnosis	Prevalence (%)
Psychogenic amnesia	6.0
Psychogenic fugue	0.2
Depersonalization disorder	2.8
Multiple personality disorder	3.0
Dissociative disorder (other)	0.2
Total	12.2

There is a consensus that dissociative fugue is much rarer than dissociative amnesia or dissociative identity disorder (prevalence, 0.2%; Ross 1991), but it appears to be more common in wartime and in the aftermath of natural disasters or other violent events.

Estimates of the prevalence of depersonalization disorder making use of modern diagnostic standards are on the order of 3% (Ross 1991). Approximately 30%–40% of individuals exposed to life-threatening danger report having experienced a transient depersonalization syndrome (Noyes and Kletti 1977). Depersonalization in adolescents can be characterized as an independent syndrome reflecting the problems of maturation and development that are typical of this period of life (Meyer 1961).

Cases of multiple personality disorder that were published in earlier years were often viewed as exotic rarities or even literary fictions, but the prevailing attitude toward this entity has changed over the last two decades, as is reflected in its inclusion in the official psychiatric diagnostic manuals as an independent diagnostic category. It is nonetheless true that immediate sociocultural influences are reflected in the "production" of the phenomenon of dissociative identity disorder (see below).

According to American epidemiologic studies, 1%–3% of the general population suffers from a dissociative identity disorder (Ross 1991). This prevalence is comparable to that of other major psychiatric disorders. However, even though a number of American authors have reported on clinical samples of several hundred patients bearing this diagnosis, the great majority of psychiatrists and psychotherapists working in inpatient psychiatric services or outpatient practice see such patients extremely rarely, or never, in the course of their professional lives. A similar situation prevails in Switzerland and in the Netherlands, as shown in the studies by Modestin (1992) and Boon and Draijer (1993), respectively.

A possible reason for the glaring numerical discrepancy between the empirical studies and the reported experiences of a large number of psychiatrists and psychotherapists is that, despite the appearance of isolated, spectacular reports in the news media, patients with dissociative identity disorders usually display much less marked manifestations in everyday clinical practice and very rarely talk about their basic,

underlying symptoms without prompting. Instead, these patients present with a multitude of other, secondary psychiatric disturbances. In typical cases, an average of 7 years elapses before the diagnosis of a dissociative identity disorder can be made by an experienced clinician (Loewenstein and Ross 1992). The available studies of hospitalized patients reveal a strong female predominance, with a ratio of approximately 9:1 (Kluft 1996).

4.3 Etiology and Pathogenesis

4.3.1 General Conceptual Approaches

There are, in essence, two conceptual approaches to an understanding of the causation of dissociative conditions:

- A complex mode of reaction to an external trauma
- Certain primary personality attributes that can predispose an individual to the occurrence of dissociation

The first approach places dissociation and the dissociative disorders in a close pathogenetic relationship with acute and PTSD, while the second approach associates them with dissociative tendencies in the normal personality and with the distribution of such tendencies in the general population and relates dissociation to other personality constructs, including hypnotizability, mental absorption, and tendency to fantasize, among others. Both approaches thus tie factors of normal psychology and trauma psychology together in a predisposition–stress model.

Dissociation as a Reaction to Trauma

Very many patients with dissociative disorders report having experienced severe psychological trauma, mostly during the years of their early development. The most serious traumas described include sexual and physical abuse, emotional neglect, and deprivation (Spiegel and Cardeña 1991). It must be kept in mind, however, that most of the data published on this topic were gathered retrospectively and may thus give a distorted picture of reality. Researchers and therapists are currently divided into two camps with respect to their attitudes toward the dilemma of "historical" versus "narrative truth" in spontaneous reports of early childhood trauma as a major cause of dissociative disorders. Those who favor the explanation of a true post-traumatic reaction (Braun 1993) are opposed by others who prefer that of a genetically predisposed personality (Paris 1998).

A problem in many studies is the frequent lack of distinction between different types of trauma, e.g. unexpected, single-event trauma versus continued or repeated exposure to trauma (Terr 1991), or the presence or absence of protective compensation in a traumatized child (Putnam 1995). A critical stance is further indicated when the rates of traumatization derived from epidemiologic studies in the general population are considered as absolute facts even though they have not been validated by detailed clinical interview, and also, conversely, when general conclusions are extrapolated from data gathered in clinically narrowly defined and generally quite small patient cohorts. The so-called false memory syndrome, which is receiving an increasing amount of attention both from the general public and from the courts, is a telling illustration of this difficulty. The relationship between trauma and dissociation does seem to be a close one, but it should not be thought of as linear and monocausal. The interrelatedness of trauma, memory, and dissociation must always be analyzed on multiple levels (Bremner et al. 1998).

Trauma and the Neurobiology of Dissociation

Dissociative processes cannot be regarded as exclusively "mental" events without any connection to cerebral information processing and the associated neurochemical changes. In the context of adjustment to a traumatic event, there are two basic modes of reaction to be distinguished. Particularly in children, these two modes can be connected to specific aspects of the individual's previous development.

An individual typically reacts to an external threat with an alarm reaction consisting of an elevation of sympathetic tone in preparation for a basic fight-or-flight response pattern. This *hyperarousal continuum* is brought about mainly by the centrally driven peripheral secretion of epinephrine and norepinephrine, by the secretion of adrenal corticotrophic hormone (ACTH) and cortisol by the hypothalamic–pituitary–adrenal (HPA) axis, and by activation of the immune system. The decisive mediating role is played by the locus ceruleus; the ventral tegmental nucleus is also an important regulator of the activity of the pontine and medullary sympathetic nuclei. Prolonged or repeated threats may lead to a sensitization of these two control centers, possibly resulting in catecholamine over-reactivity, even to less threatening stimuli.

On the other hand, a *dissociative continuum* is associated with a basic response pattern of surrender. It seems to be activated preferentially when an organized fight-or-flight response appears unlikely to succeed, or when this type of response is not yet fully developed, as in children. Just as in the hyperarousal continuum, there is initially a stress response with secretion of catecholamines and corticoids. However, in dissociation, and not in hyperarousal, there is also a

strong vagal activation. Moreover, the mesolimbic and mesocortical dopaminergic systems play an important role and, by way of the central reward systems, exert a primary influence on affect modulation. Collateral connections to the endogenous opioid system lead to a change in the perception of noxious stimuli and also to a distortion of the sense of time, place, and reality. In fact, most opiate agonists can induce dissociative reactions. Dissociation, like hyperarousal, may become sensitized, i.e. there may be a use-dependent cerebral organization of dissociation as a primary adaptive response pattern.

Both response patterns can be combined, to varying relative extents, in basic adaptive styles (Perry et al. 1995). A neurobiological interrelationship of dissociative disorder and PTSD is suggested by the fact that peritraumatic dissociative symptoms have a high predictive value for the later development of PTSD (Marmar et al. 1998).

Findings from the field of experimental neuropharmacology yield interesting insights into the neurobiology of dissociative processes. Certain substances, such as yohimbine (noradrenergic) or metachlorophenylpiperazine (m-CPP; serotonergic), indirectly induce a dissociative state in vulnerable individuals after first stimulating intense anxiety or traumatic memories. Glutamatergic neurons clearly play a decisive role in the mediation of dissociative states by the serotonergically and noradrenergically activated systems of the amygdala, hippocampus, thalamus, and cortex. It is thus not surprising that ketamine, for example, which influences glutamatergic neurotransmission, can directly induce an experience of dissociation even in normal subjects, i.e. without the need for a preceding intense anxiety reaction or traumatic memory. A possible approach to pharmacotherapy involves the use of anti-dissociative *N*-methyl-D-aspartate (NMDA) agonists (Krystal et al. 1998).

Dissociation and Dissociative Tendencies in the "Normal Personality": Distribution in the General Population

The assumption that there is a continuum extending from mild forms of dissociation in everyday life all the way to severe psychopathological dissociative states, such as a multiple personality disorder or dissociative identity disorder in extreme cases, is derived from epidemiological findings on the dissociative disorders (Kihlstrom et al. 1994). Studies have shown that normal subjects and patients with dissociative disorders differ less in terms of distinct symptom clusters than in terms of the quantitative extent, frequency, and intensity of dissociative symptoms.

The self-rating scale for dissociative experiences developed by Bernstein and Putnam (1986) (DES) is based on this conceptual approach. It assesses the frequency of dissociative phenomena in the areas of memory, cognition, consciousness, and identity. It has been shown that a high score on the DES is a highly sensitive and specific indicator of the presence of diagnostically validated dissociative disorders. The average DES score declines significantly with advancing age (Ross et al. 1991).

The adaptive function of "normal dissociation" seems to lie in the automatization of certain modes of behavior, which allows a division of attention for the performance of complex tasks. The dissociability of an individual may be defined as a function of the frequency and duration of spontaneous occurrences of a dissociated state of consciousness under natural conditions. From this perspective, a "pathological dissociation" is present when the quantitative extent of dissociability negatively interferes with necessary psychosocial adaptive behavior. On the other hand, dissociability, as measured by DES scores, also seems to be strongly influenced by psychological stressors (Putnam 1995).

Dissociation and Hypnotizability

Janet (1889) viewed hysteria and hypnosis as closely related states. He thought that both reflected the same process of dissociation. He was also convinced that only "hysterical patients" were hypnotizable. In the modern conception, hypnosis is described as a controlled and structured dissociative process characterized by total absorption, compartmentalization of experience, and elevated suggestibility (Spiegel and Cardeña 1991). In fact, the ability to be hypnotized is much more common than Janet assumed and represents a nonpathological personality trait that can be measured reliably by means of standardized psychometric instruments (e.g. the Stanford Hypnotic Susceptibility Scale, SHSS).

This trait is at its most pronounced in the early school years and normally becomes less so with increasing age and over the course of cognitive-affective maturation (Morgan and Hilgard 1973). Nonetheless, it may persist relatively unchanged into adulthood as a consequence of constitutional factors or experiences during early development. Traumatic life circumstances exert a negative influence in this respect, as they can easily overstrain the usual coping abilities and defense mechanisms of childhood. Autosuggestion and hypnosis may become one of the ways in which traumatic influences can be channeled into dissociative states (Putnam 1995).

About 10% of adults in the general population have high scores on scales of hypnotizability (Hilgard 1986). Even though individuals suffering from pathological dissociation are usually also highly hypnotizable (Frischholz et al. 1992), the converse is not true – a high degree of hypnotizability does not necessarily imply the presence of pathological dissociation.

Thus the epidemiological correlation between the DES, for example, and hypnotizability is only modest (Kihlstrom et al. 1994). Hypnotizability, as a "trait" variable, is also highly correlated with other personality traits, such as a tendency toward losing oneself and other self-altering experiences ("absorption"; Roche and McConkey 1990) or a particular inclination toward fantasy and imagination ("fantasy proneness"; Lynn and Rhue 1988; Rhue and Lynn 1989). Like hypnotizability, these personality traits do not indicate that a psychopathological condition is present, but they may be part of a complex defense or coping mechanism. The autosuggestion or autohypnosis model for dissociation should thus be understood provisionally as no more than a metaphor or clinical heuristic in need of further empirical testing (Putnam and Carlson 1998).

4.3.2 Models of Causation for Various Dissociative Symptom Clusters

The controversy concerning the connection between the dissociative disorders in general and exposure to trauma also dominates the discussion of the etiology and pathogenesis of dissociative amnesia. Practically all systematic studies and reviews reveal a high, or very high, association between this entity and multiple traumatic experiences. Coons and Milstein (1992), for example, found multiple causative factors in a series of 25 patients with severe intrapsychic stress and conflict: child abuse (60%), marital discord (24%), deviant sexual behavior (16%), suicide attempt (16%), criminal behavior (12%), death of a near relative (4%), running away (4%), or unknown (16%). A total of 72% of the patients studied had had major traumatic experiences in their own early development: sexual abuse (52%), physical abuse (40%), emotional neglect (16%), or abandonment (12%). The influence of traumatic experiences on many different aspects of memory thus rightly occupies a central position in the discussion of causal factors of dissociative amnesia.

It is important to distinguish episodic-autobiographical from semantic-procedural memory. The first system, which consists of spatiotemporally organized autobiographical experiences with a special connection to the self as participant or observer, appears to be preferentially affected in cases of dissociative amnesia (Kihlstrom 1990). This episodic memory may manifest itself in either an explicit or an implicit form. While a person with dissociative amnesia is unable to recollect a specific event intentionally in verbal form, his or her behavior may implicitly reveal an unambiguous connection to it. The same individual has no deficit of semantic-procedural memory (Schacter and Kihlstrom 1989).

It is significant that traumatically induced memory disturbances in the above sense are not uncommonly accompanied by recurring intrusive thoughts, images, dreams, and emotions that are clearly related to the original traumatic episode (Horowitz 1976). Patients with dissociative amnesia may also react with intense affect to seemingly neutral stimuli without consciously understanding their emotional meaning or being able to place them explicitly into the context of the prior trauma (Loewenstein 1996).

Clinical studies of traumatized war veterans, war prisoners, and adults who suffered massive abuse in their early development reveal that exposure to trauma and dissociative amnesia may also be associated with neurobiologically demonstrable changes in hippocampal structures, possibly as a result of excessive glucocorticoid activity. Hippocampal volume loss documented by magnetic resonance imaging (MRI) is significantly correlated with dissociation scores and with the symptom cluster of "numbing," as well as with avoidance in PTSD (Bremner et al. 1995).

The same causal factors just discussed for dissociative amnesia are of basic importance in dissociative fugue. However, the origin of dissociative fugue is thought not to lie exclusively or primarily in prior trauma; a number of other important psychodynamic factors, particularly related to premorbid psychopathology and unfavorable family background, are also thought to play a major role. Some patients show considerable degrees of separation anxiety, depressive mood disturbances, suicidal and outwardly aggressive impulses, and primitive denial tendencies. Not uncommonly, the dissociative fugue has the nature of wish fulfillment, e.g. when the patient escapes from an unbearable marital situation. Often, particularly when a crime has been committed, the suspicion of simulation may arise, i.e. a more or less conscious attempt to lessen the patient's own guilt and accountability for his or her own actions. Fatigue, sleep deprivation, and heavy alcohol consumption may be predisposing factors. A fugue may also follow mild head injury (Berrington et al. 1956; Loewenstein 1996; Riether and Stoudemire 1988).

Numerous models have been developed to explain dissociative identity disorder. Each emphasizes a different aspect of the multifactorial causation of this complex pathophysiological phenomenon. According to one pragmatic clinical model, the etiology and pathogenesis of dissociative identity disorder are a developmental sequence involving several steps (Braun 1993; Kluft 1996; Mollon 1998; Putnam 1995):

- A *capacity for dissociation* is considered essential. This is not yet a psychopathological phenomenon in itself, but rather a psychobiological variable relating to general hypnotizability that is found to a limited

extent among the general population, is at its most pronounced in the early school years, and normally becomes less so with increasing age (see above).

- Constitutional factors, in combination with *early traumatic experiences,* may result in the persistence of this capacity into adulthood. A capacity for dissociation reinforced by autosuggestion and autohypnosis is a basic defense and coping mechanism for the channeling of overwhelming traumatic experiences. It is of etiological significance in this context that a reliable history of being severely abused in childhood can be obtained in an extremely high percentage of patients with multiple personality disorder. In addition to severe sexual and, in particular, incestuous traumatization, physical abuse, and continuous injury to the narcissistic equilibrium, it seems that the experience of loss of significant others in childhood, the direct observation of the death, suicide or other self-destructive acts of near relatives, and severe physical illness involving intense and prolonged pain can also play a pathogenetic role in this disorder.
- A further factor promoting the use of the dissociative mode for the processing of traumatic experiences is a *lack of compensatory social relationships* of a consoling or supportive nature after an exposure to trauma, with the result that the trauma is encoded not as a discrete, finite event, but as an ever-present personal reality.
- *Repeated, overwhelming traumatization* that the child is unable to integrate leads, by way of dissociation, to a compartmentalization of biographical memory. Many different factors, including life experiences of diverse types, individuals of major importance in the child's world (via introjection, internalization, and identification), imaginary playmates, or external influences (encouragement of role-playing, contradictory educational demands or reward systems, identification with a dissociative parent) prepare the intrapsychic nucleus for the generation and imaginative outfitting of alter egos. The final development of the adult psychopathology of multiple personality disorder is determined not only by interpersonal therapeutic factors, but also by sociocultural ones. Because of the repeated transformation and reprocessing of the original early childhood traumatic experiences, an adult suffering from multiple personality disorder is often unable to recollect exact biographical details without contamination by confabulations and misinterpretations serving a defensive function, so that a historical objectification of the events described may be impossible.

This clinical model is thus essentially oriented toward both trauma psychology and developmental psychology. It emphasizes the relatedness of dissociative identity disorder to PTSD but also requires consideration of the special developmental psychological, i.e. cognitive and affective condition, of the multiply traumatized child. This model implies that adults without developmental disturbances who have been exposed to severe trauma generally do not develop psychopathology analogous to multiple personality disorder, but rather develop the clinical syndrome of PTSD. A change of personal identity status through indoctrination and brainwashing differs from a traumatically induced change of personality (as we have been discussing) by having its own, particular dynamics and a different symptomatic configuration (Lifton 1976). The appearance of endemic discoveries of multiple identities in the context of questionable large-scale gatherings demands a further explanatory model in which uncontrolled group-dynamic processes, ideological manipulation, religious fundamentalism, and paranoid social criticism must all be considered major factors (Spanos 1994).

4.4 Treatment

Dissociative disorders are characterized by a high degree of variability of their most prominent manifestations, a broad spectrum of intensity, variable acuity versus chronicity, and a variable measure of psychiatric comorbidity. Although epidemiologic studies have reliably documented their high prevalence and thus the correspondingly high demand for treatment, no method of treatment has been empirically tested in systematic, controlled studies. This is true both of psychotherapeutic methods and of biological (particularly psychopharmacological) interventions. Principles of therapy discussed in the literature on this topic are often of a very general nature distilled from clinical practice, though there may be a narrower focus on specific treatment problems.

There is a consensus among experts that treatment must be planned and provided largely on an individual basis. Psychotherapeutic strategies are generally preferred to biological approaches. In view of the importance of severe trauma in the etiology and pathogenesis of both the simple and, in particular, the complex dissociative disorders, stress is placed on avoiding the danger of inducing further trauma through therapeutic measures.

The treatment of a comorbid psychiatric disorder is geared toward the therapeutic standard for that particular disorder, but must also be adjusted to the specific type of therapeutic interaction that patients with dissociative disorders require. A syndrome-oriented, mainly psychopharmacological treatment of this

psychiatric comorbidity may become a priority when psychiatric emergencies arise or when access to the patient is blocked by the comorbidity, meaning that the indicated form of psychotherapy cannot be performed.

4.4.1 Dissociative Amnesia and Dissociative Fugue

The goal of rapid restoration of the full ability to remember and integrate previously dissociated memories requires a secure and reliable treatment framework and is based on a stable therapeutic alliance. Cautious history-taking accompanied by repeated reassurance and encouragement of recollection can promote this therapeutic goal. It is essential to allow the patient to set the pace of the process of recollection, thereby giving him or her a feeling of control and power (Loewenstein 1995). Active involvement by the therapist is necessary for the safety of the patient if suicidal, parasuicidal, or outwardly aggressive crises occur during the process of recollection. Temporary hospitalization must be considered in such cases. Finally, the therapist and the patient must work toward a restoration of knowledge of personal identity and of the particular life circumstances that provoked the disorder. This always requires a confrontation with the inducing trauma and a focusing on, and working through of, existing conflicts.

It must be realized that patients with dissociative fugue most probably have a more complex psychodynamic situation at the root of their disorder. A tolerant approach toward their frequent feelings of shame and guilt about sexual conflicts, social misbehavior, marital discord, financial troubles, illegal activities, etc., is often necessary. Successful treatment requires a confrontation with and gradual integration, or sometimes even simple acceptance, of modes of behavior during the fugue episode that may markedly diverge from the patient's internal moral standards and from what the patient requires of him- or herself under other circumstances (Coons and Milstein 1992).

If major affective, psychotic, obsessive-compulsive, or anxious manifestations or manifestations of PTSD or of deficient impulse control should arise during the recollection of the traumatic events that induced the dissociative state, then differential psychopharmacological interventions are definitely indicated, and the therapist must devote the necessary attention to the treatment of amnesia (Friedman 1997). As a rule, the formerly dissociated information recollected during therapy must be discussed and reprocessed repeatedly in several sessions until harmonious integration is achieved.

Hypnotic techniques or an amobarbital interview may be of assistance in the treatment of dissociative amnesia and fugue and generally leads to remission of symptoms (Loewenstein 1995). The use of these modes of treatment requires even closer attention to the general therapeutic principles discussed above.

4.4.2 Depersonalization and Derealization

The therapeutic approach to patients with primary depersonalization and derealization disorders is based on the same general principles. Treatment begins with psychoeducational instruction of the patient regarding the nature of the often very distressing symptoms, training in the recognition of possible inducing stimuli, and detailed information about the different forms of treatment available.

The patient can be taught specific techniques for the alleviation of symptoms right away (Steinberg 1995):

- *"Grounding" or reorientation techniques*: The patient is instructed, if depersonalization occurs, to make physical contact with the objects in his or her familiar environment, to touch them and concentrate on the sensations this produces, and to speak his or her own name and location and other positive features of the environment out loud, thereby producing a supportive effect, which results in increased self-confidence.
- *Distraction techniques*: Deliberate absorption in a pleasant activity, such as an intense conversation with a trusted person, can break the vicious circle of negative thoughts and emotions that generally accompanies an experience of depersonalization.
- *Controlled dissociation*: The patient is systematically instructed to raise or lower the intensity of the feeling of depersonalization voluntarily whenever an episode occurs. This allows the patient to regain a feeling of control over the dissociative symptom, which the individual usually experiences as overwhelming in the face of his or her own helplessness and powerlessness.
- *Creative visualization*: The patient is instructed to visualize with the "inner eye" some positive recollection, pleasant place, or familiar and reassuring situation as vividly as possible, thus fending off the depersonalizing experience.

Hypnotic techniques are sometimes used with success (Spiegel 1988). Cognitive strategies play a major role in the treatment of cognitive distortions, which regularly accompany traumatically induced depersonalization (Briere 1992). Modified psychodynamic techniques concentrate on depersonalization as a pathological defense mechanism leading to a fragmented perception of the self and of objective reality. A therapeutic position grounded in ego psychology and object-relation theory appears to be advantageous

(Grotstein 1981). A classical psychoanalytical treatment setting, involving a suspension of the visual modality during the therapeutic interaction, is contraindicated in most cases of chronic depersonalization (Cattell and Cattell 1974).

Extremely varied psychopharmacological strategies have been applied in the treatment of depersonalization syndromes. Although symptoms of depersonalization in the context of other primary disorders generally improve when the underlying disorder, e.g. depression or anxiety, is treated, the results obtained to date in the treatment of primary depersonalization disorder are contradictory and unconvincing. The neurobiological conception of depersonalization as a disturbance within the spectrum of obsession and compulsion (Simeon et al. 1995) and the associated hypothesis of serotonergic dysfunction suggest a role for the use of selective serotonin reuptake inhibitors. Hollander et al. (1990) reported promising initial experiences in an open study with fluoxetine and fluvoxamine. Empirically controlled studies, however, are not yet available. The same is true of the use of clonazepam (Stein and Uhde 1989) and the theoretically intriguing use of opiate antagonists, such as naltrexone, as a form of treatment for depersonalization and self-injury (Winchel and Stanley 1991).

4.4.3 Dissociative Identity Disorder

The psychotherapy of patients with dissociative identity disorder is generally time-consuming and fraught with frequent setbacks. These usually result from the therapist's attempts to identify and work through the often extremely traumatic experiences lying at the root of the disorder. It has been stated that two essential prerequisites for the successful treatment of such patients are an orientation toward the treatment model of post-traumatic disorders and an adherence to the principles of dynamic psychotherapy (Gabbard 1994).

The primary therapeutic goal of integration and unification of the individual alter egos in a single personal identity cannot always be achieved in reality. Harmonization of the internal tension between the individual personalities, without actual integration, may represent the outer limit of what therapy can achieve. In still more complicated cases, even this much is impossible, and therapy focuses on helping the patient by supporting him or her in an attempt to solve current problems and cope with the difficulties of the present living situation.

Kluft (1995) stresses the necessity of planning treatment in three basic phases, in each of which a different aspect of therapeutic technique is emphasized:

1. The phase of security (supportive interventions, ego reinforcement, empathy)
2. The phase of recollection and grief (post-traumatic treatment model)
3. The phase of connection (traditional psychodynamic treatment principles)

A safe therapeutic atmosphere and the creation of an informative and strengthening relationship based on trust – a reliable and durable therapeutic alliance – are preconditions for undertaking further therapeutic steps. It is necessary to establish contact with the alter egos that express themselves even in this early phase, to make a contract with them against premature termination of therapy, suicidal acts, self-destructive behavior, and so forth, to promote communication between the personalities, and to reduce the burden of symptoms as much as possible. Through the process of history-taking, the therapist can make acquaintance with the different personality states and begin to understand their essential characteristics, origins, and mutual relationships.

In this early phase, special symptom-oriented therapeutic strategies play an important role. They may be accompanied by a directed, highly cautious application of hypnosis, and they may be supported by differential psychopharmacotherapy. In this early stage of the therapeutic process, the therapist must not aim at forced recollection. If the therapist and patient do not succeed in establishing a secure basis for the first phase, active trauma work must not be embarked upon, as it will regularly be accompanied by the appearance of extreme affect and disorganizing recollections.

Colrain and Steele (1991) list the factors that must be considered as contraindications to the uncovering of dissociated memories in long-term psychotherapy:

- Early therapeutic stage
- An unstable therapeutic alliance
- Current or continuing abuse
- Current, acute external life crises
- Advanced age, severe physical handicaps, and/or terminal disease
- Lack of ego strength, including severe borderline and psychotic states or massive regression
- Uncontrolled, abrupt alternation of the different alter egos
- Uncontrolled flashbacks
- Severe conflicts and lack of cooperation in the alter ego system
- Severe primary alexithymia
- Temporary factors: imminent absence of the therapist, transitions in the patient's life cycle

Dealing with traumatic memories requires a high degree of empathy, a sense of the patient's ability to

deal with the associated stress, and a feeling for the maximal amount of confrontation that the patient can tolerate. A simple working off of the associated extreme affects is rarely therapeutically effective. The success of therapy depends on its being embedding in a stable therapeutic relationship, in which all of the patient's reports of experienced trauma, even if they are scarcely credible, are accepted as representing the patient's subjective reality and are taken as a kind of alarm signal communicated to the therapist. Likewise, a knowledge of the forms of possible recollections, especially visual flashbacks, somatized sensations, or blatantly staged behaviors, is essential. Recollections of this type may be based on a kernel of truth in some segment of the patient's prior biography, but they are certainly not always a simple copy of what was actually experienced, and they may indeed be considerably transformed.

The complexity of the defense process in these patients' recollection work becomes obvious when the information provided by different alter egos is contradictory, illogical, and bizarre and when confabulations fulfill the basic need for a defense mechanism rather than contributing to the discovery of the truth. In this stage of treatment, the therapist must not assume the task of clarifying the actual truth through insistent history-taking. Though the therapist must occasionally be highly active and engaged, particularly when suicidal, outwardly aggressive, or disorganized crises occur, he or she must also attempt to preserve as neutral a position as possible when receiving information from the patient. It is important both for the protection of the patient and for the maintenance of the therapist's own professionalism to prepare the patient for the possible, and even probable, appearance of unpleasant, traumatic memories and to adhere to the principle of informed uncertainty (Appelbaum and Gutheil 1991) with respect to the subjective reality that manifests itself in them.

Fascination and over-engagement on the part of the therapist often accompany the beginning of psychotherapy with such patients. This initial enthusiasm is just as often followed by repeated disappointment, exhausting setbacks, and feelings of confusion over the further course of treatment. Kluft (1996) describes the most common reaction patterns of the therapist as follows:

- Retreat into personal inaccessibility, skeptical assumption of the role of a detective for the objectification of historical truth
- The conviction that an "ordinary" therapeutic setting is deleterious to this extremely traumatized patient and that only "love" – in manifold variations, all of which violate therapeutic boundaries – can achieve a therapeutic benefit
- Assumption of the role of an advocate with the patient as client, urging the patient to pursue a legal remedy for the experienced injustice
- Counter-identification of the therapist with the patient and development of post-traumatic stress symptoms in the therapist

Highly regarded experts report that the intermediate phase of therapy is especially unlikely to proceed in a linear and predictable way. There are many obstacles along the path to an integration of the extremely different components of the patient's personality. A cohesive personal identity in relation to a biographical narrative formed by the patient in collaboration with the therapist often cannot be achieved and, when achieved, is rarely complete. The learning of new coping skills to deal with current life problems without resorting to dissociative defense mechanisms is a therapeutic goal of equal importance to the basic reconstruction of the patient's personality.

As a rule, the treatment of patients with dissociative identity disorder should follow a multimodal design. A pragmatic approach to therapy involves the application not only of techniques of cognitive analysis, but also of hypnotherapeutic strategies, temporary hospitalization, art therapy, and syndrome-oriented pharmacotherapy (Putnam and Loewenstein 1993). The few available results from follow-up studies of patients treated in this type of therapeutic setting are quite encouraging (Ellason and Ross 1997).

5 Conversion Disorders

5.1 Definition and Clinical Features

A conversion disorder involves a loss or alteration of neurological function that may cause a neurologic disease to be suspected, but that actually cannot be adequately explained in this way. A temporal relationship between the onset of physical symptoms and a psychosocial stress or intrapsychic conflict is required as part of the definition of a conversion disorder. The constellation of symptoms is not under conscious control, nor is it a culturally sanctioned response pattern, and it is not limited, in the syndromal sense, either to pain or to sexual dysfunction. The concept of conversion disorder in both ICD-10 and DSM-IV is restricted to such pseudoneurological disorders and thus differs from the symptomatologically much broader concepts found in older texts. The clinical forms of conversion disorder are characterized by pseudoneurological symptoms affecting voluntary

movement, sensation, and state of consciousness – the latter in the form of non-epileptic (psychogenic) seizures (Martin 1995).

One subtype of conversion disorder includes disturbances of motor function. The most prominent type of disturbance is a weakness or paralysis that usually affects the extremities, only rarely the neck or the trunk, and practically never the face or the tongue (Pincus 1982).

Symptoms appearing to arise from affections of the extrapyramidal system, such as tremor (Koller et al. 1989), dystonia (Marsden 1995), or dyskinesia (Williams et al. 1995), are much more difficult to evaluate. Even though the expressive nature of such disturbances, in relation to extremely problematic interpersonal interactions or psychosocial stress, may at times be unmistakable, the possibility of an underlying organic process must always be considered (Cunningham Owens 1990).

A number of characteristics of psychogenic disorders of gait and stance may be very helpful for their identification as a conversion syndrome (Brandt et al. 1994), particularly the following:

- Spontaneous fluctuation of stance and gait that may often be provoked by distraction or suggestion
- Marked slowing of movement, as in a slow-motion film
- A typical gradual increase of the amplitude of sway in the Romberg test, after an initially steady stance
- A "flat-footed walk on ice," marked by careful rolling of the foot with reduced movement in the ankles and knees, as on slippery ground
- Extremely uneconomical body posture or sudden giving way of the hip and knee joints, with catching of the body, usually without falling
- Frequent accompanying psychomotor expressive symptoms, such as a pained or strained facial expression, moaning while grasping the leg, mannered hand posture, or hyperventilation

Dysphagia ("globus hystericus"), aphonia or dysphonia, and urinary retention are other symptoms that may appear, albeit rarely (Kellner 1991).

A second subtype of conversion disorders includes disturbances of sensory function: the lack of any sense of pain may be quite impressive. Testing of the sense of touch is extremely problematical, as it is not very reliable even under normal clinical circumstances. Reported hypesthesia or anesthesia must therefore be carefully investigated; gross violation of known anatomical boundaries is a relatively reliable indicator that the disorder is psychogenic rather than neurological. Another typical finding is hemihypesthesia or hemianesthesia affecting the entire body, with strict respect of the midline (Toone 1990).

The sensory subtype also includes visual field defects or blindness as well as deafness (Newman 1993).

A significant innovation in DSM-IV is the inclusion of certain hallucinations in this diagnostic category. These misperceptions are identified as pseudohallucinations by the lack of other psychotic signs and by otherwise intact reality testing. More than one sensory modality may be affected. The perceived content of the hallucination is often of a naïve or fantastic character (Martin 1995).

A further subtype includes non-epileptic seizures; sensory deficits or symptoms may also occur. There are several typical clusters of clinical features that suggest, but do not unambiguously indicate, that a seizure is psychogenic (Savard and Andermann 1990):

- A psychogenic seizure may either start suddenly or develop gradually.
- Not uncommonly, quasi-auras involving nausea, vertigo, dyspnea, and headache may be reported to significant others, whose presence in the room obviously promotes the occurrence of a seizure. Seizures usually occur in a familiar environment at home, and very rarely at night.
- The seizure pattern itself is highly variable; it may be either uni- or bilateral, and it may include tonic, clonic, tonic–clonic, or non-purposeful high-amplitude motion components. Seizures are often accompanied by dystonic body postures, of which a dramatic, though rare example is the classical "arc de cercle"; other, striking accompanying movements include rhythmic pelvic thrusts and random patterns of movement of the upper and lower extremities, with thrusting and biting.
- The facial expression displays a broad spectrum of dramatized primary affects. Vocalizations may either precede or accompany the seizure.
- Patients are usually able to avoid painful positions or self-injury during the seizures. Self-destructive behavior may be present, however, as an expression of severe psychopathology of the personality. The rarer cases of urinary and fecal incontinence accompanying pseudoseizures are to be regarded in the same light.
- Dilated pupils indicate increased sympathetic tone, while cyanosis is the result of breath-holding.
- Geotropic eye movements are an interesting sign: passive movement of the head always causes a deviation of the eyes toward the floor and away from the examiner.
- Psychogenic seizures usually last significantly longer than epileptic seizures, for more than 2 min on average, and not uncommonly proceed to a prolonged pseudo-status epilepticus. Psychogenic sei-

zures also tend to occur more frequently than epileptic seizures, often several times a day.
- An essential requirement of the diagnosis is that the electroencephalogram (EEG) is normal during and between the seizures.

Major problems of differential diagnosis arise when a patient has both genuine epilepsy and psychogenic seizures. Psychogenic seizures are not exclusively of the major hysterical type, but may also manifest themselves as absence episodes, psychogenic syncope, fugue states, or stupor. These clinical entities illustrate the continuous transition between the conversion disorders and the dissociative disorders (Kihlstrom 1994).

Although the modern psychiatric classification systems rule out the diagnosis of a conversion disorder when sexual dysfunction or pain occurs as the single dominant symptom, these may indeed be important accompanying symptoms of conversion disorders. The same is true of a number of other somatoform disturbances of autonomic function or of sleep (Kapfhammer et al. 1992).

Supposed symptoms of conversion may instead turn out, on further investigation, to be prodromal signs of a misdiagnosed or initially not objectifiable neurological or neuropsychiatric illness. Now that sophisticated diagnostic techniques, including neuroimaging and neurophysiological testing, are available, such misdiagnoses are not likely to be as frequent as was suggested by the now classic study by Slater et al. (1965), which revealed a staggeringly high rate of misdiagnosis that was subsequently revealed during follow-up (Crimlisk et al. 1998; Mace and Trimble 1996). However, misdiagnoses are not at all rare, particularly at the time of the initial manifestation of disease (Cloninger 1987; Fishbain and Goldberg 1991; Jones and Barklage 1990; Watson and Buranen 1979).

A thorough listing of coexisting psychopathological symptoms revealed that conversion syndromes may accompany psychiatric disorders of extremely varied types; a particularly high frequency of such syndromes in patients with affective illnesses was emphasized, though they may also occur in schizophrenics (Slater et al. 1965, Ziegler and Imboden 1962). In a prospective study of neurological patients with conversion symptoms, Kapfhammer et al. (1998a) emphasized the important distinction in the differential diagnosis between a conversion disorder and a somatization disorder or artificial disorder. These diagnostic subgroups differ considerably with regard to the dynamics of disease course, psychiatric comorbidity, self-destructive motifs in illness behavior, psychosocial impairment, utilization of medical care, and associated socioeconomic cost.

5.2 Epidemiology

The collection of accurate data on the prevalence and incidence of conversion disorders is complicated by a number of methodological difficulties. These involve not only the varying theoretical conceptions of conversion and correspondingly inconstant diagnostic boundaries, but also the thoroughly heterogeneous groups of subjects and/or patients investigated, as well as the extension of the period of time in question to the subjects' entire lifetime, rather than a more reasonable restriction to the moment of investigation (Toone 1990). Obviously, prevalence figures obtained in patient populations are highly dependent on the particular outpatient or inpatient setting in which the investigation is carried out:

- Studies of lifetime prevalence performed in Scandinavia and the United States suggest that approximately 0.5% of the general population suffers from conversion symptoms (Ljungberg 1957; Weisman et al. 1976).
- It appears that only very few patients with conversion disorders present for treatment in outpatient psychiatric or psychotherapeutic practices. Stefansson et al. (1976) reported a figure of 15–20/100,000.
- Ford (1983), on the other hand, estimated the percentage of patients with conversion disorders among those in general medical facilities as being much higher than this (20%–25%).
- At the National Hospital (Queen Square), London, a specialized neurological facility, the rate of conversion disorders remained relatively constant at 0.85%–1.55% over several decades (Trimble 1981).
- In a psychiatric consultation service in a large general hospital, the frequency of conversion disorders over a period of several years was just over 4% (Kapfhammer et al. 1992, 1998a).
- The importance of sociocultural factors is revealed by the much higher prevalence figures obtained in Third World countries (Murphy 1990; Nandi et al. 1992).

Although a number of reports in the literature indicate a higher frequency of conversion disorders in subjects of lower educational and socioeconomic status, as well as in rural areas or in preindustrial societies (Folks et al. 1984; Jones 1980), skepticism is nonetheless warranted with regard to oversimplified conclusions of this type (Kapfhammer et al. 1992). Women seem to be over-represented in all studies available to date, though men very often develop conversion symptoms after accidents in the workplace or in military settings (Allodi 1974). All age-groups may be affected; the peak incidence is between the ages

of 20 and 40 (Kapfhammer et al. 1992; Tomasson et al. 1991).

5.3 Etiology and Pathogenesis

There are several models for the pathogenesis of the conversion disorders, among which psychodynamic concepts play the most important role. According to an early psychoanalytical formulation (Freud and Breuer 1895), conversion symptoms result from major conflicts of inner drives, which may be traced back to traumatic experiences of family interaction in early life. The memory of these experiences must be repressed and thereafter remains unconscious. In later life, however, the occurrence of analogous conflicts may result in their reactualization. In order to avoid a reliving of these highly affectively charged experiences, they are repressed again and then somatized. The physical symptoms represent, in symbolic expression, a compromise solution between drive impulses and defense. This conflict resolution by means of conversion constitutes the primary gain derived from the conversion symptom. The assumption of the illness role that the conversion symptom makes possible also provides the patient with a socially mediated secondary gain, which leads to a reduction of both internal and external stress. Empirical clinical observation has revealed that a subgroup of conversion patients – in particular those with psychogenic seizures – did indeed suffer a high rate of physical and sexual abuse in early childhood (Alper et al. 1993).

Over the ensuing years, there was considerable refinement and modification of the original psychoanalytic position, which was essentially a concept of hysteria. At first, drive conflicts arising from sexual traumatization at an oedipal stage of development were held to be essential for the generation of conversion states, but the range of possible conflicts became much wider thereafter, to include aggressive impulses, narcissistic regulation of self-esteem, problems of separation and individuation, and grief after loss. Ego-psychological studies showed that a fully developed symbolic representation is not present in all cases of conversion; depending on the structural level of intrapsychic processing and the depth of regression, less well developed forms of physical expression may predominate. In the current psychodynamic conception, conversion is an independent problem-solving strategy for eliminating a multiplicity of intrapsychic, interpersonal and social conflicts by means of identification with a certain illness role. This conflict-resolution mode is not connected to any particular personality structure (Mentzos 1980; Hoffmann 1996).

These modern psychodynamic conceptions are entirely compatible with the learning-theoretic approach, according to which classically conditioned and/or operant-conditioned physical reactions can be applied in the guise of symptoms of illness in order to deal with particularly stressful or conflict-laden situations:

- The importance of learning in this model becomes evident when we consider the large percentage of patients who themselves work in clinical facilities, are married to or friendly with hospital personnel, or have people with phenomenologically very similar symptoms in their immediate familial and social environment (Kapfhammer et al. 1992).
- Medico-sociological theories on the illness role and illness behavior also apply in this model context (Mechanic 1962; Pilowsky 1990). These theories also emphasize a learned pattern of behavior for eliminating emotionally stressful situations by means of excessive focusing on physical sensations and symptoms, as well as an assumption of the illness role and care-seeking behavior in medical facilities, which may take different forms in different individuals.

While psychoanalytical approaches generally presuppose a categorical unconsciousness of the motives underlying the generation of conversion symptoms, there is, in more recent theories, a dimensional conception of the patient's insight into these motives. Unconscious or automatically learned motives occupy one pole of this continuum, which, however, also includes other types that are much nearer to conscious awareness – even, at the opposite pole, motives of an interpersonal manipulative character (Ford and Folks 1985; Miller 1988).

According to an early theory that was advanced mainly by psychoanalytical authors, a hysterically structured personality was considered an important prerequisite for the development of conversion symptoms under the influence of typical conflict situations. Alarcon (1973) listed a number of descriptive features of this hysterical primary personality, i.e. histrionic behavior, emotional lability, dependence, overexcitability, egocentrism, seductive behavior, and suggestibility.

This personality type, supposedly conferring a predisposition to conversion formation, was also excellently described in the ego-psychological depiction by Shapiro (1965) of the impressionistic cognitive style, with openness to fleeting perceptual impressions, diffuse memory, and imprecise factual knowledge on one hand, and marked receptivity to unconscious symbols and affectualization on the other. From a clinical perspective, however, it must be pointed out that a hysterical personality is by no means a necessary precondition for the development of a conversion syndrome. Rather, systematic studies of personality

typology among conversion patients have revealed a multiplicity of different personalities, displaying not only histrionic, but also predominantly passive, dependent, and depressive traits (Chodoff and Lyons 1958; Folks et al. 1984; Toone 1990).

Painstaking clinical, and particularly neurological studies of patients with conversion syndromes have revealed a frequent association with coexisting major organic, primarily neurological, abnormalities (Folks et al. 1984; Marsden 1986; Merskey and Buhrich 1975; Merskey and Trimble 1979; Ron 1994; Roy 1977). Organic brain disorders may either predispose to the development of conversion disorders or serve as models of illness for the social learning process. A vivid illustration of the latter is the frequent simultaneous appearance of epileptic and non-epileptic seizures; non-epileptic seizures have been diagnosed in as many as 25% of patients with epilepsy (Ramchandani and Schindler 1993). This coexistence may be found in epileptic patients with early acquired brain damage and secondary cognitive deficits. Frontal lobe injuries with resulting impairment of attention and planning mechanisms also seem to predispose to the development of conversion disorders (Ron 1994; Spiegel 1991).

Several neurophysiological and neuropsychological hypotheses may further contribute to the understanding of conversion processes. Ludwig (1972) and Whitlock (1967) postulated that a corticofugal inhibitory mechanism for afferent impulses and a disturbance of attention function are fundamental to the generation of conversion. In accordance with the early formulation of hysteria by Kretschmer (1923) as a resorting to instinctual motor templates, including extreme forms such as the "movement storm" or "feigned-death reflex," in response to shocking affective experiences, these authors also described reactions to disaster as increasingly regressive modes of dealing with otherwise intolerable stressors. These regressive behaviors contribute to a defenseless and helpless pose and also form an obstacle to mature judgment of reality; this results in a psychological barrier being created, thus closing out the social situation or inner conflict that is perceived as dangerous. At the same time, the patient's freedom to direct his or her attention to any object desired is blocked by concentration on the physical symptom. This phenomenon is clearly related to the clinical sign of "la belle indifférence."

It may be that limited coordination of information between the two cerebral hemispheres plays a role in this process, as suggested by clinical observations of an increased frequency of conversion symptoms on the left side of the body (Bishop et al. 1978; Galin et al. 1977; Stern 1974).

Neuropsychological findings also suggest a hemispheric dysfunctionality of this type. Flor-Henry et al. (1981) demonstrated not only bifrontal changes, but also functional disturbances primarily in the nondominant hemisphere, in a group of patients with frequently recurring conversion symptoms. In their neuropsychological study, they reported a significant inhibition of verbally coded images and a simultaneous, marked affective incongruence. These experimental findings were reminiscent of psychoanalytical conceptions of an impressionistic cognitive style or of affectualization (Shapiro 1965). Observations of psychopathologically relevant changes in the presence of focal brain lesions, such as hemineglect (anosognosia or hemiplegia) with involvement of the nondominant parietal cortex, or anosodiaphoria (uncritical cheerfulness and denial of deficit) with right frontal brain damage, may provide a basis for future neuropsychological models of certain psychopathological abnormalities in conversion disorders, such as "la belle indifférence" (Cutting 1990).

5.4 Treatment

Patients with conversion disorders are very often first seen by neurologists. The first therapeutic measures should also be undertaken in this medical setting. A detailed physical examination is essential, including whatever additional diagnostic tests are indicated. It is important to communicate the findings clearly to the patient. They must be interpreted in an empathic, accepting, and not degrading manner that takes the patient's own conception of the illness into account.

Clinical experience reveals that hospitalized patients with conversion disorders quite frequently respond positively to a multiplicity of nonspecific suggestive measures (Baker and Silver 1987). When the conversion disorder takes a chronic course, patients also often learn to derive benefit from empathetically conducted physiotherapeutic exercises (Delargy et al. 1986). It remains unknown, however, whether these therapeutic measures alone can lead to persistent, long-term benefit, particularly if the underlying psychosocial or psychological problem has not been addressed.

It is clear that any psychiatric illness underlying conversion symptoms must be recognized and properly treated. In such situations, the syndrome-oriented application of differential psychopharmacotherapy is indicated (Kapfhammer 1995).

Behavior-therapeutic and cognitive-therapeutic approaches in acute clinical care have the advantage that a clearly structured treatment plan can be worked out with the patient, and the rationale for each step of the treatment can be precisely and intelligibly explained. Important elements of this process include, above all, attention to central cognitive convictions regarding the somatic symptoms, sensitization with respect to the

effects of attention and body perception, direct discussion of stress-inducing personal problems, reduction of avoidant coping strategies and promotion of more promising alternative strategies, physical activation, and reinforcement of "normal behavior." Important contact persons must then be included in the treatment process to allow the therapist to identify and alter the many forms of social reinforcement that result from and, in turn, serve to maintain the patient's illness behavior.

Lazarus (1963) reported the results of multimodally structured behavioral therapy in 27 patients suffering from conversion disorders. After an average of 14 sessions, a good or complete remission of symptoms had been achieved in 71% of the patients. Teasell and Shapiro (1993) emphasized that behavior-therapy techniques, and operant conditioning in particular, are associated with a high probability of recovery, even in chronic conversion disorders.

In a psychodynamically oriented therapeutic approach, an attempt is made to understand central conflicts and affective states as indicators of a disturbed experience of the self and the outside world and to recognize the unconscious dimension of the production of the disorder in the framework of the physician–patient relationship. Conversation with the therapist enables most patients with conversion disorders to gain a meaningful causal understanding of their condition with regard to current conflicts and psychosocial stressors. The conversion symptoms often have many different levels of reference, which can be talked through by the patient and the therapist in focal therapy, with resulting benefit (Daie and Witztum 1991; Viederman 1995). In individual cases, long-term psychoanalytic treatment can also be recommended. The indication depends on the patient's suitability for this particular treatment structure and, above all, on what the patient expects from it (Ford 1983).

Controlled studies on the treatment of conversion disorders are still lacking. Thus an overall assessment of the different types of treatment, ranging from hypnosis to family therapy, cannot provide any recommendation regarding standard modes of differential therapy.

Major changes in treatment are necessary, however, when a careful evaluation by differential diagnosis of conversion syndromes results in their reclassification as conversion disorders, somatization disorders, or artificial disorders (Kapfhammer et al. 1998a,b).

6 References

Akthar S, Brenner I (1979) Differential diagnosis of fugue-like states. J Clin Psychiatry 40: 381–385

Alarcon RD (1973) Hysteria and hysterical personality: how come one without the other? Psychiatry Q 47: 258–275

Allodi FA (1974) Accident neurosis: whatever happened to male hysteria? Can Psychiatr Assoc J 19: 291–296

Alper K, Devinsky O, Vasquez B et al (1993) Non-epileptic seizures and childhood sexual and physical abuse. Neurology 43: 1950–1953

Appelbaum PS, Gutheil TG (1991) Clinical handbook of psychiatry and law, 2nd edn. Williams and Wilkins, Baltimore

Baker JH, Silver JR (1987) Hysterical paraplegia. J Neurol Neurosurg Psychiatry 50: 375–382

Bernstein EM, Putnam FW (1986) Development, reliability and validity of a dissociation scale. J Nerv Ment Dis 174: 727–735

Berrington WP, Liddell DW, Foulds GA (1956) A re-evaluation of the fugue. J Med Science 102: 280–286

Bishop ER, Mobley MC, Farr WF (1978) Lateralization of conversion symptoms. Compr Psychiatry 19: 393–396

Boon S, Draijer N (1993) Multiple personality disorders in The Netherlands: a clinical investigation of 71 patients. Am J Psychiatry 150: 489–494

Brandt T, Huppert D, Lempert T, Dieterich M (1994) Psychogenfunktionelle Gang- und Standstörungen. In: Martinius J, Kapfhammer HP (eds) Nervenärztliche Dialoge. MMV Medizin, Munich, pp 91–101

Braun BG (1993) Multiple personality disorder and posttraumatic stress disorder: similarities and differences. In: Wilson JP, Raphael B (eds) International handbook of traumatic stress syndromes. Plenum, New York, pp 35–47

Bremner JD, Randall P, Scott TM et al (1995) MRI-based measurement of hippocampal volume in patients with combat-related posttraumatic stress disorder. Am J Psychiatry 152: 973–981

*Bremner JD, Vermetten E, Southwick SM, Krystal JH, Charney DS (1998) Trauma, memory, and dissociation: an integrative formulation. In: Bremner JD, Marmar CR (eds) Trauma, memory, and dissociation. American Psychiatric Press, Washington DC, pp 365–402

Briere J (1989) Therapy of adults molested as children: beyond survival. Springer, Berlin Heidelberg New York

**Cardeña E (1994) The domain of dissociation. In: Lynn SJ, Rhue JW (eds) Dissociation: clinical and theoretical perspectives. Guilford, New York, pp 15–31

Carlson EB, Rosser-Hogan R (1991) Trauma experiences, posttraumatic stress, dissociation, and depression in Cambodian refugees. Am J Psychiatry 148: 1548–1551

Cattell JP, Cattell JS (1974) Depersonalisation: psychological and social perspectives. In: Arieti S, Brody EB (eds) American handbook of psychiatry, vol 3, 2nd edn. Basic Books, New York, pp 766–799

Chodoff P, Lyons H (1958) Hysteria, the hysterical personality, and "hysterical" conversion. Am J Psychiatry 114: 734–740

Cloninger CR (1987) Diagnosis of somatoform disorders: a critique of DSM-III. In: Tischler GL (ed) Diagnosis and classification in psychiatry: a critical appraisal of DSM-III. Cambridge University Press, New York, pp 243–259

Colrain J, Steele K (1991) Treatment protocols for spontaneous abreactive memory work. In: Braun BG (ed) Proceedings of the eighth international conference on multiple personality disorder and dissociation. Chicago, pp 6–8

Coons PM, Milstein V (1992) Psychogenic amnesia: a clinical investigation of 25 cases. Dissociation 5: 73–79

Crimlisk HL, Bhatia K, Cope H et al (1998) Slater revisited: 6 year follow up study of patients with medically unexplained motor symptoms. Br Med J 316: 582–586

Cummings JL (1985) Clinical neuropsychiatry. Grune and Stratton, New York

Cunningham Owens DG (1990) Dystonia – a potential psychiatric pitfall. Br J Psychiatry 156: 620–634

Cutting J (1990) The right cerebral hemisphere and psychiatric disorders. Oxford University Press, Oxford

Daie MA, Witztum E (1991) Short-term strategic treatment in traumatic conversion reactions. Am J Psychother 45: 335–346

Davidson JRT, Foa EB (1993) Posttraumatic stress disorder: DSM-IV and beyond. American Psychiatric Press, Washington DC

Delargy MA, Peatfield RC, Burt AA (1986) Successful rehabilitation in conversion paralysis. Br Med J 292: 1730–1731

Ellason JW, Ross CA (1997) Two-year follow-up of inpatients with dissociative identity disorder. Am J Psychiatry 154: 832–839

Eriksson NG, Lundin T (1996) Early traumatic stress reactions among Swedish survivors of the M/S Estonia disaster. Br J Psychiatry 169: 713–716

Fishbain DA, Goldberg M (1991) The misdiagnosis of conversion disorder in a psychiatric emergency service. Gen Hosp Psychiatry 13: 177–181

Fisher C (1945) Amnestic states in war neurosis: the psychogenesis of fugues. Psychoanal Q 14: 437–468

Flor-Henry P, Fromm-Auch D, Taper M, Schopflocher D (1981) A neuropsychological study of the stable syndrome of hysteria. Biol Psychiatry 16: 601–626

*Folks DG, Ford CV, Regan WM (1984) Conversion symptoms in a general hospital. Psychosomatics 25: 285–295

Ford CV (1983) The somatizing disorders: illness as a way of life. Elsevier, New York

Ford CV, Folks DG (1985) Conversion disorders: an overview. Psychosomatics 26: 371–383

Franklin J (1990) The diagnosis of multiple personality disorder based on subtle dissociative signs. J Nerv Ment Dis 178: 4–14

Freud S (1926) Hemmung, Symptom, Angst. Gesammelte Werke, vol XIV. Fischer, Frankfurt am Main, pp 111–205

Freud S, Breuer J (1895) Beobachtung I. Frl. Anna O. (J. Breuer). Gesammelte Werke, vol I. Fischer, Frankfurt am Main, pp 221–243

Friedman MJ (1997) Drug treatment for PTSD. Answers and questions. In: Yehuda R, McFarlane AC (eds) Psychobiology of posttraumatic stress disorders. New York Academy of Sciences, New York, pp 359–371

Frischholz EJ, Lipman LS, Braun BG, Sachs RG (1992) Psychopathology, hypnotizability, and dissociation. Am J Psychiatry 149: 1521–1525

Gabbard GO (1994) Dissociative disorders. In: Gabbard GO (ed) Psychodynamic psychiatry in clinical practice. The DSM-IV edition. American Psychiatric Press, Washington DC, pp 291–325

Galin D, Diamond R, Broff D (1977) Lateralization of conversion symptoms: more frequent on the left. Am J Psychiatry 134: 578–580

Good MI (1993) The concept of an organic dissociative disorder: what is the evidence? Harvard Rev Psychiatry 1: 145–157

Grotstein J (1981) Splitting and projective identification. Aronson, New York

Hilgard ER (1986) Divided consciousness: multiple controls in human thought and action, revised edition. Wiley, New York

*Hoffmann SO (1996) Der Konversionsmechanismus. Vorschlag zur operationalen Definition eines für die Psychosomatische Medizin grundlegenden Konzepts. Psychotherapeut 41: 88–94

Hollander E, Liebowitz MR, DeCaria CM et al (1990) Treatment of depersonalization with serotonin reuptake blockers. J Clin Psychopharmacol 10: 200–203

Horowitz MJ (1976) Stress respone syndromes. Aronson, New York

Janet P (1889) L'automatisme psychologique. Alcan, Paris

Janet P (1894) État mental des hystériques. Rueff, Paris

Jones JB, Barklage NE (1990) Conversion disorder: camouflage for brain lesions in two cases. Arch Intern Med 150: 1343–1345

Jones MM (1980) Conversion reaction: anachronism or evolutionary form? A review of the neurologic, behavioral, and psychoanalytic literature. Psychol Bull 87: 427–441

Kapfhammer HP (1995) Der Patient mit Somatisierungsstörungen. In: Hippius H, Lauter H, Kapfhammer HP (eds) Psychiatrie für die Praxis 22. Der Problempatient in der Praxis. MMV Medizin, Munich, pp 35–52

Kapfhammer HP (1999a) Dissoziative Störungen. In: Möller HJ, Laux G, Kapfhammer HP (eds) Psychiatrie und Psychotherapie. Springer, Berlin Heidelberg New York, pp 1273–1302

Kapfhammer HP (1999b) Somatoforme Störungen. In: Möller HJ, Laux G, Kapfhammer HP (eds) Psychiatrie und Psychotherapie. Springer, Berlin Heidelberg New York, pp 1303–1385

Kapfhammer HP, Buchheim P, Bove D, Wagner A (1992) Konversionssymptome bei Patienten im psychiatrischen Konsiliardienst. Nervenarzt 63: 527–538

*Kapfhammer HP, Dobmeier P, Mayer C, Rothenhäusler HB (1998a) Konversionssyndrome in der Neurologie. Eine psychopathologische und psychodynamische Differenzierung in Konversionsstörung, Somatisierungsstörung und artifizielle Störung. Psychother Psychosom Med Psychol 48: 463–474

*Kapfhammer HP, Rothenhäusler HB, Dietrich E, Dobmeier P, Mayer C (1998b) Artifizielle Störungen – zwischen Täuschung und Selbstschädigung. Nervenarzt 69: 401–409

Kellner R (1991) Psychosomatic syndromes and somatic symptoms. American Psychiatric Press, Washington DC

Kihlstrom JF (1990) The psychological unconsciousness. In: Pervin L (ed) Handbook of personality: theory and research. Guilford, New York, pp 445–464

*Kihlstrom JF (1994) One hundred years of hysteria. In: Lynn SJ, Rhue JW (eds) Dissociation. Clinical and theoretical perspectives. Guilford, New York, pp 365–394

Kihlstrom JF, Glisky ML, Angiulo MJ (1994) Dissociative tendencies and dissociative disorders. J Abnorm Psychol 103: 117–124

Kluft RP (1995) Dissociative disorder. In: Gabbard GO (ed) Treatment of psychiatric disorders, 2nd edn. American Psychiatric Press, Washington DC, pp 1599–1632

*Kluft RP (1996) Dissociative identity disorder. In: Michelson LK, Ray WJ (eds) Handbook of dissociation: theoretical, empirical, and clinical perspectives. Plenum, New York, pp 337–366

Koller W, Lang A, Vetere-Overfield B et al (1989) Psychogenic tremors. Neurology 39: 1094–1099

Kopelman MD (1987) Amnesia: organic and psychogenic. Br J Psychiatry 150: 428–442

Kretschmer E (1923/1974) Hysterie, 7th edn. Thieme, Stuttgart

*Krystal JH, Bremner JD, Southwick SM, Charney DS (1998) The emerging neurobiology of dissociation: implications for treatment of posttraumatic stress disorder. In: Bremner JD,

Marmar CR (eds) Trauma, memory, and dissociation. American Psychiatric Press, Washington DC, pp 321–364
Kuch K, Cox BJ (1992) Symptoms of PTSD in 124 survivors of the Holocaust. Am J Psychiatry 149: 337–340
Lazarus AA (1963) The results of behavior therapy in 126 cases of severe neurosis. Behav Res Ther 1: 69–79
Leonhard K (1995) Aufteilung der endogenen Psychosen und ihre differenzielle Ätiologie, 7th edn. Thieme, Stuttgart New York
Lifton RJ (1976) The life of the self. Simon and Schuster, New York
Lishman WA (1998) Organic psychiatry. The psychological consequences of cerebral disorder, 3rd edn. Blackwell, Oxford
Ljungberg L (1957) Hysteria: clinical, prognostic and genetic study. Acta Psychiatr Scand 32[Suppl 112]: 1–162
Loewenstein RJ (1995) Dissociative amnesia and dissociative fugue. In: Gabbard GO (ed) Treatment of psychiatric disorders, 2nd edn. American Psychiatric Press, Washington DC, pp 1569–1597
Loewenstein RJ (1996) Dissociative amnesia and dissociative fugue. In: Michelson LK, Ray WJ (eds) Handbook of dissociation: theoretical, empirical, and clinical perspectives. Plenum, New York, pp 307–336
Loewenstein RJ, Ross DR (1992) Multiple personality and psychoanalysis. An introduction. Psychoanal Inquiry 12: 3–48
Ludwig AM (1972) Hysteria: a neurobiological theory. Arch Gen Psychiatry 27: 771–777
Lynn SJ, Rhue JW (1988) Fantasy proneness: hypnosis, developmental antecedents, and psychopathology. Am Psychol 43: 35–44
Mace CJ, Trimble MR (1996) Ten-year prognosis of conversion disorder. Br J Psychiatry 169: 282–288
Marmar CR, Weiss DS, Metzler T (1998) Peritraumatic dissociation and posttraumatic stress disorder. In: Bremner JD, Marmar CR (eds) Trauma, memory, and dissociation. American Psychiatric Press, Washington DC, pp 229–252
Marsden CD (1986) Hysteria – a neurologist's view. Psychol Med 16: 277–288
Marsden CD (1995) Psychogenic problems associated with dystonia. Adv Neurol 65: 319–326
*Martin RL (1995) DSM-IV changes for the somatoform disorders. Psychiatr Ann 25: 29–39
Mechanic D (1962) The concept of illness behaviour. J Chron Dis 15: 189–194
*Mentzos S (1980) Hysterie. Zur Psychodynamik unbewußter Inszenierungen. Kindler, Munich
Merskey H, Buhrich NH (1975) Hysteria and organic brain disease. Br J Med Psychol 48: 359–366
Merskey H, Trimble M (1979) Personality, sexual adjustment, and brain lesions in patients with conversion symptoms. Am J Psychiatry 136: 179–182
Meyer JE (1961) Depersonalization in adolescence. Psychiatry 24: 537–560
Miller E (1988) Defining hysterical symptoms. Psychol Med 18: 275–277
Modestin J (1992) Multiple personality disorder in Switzerland. Am J Psychiatry 149: 88–92
*Mollon P (1998) Multiple selves, multiple voices: working with trauma, violation and dissociation. Wiley, Chichester
Morgan AH, Hilgard ER (1973) Age differences in susceptibility to hypnosis. Int J Clin Exp Hypn 21: 78–85
Murphy MR (1990) Classification of the somatoform disorders. In: Bass C (ed) Physical symptoms and psychological illness. Blackwell, Oxford, pp 10–39
Nandi DN, Benerjee G, Nandi S, Nandi P (1992) Is hysteria on the wane? A community survey in West Bengal, India. Br J Psychiatry 160: 87–91
*Nemiah JC (1998) Early concepts of trauma, dissociation, and the unconscious: their history and current implications. In: Bremner JD, Marmar CR (eds) Trauma, memory, and dissociation. American Psychiatric Press, Washington DC, pp 1–26
Newman NJ (1993) Neuro-ophthalmology and psychiatry. Gen Hosp Psychiatry 15: 102–114
Noyes R, Kletti R (1977) Depersonalization in response to life-threatening danger. Compr Psychiatry 18: 375–384
Paris J (1998) Does childhood trauma cause personality disorders in adults? Can J Psychiatry 43: 148–153
Perry BD, Pollard RA, Blakley TL, Baker WL, Vigilante D (1995) Childhood trauma, the neurobiology of adaptation, and "use-dependent" development of brain: how "states" become "traits. Infant Ment Health 16: 271–291
Pilowsky I (1990) The concept of abnormal illness behavior. Psychosomatics 21: 207–213
Pincus J (1982) Hysteria presenting to the neurologist. In: Roy A (ed) Hysteria. Wiley, London, pp 131–144
Pope HG, Hudson, Bodkin JA, Oliva P (1998) Questionable validity of "dissociative amnesia" in trauma victims. Br J Psychiatry 172: 210–215
Putnam FW (1989) The diagnosis and treatment of multiple personality disorder. Guilford, New York
Putnam FW (1995) Development of dissociative disorders. In: Cicchetti D, Cohen DJ (eds) Developmental psychopathology. 2. Risk, disorder, and adaptation. Wiley, New York, pp 581–608
Putnam FW, Carlson BE (1998) Hypnosis, dissociation, and trauma: myths, metaphors, and mechanisms. In: Bremner JD, Marmar CR (eds) Trauma, memory, and dissociation. American Psychiatric Press, Washington DC, pp 27–56
Putnam FW, Loewenstein RJ (1993) Treatment of multiple personality disorder: a survey of current practices. Am J Psychiatry 150: 1048–1052
Ramchandani D, Schindler B (1993) Evaluation of pseudoseizures. A psychiatric perspective. Psychosomatics 43: 70–79
Rhue JW, Lynn SJ (1989) Fantasy proneness, absorption, and hypnosis: a re-examination. Int J Clin Exp Hypnosis 37: 100–106
Riether AM, Stoudemire A (1988) Psychogenic fugue states: a review. South Med J 81: 568–571
Roche S, McConkey KM (1990) Absorption: nature, assessment, and correlates. J Pers Soc Psychol 59: 91–101
Ron MA (1994) Somatization in neurological practice. J Neurol Neurosurg Psychiatry 57: 1161–1164
Ross CA (1991) Epidemiology of multiple personality disorder and dissociation. Psychiatr Clin North Am 14: 503–518
Ross CA, Joshi S, Curri R (1990) Dissociative experiences in the general population. Am J Psychiatry 147: 1547–1552
Ross CA, Joshi S, Curri R (1991) Dissociative experiences in the general population: a factor analysis. Hosp Commun Psychiatry 42: 297–301
Roth M (1959) The phobic anxiety-depersonalization syndrome. Proc Royal Soc Med 52: 587–595
Roy A (1977) Cerebral disease and hysteria. Compr Psychiatry 18: 607–609
Savard G, Andermann F (1990) Convulsive pseudoseizures: a review of current concepts. Behav Neurology 3: 133–141

Schacter DL (1986) Amnesia and crime: how much do we really know? Am Psychol 41: 286–295

Schacter DL, Kihlstrom JF (1989) Functional amnesia. In: Boller F, Grafman J (eds) Handbook of neuropsychology. Elsevier, New York, pp 209–231

Schilder P (1935) The image and appearance of the human body: studies in the constructive energies of the psyche. Kegan Paul, London

Shapiro D (1965) Neurotic styles. Basic Books, New York

Simeon D, Stein DJ, Hollander E (1995) Depersonalization disorder and self-injurious behavior. J Clin Psychiatry 56[Suppl 4]: 36–39

Sims A (1995) Symptoms in the mind. An introduction to descriptive psychopathology, 2nd edn. Saunders, London

Singer JL (ed) (1990) Repression and dissociation. University of Chicago Press, Chicago

*Slater E, Beard W, Glithero E (1965) A follow-up of patients diagnosed as suffering from hysteria. J Psychosom Res 9: 9–13

*Spanos NP (1994) Multiple identity enactments and multiple personality disorder: a sociocognitive perspective. Psychol Bull 116: 143–165

Spiegel D (1988) Dissociation and hypnosis in posttraumatic stress disorders. J Trauma Stress 1: 17–33

Spiegel D (1991) Neurophysiological correlates of hypnosis and dissociation. J Neuropsychiatry Clin Neurosci 3: 440–445

**Spiegel D, Cardeña E (1991) Disintegrated experience: the dissociative disorders revisited. J Abnorm Psychol 100: 366–378

*Spitzer C, Freyberger HJ, Kessler C, Kömpf D (1994) Psychiatrische Komorbidität dissoziativer Störungen in der Neurologie. Nervenarzt 65: 680–688

Stefansson JG, Medina JA, Meyerowitz S (1976) Hysterical neurosis, conversion types. Acta Psychiatr Scand 53: 119–138

Stein MB, Uhde TW (1989) Depersonalization disorder: effects of caffeine and response to pharmacotherapy. Biol Psychiatry 26: 315–320

Steinberg M (1994a) Systematizing dissociation: symptomatology and diagnostic assessment. In: Spiegel E (ed) Dissociation – culture, mind, and body. American Psychiatric Press, Washington DC, pp 59–90

Steinberg M (1994b) Interview guide to the structured clinical interview for the diagnosis of DSM-IV dissociative disorders. American Psychiatric Press, Washington DC

Steinberg M (1995) Depersonalization. In: Gabbard GO (ed) Treatment of psychiatric disorders, 2nd edn. American Psychiatric Press, Washington DC, pp 1633–1653

Stern DB (1977) Lateral distribution of conversion reactions. J Nerv Ment Dis 164: 122–128

Teasell RA, Shapiro AP (1993) Rehabilitation of chronic motor conversion disorders. Crit Rev Phys Rehab Med 5: 1–13

Terr LC (1991) Childhood traumas: an outline and overview. Am J Psychiatry 148: 10–20

Tomasson K, Kent D, Coryell W (1991) Somatization and conversion disorder: comorbidity and demographics at presentation. Acta Psychiatr Scand 84: 288–293

*Toone BK (1990) Disorders of hysterical conversion. In: Bass C (ed) Somatization: physical symptoms and psychological illness. Blackwell, Oxford, pp 207–234

Trimble MR (1981) Neuropsychiatry. Wiley, Chichester

*van der Hart O, Horst R (1989) The dissociation theory of Pierre Janet. J Trauma Stress 2: 399–414

Viederman M (1995) Metaphor and meaning in conversion disorder: a brief active therapy. Psychosom Med 57: 403–409

Watson CG, Buranen C (1979) The frequency and identification of false positive conversion reactions. J Nerv Ment Dis 167: 243–247

Weissman M, Meyers J, Harding P (1978) Psychiatric disorders in a U.S. urban community: 1975–1976. Am J Psychiatry 135: 459–462

Whitlock FA (1967) The aetiology of hysteria. Acta Psychiatr Scand 43: 144–162

Williams DT, Ford B, Fahn S (1995) Phenomenology and psychopathology related to psychogenic movement disorders. Adv Neurol 65: 231–257

Winchel RM, Stanley M (1991) Self-injurious behavior: a review of the behavior and biology of self-mutilation. Am J Psychiatry 148: 306–317

Ziegler FG, Imboden JB (1962) Conversion reactions. Arch Gen Psychiatry 6: 279–287

CHAPTER

7

R.A. Mayou

Somatoform Disorders

1 Unexplained Physical Symptoms

The simple descriptive term "unexplained physical symptoms" (which is the ICD-10 primary care alternative to somatoform disorder) is to be preferred to some of the other labels that have been widely used, such as "hypochondriasis" in a general sense, "functional symptoms", "functional overlay" and "somatization". The term "somatization" is widely used to describe these patients, but it is best regarded as a psychological process rather than a category. It has been variously defined; one widely quoted definition is "a tendency to experience and communicate psychological distress in the form of somatic symptoms and to seek medical help for them" (Lipowski 1988). This is unsatisfactory in that it contains several distinct elements and it also implicitly assumes a process in which psychological distress is expressed as bodily symptoms. The term "somatization" is currently used in at least three different ways: (1) high levels of medically unexplained symptoms, (2) somatic preoccupation or worry about illness and (3) the somatic presentation of traditional psychiatric disorders (Kirmayer and Robbins 1991).

Although the association with a range of psychiatric disorders has repeatedly been shown, patients with medically unexplained symptoms have never been satisfactorily accommodated in either medical or psychiatric classifications. As a result, physicians and psychiatrists have used different terms in diagnoses. Physicians have tended to use "syndrome descriptions" which reflect in their special interests, e.g. irritable bowel, fibromyalgia, chronic pelvic pain. This has recently been usefully modified by the development of operational syndrome definitions such as those used for chronic fatigue syndrome (Wessely et al. 1998) and irritable bowel.

Psychiatrists, on the other hand, have focused on the associated psychological symptoms and have tried to use standard psychiatric diagnostic categories together with the recent and still ill-defined group of somatoform disorders. This grouping was introduced in DSM-III for conditions in which physical complaints are more conspicuous than psychological ones and is now widely used. In ICD-10, it is grouped as a part of "neurotic, stress-related and somatoform disorders" (F40–F48) and is defined as follows: "the main feature of somatoform disorders is repeated presentation of physicals symptoms, together with a persistent request for medical investigations, in spite of repeated negative findings and reassurances by doctors that the symptoms have no physical basis". It includes several subcategories with traditional names as well as new groupings; it is unfortunate that there are considerable differences between ICD-10 and DSM-IV.

In everyday clinical practice, it is best to formulate clinical problems according to the physical syndrome, course and disability and the presence of any associated psychiatric disorder or disorders.

1.1 Epidemiology

In the general population, unexplained physical symptoms severe enough to cause distress and some limitation are extremely frequent. For example, a North American report found that 41% of subjects reported back pain in the previous 6 months, 26% headache, 17% abdominal pain, 12% chest pain and 12% facial pain. Many of these complaints were long-standing, recurrent and associated with some limitation of activities. As in other studies, the greater the number of physical complaints, the higher the proportion suffering from current and lifetime psychiatric disorder (Von Korff et al. 1988). There are few cross-cultural differences in overall prevalence, although patterns of types of symptoms vary from country to country (Kirmayer et al. 1994; Simon and Von Korff 1991).

In primary care, at least a fifth of all new presentations are of unexplained physical symptoms (Goldberg 1995). These are of many types, but among the commonest are abdominal pain, chest pain, musculoskeletal pains, fatigue and headache (Kroenke and Mangelsdorff 1989). Most of the symptoms are transient, but a proportion are persistent and associated with significant disability and very considerable use of medical resources. Medically unexplained symptoms are frequent in many general hospital outpatient clinics and especially so in the selected population referred to pain clinics. Follow-up studies suggest that negative investigation fails to reassure many clinic attenders and that this is especially so in those assessed as having high health anxiety. Non-specific diagnoses are most frequent among general hospital inpatient discharge diagnoses (Mayou et al. 1991). In a series of Danish case register studies, Fink has documented the extensive medical and surgical resources used by a small number of people who have multiple admissions to hospital for which no physical diagnosis is recorded (Fink 1995).

1.2 Aetiology

Aetiology has been viewed from various standpoints – psychodynamic, psychiatric disorder, behavioural,

cognitive. These explanations are not mutually exclusive; it is probable that the aetiology is often multicausal (Kellner 1991; Barsky and Klerman 1983). Current models emphasize the importance of attribution or cognitive interpretation of bodily sensations. This interpretation may be affected by cultural and illness belief factors or by personality and other individual psychological factors (Mayou et al. 1995), as illustrated in Fig. 1.

The model implies that pathophysiological mechanisms cause bodily perceptions which are then interpreted and that subjective symptoms and disability are determined by a process of attribution, or cognitive interpretation. This in turn is determined by the patient's recent and lifetime illness experience and by aspects of personality and current mental state. Once symptoms have occurred, maintaining factors will inevitably include the reactions of other people, family, friends, acquaintances and especially doctors, and secondary psychological factors, such as anxiety, which may reinforce the underlying pathophysiological processes and the attribution of meaning. There is very considerable evidence about each aspect of this model, and it is useful to examine these in turn.

A very large number of possible minor pathological and physiological causes have been described, and they include the physical consequences of psychiatric disorder, especially anxiety (Sharpe and Bass 1992). These include the following:

1. Major pathology
2. Minor pathology
3. Physiological processes
   a) Sinus tachycardia and benign minor arrhythmias
   b) Effects of fatigue
   c) Hangover
   d) Effects of overeating
   e) Effects of prolonged inactivity
   f) Autonomic effects of anxiety
   g) Lack of sleep

Recent research has consistently shown evidence that illness experience, remote and recent, appears to predispose to and precipitate medically unexplained symptoms:

- Childhood illness
- Family illness and consultation in childhood
- Childhood consultation and school absence
- Physical illness in adult life
- Experience and satisfaction with medical consultation
- Illness in family and friends
- Publicity in television, newspapers, etc.
- Knowledge of illness and its treatment

It is not surprising, and indeed may be even appropriate, that past experience should affects subjects' interpretation and concern about symptoms which may be medically minor but may, in fact, be of considerable individual significance.

There has been extensive research on understanding the role of individual perception and interpretation of symptoms in experimental studies in populations with and without physical disorder (Pennebaker and Watson 1991) and in clinical studies (Barsky 1992; Mayou et al. 1995). These have resulted in health belief, attribution and cognitive models that combine the roles of personality variables, attention and health beliefs (Cioffi 1991). They result in two separate, but related, aspects: amplification/minimization and misinterpretation. Cognitive behavioural formulations emphasize the central significance of health anxiety and of the reinforcing effects of its physiological, cognitive, affective and behavioural consequences (Salkovskis and Clark 1993). It is evident there are no real differences in the psychological mechanisms of symptom perception whether the underlying cause is attributable to major pathology or to physiological processes or trivial pathology. The aetiology of physically unexplained symptoms differs from that of the impact of physical illness only in the proportion of causation attributable to psychological variables.

There is very considerable complexity in a psycho-

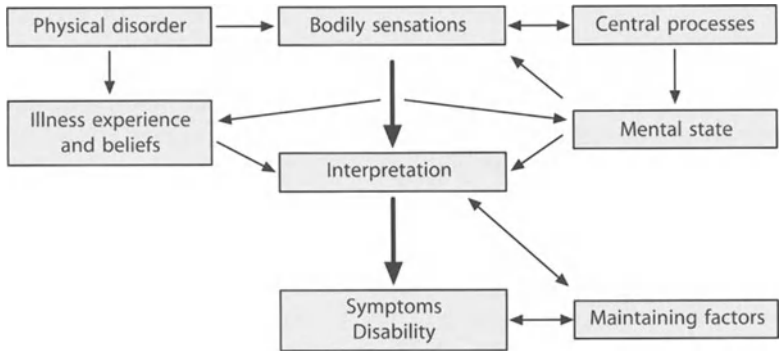


Fig. 1. Model of the aetiology of unexplained physical symptoms

logical and psychiatric literature which uses varying terms, each of which has been used in many different ways. They refer to:

- Vulnerable personality – individuals differ widely in their tendency to health anxiety about minor bodily sensations and a small extreme subgroup may be diagnosable as suffering from a hypochondriacal personality disorder.
- Psychological reaction to acute or chronic stresses – the general worry and stress may result in bodily symptoms of autonomic arousal and in a general increase in health anxiety and a tendency to misinterpretation.
- Psychiatric disorder – although not invariably associated with disabling and persistent unexplained symptoms, it is usual when symptoms are severe, multiple or recurrent and accompanied by disability and disease conviction.

Psychiatric disorder is important in the aetiology of chronic unexplained symptoms, but there are many people whose physical symptoms are due to psychological processes but who do not suffer from diagnosable psychiatric disorder. They may benefit from psychological or psychiatric treatment. Many have sub-threshold disorders (Janca et al. 1995). The uncertainties that have always surrounded the conditions now collectively referred to as somatoform disorders has meant that there has been great variation between national systems of classification, criteria and terminology.

There have been numerous studies of the association between psychiatric disorder and unexplained symptoms. There is a strong association between the number of physical symptoms and psychiatric disorder. The range of diagnosis is wide. In primary care, adjustment disorder is the commonest diagnosis, but affective and somatoform disorders are also frequent. The WHO Primary Care Study suggests remarkably similarities in all the participating countries (Üstün and Sartorius 1995). In secondary care, affective disorder and panic and other anxiety disorders predominate but, as in other medical settings, there is considerable co-morbidity with somatoform disorders.

A Canadian study which examined patterns of somatization in 685 patients attending family medicine clinics (Kirmayer and Robbins 1991) found that 26% met criteria for at least one of three definitions of somatization: (1) numerous unexplained physical symptoms, (2) illness worry or hypochondriasis, and (3) somatic presentations of major depression and anxiety. Most patients met criteria for only one of these three types of somatization, suggesting three distinct groups.

Once established, the symptoms and associated distress may be perpetuated by secondary physiological, behavioural or cognitive processes as well as by a number of social factors, including the reaction of others. There is substantial evidence that the behaviour of doctors may make problems worse by failing to provide explanations, by giving information which is inconsistent or ambiguous or by over-investigation. Simple reassurance following negative investigation is often ineffective (Howard and Wessely 1996), especially in people who are prone to suffer health anxiety as a feature of their personality.

1.3 Assessment

Psychological factors should always be considered from the beginning of the assessment in any patient with persistent and unexplained physical symptoms. Early assessment of psychological and social factors and awareness of the possibility of negative physical investigation enables the doctor to pursue physical and psychological care in parallel. The general principles of assessment are as follows:

- Consider psychological factors from outset
- Appropriate physical investigation to exclude physical cause
- Clarify psychological and physical complaints
- Understand patient's beliefs and expectations
- Identify depression or other psychiatric disorder
- Identify psychosocial problems

Past history may provide evidence of previous episodes of unexplained physical symptoms or of a psychiatric disorder. It is important to enquire about mood and anxiety symptoms and also to question patients about their beliefs about the nature and meaning of their symptoms (Marple et al. 1997).

There is a tendency for non-specialists to pursue excessive physical investigation, often in response to demands from patients and their families. This should be avoided as it often creates problems becoming less and less effective and leading to further requests for investigation and reassurance. Investigations should be chosen for good medical reasons and, if negative, the results need to be clearly explained and discussed. It has repeatedly been found that straightforward reassurance that findings are negative is only transiently effective in those who have high usual levels of health anxiety. Indeed, repeated reassurance may exacerbate problems.

On completion of investigations, it should be possible to make a formulation of the causes in terms of the aetiological model in Fig. 1, which can be the basis of treatment, whether this is simple explanation and reassurance or specialist psychological treatment. Treatment of any underlying physical causes and modification of any of the evidence in the model is

valuable, but the central themes of treatment are usually the management of psychiatric disorder and the modification of inappropriate interpretation.

1.4 Treatment

Improved treatment of unexplained, somatoform, symptoms has two elements. In view of the very large numbers of patients who are disabled and who use very considerable amounts of medical care, the main emphasis will need to be on simple ways of improving general medical care:

1. Measures to improve recognition, explanation, advice and avoidance of iatrogenic problems in routine medical care; good early treatment by non-specialists can prevent unnecessary investigation and long-term iatrogenic disability.
2. The role of specialist services in assessment and treatment and also supervision and training of other disciplines involved in care.

Consideration of the interaction of minor physical and psychological factors from the outset enables the doctor to avoid all the difficulties that arise from presenting patients with sudden change from an apparently physical to a psychiatric explanation. The doctor can show that the presenting symptoms are accepted as real, familiar and deserving of treatment. Reassurance that there is not a serious or sinister underlying physical disorder must be accompanied by an explanation and specific advice which meets the patient's anxieties. With the majority of patients in primary care, it is unnecessary to raise issues of psychiatric diagnosis, although it is often important to discuss the ways in which everyday psychological factors have an important role in *all* medical problems.

Follow-up review and discussion and the involvement of a close relative can be useful. Goldberg et al. (1989) described a model of what they refer to as "reattribution", which they have used successfully to train primary care doctors to manage unexplained physical symptoms. This has three important elements: making the patient feel understood, changing the agenda and making the link to psychological distress.

The general principles of treatment can be summarized as follows:

- Emphasize symptoms are real and medical care is appropriate
- Minimize/control physical care
- Offer explanation and discuss
- Discuss role of psychological factors in all medical care
- Treat primary psychiatric disorder
- Agree treatment plan

In the past, referral to psychiatrists for patients with persistent disabling symptoms has often been unhelpful, because many patients do not suffer from the types of psychiatric disorder seen in mental illness services. However, there is very considerable evidence that standard treatments of anxiety and depression and psychological treatments especially adapted for particular clinical problems can be highly effective. Unfortunately, they are neither widely used nor available at present. The main reasons for referral to psychiatrists or to other specialists are:

1. For help in diagnosis
2. For comprehensive assessment and advice on management
3. To provide specialist treatment

Since many patients find psychiatric explanations unacceptable and are reluctant to attend clinics which they see as treating mental illness, it is essential that referring doctors have a good working relationship with the psychiatrist or psychologist and are willing to explain the nature of the referral to the patient. Patients who have confidence in their primary care doctors or physical specialists are much more likely to accept referral if it is seen as involving a colleague of their own doctor and if they are assured that there condition is seen as real, familiar and deserving of medical treatment.

1.5 Specialist Treatments

The more fully that specialist treatment can be integrated with routine physical care, the more likely patients are to find the treatment acceptable and helpful. The general principles and procedures for specialist treatments need to be adapted and overall treatment plans formulated to take account of the presenting symptoms. For example, cognitive behavioural treatment has been shown to be effective in the treatment of a number of syndromes, including general medical symptoms (Speckens et al. 1996), chronic fatigue (Sharpe et al. 1996), non-cardiac chest pain (Mayou et al. 1997) and irritable bowel. The detailed treatments are directed towards the rather different symptoms, limitations and cognitions of each of these common problems. The therapist must be familiar with the medical issues and with the particular procedures, self-help materials and with other elements of treatment that have been shown to be effective.

The specialist, with knowledge of both physical and psychological aspects of somatoform symptoms, is better placed than the primary care doctor or general hospital physician to assess aetiology and to provide

formulations to the patient of the causes of the symptoms and the way they are being managed. Many patients are reassured by meeting a confident, sympathetic doctor who asks about their worries, accepts the symptoms as being real and deserving of treatment, finds them familiar and is able to discuss practical treatment plans.

Major depressive illness is an important cause of unexplained physical symptoms, and standard antidepressant treatment is required. Patients who are sceptical of the importance of psychological causes of their complaints are often reluctant to take antidepressant medication and are worried about its side-effects. The general principles of all treatment – careful explanation, the involvement of relatives, the support of other doctors – are fundamental to dealing with such concerns. It is also important to choose an antidepressant whose side-effects are least likely to cause problems and to provide the patient and family with a complete explanation.

Although, in general, antidepressants can only be expected to be effective where there are psychological symptoms of depressive disorder, there is evidence that a number of pain syndromes in which depression is not conspicuous respond to antidepressants.

Cognitive behavioural treatment adapted from methods which have been shown to be successful for anxiety disorders have now been demonstrated to be effective in medical outpatients (Speckens et al. 1996) with a number of specific types of physical symptoms, such as some chronic pain, non-cardiac chest pain (Mayou et al. 1997) and irritable bowel, as well as with the somatoform syndromes such as hypochondriasis (Warwick et al. 1996; Clark et al. 1998).

Individual and group psychotherapies have been widely used. They appear to be most effective when directed towards helping the patient establish an understanding of the origin of the symptoms and the ways in which changes in understanding and behaviour can lead to improvement (Guthrie 1995).

In the minority of patients with long histories of multiple symptoms, effectiveness of specific treatments is limited. Management has to be directed towards simplifying medical treatment so as to provide a consistent plan which can be agreed with the patient by a single doctor. It is frequently possible to very considerably reduce medical resources being used, to avoid the complications of excessive and contradictory care by a number of people and to achieve the improvement of the functional status and well-being of patients and their families. The techniques are considered more fully below in the section on somatization disorder.

With complicated chronic pain and other symptoms, treatment may involve individualized programmes of physical, psychological and social elements. It is difficult to evaluate such programmes, but they do appear to be helpful in outpatient and also in inpatient treatment (Main and Benjamin 1995).

2 Somatoform Disorders

Somatoform disorders were introduced as a category in DSM-III as a means of bringing together a number of traditional categories with recently proposed new groupings. The classification was seen as provisional, but has been little changed in DSM-IV. ICD-10 included a similar category but with significant differences in the sub-categories, which are shown in Table 1.

Although both classifications are intended to be provisional, they are often treated as having greater validity than was claimed by those who devised the criteria. The whole concept requires critical review (Murphy 1990; Janca et al. 1995). A major difficulty is the lack of applicability to a number of non-Western countries who do not share the mind/body separation which characterizes Western medicine (Kirmayer et al. 1994; Lee 1997). There are other significant problems:

1. There is considerable overlap with other major psychiatric disorders (especially anxiety).
2. The clinical descriptions of the specific psychiatric disorders are largely derived from hospital-based experience and are not readily applicable to the large number of people with functional complaints

Table 1. Categories of somatoform disorders in ICD-10 and DSM-IV

ICD-10	DSM-IV
Somatization disorder	Somatization disorder
Undifferentiated somatoform disorder	Undifferentiated somatoform disorder
Hypochondriacal disorder	Hypochondriasis
Somatoform autonomic dysfunction	No category
Persistent pain disorder	Pain disorder associated with psychological factors (and a general medical condition)
Other somatoform disorders	Somatoform disorders not otherwise specified
No category	Body dysmorphic disorder
No category	Conversion disorder
Neurasthenia	No category

in the community and primary care. Many of these have sub-threshold emotional problems.
3. Disabling, medically unexplained symptoms can occur in the absence of psychiatric diagnosis (especially in primary care).
4. Diagnostic criteria are based on a mixture of principles – aetiology, symptom counts, consultation, response to medical treatment.
5. The more precise sub-categories are only appropriate for a minority of subjects.
6. Many patients are classified in residual categories based on unexplained presenting symptoms.

The remainder of this chapter describes the main categories of somatoform disorder in ICD-10. The ICD category of neurasthenia is described in a separate chapter (Chap. 8, this volume), and conversion disorder, which is included with the somatoform disorders in DSM, is also discussed elsewhere in this book.

There have been few major epidemiological studies of somatoform disorder or of its sub-categories. General population studies of somatoform disorders have focused on patients with multiple symptoms (Simon and Von Korff 1991). Research in primary care has found somatoform disorders to be frequent among consecutive attenders, and there is often co-morbid affective or anxiety disorder. The most detailed evidence is provided by the WHO Primary Care Study, which has provided findings from centres in 14 countries (Üstün and Sartorius 1995). In contrast to the frequent claims that somatoform disorders are most frequent in non-Western cultures, it is apparent that there are relatively few differences between the wide range of countries in this project (Gureje et al. 1997a).

Table 2 summarizes findings from a Danish primary care study which demonstrates the marked differences in using ICD-10 and DSM-IV (Fink et al. 1999). DSM-IV is dominated by undifferentiated somatoform disorder and somatoform disorder not otherwise specified (NOS); in contrast, in ICD-10, undifferentiated somatoform disorder is less common and the highest prevalence is for the descriptive category of autonomic dysfunction. In both systems, the more precise categories of somatization disorder and hypochondriasis are relatively uncommon. Co-morbid psychiatric disorder was present in about half the cases.

Table 2. Somatoform disorders in a Danish primary care study of consecutive attenders (%) (adapted from Fink et al. 1999)

	ICD-10	DSM-IV
Somatization disorder	6	1
Undifferentiated somatoform disorder	7	27
Pain disorder	8	8
Hypochondriasis	4	4
Autonomic dysfunction	14	–
Neurasthenia	4	–
Conversion disorder	–	3
Somatoform disorder NOS	–	27

2.1 Hypochondriasis

The history of the term "hypochondriasis" extends over many centuries. In the twentieth century, it has had a relatively general lay meaning and a more precise psychiatric meaning, which has usually emphasized an excessive preoccupation with having a physical disorder. Hypochondriacal concerns occur in many psychiatric disorders, but there has been considerable argument about whether there is a primary syndrome. There is substantial overlap with anxiety, depression and somatization disorder. Definitions emphasize the cognitive and behavioural elements of misattribution of bodily sensations, disease conviction and reassurance-seeking (Côté et al. 1996).

DSM-III attempted to provide definitive criteria, and these remain largely unchanged in DSM-IV; ICD-10 is similar (see Appendix A). The main problem in diagnosis is that the criteria depend heavily upon the patient having sought help and having been appropriately investigated and reassured. In fact, there is evidence that many patients with unexplained physical symptoms have not had any convincing explanation and that reassurance may have been ambiguous or, indeed, even anxiety-provoking.

Hypochondriasis has been said to be relatively common in medical outpatient clinics with considerable psychiatric co-morbidity (Barsky et al. 1992). The WHO Primary Care Study reported that ICD-10 hypochondriasis was rare, with a prevalence of 0.8%, but that a less restrictively defined disorder had a frequency of 2.2% (Gureje et al. 1997b). Patients with this abridged hypochondriasis often had co-morbid major depression and generalized anxiety disorder and described considerable psychiatric ill-health and functional disability. There were no differences in rates between the 14 countries in the study. The key factor in the proposed abridged diagnosis is removal of the criterion of "persistent refusal to accept medial reassurance". The findings of this large study are consistent with other evidence that our current criteria are inadequate to define a syndrome characterized by health anxiety or disease conviction. The wider concept is also supported by accounts of *sub-threshold* and *transient* hypochondriasis in medical populations (Barsky et al. 1990).

The aetiology of hypochondriasis has generally been discussed in the terms discussed above for medically unexplained symptoms. However, three issues deserve particular attention:

1. *Is hypochondriasis a depressive disorder?* Kenyon (1964), in a widely quoted, large case-note study of patients diagnosed in a specialist psychiatric hospital, concluded that hypochondriasis was always secondary to another syndrome, usually depression. More recent research on representative populations has consistently supported the idea of a primary syndrome characterized by bodily preoccupation and conviction of disease.
2. *Is hypochondriasis a personality disorder?* It has been argued that hypochondriasis is often a lifelong problem of attitudes and behaviour (Tyrer et al. 1990). Although this may well be true for a subset of patients, it is probably more common for the syndrome to be episodic, with clear onset in adult life.
3. *Cognitive behavioural explanations.* Salkovskis and Clark (1993) have examined the role of cognitive explanations of panic disorder and hypochondriasis and conclude that both disorders result from a persistent tendency to misinterpret bodily changes as indicating catastrophic harm. Despite the overlap between the two syndromes, they suggest there are differences in terms of the perceived imminence of the catastrophe, the role of safety-seeking and the nature of some of the beliefs and assumptions upon which misinterpretations are based. Overall, the cognitive explanations fit in well with the concept of hypochondriasis as a subgroup of the whole spectrum of anxiety disorders in which health anxiety is especially prominent.

Treatment of hypochondriasis follows the general principles described earlier in this chapter. Associated depression should be treated with antidepressants. The disease conviction and misinterpretations seem to respond to a variety of individual or group psychotherapy or behavioural treatments. Although these treatments appear at first sight rather varied, they share the common aim of trying to enable the patient to have a clearer understanding of the origin of the symptoms and to answer specific worries and anxieties. This informed reassurance and explanation appears to be substantially more effective than straightforward reassurance that findings, examinations and investigations are all negative.

There have been several uncontrolled evaluations of treatment on populations of patients satisfying diagnostic criteria for hypochondriasis. Warwick and colleagues (1996) have reported an inpatient randomized controlled trial of intensive cognitive behavioural treatment. There were significant benefits in the treated group compared with those on the waiting list, benefits which were maintained at 3 months.

Anxiety and avoidance due to the fear of contracting a disease, e.g. human immunodeficiency virus (HIV), is know as *disease phobia* (Bianchi 1971) or illness phobia. There is no agreement as to whether this is best classified as a subgroup within hypochondriasis or within specific phobias. It is probably very similar to that of other specific phobias. Fears on injury, blood and dental treatment are common examples.

Dysmorphophobia, excessive concern about bodily appearance, is included with hypochondriasis in ICD-10, but is classified as a separate somatoform disorder, body dysmorphic disorder, in DSM-IV (Phillips and Hollander 1996). The ICD formulation has the advantage of emphasizing the probable similarity in psychological aetiology to those described for hypochondriasis, while DSM emphasizes the descriptive syndrome.

The epidemiology is unclear. The condition is uncommon, but probably more frequent than case reports and highly selected series have suggested. The rather few reports from less selected populations give a truer picture of a range of concern about appearance. Concern about appearance may be a symptom of a number of psychiatric disorders, including schizophrenia and delusional disorders.

Patients with dysmorphophobia often feel that they have not been taken seriously by doctors, and their feelings of grievance exacerbate their distress. Their complaints deserve careful assessment and collaboration between physicians and surgeons on the one hand and psychiatrists and psychologists on the other. In this way, it is possible to identify those who may respond to surgery or other physical treatment and those for whom specific psychiatric care may be useful.

Treatment is often difficult. Antidepressant medication can be effective in those with symptoms of depression. Plastic surgery or other physical treatment is often effective in those who have relatively precise dissatisfactions and realistic expectations of what may be achieved. Such treatments are ineffective and, indeed, are contraindicated in those who have vague complaints, especially those who have concerns about several parts of the body and who have histories of previous failed treatment. Cognitive behavioural treatment has been reported to be successful in a small number of subjects.

2.2 Somatization Disorder

Somatization disorder was introduced in DSM-III to describe a chronic condition characterized by numerous changing physical symptoms over a period of

many years. The precise formulation arose from a programme of research on a syndrome that was originally alleged to be a form of hysteria and was given the name Briquet's syndrome (Guze et al. 1986). The syndrome is now rightly separated from hysteria, but has somewhat different criteria in DSM-IV and ICD-10 (see Appendix B). There is a high degree of psychiatric co-morbidity, and personality disorder is common. The severity and the very considerable chronicity required to make the diagnosis suggest that there may well be advantages in considering the syndrome as a form of personality disorder (Speckens et al. 1995).

The definition depends on an entirely arbitrary cut-off based on symptom counts. While the syndrome defined in this manner has high reliability, there is little evidence that it has true validity and that it can be separated from similar syndromes with somewhat fewer physical complaints or with different patterns of complaints. It has been argued that there would be advantages in a broader definition (abridged somatization disorder). The prevalence of this suggested disorder is considerably greater than somatization disorder and may have some provisional practical utility (Escobar et al. 1989). However, there appears to be little justification for its arbitrary separation from undifferentiated somatoform disorder. As originally defined, somatization disorder is uncommon, between 0.1% and 0.4% in the United States (Simon and Von Korff 1991). It has a similar prevalence in all countries and cultures (Gureje et al. 1997a).

It is generally agreed that treatment of somatization disorder is difficult and that it is probably best directed towards control of symptoms and medical care rather than cure. Smith and colleagues (1986, 1995) have demonstrated in a series of important studies that provision of very simple advice to doctors treating patients with somatization disorder which suggested controlling medical care appeared to be effective in reducing consultation and health care costs, although there was little effect on their measures of patient outcome. It seems unlikely that such a simple procedure will be adequate in the majority of patients. However, it does illustrate the importance of controlling medical care, offering regular appointments with a single physician and trying to achieve more positive and instructive aims.

Clinical experience suggests that the following measures are beneficial to patients and their families and can improve the satisfaction of those treating them and reduce health costs:

- Simplify medical care
- Treat any co-existing psychiatric disorder
- Clarify psychosocial problems and tasks with patient
- Offer to see patient at regular, fixed intervals
- Negotiate treatment plan with patient and family
- Consider written contact

2.3 Undifferentiated Somatoform Disorder

Many patients present with features of more than one somatoform disorder and do not easily fit into any single category. Those who have multiple unexplained symptoms but are sub-threshold for somatization disorder are included in undifferentiated somatoform disorder. The very large size of this group indicates the failure of the sub-classification of somatoform disorders.

This has resulted in proposals for the abridged somatization disorder described above or for a new category of multi-somatoform disorder (Kroenke et al. 1997). Alternatively, it may be argued that categorical definition is unhelpful and that it would be preferable to look in terms of dimensions of symptoms, behaviour, attributions and disability (Janca et al. 1995). The lack of validity for a large residual category means that treatment must follow the general principles outlined earlier in this chapter, and expectations of success would be appropriately modest in those with chronic multiple or recurrent symptoms.

2.4 Somatoform Autonomic Dysfunction

Unlike DSM-IV, ICD-10 has a sub-category for those disorders whose symptoms appear to arise from overactivity of the autonomic nervous system. There is a sub-division according to the bodily system which is predominantly affected. Four features should be present:

1. Persistent troublesome symptoms and autonomic arousal
2. Additional symptoms referred to a specific organ or system
3. Preoccupation and distress about the possibility of serious disorder that does not respond to reassurance
4. No evidence of significant disturbance of structure or function of the organ or system

Although it is alleged that this syndrome can be differentiated from generalized anxiety disorder and from hypochondriasis, clinical experience suggests that this is not so. Epidemiological studies have found it to be a relatively large but residual category corresponding to some extent to the residual categories of DSM.

2.5
Persistent Pain Disorder

This category (referred to as "pain disorder associated with psychological factors" in DSM-IV) depends upon preoccupation with pain which causes significant distress or impairment and which appears to be due in substantial part to psychological factors (King and Strain 1996). The criteria require no judgement about the aetiological role of the psychological factors and is therefore a diagnosis of exclusion of other primary psychiatric disorder and of medical explanations. It describes a single diagnostic presenting problem rather than a psychiatric syndrome in any conventional sense.

While the category is an admission of our current lack of understanding, it may have some operational value in clinical practice. However, it separates consideration of patients who attend pain clinics for predominantly musculo-skeletal complaints from all other somatoform symptoms.

Treatment of chronic pain must reflect the multiple factors which are likely to be important in the formulation of aetiology in any individual patient. Treatment plans should include management and underlying physical problems, treatment of associated depression or other primary psychiatric disorder and cognitive behavioural measures to modify pain behaviour and perception and reduce disability. It is generally important to involve close relatives. More severe and chronic pain problems may require multidisciplinary treatment programmes (Wilson and Gil 1996; Main and Benjamin 1995).

2.6
Current Status

Review of the overall category and of its component sub-categories, suggests that somatoform disorders should be seen as a provisional operational classification with uncertain boundaries and sub-categories (Janca et al. 1995). They are in many ways residual descriptive categories for those who do not have conspicuous psychological symptoms. Present specific categories derive from historical formulations as much as from research and clinical experience. A worryingly high proportion of subjects are classified in non-specific or undifferentiated categories.

Increasing awareness of the inadequacies of our current concepts of somatoform disorder should not lead us to underestimate the value of having had provisional operational criteria applicable to very large numbers of people suffering disabling clinical problems. The present criteria, recent research and very considerable clinical experience point the way to improved classification. It is probable that this should be multidimensional so as to cover the type of clinical syndrome, the nature of underlying beliefs and health anxiety, course and severity.

Appendix A: ICD-10 Research Diagnostic Criteria for Hypochondriacal Disorder

A. *Either* (1) persistent belief, of at least 6 months duration, of presence of a maximum of two serious physical disorders; *or* (2) persistent preoccupation with a prejudged definite disfigurement.
B. Preoccupation with the belief and symptoms causes persistent distress and interference with personal functioning.
C. Persistent refusal to accept medical reassurance.
D. Exclusion of schizophrenia and mood disorder.

Appendix B: ICD-10 Research Diagnostic Criteria for Somatization Disorder (Abbreviated)

A. A history of at least 2 years' complaints of multiple and variable physical symptoms that cannot be explained by any detectable physical disorders.
B. Preoccupation with the symptoms causes persistent distress and leads the patient to seek repeated (three or more) consultations or sets of investigations ... In the absence of medical services ... there must be persistent self-medication or multiple consultations with local healers.
C. There is persistent refusal to accept medical reassurance that there is no adequate physical cause for the physical symptoms.
D. There must be a total of six or more symptoms from a list, with the symptoms occurring in at least two separate groups. [List of fourteen symptoms categorized as gastrointestinal, cardiovascular, genitourinary, skin and pain]

3 References

Barsky A (1992) Amplification, somatization, and the somatoform disorders. Psychosomatics 33: 28–34

Barsky AJ, Klerman GL (1983) Overview: hypochondriasis, bodily complaints, and somatic styles. Am J Psychiatry 140: 273–283

Barsky AJ, Wyshak G, Klerman GL (1990) Transient hypochondriasis. Arch Gen Psychiatry 47: 746–752

Barsky AJ, Wyshak G, Klerman GL (1992) Psychiatric comorbidity in DSM-III-R hypochondriasis. Arch Gen Psychiatry 49: 101–108

Bianchi GN (1971) Origins of disease phobia. Aust NZ J Psychiatry 5: 241–257

Cioffi D (1991) Beyond attentional strategies: a cognitive-perceptual model of somatic interpretation. Psychol Bull 109(1): 25–41

Clark DM, Salkovskis PM, Hackman A, Wells A, Fennell M, Ludgate J, Ahmad S, Richards HC, Gelder M (1998) Two psychological treatments for hypochondriasis. Br J Psychiatry 173: 218–225

Côté G, O'Leary T, Barlow DH, Strain JJ, Salkovskis PM, Warwick HMC, Clark DM, Rapee R, Rasmussen SA (1996) Hypochondriasis. In: Widiger TA, Frances AJ, Pincus HA, Ross R, First MB, Wakefield Davis W (eds) DSM-IV sourcebook, vol 2. American Psychiatric Association, Washington DC, pp 933–947

Escobar JI, Rubiio-Stipec M, Canino G, Karno M (1989) Somatic symptom index (SSI): a new and abridged somatization construct. Prevalence and epidemiological correlates in two large community samples. J Nerv Ment Dis 177: 140

Fink P (1995) Psychiatric illness in patients with persistent somatisation. Br J Psychiatry 166: 93–99

Fink P, Sorensen L, Engberg M, Holm M, Munk-Jorgensen P (1999) Somatization in primary care. Prevalence, health care utilization, and general practitioner recognition. Psychosomatics 40: 330–338

Goldberg D (1995) Epidemiology of mental disorders in primary care settings. Epidemiol Rev 17(1): 182–190

*Goldberg D, Gask L, O'Dowd T (1989) The treatment of somatization: teaching techniques of reattribution. J Psychosom Res 33: 689–695

*Gureje O, Simon GE, Üstün TB, Goldberg DP (1997a) Somatization in cross-cultural perspective: a World Health Organization study in primary care. Am J Psychiatry 154(7): 989–995

Gureje O, Üstün TB, Simon GE (1997b) The syndrome of hypochondriasis: a cross-national study in primary care. Psychol Med 27: 1001–1010

Guthrie E (1995) Treatment of functional somatic symptoms: psychodynamic treatment. In: Mayou R, Bass C, Sharpe M (eds) Treatment of functional somatic symptoms. Oxford University Press, Oxford, pp 144–160

Guze SB, Cloninger CR, Martin RL, Clayton PJ (1986) A follow-up and family study of Briquets' syndrome. Br J Psychiatry 149: 17–23

Howard LM, Wessely S (1996) Reappraising reassurance – the role of investigations. J Psychosom Res 41(4): 307–311

Janca A, Isaac M, Costa e Silva JA (1995) World Health Organization international study of somatoform disorders – background and rationale. Eur J Psychiatry 9(2): 100–110

Kellner R (1991) Psychosomatic syndromes and somatic symptoms. American Psychiatric Press, Washington DC

Kenyon FE (1964) Hypochondriasis: a clinical study. Br J Psychiatry 110: 478–488

King SA, Strain JJ (1996) Somatoform pain disorder. In: Widiger TA, Frances AJ, Pincus HA, Ross R, First MB, Wakefield Davis W (eds) DSM-IV sourcebook, vol 2. American Psychiatric Association, Washington DC, pp 915–931

**Kirmayer LJ, Robbins JM (1991) Three forms of somatization in primary care : prevalence, co-occurence, and sociodemographic characteristics. J Nerv Ment Dis 179: 647–655

Kirmayer LJ, Young A, Robbins JM (1994) Symptom attribution in cultural perspective. Can J Psychiatry 39: 584–595

*Kroenke K, Mangelsdorff D (1989) Common symptoms in ambulatory care: incidence, evaluation, therapy and outcome. Am J Med 86: 262–266

Kroenke M, Spitzer RL, deGruy III FV, Hahn SR, Linzer M, Williams JBW, Brody D, Davies M (1997) Multisomatoform disorder: an alternative to undifferentiated somatoform disorder for the somatizing patient in primary care. Arch Gen Psychiatry 54: 352–358

Lee S (1997) A Chinese perspective of somatoform disorders. J Psychosom Res 43(2): 115–119

Lipowski ZJ (1988) Somatization: the concept and its clinical application. Am J Psychiatry 145: 1358–1368

Main CJ, Benjamin S (1995) Psychological treatment and the health care system: the chaotic case of back pain. Is there a need for a paradigm shift? In: Mayou R, Bass C, Sharpe M (eds) Treatment of functional somatic symptoms. Oxford University Press, Oxford, pp 214–230

Marple RL, Kroenke K, Lucey CR, Wilder J, Lucas CA (1997) Concerns and expectations in patients presenting with physical complaints. Arch Intern Med 157: 1482–1488

Mayou R, Bryant B, Sanders D, Bass C, Klimes I, Forfar C (1997) A controlled trial of cognitive behavioural therapy for non-cardiac chest pain. Psychol Med 27: 1021–1032

Mayou RA, Seagroatt V, Goldacre M (1991) Use of psychiatric services by patients in a general hospital. Br Med J 303: 1029–1032

**Mayou RA, Bass C, Sharpe M (1995) Treatment of functional somatic symptoms. Oxford University Press, Oxford

Murphy MR (1990) Classification of the somatoform disorders. In: Bass C (ed) Somatization: physical symptoms and psychological illness. Blackwell, Oxford, pp 10–39

Pennebaker JW, Watson D (1991) The psychology of somatic symptoms. In: Kirmayer LJ, Robbins JM (eds) Current concepts of somatization: research and clinical perspectives. American Psychiatric Press, Washington DC, pp 21–35

Phillips KA, Hollander E (1996) Body dysmorphic disorder. In: Widiger TA, Frances AJ, Pincus HA, Ross R, First MB, Wakefield Davis W (eds) DSM-IV sourcebook, vol 2. American Psychiatric Association, Washington DC, pp 949–960

Salkovskis PM, Clark DM (1993) Panic disorder and hypochondriasis. Adv Behav Res Ther 15: 23–48

Sharpe M, Bass C (1992) Pathophysiological mechanisms in somatization. Int Rev Psychiatry 4: 81–97

Sharpe M, Hawton K, Simkin S, Surawy C, Hackman A, Klimes I, Peto T, Warrell D, Seagroat V (1996) Cognitive behaviour therapy for the chronic fatigue syndrome: a randomised controlled trial. Br Med J 312: 22–26

**Simon GE, Von Korff M (1991) Somatization and psychiatric disorder in the NIMH epidemiologic catchment area study. Am J Psychiatry 148: 1494–1495

Smith GR, Hanson RA, Ray DC (1986) Patients with multiple unexplained symptoms. Arch Intern Med 146: 69–72

Smith G.R, Rost K, Kashner TM (1995) A trial of the effect of a standardized psychiatric consultation on health outcomes and costs in somatizing patients. Arch Gen Psychiatry 52: 238–243

Speckens AEM, Van Hemert AM, Spinhoven P, Hawton KE, Bolk JH, Rooijmans HGM (1995) Cognitive behavioural therapy for medically unexplained physical symptoms: a randomised controlled trial. Br Med J 311: 1328–1332

Speckens AEM, Van Hemert AM, Bolk JH, Rooijmans HGM, Hengeveld MW (1996) Unexplained physical symptoms:

outcome, utilization of medical care and associated factors. Psychol Med 26: 745–752

Tyrer P, Fowler-Dixon R, Ferguson B, Kelemen A (1990) A plea for the diagnosis of hypochondriacal personality disorder. J Psychosom Res 34: 637–642

*Üstün TB, Sartorius N (1995) World Health Organisation. Mental illness in general health care. An international study. Wiley, Chichester

Von Korff M, Dworkin SF, Le Resche LL, Kruger A (1988) An epidemiologic comparison of pain complaints. Pain 32: 173–183

Warwick HMC, Clark DM, Cobb AM, Salkovskis PM (1996) A controlled trial of cognitive-behavioural treatment of hypochondriasis. Br J Psychiatry 169: 189–195

**Wessely S, Hotopf M, Sharpe M (1998) Chronic fatigue and its syndromes. Oxford University Press, Oxford

Wilson JJ, Gil KM (1996) The efficacy of psychological and pharmacological interventions for the treatment of chronic disease-related and non-disease-related pain. Clin Psychol Rev 16(6): 573–597

CHAPTER

8

S. Wessely

Neurasthenia

The first difficulty felt by anybody who enters on the study of neurasthenia is caused by the wide divergence of views held by the medical authorities both as to its nature and causation. (Cobb 1920)

1 Historical Development

Neurasthenia was originally described by the New York neurologist George Beard in 1869 (Wessely 1996), although some believe that the unfashionable midwest psychiatrist Van Deusen has an equal claim. No matter, over the next 30 years the concept enjoyed extraordinary, albeit not universal success across Europe and America. Bumke (1925) wrote that there was no instance in the history of medicine of a label having the impact of neurasthenia. Hundreds of journal articles and textbooks appeared between 1870 and 1900, and numerous doctors and clinics devoted themselves to the care of neurasthenic patients, not least because as an illness of the newly successful middle classes, it was associated with substantial income.

However, by the early part of the twentieth century, neurasthenia was in retreat. There were a number of reasons for this. Its pathological basis was called into question by the new science of neurophysiology, and it was no longer a socially exclusive illness. The interest shown by neurologists and physicians waned, and as more and more were convinced by the role of psychological factors, this interest was transferred to psychiatry, with the inevitable loss of prestige and revenue. Treatment shifted; in place of the much hyped "rest cure" (*Rastkur*) came the realisation that not only did enforced rest fail to cure, it probably was itself a cause of disability. The new psychiatric classifications of depression and anxiety gained credibility as rapidly as neurasthenia lost it (see Wessely et al. 1998).

It is ironic that neurasthenia has fallen into disrepute in the country of its birth, the United States, while it has taken hold in China and South Asia and has never left the former Soviet Union. Its history in China is particularly instructive. It was introduced into Chinese thinking probably in the 1920s, almost certainly due to the influence of American medicine. It was called *shenjing shuairuo* (nerve weakness) and was successful probably because the concept of depletion of nerve energy was compatible with the older tradition of Chinese medicine (Lee 1998). Its apogee was reached under the Communist regime, in the 1950s and 1960s. Seminal studies by Arthur Kleinman in the 1970s (Kleinman 1982) suggested that this was now the principal socially acceptable way for people to express psychological distress - hidden under a somatic disguise.

Since the end of the Cultural Revolution, the changing political and economic climate has, however, led to a decline in the use of the term "neurasthenia". Something similar can be observed in Japan, where the concept of neurasthenia was extremely popular until recently and was largely applied to patients who would now be diagnosed as anxiety disorders or hypochondriasis (Kitanishi and Kondo 1994). Another usage in Japan is as a euphemism for schizophrenia (Munakata 1989), since Japanese society has a particular abhorrence of serious mental illness.

2 The Neurasthenia Syndrome

What was, and what is, neurasthenia? The answer depends upon when and where the question is asked. A Victorian physician and follower of Weir Mitchell or Beard would answer that it is an organic disease of the nervous system, resulting from a system giving way under the pressures of modern life. A psychiatrist following the lead of Bleuler and Kraepelin would say it is a mild form of depression. A modern Western psychiatrist would hardly recognise the word, while a primary care doctor in the modern non-English speaking world would reply that it is a stress-related disorder characterised by fatigue but with little or no depression. Finally, some modern doctors would say it is an old term for the condition that we now call "chronic fatigue syndrome" (CFS) and whose physical aetiologies we are rediscovering after years of psychiatric neglect.

DSM-IV, the bible of American psychiatry, ignores neurasthenia and treats it as a "waste basket category", largely the result of failure to diagnose more conventional diagnoses. In contrast, ICD-10, more avowedly international, allows neurasthenia a place (F48.0). Confusion still remains over the status of "post-viral fatigue syndrome", which, under a virtually identical case definition, is listed as a neurological disorder (G93.3). Subsequent clarifications by the World Health Organization (WHO) made it clear this was a political compromise to allow all doctors to code for a fatigue syndrome, whether or not they considered it a psychiatric or neurological condition.

Neurasthenia is defined in ICD-10 (F48.0) as follows:

A. It must include one of the following:
   - Persistent and distressing complaints of feelings of exhaustion after minor mental effort
   - Persistent and distressing complaints of fatigue and bodily weakness after minor physical effort

B. At least one of the following symptoms must be present:
   - Muscle aches and pains
   - Dizziness
   - Tension headaches
   - Sleep disturbances

- Inability to relax
- Irritability

C. Symptoms are not relieved by rest, relaxation or entertainment
D. Minimum duration 3 months
E. Not associated with organic brain syndromes, mood, panic or generalised anxiety disorders

Hence there are many neurasthenias, but one thing every conceptualisation of neurasthenia has in common is that it is an illness of fatigue and fatigability. At the heart of the neurasthenia is the complaint of fatigue, precipitated by minor physical or mental effort. This fatigue is a subjective experience, one inaccessible to direct measurement. As such, it should be distinguished from the neurologist's use of fatigue as an objective failure to sustain force (as in myasthenia gravis) or the neuropsychologist's use of fatigue as a failure of mental performance. As the Victorians knew, and we are rediscovering, while a neurasthenic patient complains of abnormal physical and mental fatigue, objective tests of muscle or brain are essentially normal. Instead, what is being experienced is a sense of increased effort – that everyday tasks can be performed, but at the expense of an unpleasant sensation of increased physical or cognitive effort (Lawrie et al. 1997).

Neurasthenic patients clearly experience increased effort in everyday physical and mental tasks, reflected in sense of painful muscle exertion, and painful cognitive processing. This increased effort is not the result of increased neuromuscular or metabolic demands (a Victorian concept), nor does it result in any substantial decline in actual muscle or cognitive performance. The result is a mismatch between patients' evaluation of their physical and mental functioning and the external evidence of any consistent deficits. The basis of this disorder of effort must remain speculative, and the perception of effort is a complex topic, but I assume it is because the sufferer needs to devote more attention, or even energy, to processes that the rest of us perform automatically, be it muscular exertion or mental concentration (Lawrie et al. 1997).

3 Prevalence

Chronic fatigue is common – indeed, it is second only to headache in most community services. Wherever one looks, between 20% and 40% of the population answer yes to the question "do you feel tired all the time?". Women invariably out number men. Thus a U.K. study finding that 20% of men and 30% of women complain of being tired all the time, all day and every day for the past month, can be taken as typical of a large literature (Wessely 1995). Chronic fatigue is relatively unaffected by age – it is rare before puberty but is, if anything, more common in younger age-groups.

Because neurasthenia is defined as more than just chronic fatigue, it is less common. Exact prevalences are crucially dependent on definition. Hence on the Swedish island of Lundby, deservedly a mecca for psychiatric epidemiologists, the lifetime prevalence of "fatigue syndrome" was 33% for women and 21% for men. The point prevalence of neurasthenia in Zurich, as reported by Jules Angst, is lower, but still substantial. Again, women were affected more commonly than men. If more stringent criteria are used, the prevalence naturally falls, particularly when co-morbid psychiatric disorder is excluded (see below), but it is worth remembering that there is no study that reports that neurasthenia is less common than such familiar disorders as schizophrenia or obsessive compulsive disorder. Neurasthenia is a public health issue.

Perhaps the best estimates of the prevalences of pure neurasthenia comes from the WHO Primary Care Study (Ustun and Sartorius 1995). The prevalence of neurasthenia ranged from a high of 3.7% (Manchester) to a low of 0.3% (Athens and Bangalore). Whether or not these represent true differences or alternatively differences in illness behaviour, recognition and methodology cannot be answered; however, what is clear is that there are relatively large numbers of people presenting in primary care who fulfil the diagnostic criteria for neurasthenia – indeed, it is the third most common mental disorder encountered in that setting (Ustun and Sartorius 1995).

Simply because there is an operational definition of a condition gives no information on its nosological status. The principle criticisms of the neurasthenia concept come from two sources. The first is the relationship between neurasthenia and the symptom of fatigue. The second is the relationship between neurasthenia and other psychiatric disorders. As already described, wherever it has been studied, chronic fatigue emerges as one of the most common symptoms experienced in the community. Like all symptoms, it is dimensionally, and not categorically, distributed. Fatigue is part of the human condition, as is chronic fatigue. If we study the distribution of fatigue-related symptoms in the population, a normal distribution is found (Pawlikowska et al. 1994). There is no point of rarity, and no clear division between the normal and the abnormal. An appropriate analogy would be with blood pressure – we all know that there is no unique illness called "high blood pressure", nor a diastolic

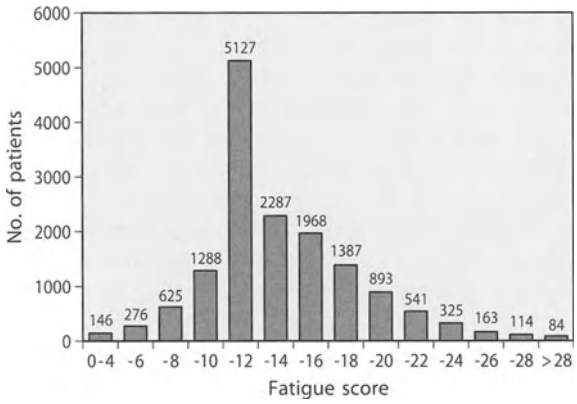


Fig. 1. The distribution of fatigue in the community. (From Pawlikowska et al. 1994)

reading which separates normal from abnormal. Instead, there is a gradient of risk, with no clear delineation between normal and high. Yet that does not invalidate the importance of high blood pressure, nor its association with morbidity, and the need for doctors to treat it. The same is true of fatigue/neurasthenia (Fig. 1).

4 Aetiologies

We have not really progressed since Kraepelin divided neurasthenia into congenital and acquired. Hence on the one hand we have a long series of studies claiming that neurasthenia results from some form of toxic insult, be it infection, trauma or chemical insult. Neurasthenic syndromes have been described in those recovering from Epstein Barr virus infection, people with prolonged influenza, those exposed to organic solvents or organophosphate pesticides, patients with mild head injuries and in Vietnam veterans possibly exposed to Agent Orange, among others. All of these suggest that neurasthenia may be a non-specific response to insult. Longitudinal studies in which data is collected prior to the onset of illness are rare and are confined to the field of post-infectious neurasthenia, but those that have been conducted suggest that prior psychological symptoms increase the vulnerability to subsequent neurasthenia after an infective trigger.

This model, in which pre-exposure to psychological or social variables explains some of the post-exposure outcome, is not unique to the relationship between insults such as viral infections and the development of neurasthenia/chronic fatigue. We found the same to apply to viral meningitis, in which the contribution of the precipitating infection was more substantial, but the combination of previous psychiatric disorder and viral meningitis was particularly associated with the onset of CFS. Similar results have been presented to explain the transition from acute to chronic back pain, from herpes zoster to post-herpetic neuralgia, from acute gastroenteritis to irritable bowel syndrome, from head injury to the post-concussional syndrome, from trauma to post-traumatic stress disorder (PTSD) and to explain the development of post-surgical fatigue (see Wessely et al. 1998).

5 Neurasthenia and Other Psychiatric Disorders

There are few psychiatric disorders that are not associated with chronic fatigue. It would be tedious to list all the studies that show that fatigue is a common, and often the most common, symptom of depressive disorders – it is simpler to state that I know of no exception. Fatigue is also part and parcel of all of the anxiety disorders, is frequent in eating disorders and can be associated with virtually any psychotic or organic mental disorder. Given that much the same comment can be made about the role of fatigue in medical disorders, it is for that reason that fatigue is not a subject which attracts much interest among doctors, who are aware of its non-specificity, and that the presence of fatigue is not helpful in the matter of diagnosis. It is interesting to note that the general public sees things very differently – for them, fatigue is a very important symptom indeed (Dohrenwend and Crandell 1970).

Given the importance of fatigue in neurasthenia and its ubiquity elsewhere, virtually every study notes the overlap between neurasthenia and the common psychiatric disorders. For example, in Chinese Americans, the annual prevalence of neurasthenia was 6.4%, but fell to 3.6% when co-morbid conditions were excluded (Zheng et al. 1997). In Australian primary care, the prevalence of fatigue syndrome was far higher (25%), but again two thirds were classified as cases of psychiatric disorder (Hickie et al. 1996). In the Zurich studies, 79% of the cases of neurasthenia were associated with depression and/or anxiety (Merikangas and Angst 1994). Much the same can be said of CFS – a considerable proportion of patients, especially in specialist settings, also fulfil criteria for recognised psychiatric disorders, but invariably somewhere between a third and a half do not. If we follow Occam's Razor and try not to create more categories than is necessary, it remains true that a category resembling neurasthenia is essential.

The relationship between neurasthenia and psychiatric disorders appears to be predictable, whether or not it is neurasthenia or CFS that is studied. In general, neurasthenic subjects experience more psychiatric

distress than normal subjects, but less than those with well-defined psychiatric disorders such as depression. On the other hand, they usually score as highly, and occasionally higher, on measures of somatic symptoms. Much the same applies to CFS.

Longitudinal studies confirm the overlap between neurasthenia and psychiatric disorders, but also show that the two are independent. Fatigue is certainly a well-known prodrome to depression (Dryman and Eaton 1991), but studies from London, Zurich and Australia suggest that "pure" fatigue syndrome is not particularly associated with subsequent depression; in other words the interchange between chronic fatigue/neurasthenia and depression is due to the large contribution made by co-morbidity – neurasthenia/fatigue and psychiatric disorder in the same person – but a pure neurasthenic syndrome is relatively stable and is not inevitably followed by depression.

The links between depression and neurasthenia are a partial explanation for the over-representation of women in studies of fatigue, neurasthenia or CFS. However, although one or two studies find the excess in women to be entirely accounted for by depression, most do not. Adjustment for mood disorder reduces, but does not eliminate, the excess in women, which is more likely to be related to some common social and biological confounders.

In conclusion, there is now persuasive evidence that neurasthenia is associated with, but not the same as, depression or anxiety. There is no evidence that it is a discrete disorder, however, with a clear division between normal and abnormal fatigability; it is instead a question of degree.

6
Chronic Fatigue Syndrome

Neurasthenia had largely vanished by the 1960s from Western medicine and psychiatry, and with it the interest in fatigue states. These were revived in the 1980s by the emergence in the English-speaking world of what became known as CFS, an illness to all intents and purposes indistinguishable from neurasthenia as originally proposed by Beard and Mitchell. Given the close overlap between the concepts, it is not surprising that 97% of those attending a CFS clinic in Wales fulfilled criteria for neurasthenia (Farmer et al. 1995). Likewise, a soon to be published comparison of neurasthenia and CFS in Chinese American and Causasian inhabitants of Los Angeles showed that neurasthenia in the ethnic Chinese and CFS in the Caucasians were strikingly similar, although the explanations each group gave for their illness could hardly be more different (Lin et al. 1996). Nevertheless, one should beware of equating all the diagnoses of neurasthenia with CFS; as currently used in South Asia, for example, neurasthenia is not the same as CFS and is more characterised by complaints such as headache, insomnia and tension (Lee 1998).

The emergence of CFS in the last decade has been accompanied by an upsurge of interest in the biological basis of the syndrome, an interest not seen when neurasthenia was seen as a solely "psychiatric condition". Much of this activity has not led to any consistent insights, but one area relevant to this contribution has been comparisons between the neurobiological associations of the major psychiatric disorders and those of CFS. These have emphasised the similarities between, for example, major depression and CFS, but also the differences. Hence the sleep abnormalities commonly found in severe depression are not normally encountered in the "pure" fatigue syndromes. Likewise, the well-known pattern of an overactive hypothalamic–pituitary–adrenal axis, with high cortisol levels which do not drop after dexamethasone challenge, has not been found in CFS, and instead a mild pattern of hypocortisolaemia has been reported by most, but not all, groups. As in major depression, the significance of these observations remains unclear, partly reflecting difficulties with sample choice and size and also the fact that major depression itself is a far from homogenous concept.

7
Management

When neurasthenia burst onto the European medical scene at the end of the nineteenth century, the first treatment proposed was the Weir Mitchell rest cure – sufferers were treated with diet, massage, rest and isolation from work and family. The readers of this volume may like to note that it was in Germany that the rest cure found its most ready acceptance. Playfair's book was available in German in 1883, only a year after its publication in English, and a German edition of Weir Mitchell was available by 1886, reviewed by Freud in the following year. Large numbers of "retreats", private clinics and rest homes appeared across Central Europe between 1880 and 1900 (see Wessely et al. 1998). It was financially vital to the neurologist, since, as one of them wrote in 1894 (Ziemssen 1894), the neurologist should not "undertake a thoroughgoing course of this sort of treatment unless in a private institution". Fortunately, the author continued, "we have in Germany an abundance of good private institutions." The rest cure became the most commonly used treatment for nervous disorder across Central Europe and America. It "provided the

raison d'être for the clinic, since isolation could not by definition be procured at home, nor could the expensive apparatus of electrotherapy". As Shorter (1990) points out, "physicians in these competitive, profit-making clinics were happy to comply with the patients' desire for face saving (organic) diagnoses, and made great use of such expressions as ... chronic fatigue and neurasthenia".

However, several reasons can be identified for the fall of the rest cure, and ultimately of the diagnosis neurasthenia itself. The illness was no longer the exclusive province of the rich and successful. The rest cure not only did not work, it seemed to make people worse, and even to condemn them to the life of a chronic invalid. When it did work, this was more due to the charisma of the doctor than to any inherent property of the treatment. Neurasthenia itself shifted from being seen as an organic disease of the central nervous system to a psychological disorder best managed by the new analytically inspired psychiatry. The rest cure disappeared.

It was revived under the auspices of the emergence of CFS in the 1980s. Popular and even some professional texts praised the virtues of rest as the best form of management and warned of the dangers of excessive exercise, where "excessive" could mean virtually any activity involving more than minimal effort. Common sense and a knowledge of physiology could have pointed out the dangers of this approach, but sadly it was several years before this indeed happened. In the next section, I shall outline our new thinking on the subject, another example of "back to the future".

8
Cognitive Behavioural Approaches to Chronic Fatigue Syndrome/Neurasthenia

In this section, I will outline in a little more detail what I see as the most promising practical approach to the problem of neurasthenia/chronic fatigue, drawing in particular on the cognitive behavioural model of symptoms and the use of cognitive behavioural therapy (CBT) in treatment.

The treatment of CFS has been a mess, although signs of a return of common sense are now visible. For many years, sufferers were given no advice or help except exhortations to rest, "live within your limits" and wait for the medical breakthrough. Not only was there no evidence that this is helpful, but such advice ignored a vast literature stretching back many years concerning its adverse effects. Perhaps in consequence, a series of outcome studies made gloomy reading (Joyce et al. 1997).

However, the situation may now be improving. Firstly, the rather dispiriting prognostic studies all refer to patients seen in specialist centres. In these studies, the strongest predictor of poor outcome was the strength of attribution of symptoms to a solely physical cause. However, we have shown that patients with CFS in primary care are not characterised by such intense illness beliefs and also do not correspond to the "Yuppie flu" stereotype. It is plausible that such patients have a far better prognosis. Secondly, outcome studies refer to the situation without treatment, perhaps because many suggested either that no treatment was possible or alternatively promoted ineffective treatments (Wilson et al. 1994).

Several groups have suggested an alternative model for understanding CFS (Wessely et al. 1991; Surawy et al. 1995; Sharpe et al. 1997). At the heart is the message that whatever triggers CFS may not perpetuate it. For example, an ordinary viral infection may *precipitate* fatigue which, for the majority of the population, is resolved when a normal recovery is made. However, on rare occasions, the presence of *perpetuating* factors (such as psychosocial stressors, rapid deconditioning, failure to rest adequately or concurrent depression) may delay or impede recovery. Fatigue then becomes chronic, persisting long after the departure of the original trigger and maintained by new variables. These include the following:

1. The effects of inactivity: Prolonged rest is effective in the short term, but counterproductive in the medium and long term. With the passage of time, more symptoms and greater fatigue will continue to occur at progressively lower levels of exertion. Inactivity therefore sustains symptoms and increases sensitivity to them. Several studies have shown that coping with symptoms by avoidance behaviours is associated with worse disability.
2. Inconsistent activity: Many sufferers that we encounter are not persistently inactive (although some end up that way). Instead, excessive or prolonged rest is followed by a burst of activity, which, compared to the preceding level of inactivity, is often "too much, too soon". This pattern may also be reinforced by the sense of frustration often encountered in sufferers and perhaps also by pre-existing personality and lifestyle factors. Many patients have attempted sudden increases in activity and find that they culminate in exhaustion, for which the inevitable response is further rest. This "stop and start" pattern means that, while extremes of disability are often avoided, sufferers are unable to build up a sustained level of recovery. This pattern often leads to the characteristic complaint of CFS sufferers that any activity must be "paid for" later by further pain and fatigue. Delayed fatigue

and myalgia are well-recognised physiological phenomena that occur between 24 and 48 h after any exertion in excess of a person's current (and not previous) fitness.

3. Illness beliefs and fears about symptoms: These can influence disability, mood and behaviour in any illness. In CFS, unhelpful and inaccurate illness beliefs, reinforced by occasionally ill-informed media coverage, include fear that any activity which causes an increase in fatigue is damaging or impossible, that "doing too much" causes permanent muscle damage, that CFS is due to a persistent virus or progressive immune disorder, that CFS is irreversible or untreatable and that rest is the only remedy.
4. Symptom focusing: Increased symptom focusing is also noted in CFS (Ray et al. 1995). Concern about the meaning and significance of symptoms (which are often interpreted as "warning signals") is heightened by the unpredictable nature of CFS. Increased concern leads to heightened awareness, selective attention and "body watching", which can then intensify both the experience and perceived frequency of symptoms, thereby confirming illness beliefs and reinforcing illness behaviour.
5. Emotional consequences: Of whatever cause, depression and anxiety are strongly associated with fatigue and muscle pain, impaired memory and concentration, and reduced activity.

I and many of my colleagues have suggested that CFS may be better understood (and hence treated) by focusing on these and other possible perpetuating factors and the many ways in which they interact in self-perpetuating vicious circles of fatigue, behaviour, beliefs and disability (Wessely et al. 1991; Sharpe and Chalder 1994; Surawy et al. 1995). This is illustrated in Fig. 2.

Such models have been associated with the development of two styles of rehabilitation. The first, based on a simple exercise avoidance/deconditioning model, is graded exercise. Recent clinical trials indeed confirm this is both safe and effective (Fulcher and White 1997; Wearden et al. 1998). However, the benefits of exercise are not related to increasing physical fitness and instead may be more linked to improved confidence and self control – in other words the gains may be behavioural rather than physiological. Exercise can also be difficult to "sell" to the patient and, if carried out in an over-enthusiastic or uncontrolled fashion, can be counterproductive – patients who exercise too aggressively will develop fatigue and myalgia, with a consequent reactivation of illness fears and loss of confidence in the physician.

The other treatment strategy is CBT. This does involve specific exercise regimes, but is a more collaborative and cautious method of exploring the role of conscious beliefs in determining behaviours and of collaborating on a series of behavioural experiments aimed at reducing the inconsistent activity patterns so characteristic of CFS, before embarking on a behavioural programme to reduce avoidance behaviours and gradually increase activities. Two randomised controlled trials confirmed the efficacy of this approach, which was also associated with better compliance than simple exercise programmes (Sharpe et al. 1996; Deale et al. 1997).

In conclusion, the exact form of rehabilitation for chronic fatigue/neurasthenia remains unclear and is probably not a crucial issue anyway. Exercise programmes are safe and effective, but compliance may be a problem. More behavioural programmes, emphasising consistency in activity rather than exercise per se, may be less of a high-risk strategy, particularly if combined with some form of cognitive intervention

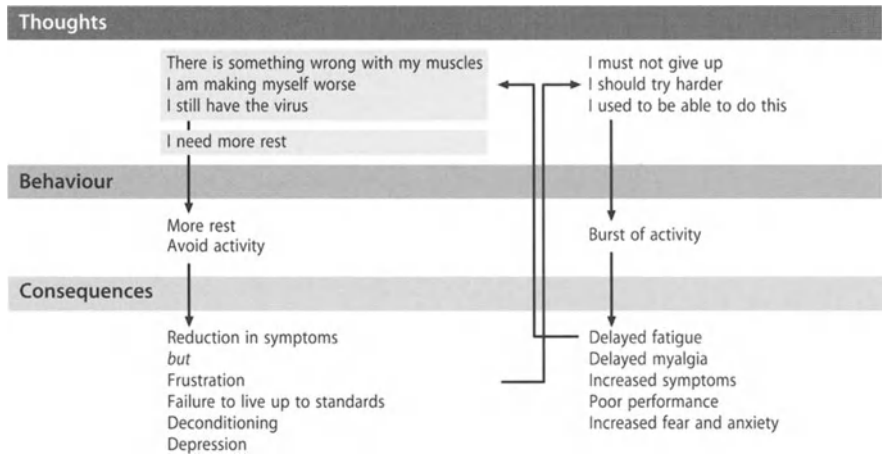


Fig. 2. Perpetuation of chronic fatigue syndrome

designed to provide a rationale for treatment. The aim is to reduce the stimulus-driven method of coping adopted by most chronic sufferers, in which the day to day, and even hour to hour, variation of symptoms guides what activity is undertaken, by a programme in which rest, sleep and activity are managed by timetable, and not in response to symptoms. Improvement, in which activity slowly increases and fatigue reduces, will happen, but may take weeks and months (Sharpe et al. 1996; Deale et al. 1997). Overall, providing psychological support, reducing disability and encouraging activity is the core of good clinical management (Wilson et al. 1994). In contrast, we can be clear that, of all the myriad of treatments proposed for chronic fatigue/neurasthenia, prolonged bed rest is one of the few that can be guaranteed not to work.

9 Drugs and Neurasthenia

The role of pharmacotherapy remains less clear (Wessely et al. 1998). At present it seems safe to assume that where depression appears to be a prominent feature, antidepressant are indicated. However, in the case of "pure" neurasthenia, the evidence is equivocal. Several good trials have failed to show benefit from selective serotonin re-uptake inhibitors. Whether or not this represents sample selection, or alternatively a defect in serotonin metabolism that is the opposite of that found in major depression, remains unclear. Clinical experience and some controlled trials from the related field of fibromyalgia suggest that tricylic antidepressants, given in doses not thought to be effective in patients with mood disorder, may help sleep, fatigue and pain and should be encouraged. Finally, there is preliminary evidence of benefit from monoamine oxidase inhibitors, something which also acts as a timely reminder of the overlap between neurasthenia and atypical depression.

10 Conclusion

I conclude that, despite all its difficulties, the concept of neurasthenia remains a valid one for psychiatrists and epidemiologists alike (Hickie et al. 1997). However, I doubt that there is a single cause of neurasthenia, any more than there is a single cause of high blood pressure. I doubt that one factor, be it psychosocial, viral or immunological, will satisfactorily explain anything other than a minority of cases. Instead, like all psychiatric disorders, it is also certainly multifactorial. New evidence suggests that chronic fatigue/neurasthenia has a genetic basis, which is not unexpected. There remains a strong clinical case for assuming that exposure to certain infections, such as the Epstein Barr virus, play a role in initiating symptoms. It is tempting to assume, although difficult to prove, that differences in the way in which the body deals with such agents and the ability to mount and control an efficient immune response may play a part.

Regardless of the results of further neurobiological inquiries, there can be no doubting the important role played by psychological, social and cultural factors in the aetiology of neurasthenia. Previous experience clearly increases vulnerability to fatigue syndromes. Whether or not coincident life stress also plays a part remains uncertain, but it would be surprising if it did not. Subsequent events certainly play a crucial role in determining the eventual outcome. The person's own emotional make-up and coping strategies are important, as are the reactions of family, friends and the medical profession. It is variations at this sociocultural level that account for the many different interpretations and usages of the term "neurasthenia" across the globe.

11 References

Bumke O (1925) Die Revision der Neurosenfrage. Munch Med Wochenschr 72: 1815–1819

Cobb I (1920) A manual of neurasthenia (nervous exhaustion). Balliere, Tindall and Cox, London

Deale A, Chalder T, Marks I, Wessely S (1997) A randomised controlled trial of cognitive behaviour versus relaxation therapy for chronic fatigue syndrome. Am J Psychiatry 154: 408–414

Dohrenwend B, Crandell D (1970) Psychiatric symptoms in community, clinic and mental hospital groups. Am J Psychiatry 126: 1611–1621

Dryman A, Eaton W (1991) Affective symptoms associated with the onset of major depression in the community; findings from the US National Institute of Mental Health Epidemiologic Catchment Area Program. Acta Psychiatr Scand 84: 1–5

Farmer A, Jones I, Hillier J, Llewelyn M, Borysiewicz L, Smith A (1995) Neuraesthenia revisited: ICD-10 and DSM-III-R psychiatric syndromes in chronic fatigue patients and comparison subjects. Br J Psychiatry 167: 503–506

Fulcher K, White P (1997) Randomised controlled trial of graded exercise in patients with chronic fatigue syndrome. Br Med J 314: 1647–1652

Hickie I, Koojer A, Hadzi-Pavlovic D, Bennett B, Wilson A, Lloyd A (1996) Fatigue in selected primary care settings: sociodemographic and psychiatric correlates. Med J Aust 164: 585–588

*Hickie I, Hadzi-Pavlovic D, Ricci C (1997) Reviving the diagnosis of neurasthenia. Psychol Med 27: 989–994

Joyce J, Hotopf M, Wessely S (1997) The prognosis of chronic fatigue and chronic fatigue syndrome: a systematic review. Q J Med 90: 223–233

Kitanishi K, Kondo K (1994) The rise and fall of neurasthenia in Japanese psychiatry. Transcult Psychiatr Res Rev 31: 137–143

Kleinman A (1982) Neurasthenia and depression: a study of somatization and culture in China. Cult Med Psychiatry 6: 117–190

*Lawrie S, MacHale S, Power M, Goodwin G (1997) Is the chronic fatigue syndrome best understood as a primary disturbance of the sense of effort? Psychol Med 27: 995–999

*Lee S (1998) Estranged bodies, simulated harmony, and misplaced cultures: neurasthenia in contemporary Chinese society. Psychosom Med 60: 448–457

Lin K, Cheung F, Zheng Y, Weiss M, Nakasaki G, Ren Y (1996) A cross cultural study of neurasthenia and CFS in LA. Xth World Congress of Psychiatry, Madrid, 23–28 August 1996, p 184

Merikangas K, Angst J (1994) Neurasthenia in a longitudinal cohort study of young adults. Psychol Med 24: 1013–1024

Munakata T (1989) The socio-cultural significance of the diagnostic label "neurasthenia" in Japan's mental health care system. Cult Med Psychiatry 13: 203–213

Pawlikowska T, Chalder T, Hirsch S, Wallace P, Wright D, Wessely S (1994) A population based study of fatigue and psychological distress. Br Med J 308: 743–746

Ray C, Jeffries S, Weir W (1995) Coping with chronic fatigue syndrome: illness responses and their relationship with fatigue, functional impairment and emotional status. Psychol Med 25: 937–945

Sharpe M, Chalder T (1994) Management of the chronic fatigue syndrome. In: Illis L (ed) Neurological rehabilitation. Blackwell, Oxford, pp 282–294

Sharpe M, Hawton K, Simkin S, Surawy C, Hackmann A, Klimes I, Peto T, Warrell D, Seagroatt V (1996) Cognitive behaviour therapy for chronic fatigue syndrome; a randomized controlled trial. Br Med J 312: 22–26

*Sharpe M, Chalder T, Palmer I, Wessely S (1997) Chronic fatigue syndrome: a practical guide to assessment and management. Gen Hosp Psychiatry 19: 195–199

Shorter E (1990) Private clinics in Central Europe 1850–1933. Soc Hist Med 3: 159–195

Surawy C, Hackmann A, Hawton K, Sharpe M (1995) Chronic fatigue syndrome: a cognitive approach. Behav Res Ther 33: 535–544

Ustun T, Sartorius N (eds) (1995) Mental illness in general health care: an international study. World Health Organization/ Wiley, Chichester

Wearden A, Morriss R, Mullis R, Strickland P, Pearson D, Appleby L, Campbell I (1998) A double-blind, placebo controlled treatment trial of fluoxetine and a graded exercise programme for chronic fatigue syndrome. Br J Psychiatry 172: 485–490

Wessely S (1995) The epidemiology of chronic fatigue syndrome. Epidemiol Rev 17: 139–151

Wessely S (1996) Neurasthenia and chronic fatigue. In: Porter R, Berrios G (eds) The history of psychiatry. Athlone, London, pp 509–532

Wessely S, Butler S, Chalder T, David A (1991) The cognitive behavioural management of the post-viral fatigue syndrome. In: Jenkins R, Mowbray J (eds) Postviral fatigue syndrome. Wiley, Chichester, pp 305–334

*Wessely S, Hotopf M, Sharpe M (1998) Chronic fatigue and its syndromes. Oxford University Press, Oxford

*Wilson A, Hickie I, Lloyd A, Wakefield D (1994) The treatment of chronic fatigue syndrome; science and speculation. Am J Med 96: 544–549

Zheng Y, Lin K, Takeuchi D, Kurasaki K, Wang Y, Cheung F (1997) A epidemiological study of neurasthenia in Chinese Americans in Los Angeles. Compr Psychiatry 38: 249–259

Ziemssen H (1894) Neurasthenia and its treatment. In: Clinical lectures on subjects connected with medicine and surgery. New Sydenham Society, London, pp 53–86

CHAPTER

9

M.J. Kelleher, H.S. Keeley, D. Chambers, P. Corcoran

Suicide

1 Introduction

It is true to say that suicide is a feature of the human condition and has been present, although usually as a rare phenomenon, in all societies and at all times. Suicide and parasuicide together may be referred to as "suicidal behaviour". Parasuicide or deliberate self-harm will be dealt with in the next chapter, but the two topics overlap considerably and, to avoid repetition, issues which may be relevant to both areas may only be dealt with in one.

2 Definition

Defining suicide is by no means as easy as it would first appear. Deaths may be divided into natural and unnatural, or internal and external causes, which is the division used for statistical purposes. How each individual death is categorised depends on many factors, including who is making the decision, in what circumstances, for what purposes, the subjective definition of the words used and how carefully intention is assessed. Many definitions have been used in the past, and the one given below is that recommended by the World Health Organization (WHO 1992):

Suicide is an act with a fatal outcome, which the deceased, with the knowledge and expectation of a fatal outcome, had himself planned and carried out with the object of bringing about the changes desired by the deceased.

The definition of suicide has three elements; *intention,* potential *lethality* and *fatality* of outcome. Only the last is certain but, even here, doubt may creep in if the person lives for a long time before eventually dying, e.g. on a life support machine. Lethality may also give rise to confusion, as in the case of someone who dies from an overdose of a small amount of medicine to which he or she is particularly sensitive. However, establishment of intention causes the greatest difficulty.

3 Legal Definitions and Statistics

A narrow legal definition of intention requires evidence in words, either written (as in a suicide note) or spoken to another, who is prepared to give evidence. If the coroner hears a case before a jury, he may seek such evidence. The decision made is to be in keeping with the criminal law, i.e. beyond reasonable doubt. As many cases cannot match this level of certainty, they are unsatisfactorily left as *undetermined* as to whether death was accidental or purposefully intended. This is the situation as it applies in England and Wales. Other jurisdictions use a balance of probabilities approach which, drawing on evidence from a number of sources, allows a verdict of suicide to be made for statistical purposes even in the absence of proof beyond reasonable doubt. In the Republic of Ireland, where this approach is adopted, the ratio of undetermined death to suicide has fallen to 0.04 for both genders. The male and female ratios in England and Wales have risen to 0.46 and 0.71, respectively (Kelleher et al. 1996). In an examination of 16 countries between the years 1974 and 1991, Neeleman and Wessely (1997) found that the ratio of undetermined death to suicide had increased in only two countries besides England and Wales, namely, Belgium and New Zealand. This situation has important implications for research and for meeting national health targets.

4 Variations in Suicide Rates

At present, about 1% of deaths in Britain and Ireland are due to suicide. In the younger age-groups, the relative proportion of these deaths is much higher because deaths due to other causes are so infrequent. In many Western countries, suicide and road traffic accidents are the most common causes of death in men aged 15–24 years (WHO 1997).

Few, if any, administrations over-estimate their suicide rates. The tendency is to under-estimate. This may occur at a local level where obvious suicides are sometimes misclassified in an attempt to spare the feelings of the bereaved. Defective bureaucratic procedure can also cause misclassification at a national level and at an international level, if the returning of statistics to the WHO is given a low priority. Variation in official recording practices is probably the most important single reason why suicide rates vary between countries and over time. Douglas (1967) questioned and doubted what could reliably be gleaned from such figures. However, most researchers today accept that official statistics are of value in estimating likely differences or trends in rates, especially if examined over time and with the proportions of alternative verdicts taken into account (Sainsbury and Jenkins 1982).

5
Religious and Civil Sanctions

There have been great variations in the religious attitude to suicide in different European countries. Christian churches have strongly condemned suicide since the time of Augustine, who, as a part of his ideological war against the Donatists, stated that the suicide of Judas was a more heinous crime than the betrayal of Christ. This contrasts with the Old Testament, which reports suicides in a very matter-of-fact way, without condemnation (Barraclough 1992). From the time of Augustine to the Reformation, the suicide rate in Europe is considered to have been very low. This was so among both Christians and Jews. Islam subsequently adopted the condemnatory Christian and Jewish attitude to suicide (Umri 1987). After the Reformation, the suicide rates began to rise, which has been attributed to the loss of an all-embracing religious purpose which "inspired all aspects of life" (Masaryk 1970). The effect of such influences can be seen in the official attitude to suicide. France decriminalised suicide at the time of the Revolution, but it was not decriminalised in the United Kingdom until 1963 and in the Republic of Ireland in 1993. However, long before these dates, civil and religious sanctions had been effectively abandoned.

The greatest probable effect of official condemnation, whether religious or civil, is to increase the morbidity and suffering of the bereaved as well as obfuscating the magnitude of the problem, thereby making research and prevention more difficult. It has been found that countries with religious sanctions against suicide return significantly lower rates of suicide to the WHO for both males and females. Countries with religious sanctions also return less complete data, strongly suggesting the likelihood of under-recording (Kelleher et al. 1998a). However, religion may have a genuinely inhibitory effect on the expression of suicide. Durkheim found that Catholics and Jews had lower rates of suicide than Protestants, which he explained in terms of *integration* (Durkheim 1952). More recently, Stack and Wasserman (1993) found that frequency of church attendance would appear to be the discriminating factor rather than religious allegiance itself. From a clinical and sociological point of view, in risk assessment, meeting others and forming social bonds is more important than the specific church that the person attends.

The effect of religion is as relevant to children as to adults. A recent South African study revealed that none of the Muslim children in a multi-denominational school had reported experiencing suicidal behaviour among family, friends or peers, whereas such experiences were reported by children from the other denominations, namely Hindu, Christian and traditional African (Pillay 1995).

6
Cultural Issues

Examination of the suicide rates of the 27 European countries which report mortality statistics to the WHO reveals that 23 of them experienced an increase in suicide over the period from 1960 to 1985. There were relatively large increases in Northwest and Central Europe, while rates decreased in Mediterranean countries (Diekstra 1989). This strongly suggests the influence of socio-cultural difference and change. The values associated with each culture may reflect differently on suicide, in that the strength of condemnation may vary.

European countries with a higher quality of life have higher suicide rates than countries that are not as stable economically and indeed have large areas characterised by poverty (Inglehart 1990). By quality of life, we refer not only to economics but also to psychological well-being as measured by "values and attitudes" studies (World Values Study Group 1994).

This may be explained in terms of a greater sense of community and stronger social bonds in countries characterised by widespread economic disadvantage. In plain terms, it is not as bad to be oppressed or depressed in areas with a strong sense of community, sharing the same type of problems, as in countries whose populations enjoy greater wealth and general well-being. The isolation and sense of hopelessness that accompanies marginalisation may be accentuated in such societies.

However, those who have claimed that the rising suicide rate is an inevitable by-product of industrialisation have yet to explain the recent reduction in the male suicide rate in Japan and Finland. Variations in rates are undoubtedly associated with different stages in the economic development of a society, but factors such as national, cultural and ethnic identity are also crucial in determining the rate of suicide.

7
Youth and Developmental Factors

Rates of suicidal behaviours show a distinct escalation at adolescence. Developmental changes that occur at this time partly explain how this may come about. Erikson (1963) describes this period as a stage concerned with the formation of identity. Physical,

sexual, social and emotional developments bring the adolescent further into the social arena of the adult with associated new responsibilities. From a cognitive perspective, cognitive errors and distorted thinking, as in Beck's concept of the triad of negative thoughts about self, others and the future, are central in the genesis of suicidal ideation. Developments in adolescent thought processes mean that they tend to think more in hypothetical terms and less in terms of concrete and immediate experience, i.e. they are less inclined to accept things as they are and more to consider how they could be. As the adolescent struggles through the developmental phases inherent in the formation of an identity, he or she may develop depressive symptoms, especially hopelessness, which has been found to be a powerful predictor of suicidal behaviour (Beck et al. 1985).

Adolescents have been described as being caught between two worlds, their need for independence conflicting with dependency needs and a desire to be part of a family. Richman (1986) has developed a family systems approach to the treatment of suicidal adolescents. This approach is based on the premise that disturbance in the family structure and an inability to accept change or contain crises promote suicidal acting out. This and other similar theories stress the importance of parental psychopathology and other, often unconscious, dynamics within a pathogenic family system that may result in a child being driven to self-destruction (Berman and Jobes 1991).

This is also a key time for risk-taking behaviour of all types and, in particular, experimentation with alternative life-styles and behaviour. Adolescents' thoughts also tend to be egocentric, and they believe that their intentions, motives, behaviour and personality are scrutinised by the people around them, which may lead them to rebel (Mussen et al. 1990). Youth subculture is associated with groups who reject mainstream opinions, embracing instead anti-establishment art forms and modes of expression. The mods, rockers and punks of previous decades were examples of subcultures whose deviance was reinforced by media portrayals and the creation of moral panic among other elements of society. The glorification of death in general and of suicide, in particular of iconic figures in these subcultures, may have a potentially destructive effect on their members. Media representations of youth culture suggest that negativity and ambivalence with regard to life and death is encouraged.

Homosexuals and those with gender identity disorders are subgroups at increased risk of suicide. A majority of suicide attempts by homosexuals occur during their youth, and gay youth are said to be two to three times more likely to attempt suicide than other young people, possibly as a result of peer rejection and social isolation. In the United States, they comprise 30% of completed youth suicide annually (Gibson 1989).

8
Age and Gender Variations

For most of the Western world, the female suicide rate is about one third that of males. There are a few countries, however, where the female rate actually exceeds the male – a significant example being the People's Republic of China (Canetto 1997). Ethnic and regional studies within different countries reveal other exceptions. For example, among the Maring people of Papua New Guinea, suicide is only associated with women, whereas among the Gainj of Papua New Guinea, it is confined to married women (Healy 1979). In the Deganga region of India, more than two thirds of suicides are women (Banerjee et al. 1990). Death from suicide among the Aguaruna of the Peruvian Amazon region is especially common among women and young men (Brown 1986). In England and Wales, mortality by suicide among immigrants of Indian origin is higher among women than men (Raleigh et al. 1990). When mortality data are examined by age and gender, we find additional exceptions to the preponderance of males among the casualties of suicide. For example, for those aged 15–24 years, female suicide mortality rates exceeded rates in males in several Asian and South American countries (Barraclough 1988). It is important to recognise that international comparisons of gender ratios can only be considered to be reliable if we assume that recording practices within each jurisdiction are consistent for both males and females. This may not be a valid assumption and should be borne in mind when weighing up the evidence.

Again, as with variation in overall international rates, socio-cultural influence is suggested by these gender variations. In most of these countries, women occupy a subordinate position in the social structure, with poor life chances. For young females, particularly in countries of low economic development, the tendency to turn to prostitution and use of illicit drugs is dangerously high. This becomes even more of a risk in countries where the sex industry has become accepted as economically necessary. The suicide rate of young females is, in turn, adversely affected.

The previous examples suggest that women in particular social and economic circumstances may be equally affected by suicide, but the fact remains that, in general, there are protective factors which are more relevant to women, particularly in the West. These factors include a tendency to prioritise family and child-related concerns, which has been found useful in

differentiating between suicidal ideators and parasuicides (Linehan et al. 1983).

Indeed, in keeping with the idea of protective factors for females, socio-cultural influence in the Republic of Ireland is suggested by a ratio of male to female suicide of 7:1 for 15- to 24-year-olds, and recent figures indicate that this ratio is increasing. This may be related to improved social conditions for young females in Irish society, which is reflected, for example, in the fact that in 1995/1996 there was a majority of females in third-level education for the first time in the history of the state.

Studies focusing on female participation in the labour force (FPLF) have made similar findings. It would appear that in countries where FPLF is relatively new and not readily accepted, male suicide rates are adversely affected. The explanation offered is that FPLF conflicts with traditional cultural values by undermining the cultural interpretation of masculinity. The point is also made that males are more likely to have their wife as sole confidante (McGrath et al. 1990; Veroff et al. 1981), and thus they suffer a sense of loss if their wife decides to participate in the labour force (Stack 1998).

Increased FPLF has implications for female suicide rates as well. Professional women tend to acquire the suicide rates of their male colleagues, perhaps a reflection of the increasing similarity of their roles (Boxer et al. 1995). In addition, women in the past were conservative in their choice of method of suicide, many dying as a result of overdose. In recent times, they are inclined to choose more violent methods such as hanging, which was so rare that in the past pathologists were instructed to regard such an event as murder until proven otherwise (M. Bolster, personal communication).

A further factor which may be seen as protective for females is the ability to cope with and accept emotions and feelings better than males. In association with cultural stereotyping, men may be less aware than women of their emotional distress. Even if they are aware, they may be less likely to express it either to themselves or to others, and this may have a fatal outcome. While this ability may have some physiological basis, it is encouraged and maintained by social reinforcement and the resultant phenomenon of the "self-fulfilling prophecy".

Age, as a correlate of suicide, is equally significant in suicide research. Recent years have seen a change in the age profile of suicide victims in most European countries. Traditionally, the old were more at risk than the young. The profile of the elderly, retired, socially isolated man who is precipitated into suicide by some kind of loss event, usually bereavement, still holds true to some extent (Kelleher et al. 1997a). However, many countries, including England and Wales (Charlton et al. 1992), Norway (Rettorstøl 1993) and Ireland (Kelleher et al. 1995), have seen a marked increase in suicide among young men in the last 30 years or so. In Ireland, the suicide rate in 15- to 24-year-old males increased by over 80% (from 13.5 to 24.5 per 100,000) between 1989 and 1996. As changes in recording practices have had a limited effect, the rise is most likely to be due to social and attitudinal changes. The exact identification of these changes is a different matter.

None of the specific explanations that have been postulated for the increase in Ireland appear convincing. Change in illness patterns is unlikely to be sufficient, because there is no evidence that psychological illnesses such as depression have increased sufficiently in either frequency or severity to explain the increase in young male suicide. The rise in substance abuse has been suggested as an explanation of this trend (Neeleman and Farrell 1997), but this appears to be unlikely when we consider the reported increase in rural areas (Dudley et al. 1992; Kelleher et al. 1997b), where hard drug abuse is less common (An Garda Síochána 1996). It should again be emphasised that the rise in suicide rates among young males is part of an international trend. Most industrialised countries, with the exception of Japan, are reporting such changes. The matter must therefore be viewed as being an international problem requiring cross-national, as well as local, research and co-operation, if it is to be adequately addressed.

9
Psychological Autopsy Studies

One of the best tools available for the investigation of suicide is the psychological autopsy study. It involves investigating the suicide victim's background by means of interviews with his or her close relatives, medical doctors and others who would have been in contact. In the case of young people, peers are often not contacted because of the fear of inducing modelling behaviour and because of difficulties with consent. Various such studies have been done, recently by Foster et al. (1997) and in the past by Barraclough et al. (1974), among others. These have identified 90% or more of suicides as being mentally ill. Depression is the commonest illness found and is present in about two thirds of cases. It is particularly common among women, present in over 80% of cases in an Irish study (M.J. Kelleher et al., unpublished work). Alcohol is related to both depression and suicide. Some history of alcohol abuse is present in over half of the male suicides in the above Irish study. It may be used in

moderate amounts in the run-up to the act of suicide, particularly by young men. The consequences of many years of heavy drinking predispose to suicide, particularly in middle-aged and older men.

10
Study of Suicide Notes

Another potentially useful tool to help us understand the psychological processes underlying a person's decision to die by suicide is the systematic study of suicide notes. While Schniedman (1996) acknowledges that the study of suicide notes provides valuable insights into the suicidal mind, he stresses that it is important that they be set in the context of the person's background and experiences. This is particularly important when we consider notes left by younger people who die by suicide, which tend to be more centred on the here and now and often show signs of "magical thinking". Posner et al. (1989) analysed the notes left by 17 individuals who died by suicide between the ages of 10 and 20 years from a psychoanalytical perspective and found that the most prominent themes were those of love, ambivalence and aggression turned inward. A reason that was often given for the suicide was that of a feeling of being a bad person.

11
Suicide and Physical Illness

There is evidence that certain physical illnesses carry a higher risk of suicidal behaviour than others. Much of this association is undoubtedly due to high co-morbidity rates between these illnesses and depression, especially with neurological conditions such as multiple sclerosis and parkinsonism, and has been documented with certain types of stroke. A high risk of suicide also occurs in conditions such as Huntington's chorea and the muscular dystropies, where there is a long, chronically progressive downward course of the illness with no available hope of cure. Suicidal behaviour also occurs in the context of a diagnosis of acquired immunodeficiency syndrome (AIDS) or related illnesses and is part of the complex physical and psychological changes that happen in this situation.

There are other less dramatic associations, however, and the presence of chronic pain, especially back pain, can be a factor in suicidal behaviour, either by itself or in combination with other stresses. It is often easy to forget to ask about depressive symptoms or to accept them as "natural" in circumstances where the person is chronically disabled. However, they are treatable and their alleviation can be a great relief to the patient. The presence of depressive symptoms is also associated with requests for euthanasia, and adequate support and treatment of the person may encourage them to re-examine their situation in a more positive light.

12
Suicide and Mental Illness

12.1
Depression

Depressive illness is, as stated above, the commonest diagnosis among women who kill themselves. Unlike men, many of these women will have been formally diagnosed and treated for their depression. The greater willingness on the part of women to identify and accept treatment for their psychological distress may be part of the reason why their rates have not risen in most countries. However, advances in medical treatment for depression up to now do not seem to have affected the overall rise in suicide rates in developed countries, although this situation may now be changing (Montgomery et al. 1995). In cases of suicide, it is common to find anxiety co-morbid with depression, and a subgroup of adolescent suicide cases has been described who show evidence of anxiety, perfectionism and distress at times of change (Shaffer 1988).

12.2
Schizophrenia

Among mental illnesses, one of the most devastating is schizophrenia. It is sometimes not fully appreciated that schizophrenia has a comparatively high association with suicide occurring in between 10% and 15% of patients. In fact, suicide is the commonest cause of death in this condition (Mortenson and Juel 1993). Early age of onset, maleness, severity of condition, negative symptoms, return of insight and loss of status are all contributing factors. The commonest correlate, however, is the concurrent presence of depression (Schuwall and Siris 1994). The increased use of new atypical anti-psychotic medications may be associated with a decrease in suicidal behaviour in this group.

12.3
Personality Disorder

Personality disorder, particularly when associated with aggressiveness and impulsivity, may culminate in

suicide. The commonest types of personality disorder associated with suicidal behaviour are those of dysocial, borderline and impulsive types (Kaplan and Sadock 1991). These types of personality problems are often associated with very dysfunctional backgrounds. The association between borderline personality disorder and a reported history of having been sexually abused in childhood has been noted in the past (Waller 1993). Indeed, it has been emphasised that a history of child sexual abuse should be regarded as one element in a matrix of adverse experiences that increase the individual's vulnerability to greater levels of psychopathology and suicidal behaviour (Mullen et al. 1993). The dynamics of this include both a need for and an inability to maintain supportive relationships associated with unrealistic demands and an explosive, disproportionate response when let down.

12.4 Substance Abuse

Drugs, whether prescribed or illicit, may have an association with depression, suicidal behaviour and suicide. Some hypotensive agents, including calcium channel blockers and some anti-parkinsonian drugs, may cause depression and give a sense of hopelessness. The use of street drugs, including ecstasy, amphetamines, cocaine and heroin, may be followed, on occasion, by parasuicidal acts and, less frequently, by suicide itself. Hawton et al. (1993) found that substance misuse is the best predictor of eventual suicide. A recent article suggests that the rise in young male suicide may be associated with higher levels of substance misuse (Neeleman and Farrell 1997).

12.5 Suicide Among Recently Discharged Psychiatric Patients

In recent years, it has been noted that there is a substantial increase in suicide among those recently discharged from psychiatric hospitals, greater in the first few weeks but remaining elevated for several months (Goldacre et al. 1993). Many factors may contribute to this, including worsening of illness, dissolution of support structures, inability to get or retain a job and return to previous bad habits such as non-compliance with treatment and abuse of drugs or alcohol. In general, the problem seems to lie in the inability to function adequately in society following a psychiatric admission. Studies indicate that suicides occurring both during hospital admission and in trial discharges and early post-discharge outpatient treatment are increasing, but are still relatively rare occurrences and as such are difficult to predict. More active treatment of high-risk patients may therefore help. Questions regarding procedures on discharge assessment will be addressed later.

13 Biochemical Theories

Many biological theories have been advanced in explaining suicidal behaviour. Among the proposed substrates, the most enduring links have been found with the brain neurotransmitter serotonin. Serotonin abnormalities have been linked to suicidality across psychiatric diagnoses (van Praag 1991). Abnormalities in the serotinergic system have been found in the frontal cortex of suicide victims (Mann et al. 1986). It has also been found that those suicide attempters who eventually died by their own hand had a markedly lowered level in their cerebrospinal fluid (CSF) of the serotonin by-product 5-hydroxy-indole-acetic acid (Nielsen et al. 1994). The link between impulsivity and aggression has been highlighted by, among others, Asberg et al. (1986) and Coccaro (1989), who suggested that reduced central serotonin activity might predispose to a lowered threshold for aggressive responses to noxious stimuli rather than for unprovoked aggressive behaviour. Suicide, especially by violent means, is viewed in this context as a form of auto-aggression (Brent et al. 1993). Abnormal platelet serotonin aggregation may identify a subgroup of suicide attempters at high risk of repeating and perhaps of eventual death by their own hand. This is only one potential marker, and many others, including the levels of cholesterol and the soluble receptor of the human lymphocyte growth promoter interleukin-2, have also been advanced (Engelberg 1992; Nassberger and Traskman-Bendz 1993).

14 Genetics

If there is a chemical basis to suicidal behaviour, then these are likely to be reflected in genetic differences between cases and controls. There is evidence that this is so for suicide from twin and adoption studies (Roy et al. 1995; Schulsinger et al. 1979; Wender et al. 1986). Family studies of suicide attempters also show that their relatives have an increased frequency of suicide attempts (Egeland and Sussex 1985; Brent et al. 1996). Candidate genes investigated so far have been those associated with serotonin, principally polymorphisms of the rate-limiting enzyme tyrosine hydroxylase (Abbar et al. 1995; Nielsen et al. 1994). On the basis

of a detailed review of the literature, Brent et al. (1994) concluded that there was a strong family relationship between the transmission of suicidal behaviours and the tendency to impulsive aggression. However, results so far have been conflicting, and it is not currently possible to determine the exact manner in which such biological factors might alter a person's predisposition towards suicidal behaviour. What is likely is that this research will not result in a simple blood test dictating a management choice, but rather that biological factors will be added to psychosocial variables, rather as family history is considered at present, when it comes to the assessment of future suicidal risk. Some researchers have attempted to form models incorporating biological, social and psychological factors in order to explain the process or processes.

15 Prevention

A major problem in suicide prevention is the non-usage of services by those at risk. This is particularly so in males, where one study found that only half had been seen by a general practitioner or a psychiatrist in the year before death compared to 86% of the females (M.J. Kelleher et al., unpublished work). It has been proposed that suicide awareness programmes in schools and colleges would help prevent suicide in the young. Opinions differ on this, however, some fearing that such programmes are at best ineffective and at worst might be associated with an increase in suicidal behaviour. The wisest approach in the present state of knowledge is one of case identification within schools, either through the better education of teachers in such endeavours or perhaps, though less convincing, through surveys of the student population. The effectiveness of postvention support programs in sensitive settings, such as schools, has become a contentious issue and great care is needed to prevent glorification of the victims, which may only encourage further modelling.

The Defeat Depression Campaign, instituted by the Royal College of Psychiatrists in the United Kingdom, is an attempt to increase public awareness of psychiatric illness in general and depression in particular. It may have, as one outcome, a reduction in suicide by those who attend general practitioners and medical services. This was the experience reported in the Gotland study, where general practitioners were given extra training in the signs and symptoms of depression (Rutz et al. 1989). In this case, the effect was only temporary; however, more recent research has suggested that advances in the pharmacological treatment of depression may result in a fall in rates of suicide (Isaacson and Bergman 1998).

Prevention, in the short-term at least, will be enhanced by limiting access to methods. Such limiting would include the barring of hand-guns and restrictions on gun licences in general; the change from coal to natural gas; the restriction of the availability of poisons; the limitations of over-the-counter purchase of medicines which may be harmful in overdose; and better prescribing habits by doctors.

16 Response

The aftermath of suicide may be devastating for relatives. Only in recent years has it been appreciated how great the trauma can be. Such bereaved people may need counselling, and the Canadian Task Force (1994) recommends that, to succeed, bereavement postvention programmes should be part of a comprehensive package that includes prevention and intervention components, affiliated with the existing mental health services. However, suicide survivors are no more likely to have a complicated or pathological bereavement reaction than others who were bereaved unexpectedly or tragically (Murray-Parkes 1987). Nevertheless, the grief response differs because of the social issues associated with a suicide death. A search for meaning and crisis of values is a common part of all bereavement reactions but especially occurs after a suicide. Another difficulty is the lack of social supports perceived by suicide survivors. Sometimes this is a real problem associated with stigma, but it can happen that the bereaved fail to identify potential sources of support, simply because they feel that the pain of their loss is too great to be helped by others. Self-help groups can be invaluable in this situation, and allied with professional involvement, they can help the bereaved to integrate their experiences into a comprehensive framework.

Death by suicide is usually perceived as violent, regardless of the method used, and the act may have been directed aggressively against the bereaved person. Survivors can have difficulty with their feelings of anger and guilt that may be denied or displaced onto services or people who are seen as having failed to prevent the death. Such reactions can also lead to the development of secrets or myths that distort communication within the family and set the seeds for later difficulties, including further suicidal behaviour. The other, most striking, reaction seen with suicide survivors relates to their need to search for clues to the reason for the suicide. This preoccupation with the

question "why" may never completely leave the bereaved. If there is a strong element of denial, then the survivor may look for evidence that the death was not a suicide. Sometimes illness supervenes, requiring treatment in its own right. The prescription of medication, however, is not a panacea and should be limited to specific indications. It may, perhaps rightly, be construed by the bereaved as a distancing or a medicalisation of what are often normal features of the grief process.

Certain groups of survivors have particular needs and reactions. For example, children may be "protected" from the truth about a suicide, but this may lead them to develop disturbing fantasies about death and the deceased. They may also be unable to express their guilt, anger and worries. These worries often centre around the idea that they are somehow responsible for the death. They should be encouraged to attend funeral rituals and to discuss their feelings in an open manner, using age-appropriate language. Adolescents can react to a suicide with intense feelings of guilt, and this can impair their development of self-esteem, leading to difficulties, e.g. with substance abuse. They are also very vulnerable to modelling behaviour and may decide, through identification with the deceased, that they are "fated" to die in the same manner. People with learning disabilities also form a discrete group whose needs may not be adequately recognised and met, especially by care staff, who may not recognise the contribution of bereavement to behavioural problems (Hollins and Esterhuyzen 1997).

Sometimes, professional carers are sorely affected (Grad 1996). Many, especially general practitioners who are more isolated and those in a psychotherapeutic relationship with the deceased, have significant bereavement reactions which they may find difficult to express. Each should be given an opportunity to discuss and effectively deal with their responses. Nothing should be said or done that would unnecessarily add to their distress. However, if the suicide occurs within a service, the matter must be investigated. This should be done compassionately and efficiently. The scape-goating of individuals ought not to be the aim, although, sometimes, blame may have to be attached where duty clearly is not honoured. More frequently, the system itself requires review so that similar events might be avoided in the future.

In the aftermath of suicide, evidence may have to be given before a coroner's court. This is traumatic for both the survivors and the professional involved. For the survivors, it is often the first occasion on which they will have had contact with the court system. Unfamiliarity with the system and fear of ridicule may prevent them from participating properly in the process that would otherwise provide an opportunity to clarify some outstanding issues. For the professionals, reports have to be written and sometimes evidence given in court. Preparation for this should be a mandatory part of training for all health care workers. Despite the legalities, suicide is, ultimately, a matter of choice and, as stated earlier, no longer a criminal offence. By the same token, those who look after the suicidal should not have to work in fear of being branded professionally negligent in the event of a suicide, provided they have taken reasonable care. The prevention of suicide should not be seen as carrying the same burden of responsibility as a death directly caused by medical neglect.

17
Euthanasia and Physician-Assisted Suicide

The moral dilemma regarding physician-assisted suicide (PAS) and euthanasia is being increasingly debated. In euthanasia, the doctor ends the patient's life, whereas in PAS, the patient ends his or her own life with either advice or medications provided by the physician. In the Netherlands, where the practice is illegal, but not prosecuted provided specific criteria are fulfilled, women are more likely to choose euthanasia and men to die by PAS (van der Maas et al. 1996).

An unresolved question is whether euthanasia as a way of dying will become more popular in the years ahead. For a short time, the Northern Territories of Australia legalised PAS. The Australian Senate repealed this law in March 1997. In July 1997, the American Supreme Court ruled that citizens do not have a positive right to demand assisted suicide (Churchill and King 1997). A survey of the 51 member countries of the International Association for Suicide Prevention disclosed that in 12 countries, out of 49 respondents, active euthanasia was reported as occurring and in 29 countries, passive euthanasia occurred (Kelleher et al. 1998b). While some regard the distinction between active and passive euthanasia as being blurred or non-existent, there is an important clinical difference between discontinuing a treatment or failing to start one (passive euthanasia) and giving an injection calculated to stop the heart (active euthanasia). Psychiatrists individually and their governing bodies collectively will, in time, have to address the implications of these difficult subjects. In the Netherlands, it is now acceptable practice to terminate life on the grounds of psychological pain and suffering. It is not acceptable anywhere else.

18 Conclusion

Suicidal behaviour, whether fatal or non-fatal, is a fact of life that must not be glamorised, ignored, stigmatised or underestimated. Everyone is a potential victim, either directly or indirectly through contact with a suicide or suicide attempter. There already exists an international body of researchers who are working together in an attempt to better understand the phenomenon of suicide. The ultimate aim of this research is the formulation of effective prevention programmes that will reduce the suicide rate internationally or at the very least halt the almost universal increase in suicide.

Any problem faced by society must be brought up from the depths of social stigmatisation in order for the key issues to be dealt with on local, national and international levels. It is easy to turn a blind eye to the adversities of problems such as suicide, but as long as individuals and societies do so, a clear strategy for prevention will prove difficult to execute. The important issue is to remain focused and to continue to cooperate in the international attempt to prevent suicidal behaviour.

19 References

Abbar M, Courtet P, Amadeo S, Caer Y, Mallet J, Baldy-Moulinier M, Castelnau D, Malafosse A (1995) Suicidal behaviours and the tryptophan hydroxylase gene. Arch Gen Psychiatry 52: 846–849

An Garda Síochána (1996) Annual report. Irish Government Publications Office, Dublin

Asberg M, Nordstrom P, Traskman-Bendz L (1986) Cerebrospinal fluid studies in suicide: an overview. Ann NY Acad Sci 487: 243–255

Banerjee G, Nandi D, Nandi S, Sarkar S, Boral GC, Ghosh A (1990) The vulnerability of Indian women to suicide: a field study. Ind J Psychiatry 32: 305–308

*Barraclough B (1988) International variation in the suicide rates of 15–24 year-olds. Soc Psychiatry Psychiatr Epidemiol 23: 75–84

*Barraclough B (1992) The Bible suicides. Acta Psychiatr Scand 86: 49–64

*Barraclough B, Bunch J, Nelson B, Sainsbury P (1974) A hundred cases of suicide: clinical aspects. Br J Psychiatry 125: 355–373

*Beck AT, Steer RA, Kovacs M, Garrison B (1985) Hopelessness and eventual suicide: a ten-year prospective study of patients hospitalised with suicidal ideation. Am J Psychiatry 142: 559–563

Berman AL, Jobes DA (1991) Adolescent suicide: assessment and intervention. American Psychological Association, Washington, DC

Boxer PA, Burnett C, Swanson N (1995) Suicide and occupation: a review of the literature. J Occupat Environ Med 37(4): 442–452

Brent DA, Johnson B, Bartle S, Bridge J, Rather C, Matta J, Connolly J, Constantine D (1993) Personality disorder, tendency to impulsive violence, and suicidal behaviour in adolescents. J Am Assoc Child Adolesc Psychiatry 32: 69–75

Brent DA, Perper JA, Moritz G, Liotus L, Scheers J, Balach L, Roth C (1994) Familial risk factors for adolescent suicide: a case control study. Acta Psychiatr Scand 89: 52–58

Brent DA, Bridge J, Johnson BA, Connolly J (1996) Suicidal behaviour runs in families: a controlled family study of adolescent suicide victims. Arch Gen Psychiatry 53: 1145–1152

Brown MF (1986) Power, gender and the social meaning of Aguarana suicide. Man 21: 311–328

Canadian Task Force on Suicide (1994) Suicide in Canada: update of the Report of the Task Force on Suicide in Canada. Ministry of National Health and Welfare

*Canetto SS (1997) Gender and suicidal behaviour: theories and evidence. In: Maris RW, Silverman MM, Canetto SS (eds) Review of suicidology. Guilford, New York, pp 138–167

Charlton J, Kelly S, Dunnell K, Evans B, Jenkins R, Wallis R (1992) Trends in suicide deaths in England and Wales. Popul Trends 69: 10–16

Churchill LR, King NMP (1997) Physician-assisted suicide: euthanasia or the withdrawal of treatment. Br Med J 315: 137–138

Coccaro E (1989) Central serotonin and impulsive aggression. Br J Psychiatry 155[Suppl 8]: 52–62

Diekstra RFW (1992) Epidemiology of suicide: aspects of definition, classification and preventive policy. In: Crepet P, Ferrari G, Platt S, Bellini M (eds) Suicidal behaviour in Europe: recent research findings. Libbey, Rome, pp 15–44

Douglas J (1967) The social meanings of suicide. Princeton University Press, New Jersey, pp 163–231

Dudley M, Walters B, Kelk N, Howard J (1992) Youth suicide in New South Wales: urban/rural trends. Med J Aust 156(2): 83–88

Durkheim E (1952) Suicide: a study in sociology (translated by Spaulding J, Simpson G). Routledge and Kegan Paul, London, pp 208–209

Egeland JA, Sussex JN (1985) Suicide and family loading for affective disorders. JAMA 254: 915–918

Engelberg H (1992) Low serum cholesterol and suicide. Lancet 339: 727–729

Erikson E (1963) Childhood and society. Imago, London

*Foster T, Gillespie K, McClelland R (1997) Mental disorders and suicide in Northern Ireland. Br J Psychiatry 171: 447–452

Gibson P (1989) Gay male and lesbian youth suicide: report of the US Secretary's Task Force on Youth Suicide, vol 3. US Government Print Office, pp 110–141

Goldacre M, Seagroatt V, Hawton K (1993) Suicide after discharge from psychiatric in-patient care. Lancet 342: 283–286

Grad O (1996) Suicide: how to survive as a survivor. Crisis 17(3): 136–142

*Hawton K, Fagg J, Platt S, Hawkins M (1993) Factors associated with suicide after parasuicide in young people. B Med J 306: 1641–1644

Healey C (1979) Women and suicide in New Guinea. J Soc Anal 2: 89–107

Hollins S, Esterhuyzen A (1997) Bereavement and grief in adults with learning disabilities. Br J Psychiatry 170: 497–501

Inglehart R (1990) Culture shift in advanced industrial society. Princeton University Press, New Jersey, pp 244–245

Isaacson G, Bergman U (1998) Does increased use of antidepressants reduce suicide rates? Presented at the 7th European Symposium on Suicide and Suicidal Behaviour, Ghent, Belgium, 10th September

Kaplan HI, Sadock BJ (1991) Synopsis of psychiatry, 6th edn. Williams and Wilkins, Baltimore

Kelleher MJ, Corcoran P, Keohane B (1995) Suicide, road traffic and cancer deaths among the young in Ireland. Ir Med J 88(3): 96–98

Kelleher MJ, Corcoran P, Keeley HS, Dennehy J, O'Donnell I (1996) Improving procedures for the recording of suicide statistics. Ir Med J 89(1): 14–15

Kelleher MJ, Keohane B, Corcoran P, Keeley HS (1997a) Elderly suicides in Ireland. Ir Med J 90(2): 72–74

Kelleher MJ, Keeley HS, Corcoran P (1997b) The service implications of regional differences in suicide rates in the Republic of Ireland. Ir Med J 90(7): 1–7

Kelleher MJ, Chambers D, Corcoran P, Williamson E, Keeley HS (1998a) Religious sanctions and rates of suicide worldwide. Crisis 19(2): 78–86

Kelleher MJ, Chambers D, Corcoran P, Keeley HS, Williamson E (1998b) Euthanasia and related practices worldwide. Crisis 19(3): 109–115

Linehan M, Goodstein JL, Nielsen SL, Chiles JA (1983) Reasons for staying alive when you are thinking of killing yourself: the reasons for living inventory. J Consult Clin Psychol 51(2): 276–286

Mann JJ, Stanley M, McBride A, McEwen BS (1986) Increased serotonin and β-adrenergic receptor binding in the frontal cortices of suicide victims. Arch Gen Psychiatry 43: 954–959

Masaryk TG (1970) Thomas G Masaryk: suicide and the meaning of civilisation (translated by Weist WB, Batson RG). University of Chicago Press, Chicago, pp 126

McGrath E, Keita GP, Strictland B, Russo NF (1990) Women and depression. American Psychological Association, Washington, DC

Montgomery SA, Dunner DL, Dunbar GC (1995) Reduction of suicidal thoughts with paroxetine in comparison with reference antidepressants and placebo. Eur Neuropsychopharmacol 5: 5–13

Mortensen PB, Juel K (1993) Mortality and causes of death in first admitted schizophrenic patients. Br J Psychiatry 163: 183–189

Mullen PE, Martin JL, Anderson JC, Romans SE, Herbison GP (1993) Child sex abuse and mental health in adult life. Br J Psychiatry 163: 721–732.

Murray-Parkes C (1998) Bereavement: studies of grief in adult life. Penguin, London

Mussen PH, Conger JJ, Kagan J, Huston AC (1990) Child development and personality. Harper and Row, New York, pp 584

Nasberger L, Traskman-Bendz L (1993) Increased soluble interleukin-2 receptor concentrations in suicide attempters. Acta Psychiatr Scand 88: 48–52

Neeleman J, Farrell M (1997) Suicide and substance misuse. Br J Psychiatry 171: 303–304

Neeleman J, Wessely S (1997) Changes in classification of suicide in England and Wales: time trends and associations with coroners' professional backgrounds. Br J Psychol Med 27: 467–472

Nielsen DA, Goldman D, Virkkunen M, Tokola R, Rawlings R, Linnoila M (1994) Suicidality and 5-hydoxyindoleacetic acid concentration associated with a tryptophan hydroxylase polymorphism. Arch Gen Psychiatry 51: 34–38

Pillay BJ (1995) Perceptions of suicide: a study at a primary school. In: Schlebusch L (ed) Suicidal behaviour; proceedings of the third South African conference on suicidology. University of Natal, Durban, pp 163–171

Posner JA, La Haye A, Chiefetz PN (1989) Suicide notes in adolescence. Can J Psychiatry 34: 171–176

Raleigh VS, Bulusu L, Balarajan R (1990) Suicides among immigrants from the Indian sub-continent. Br J Psychiatry 156: 46–50

Rettorstøl N (1993) Suicide: a European perspective. Cambridge University Press, Cambridge, pp 75–91

Richman J (1986) Family therapy for suicidal people. Springer, Berlin Heidelberg New York

Roy A, Segal NL, Sarchiapone M (1995) Attempted suicide among living co-twins of twin suicide victims. Am J Psychiatry 152(7): 1075–1076

*Rutz W, von Knorring L, Walinder J (1989) Frequency of suicide on Gotland after systematic postgraduate education of GPs. Acta Psychiatr Scand 80: 151–154

Sainsbury P, Jenkins JS (1982) The accuracy of officially reported suicide statistics for purposes of epidemiological research. J Epidemiol Commun Health 36: 43–48

*Schniedman ES (1996) The suicidal mind. Oxford University Press, New York

Schulsinger F, Kety SS, Rosenthal D, Wender PH (1979) A family study of suicide. In: Schou M, Stromgren E (eds) Origin, prevention and treatment of affective disorders. Academic, London, pp 277–287

Schuwall M, Siris SG (1994) Suicidal ideation in post psychotic depression. Compr Psychiatry 35: 132–134

*Shaffer D (1988) The epidemiology of teenage suicide: an examination of risk factors. J Clin Psychol 49[Suppl 9]: 36–39

Stack S (1998) The relationship of female labour force participation to suicide: a comparative analysis. Arch Suicide Res 4(3): 249–261

Stack S, Wasserman I (1993) The effect of religion on suicide ideology: an analysis of the networks perspective. J Sci Stud Religion 31(4): 457–466

Umri J (1987) Suicide or termination of life. Islam Compar Law Q 7(2): 136–145

van der Maas PJ, van der Wal G, Haverkate I, de Graaff CLM, Kester JGC, Onwuteaka-Philipsen BD, van der Heide A, Bosma JM, Willems DL (1996) Euthanasia, physician-assisted suicide and other medical practices involving the end of life in the Netherlands, 1990–1995. N Engl J Med 335: 1699–1705

*van Praag HM (1991) Serotinergic dysfunction and aggression control. Psychol Med 21: 15–19

Veroff J, Duvan E, Kulka R (1981) The inner Americans: a self portrait from 1957–1976. Basic Books, New York

Waller G (1993) Sexual abuse and eating disorders: borderline personality disorders as a mediating factor. Br J Psychiatry 162: 771–775

Wender PH, Kety SS, Rosenthal D, Schulsinger F, Ortmann J, Lunde I (1986) Psychiatric disorders in the biological and adoptive families of adopted individuals with affective disorders. Arch Gen Psychiatry 43: 923–929

WHO (1992) The ICD classification of mental and behavioural disorders: clinical descriptions and diagnostic guidelines (CDDG). World Health Organization, Geneva

WHO (1997) World health statistics annual. World Health Organization, Geneva

World Values Study Group (1994) World values survey, 1981–1984 and 1990–1993 (computer file) ICPSR version. Institute for Social Research (producer)/Inter-university Consortium for Political and Social Research (distributor), Ann Arbor

CHAPTER

10

M.J. Kelleher, H.S. Keeley, M. Lawlor,
D. Chambers, C. McAuliffe, P. Corcoran

Parasuicide

1 Introduction

Having discussed suicide in the previous chapter, the present chapter addresses the related topic of parasuicide, which is being increasingly recognised as a major public health problem. Parasuicide is one of the strongest risk factors for both repetition and eventual suicide and, as such, the topic requires serious consideration. Given that the suicide attempt rate is at least ten to 15 times higher than the actual suicide rate (Böhme 1994), the scale of the problem has important implications for the provision of health services and thus needs to be addressed as a priority in its own right. Parasuicide has been identified as one of the Health For All targets by the World Health Organisation (WHO Regional Office for Europe 1985).

2 Definition and Nomenclature

Because of the overlap between those who attempt and those who complete suicide, there is a problem of definition and nomenclature in suicidology.

Stengel (1952) identified epidemiological differences between people who died by suicide and the survivors of an apparent suicidal act. Stengel believed that suicidal intent was essential in both groups, so that those who survived were failed suicides (Stengel and Cook 1958). Subsequently, this notion that a degree of suicidal intent was evident in all suicidal behaviour was contested. Kessel and Grossman (1965) proposed the terms "deliberate self-poisoning" and "deliberate self-injury" to replace "attempted suicide" as it was recognised that "most attempted suicides had performed their act in the belief that they were comparatively safe". Although these terms assumed that a desire to die was not present, they did imply that the behaviour was not accidental. This terminology of deliberate self-poisoning and deliberate self-harm was rejected by Kreitman and Philip (1969), as it neglected the "very real association that exists between attempted suicide and completed suicide". Kreitman and Philip (1969) introduced the concept of "parasuicide" to describe "an event in which the patient simulates or mimics suicide, in that he is the immediate agent of the act which is actually or potentially harmful to himself". Kreitman stated that this type of patient was not usually addressing himself to the task of self-destruction. These terms, i.e. deliberate self-harm and parasuicide, are sometimes used to describe a similar population in clinical practice. In this chapter, we refer to parasuicide as defined in the tenth version of the International Classification of Diseases (World Health Organisation 1992):

> An act with nonfatal outcome, in which an individual deliberately initiates a non-habitual behaviour that, without intervention from others, will cause self-harm, or deliberately ingests a substance in excess of the prescribed or generally recognised therapeutic dosage, and which is aimed at realising changes which the subject desired via the actual or expected physical consequences.

The definition includes acts that are interrupted before the actual self-harm occurs, but excludes acts of self-harm by people who for whatever reason are unable to understand the meaning or consequences of their action. This term was introduced because of the difficulties associated with the retrospective assessment of the intent of a particular attempt and the need for an all-inclusive general term for epidemiological studies and statistical analysis. It encompasses a diverse population which includes a small but significant proportion of failed suicides who are usually older and more often choose potentially fatal methods. From a clinical point of view, it is often both possible and necessary to distinguish between a failed suicide and a minor episode of deliberate self-harm with low intent. This distinction is one of the most important functions of the psychiatric assessment and carries both service and treatment implications.

3 Epidemiology of Parasuicide in Europe

No country in the world has national figures for parasuicide. In 1986, the WHO Regional Office for Europe appointed a Working Group to consider the possibility of establishing a collaborative European study of attempted suicide. The WHO/EURO Multicentre Study of Parasuicide was proposed and started in 1989 in 16 research centres across Europe. Given the scale of the problem and the logistics involved in national-level studies, it was decided to adopt a multicentre approach whereby the research would be carried out in defined catchment areas. At present, there are 26 centres involved, and this number continues to grow as new centres are established in Eastern Europe.

The study covers two broad areas of research – the monitoring of trends in the epidemiology of parasuicide (the Monitoring Study) and follow-up investigations with a view to predicting future suicidal

behaviour (the Repetition Prediction Study). The Monitoring Study has four main purposes (Platt et al. 1992):

1. To use local case-registers to monitor parasuicide in defined catchment areas
2. To estimate the true incidence of medically treated parasuicide using standardised definitions and case-finding criteria
3. To identify sociodemographic risk factors significantly associated with parasuicide
4. To ascertain variations in patterns of treatment

As a prerequisite to the study, the catchment areas had to be described in detail. This work identified wide variation between the centres on three levels: (1) the structure and provision of health services, (2) social, cultural and demographic factors and (3) the extent to which they are representative of their respective countries (Bille-Brahe et al. 1996). Despite these limitations, the use of standardised methodologies across a large number of centres (4,591,000 total population) has meant that the WHO/EURO Study constitutes the largest pool of specific data about suicide attempts in the world.

Figure 1 indicates the average person-based suicide attempt rates from the participating centres for 1989–1992. There is wide-ranging fluctuation in rates between the centres. While the centres from Southern Europe are associated with low rates, it is difficult to make further generalisations. Based on these data, an "average" European rate of parasuicide for persons aged over 15 years would be 140 and 193 per 100,000 for males and females, respectively (Schmidtke et al. 1994b).

Average age-specific rates for the period 1989–1992 are shown in Fig. 2. For each age-group, the female rate exceeds that of the males. Whereas the highest rate among females is for those aged 15–24 years, the male rate peaks among those aged 25–34 years.

For the period 1989–1993, a total of 22,655 episodes made by 17,486 people were recorded. The commonest method used was self-poisoning (80% of females and 64% of males). Over 40% of the individuals had made previous attempts, and an average of 15% made further attempts during the study period. The female to male ratio, previously thought to be in the region of 2:1, has been diminishing over the years (Schmidtke et al. 1994a). In Ireland, it has been close to parity for a number of years, whereas the Helsinki Centre reports rates in males that are slightly higher than those in females (Lawlor et al. 2000).

Important differences emerged between the attempters and the local general populations. The category with the highest proportion of attempters over the whole study period was that of single people. The rates of being single do diminish with increasing age, as might be expected, but 23% of those older than 25 years were divorced or separated. As a corollary, an

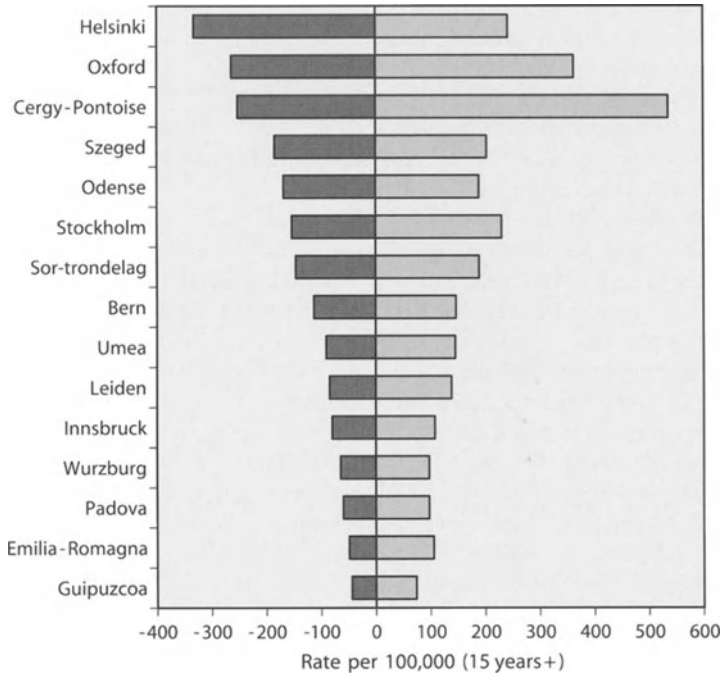


Fig. 1. Average person-based suicide attempt rates for the participating centres, 1989–1992. *Dark boxes*, males; *light boxes*, females. (Reprinted from Kerkhof et al. 1994 with the editor's permission)

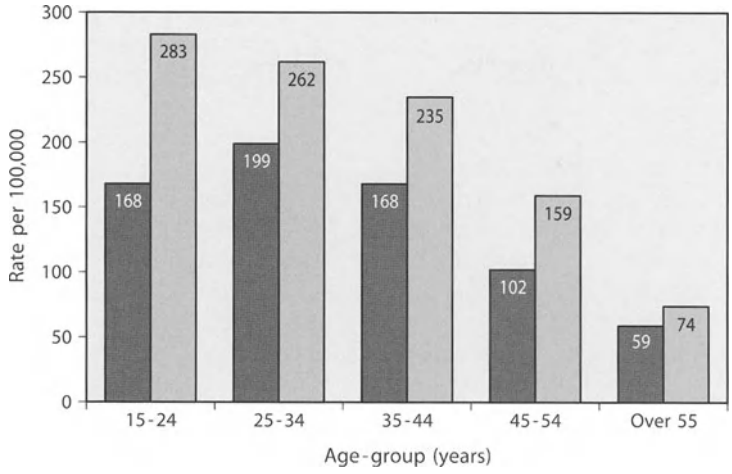


Fig. 2. Average age-specific parasuicide rates, 1989–1992. Data from the WHO/EURO Parasuicide Monitoring Study participating centres, 1989–1992. *Dark boxes*, males; *light boxes*, females. (Kerkhof et al. 1994)

average of 27% of the males and 20% of the females were usually living alone. Another interesting result was that 9% of the males and 6% of the females had changed their living conditions shortly before the attempt. This was usually a change from a socially stable living situation to living alone or in an institution (Schmidtke et al. 1994b).

4 Social Variables

It has been indicated that there is a strong independent association between suicidal behaviour and unemployment (Platt 1984). Compared with the general population, suicide attempters more often belong to the social categories associated with social destabilisation and poverty (Schmidtke et al. 1996). The rate of parasuicide in a particular area is socially determined and is greatest in urban areas of social deprivation (Morgan et al. 1975; Daly et al. 1986; Kelleher et al. 1996). There are many inter-related sets of disadvantages in these high parasuicidal areas, including overcrowding, poor education, high unemployment rates and low rates of ownership of domestic property. However, the best predictor of the rate of parasuicide in an area is the local unemployment rate which, of course, is correlated with the other measures of disadvantage (Kelleher et al. 1996).

The length of educational training required to obtain full-time employment has increased in Western Europe, so that a third-level qualification is required in many cases. This inflation of the credentials required in order to work contributes to the difficulties encountered by early school-leavers. Remaining in education can be protective against suicidal behaviour (Kelleher 1998a). Data from the WHO/EURO Monitoring Study indicates that 57% of parasuicides had only a low level of formal education, whereas only 12% had completed education to the highest level (Schmidtke et al. 1996). Given the relationship between unemployment and suicidal behaviour, the effect of educational disadvantage greatly increases the likelihood that a person will be unemployed for most of their potentially productive lives and therefore be at greater risk of engaging in suicidal behaviour. It is important to recognise that the impact of being out of work is not just due to material deprivation. The unemployed, especially married men, are denied the opportunity to undertake roles which are deemed appropriate by society and are excluded from valued categories of experience associated with employment. Social support, whether emotional understanding or practical aid, will help to reduce distress, but it is important that it is not viewed as a panacea. In the absence of measures intended to increase the individual's sense of control over their own life, social interventions may have the effect of reducing the patient's autonomy even further, which has negative results for both those giving and receiving aid (Whelan 1993).

5 Personal Risk Factors

While the precipitant for a suicide attempt is often an inter-personal difficulty, contrary to previous opinions, parasuicidal patients have had multiple personal

handicaps with which to cope (Keeley et al. 1999). In general, parasuicidal patients also have poor coping skills, which would make them more vulnerable to the effects of potential stressors (Linehan et al. 1987). A person who attempts suicide tends to give a story of many problems building up, one upon another, leading to the perception that they cannot be resolved. These may include a strong family history of mental illness and addictive behaviour, childhood stresses, including physical and sexual abuse, poor education and leaving school early, especially if male. A recent study which followed up a cohort of deliberate self-poisoners in Ireland (Kelleher et al. 1999) found that deliberate self-poisoning was associated with high levels of individual psychopathology, as shown in Table 1. Personal psychopathology interacts with the social pressures, including relationship difficulties and loss events, leading to a perceived absence of any choice other than suicidal behaviour.

Furthermore, a sense of powerlessness increases distress both by its demoralising effect and also by inhibiting adequate coping. There are indications from the European data presented above that a reduction in social stability, e.g. being thrown out of home, is associated with suicide attempts. The absence of supportive relationships and overwhelming social stresses are negative prognostic signs. A diagnosis of schizophrenia or substance misuse can be seen as indicating a high risk of repetition, especially for men (Roy et al. 1984; Deykin and Buka 1994). Other authors have attempted, with varying degrees of success, to form models incorporating biological, social and psychological factors in order to explain the process of suicidal behaviour (Maris et al. 1992). One possible model, based on Irish data, is shown in Fig. 3.

Table 1. Personal psychopathology of deliberate self-poisoners (Kelleher et al. 1999)

Diagnosis	Males (*n*)	(%)	Females (*n*)	(%)
Psychoses	3	4	7	6
Personality disorder	29	38	31	28
Alcohol abuse	53	70	20	18
Substance abuse	21	28	12	11
Other	9	12	13	12
Adjustment reaction	7	9	35	32
Affective disorder	25	33	54	49
Anxiety states	3	4	1	1
Affective disorder with anxiety states	5	7	12	11

5.1 Biological Factors

There is also evidence that there is a chemical basis to suicidal behaviour. Much of this information is dealt with in the suicide chapter, but a brief synopsis will be given here. One of the most robust findings in biological psychiatry is that discovered by Asberg et al. (1976), who found that those suicide attempters who eventually died by their own hand had a markedly lowered level in their cerebrospinal fluid (CSF) of the serotonin by-product 5-hydroxy-indole-acetic-acid (Asberg et al. 1986; Nielsen et al. 1994). This was felt to correlate with Mann's finding of abnormalities in the serotinergic system in the frontal cortex of suicide victims (Mann et al. 1986). These results are also reflected in more recent genetic work. A detailed review of the literature by Brent et al. (1994) concluded that there was a strong family relationship between the transmission of suicidal behaviours and the tendency to impulsive aggression. Family, twin and adoption studies (Roy et al. 1995; Schulsinger et al. 1979; Wender et al. 1986) indicate that the relatives of suicide attempters have an increased frequency of suicide attempts (Egeland and Sussex 1985; Brent et al. 1996).

5.2 Suicidal Ideation

Because suicidal thoughts, at a superficial level at least, may occur in many young people, they cannot be considered either statistically abnormal or reflecting mental illness in the absence of other evidence. One study from an Australian university found that 62% of students reported suicidal ideation, while 7% reported past attempts (Schweitzer et al. 1995). Other studies also indicate high rates of ideation among students (Mishara et al. 1976; McKey et al. 1993; McAuliffe 1998). It is important, however, to distinguish between a person's consideration of self-harm as *an* option – which may enhance their sense of control and freedom – and a person who is perturbed and feeling constricted, who views self-harm as the *only* option (Shneidman 1996).

Thus suicidal ideation becomes significant in young people when acting on these thoughts seems to be the only solution. It is at this point that there is a serious risk of attempting suicide. Across diagnostic categories, where ideation is concomitant with high levels both of hopelessness about the future and helplessness in coping with it, suicide risk is particularly high. Also

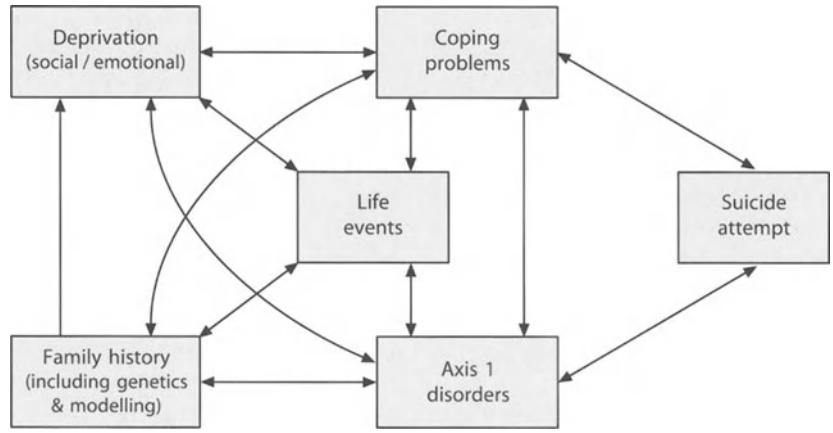


Fig. 3. Social and familial pathways to suicidal behaviour

in the absence of psychopathology, ideation with a plan is an important risk factor for completed suicide. Brent et al. (1993) found that adolescent suicide victims were significantly more likely to indicate suicidal ideation with a plan in the week prior to their deaths. In fact, when compared with demographically matched controls, 77% of the suicide victims in their study were known to have had ideation compared with none of the control group. Ideation is also a key precedent to parasuicide. In one prospective community study of adolescents (Andrews and Lewinsohn 1992), 87% of those who had a history of attempted suicide also reported prior suicidal ideation.

Linehan et al. (1987) argue that suicide ideators as a control group are "clinically significant" because of the following:

1. Suicides and parasuicides emerge from a population of suicide ideators.
2. Clinicians have to distinguish between patients who express suicidal intention but do not go on to parasuicide, and those who will engage in parasuicidal behaviour.

In summary, although there are problems in extrapolating from suicide ideators to parasuicides, ideation provides a useful preliminary measure of suicide intent and risk. In terms of conversion, where suicidal thoughts progress from conceptualising suicide as one solution to the only option, planning may become a logical step and therefore enquiring about preparation for the act remains a useful assessment tool.

5.3 Psychiatric Illness

Psychiatric diagnoses were collected in five centres of the WHO/EURO Monitoring Study. These clinical diagnoses were made by psychiatrists using Chap. V of the ICD-9 system of classification. Only 33% of attempters were given a diagnosis in these centres. It is unclear why the remaining 67% were not diagnosed. Obviously, cases that did not meet the ICD-9 criteria were not accorded a formal psychiatric diagnosis, but it is likely that variation in access to psychiatric assessment was a major factor. Twenty-one per cent of both sexes were given the diagnosis of "adjustment disorder or abnormal reaction to stress". This indicates that the most relevant precipitant to the episode was a response to an external stress, rather than personal psychopathology. The next most common diagnosis, for men, was abuse of substances, including alcohol, at 18%; the second commonest diagnosis for females was "neuroses and personality disorders without neurotic depression" at 22% and "neurotic depression" at 21% (Schmidtke et al. 1994b).

For some attempters, the prognosis is abysmal, with many repeating the behaviour (Kelleher et al. 1999). A total of 15%–20% of those who have a history of deliberate self-harm will repeat the behaviour (De Moore and Robertson 1996), and about 1% of adults go on to take their lives within 12 months of an attempt (Hawton and Fagg 1988). The psychiatric diagnoses for those who had made more than one suicide attempt in the WHO/EURO study (Schmidtke et al. 1994b) differed from the first-evers. More male repeaters had a diagnosis of addiction (23% vs. 14%), and there were more diagnoses of schizophrenia (14% vs. 4.5%). The most significant difference for females was that "adjustment disorder" was diagnosed in 30% of first-evers and in 17% of repeaters.

Given that a significant number of parasuicides (especially males) present with a diagnosis of "addiction", it is important to examine the link between suicidal behaviour and alcohol misuse. Alcohol dependence plays a causal role in 25% of all suicide cases

(Murphy and Wetzel 1992). Among adolescents, alcohol misuse has been shown to predict seriousness of suicidal attempts (Beautrais et al. 1996) and completed suicide (Shaffer et al. 1996). Alcohol may lower brain serotonin, thereby disposing to suicide by reducing impulse control (Sellers et al. 1992). Fombonne (1998) demonstrated a linkage between the rapid rise in young male suicidal behaviour and alcohol misuse over a 20-year period in a sample of London adolescents. The link between alcohol usage, altered serotonin neurotransmission and suicidal behaviour may be an important mechanism in the induction of parasuicide.

The previous chapter discussed the relationship between suicide and mental illness, in particular the relationship between depression, schizophrenia, personality disorder, substance abuse and suicide. Again, it is worth stating that the use of street drugs such as ecstasy, amphetamines, cocaine and opiates may be followed by parasuicide. This may involve the serotonergic brain system (Steele 1994). There is a strong relationship between severity of substance misuse (in terms of the drug chosen and method of administration) and risk of non-fatal suicidal behaviour (Garrison 1993). Substance misuse is associated with a seven-fold greater risk of attempting suicide (Deykin and Buka 1994). It is also more likely to lead to a more lethal outcome in suicidal behaviour, as more violent methods are chosen (Neeleman and Farrell 1997). According to the WHO/EURO study, the relationship between personality disorder and parasuicide is most pronounced in females, especially those who repeat. Most studies dealing with personality disorder and suicidality show a high degree of these disorders in people who commit self-destructive acts, particularly the borderline and impulsive subtypes. It is also likely that personality disorder interacts with other comorbid factors, such as substance misuse, depression and alcoholism (Casey 1989).

Although suicide is the leading cause of death in schizophrenia, with lifetime estimates of 10%, it may not be commonly realised that 40%–55% of individuals with schizophrenia attempt suicide (Roy et al. 1984). In those people with schizophrenia who attempt suicide, males are as common as females (Wilkinson and Bacon 1984). Less is known about the specific risk factors associated with suicidal behaviour in schizophrenia, as most research has focused on risk factors in depression. Suicide attempts occur early in the course of schizophrenia. Given that the short-term suicide risk varies over the course of schizophrenia, and that it is extremely difficult to predict the long-term risk, parasuicide in schizophrenia must be rigorously assessed.

Barraclough et al. (1974) found a current diagnosis of depression in 70% of completed suicides. Issacson and Rich (1997) found that depression occurred in at least half of all deaths by suicide. In another study, Foster et al. (1997) found an overall rate of depression of 36% with over double the incidence among women (60%) as among men (29%). In the WHO/EURO study, neurotic depression was diagnosed according to ICD-9 criteria in 21% of all women who attempted suicide. Kelleher et al. (1999) also stressed the relationship between deliberate self-poisoning and affective disorder with 33% of male ($n = 83$) and 54% of female ($n = 111$) self-poisoners presenting with a diagnosis of affective disorder (see Table 1). While suicidal ideation is a recognised symptom in depressive disorders, parasuicides appear to represent a different sociodemographic population to completed suicides, although a degree of overlap exists.

6
Suicidal Behaviour in Institutions

6.1
Psychiatric Hospitals

Given the high prevalence of individual psychopathology which is obviously present in psychiatric hospitals and in spite of the best efforts of medical and nursing staff, suicidal behaviour does occur in psychiatric hospitals. When it does happen, it sometimes occurs in clusters (Taiminen and Helenius 1994). The factors that discriminate between the suicidal and non-suicidal in hospital populations may differ from community ones. Self-burning is one type of self-harm that appears to be more common among in-patients, especially females with a history of personality disorder (O'Donoghue et al. 1998). The current ethos of open door units and increased patient autonomy has the effect of making it more difficult to restrict access to inflammable clothing and matches, etc. The situation regarding the frequency of suicide among recently discharged psychiatric patients has already been discussed in the previous chapter. Briefly, it has been noted that there is a substantial increase in suicide among those recently discharged from psychiatric hospitals, which is greater in the first few weeks but remains elevated for several months (Goldacre et al. 1993). The problem seems to lie in the inability of the individual to function adequately in society following a psychiatric admission. The difficulty in returning to full social functioning is associated with social stigma and perceived loss of a meaningful role in society.

6.2 Prisons

A more detailed description of prison psychiatry is outlined in Chap. 17 (Vol. 1, Part 2) and the following is therefore only a brief description regarding suicidal behaviour in prisons.

Official concern was expressed at the high and apparently increasing rate of suicidal behaviour in prisons in the United Kingdom and Ireland in the early 1990s. There are some indications that the increase has levelled off, but the U.K. rates are five to six times higher than in the general population (Dooley 1990). To some extent, this is to be expected as the population in prison mainly consists of young men who are all undergoing major stresses just by virtue of their incarceration. There is also a high incidence of substance abuse and dysfunctional family background within the prison population, all of which would predispose to difficulties in coping with stress. A U.K. report found that suicide attempts in prison were more likely to occur in the early stages of imprisonment. Over 25% of suicides occurred within 1 month and over 50% within 3 months of incarceration. The majority of prison suicides were by individuals on remand (Crighton and Towl 1997). Procedures mainly aimed at increasing communication both within prison and between prisoners and the outside world have been introduced in an effort to reduce the degree of helplessness felt by prisoners (Dooley 1994). Policies aimed at increasing the proficiency of prison staff in dealing with the problem have also been instituted, e.g. by establishing suicide awareness groups within the prison.

6.3 Schools

Clusters of suicidal behaviour among young people in a particular area or school are most often attributed to the Werther or contagion effect. The term "Werther effect" was coined by Phillips (1974), based on the character of a young man in Goethe's 1774 novel *The Sorrows of Young Werther* who killed himself. This effect may result from media accounts of a celebrity suicide or personal ties and emotional identification with a person who attempts or dies by suicide. For instance, much has been made of copycat suicides following the death of the rock singer Kurt Cobain. It is likely that those already at risk are most likely to be affected and that the contagion tends to influence their choice of method and the setting. This can lead to an apparent increase in completed suicide if the attempter changes to a more lethal method, e.g. hanging or guns, in imitation of the celebrity. Clusters may also surround the unexpected death of a high-profile student in a school, and again those already emotionally vulnerable are most at risk (Gould et al. 1990). Prevention of such clusters may best be managed by combining case identification of those most at risk with general measures involving promotion of mental health status and reduction of stigma and bullying, rather than by specific suicide intervention packages (Kelleher 1998b). The latter programmes may have the effect of glamorising suicidal behaviour and have been found, in some cases, to result in outbreaks of deliberate self-harm among students (Callahan 1996). The issue of school postvention programmes will be further discussed later.

7 Intervention

On first consideration, there appears to be a continuum from suicidal ideas, whether fleeting, recurrent or more persistent; to impulsive and more detailed planning; to attempts of varying severity; culminating in suicide itself. However, this is not necessarily true, and many people may make a serious attempt without ever having previously thought about suicide. Undoubtedly, others have given prior consideration to the decision, sometimes on and off, over many years. No one knows, however, what the gateway factors are from one level to the next.

Presumably there are a mixture of inducing and inhibiting issues. In such an apparently incoherent field, chaos theory may be invoked as an explanatory hypothesis. Reid (1998) indicates that chaos theory can be relevant to psychiatry. In particular, he describes a model of the possible effects of psychosocial stressors in bipolar illness. Perhaps this model can be applied to the understanding of parasuicidal behaviour. Chaos theory refers to the seemingly random type of variability that can arise in a non-linear system. Although appearing random, the behaviour of an object in such a system can be modelled. Parasuicidal patients inhabit a vulnerable psychosocial system, as indicated from the WHO/EURO Monitoring Study. In such a vulnerable and unstable psychosocial system, a stressor, if it has sufficient amplitude, may drive the system into sustained chaotic dynamics. A stressor may therefore produce a long-lasting effect, with subsequent episodes apparently not linked to any precipitant and appearing to rise spontaneously. This might help to explain the apparent reduced importance of external stressors as precipitants in subsequent episodes, noted in the WHO/EURO study (Schmidtke et al. 1994b).

7.1 Assessment

Intervention after an act of parasuicide presents a host of problems that must be addressed. The policy of the Royal College of Psychiatrists (1996) in the United Kingdom is that such cases should be psychiatrically assessed, particularly if the event results in admission to a general hospital. Table 2 gives a schema for some essential elements in the assessment of suicidal risk following an act of parasuicide.

In general, assessment must include the diagnosis and establishing the reason for the attempt; severity of illness if present; likelihood of repetition of suicidal behaviour in the short term; discussion with an informant if available; decision with regards to disposal once safe to discharge from general hospital, which may be as psychiatric in-patients, psychiatric out-patients or referral back to a general practitioner.

Assessment of suicidal risk that is based on risk factors alone is not sufficient. Risk assessment in suicidal patients involves determining the balance between risk and protective factors (Appleby 1997). Ringel's pre-suicidal syndrome (1976) may also inform assessment. First described in 1958, there are three distinct elements, namely, constriction, inhibited aggression turned towards the self and suicidal fantasies which relate to a specific state of mind that may lead to suicidal acts. Knowledge of the above factors can facilitate earlier awareness of serious suicidal intent, rendering earlier intervention possible. The use of validated psychological scales may aid in decision-making. Beck's Suicide Intent Scale is a 15-item questionnaire, which is administered by the clinician to an individual following an act of attempted suicide. It has been consistently validated as a measure of the seriousness of intent to die (Beck et al. 1974a; Beck and Lester 1976) and provides a means of assisting a clinician in assessing suicidal risk. Hopelessness (i.e. a negative

Table 2. Assessment of risk in parasuicides

Task	Questions	Indicators of increased risk
Establish the patient's suicide intent at the time of the attempt	Did he or she wish to die? Did he or she plan the act? In choosing the method, did he or she expect to die? Did he or she seek help following the act?	Evidence of planning, e.g. selecting isolated location Evidence of a final act, e.g. suicide note Method with a high perceived lethality Absence of help-seeking following the act
Determine if there is a continued risk of suicide; he or she should be asked direct questions	Does he or she still express a wish to die? How does he or she see the future? What has changed for him or her? Do risk factors still outweigh protective factors?	Hopelessness Lack of remorse Ambivalence about the future Past history of suicidal behaviour
Establish if any psychiatric illness is present; he or she may require formal psychiatric evaluation	Does he or she have a history of psychiatric illness? Is he or she being treated for psychiatric illness? If yes, could his or her treatment regime/compliance be improved? Does he or she have a family history of psychiatric illness?	For depression, presence of agitation, insomnia For alcoholism, presence of social isolation, depressed mood For personality disorder, presence of sociopathic or borderline subtypes For schizophrenia, being male, young, chronically ill For neuroses, presence of panic attacks
Establish the coping skills/resources available to him/her	Did he or she experience a recent loss? Does he or she have interpersonal difficulties? Does he or she live alone? Does he or she have family or community support? Is he or she physically ill? Does he or she have financial, employment, housing or legal worries?	Bereavement/loss Marital/relationship difficulties Older age, male sex Social isolation Physical/psychiatric illness Unemployment Criminal record
Decide on the level of intervention	Have the medical/surgical complications of the act been treated? Does he or she require psychiatric treatment? Will he or she accept treatment voluntarily? Does he or she require compulsory admission? If he or she is being discharged, has an emergency contact number been provided?	Compulsory admission In-patient treatment Poor compliance

view of the future) is the cognitive feature which is most consistently related to suicidal ideation, intent and completion in adults (Beck et al. 1990). The Beck Hopelessness Scale (Beck et al. 1974b) is a 20-item self-report questionnaire which assesses the level of "negative view of the future" held by the patient. It has a high false-positive rate, which needs to be kept in mind (Beck et al. 1990). It has to be stressed, however, that these scales act as an aid to assessment and do not substitute for a detailed psychiatric evaluation.

In high-risk cases, admission to hospital is the preferred option. Compulsory admission may have to be considered. Ideally, this is best avoided as it may mean prevention in the short term but may result in hostility towards the service should suicidal ideas return later. Hospitals should have a protocol for managing suicidal patients that maximises observation while minimising intrusion. The level of supervision must be graduated to the degree of perceived risk. Good therapeutic relationships at nursing and doctor level are essential.

7.2 Aftercare

Aftercare, i.e. after acute medical treatment, varied considerably across the WHO/EURO centres. Perhaps surprisingly, the commonest type of treatment offered between all centres was in-patient (50% of females and 53% of males), then out-patient treatment (19% of males and 22% of females), and no treatment was recommended for 16% of both sexes. In-patient treatment was recommended slightly more often for those who made more than one attempt, whether male or female. It must be pointed out, however, that this contradicts more recent findings which suggest that out-patient treatment is the more likely next stage of care. For example, Hawton et al. (1997), in an examination of deliberate self-harm in the Oxford region of England, report that in-patient care accounted for only 10% of the aftercare provided following assessment ($n = 738$). It is suggested here that the reason for such a discrepancy between studies may relate to the demographic characteristics of the samples in question. Furthermore, variations in service provision, i.e. supply factors, may determine the type of service provided to parasuicidal individuals. As one example of such differences, acute assessment is carried out by trained social workers in the United Kingdom, whereas it is performed by psychiatrists in Ireland. The absence of proven interventions (as discussed below) may account for such variations in clinical practice.

Because the problems of individual parasuicidal patients may be multiple, a team approach is essential (Kelleher et al. 1999). To this end, referral for assessment to a variety of mental health professionals, including social workers, psychologists and nursing personnel, could be made at the discretion of the consulting clinician, taking into account the individual circumstances.

The social needs of some parasuicidal patients may be enormous. Often low stress tolerance is conjoined to situations that are themselves very stressful. The parasuicidal person may be the only outward sign of a very dysfunctional domestic unit with other disturbed members, including dependent children. Social work intervention may be essential to unravel the social and other problems, including, where appropriate, the provision of case work care (Gibbons et al. 1978). However, it may be that the effect from social work intervention is stronger in females than males. There is a strong case to be made for the attachment of a social worker to each accident and emergency department so that all those presenting with deliberate self-harm will have the opportunity for assessment.

In a study examining the experience of suicide in the lives of 100 college students, Keeley and Kelleher (1998) found that friends and family were most popular as sources of help, but perhaps more significantly almost 50% did not know where they could turn. In response to a more specific question on the type of organisation they thought would provide acceptable help, 47% of the students suggested the Samaritans, while, with the exception of one student, medical services were never suggested. Three per cent mentioned the college services, but 40% were unable to make any suggestions. It appears from this information that the statutory services have a marketing problem when it comes to making their services acceptable to the young.

7.3 Specific Interventions

The efficacy of cognitive behavioural therapy has been demonstrated in selected populations in several research studies, some of which are ongoing (Linehan et al. 1987; McLeavey et al. 1987). Its place in routine clinical practice is not yet assured. Part of the difficulty is the various exclusion criteria that apply and, even more importantly, inadequate compliance by the target population. Other proposed intervention treatments include the Green Card study, as suggested by Morgan et al. (1993). This involves giving suicide attempters a card which increases their awareness of the extent of services to which they might have access without having to harm themselves. This approach has been piloted and has been found to be very acceptable to the patients, but it suffers from problems with acceptance

by clinicians and the lack of a sufficiently large sample to give a proper statistical indication of its benefits. A systematic review of the effectiveness of controlled intervention trials in deliberate self-harm has been published recently, and it concludes that there is a need for larger trials of treatment. The only significantly effective treatments identified in the review were depot flupenthixol in multiple repeaters and dialectical behaviour therapy (Hawton et al. 1998). This report confirms the high degree of uncertainty as to which forms of treatment are most effective.

7.4 General Practice

Management of parasuicide exclusively by general practitioners is considered unusual in Europe (Platt et al. 1992; Fitzsimons et al. 1997). The extent to which it happens depends on the structure of the health services in the country under study. Estimates range from 7% in Ireland (Fitzsimons et al. 1997) to 13% in France (Platt et al. 1992) and 20% in England and Wales (Kennedy and Kreitman 1973). There is some evidence that cases that are managed by the person's general practitioner, with or without the additional hospital services, may have a lower repetition rate than those managed by the hospital services alone. It is likely that general practitioner treatment might reduce the high level of non-compliance with aftercare of suicide attempters, whether because of their increased understanding of the person's social situation or because of reduced stigma and ease of access. Such a change in treatment would require increased resourcing and training of general practitioners, including the provision of community-based psychiatric back-up, but could have positive implications for hard-pressed casualty and other general hospital departments.

7.5 Voluntary Services

It is the voluntary services that are often the first to recognise a particular social need. The policy of the Samaritans and other "befriender" groups is that the volunteer does not influence or intervene in people's decisions to take their own lives unless specifically asked to do so (Befrienders International 1995). They therefore provide emotional support as opposed to professional directive counselling, unless the client requests referral. This implies limitations in the Samaritan's role, but on the other hand it also means that some clients may be more likely to contact somewhere where they will not be forced to do something that they do not want – even if it is perceived as good treatment and ensures that a separate identity is maintained.

The 12 member states of the European Union were surveyed in a study by Befrienders International (1995). The number of contacts per suicide varied widely between centres, with the United Kingdom/Ireland having the highest and Denmark the lowest. This may be related to the number and accessibility of centres. Germany, next to Denmark, has the second lowest number of contacts per suicide, but unlike Denmark it cannot be accused of having fewer prevention centres. A possible explanation offered is that the wide variety of services available can lead to reduced clarity about their perceived function. Numerous centres in Denmark have become professionalised, and this may discourage people who do not believe they have a psychiatric problem or fear being diagnosed with one. This last feature is particularly important to the nature of the service offered by the Samaritans and might indicate a need for befriender organisations to retain their independence from the statutory services if they are to continue to appeal to younger age-groups.

7.6 Helplines

Third-party callers provide a potential avenue for reaching high-risk groups who do not directly approach health services, particularly males, and young parasuicides who do not want to be referred for more specialist care. It was found that in the first year of a hotline launched in Geneva (Perret-Catipovic 1999), most callers were family members or friends of young suicidal people. There were equal numbers of calls concerning potentially suicidal males as females, although the parasuicide ratio in Switzerland is four females for every one male. In recognition that, in addition to mental health professionals, many other people have a role to play in preventing suicide, the gatekeeper intervention training programme was developed in Alberta, Canada, for the purpose of developing the intervention skills of a large range of professionals, volunteers and people who by virtue of their work have the opportunity to help people in distress. It has been evaluated positively in several studies (Bagshaw 1988; Crookall and McLean 1986; Tierney 1988).

7.7 Prediction of Repetition

As previously stated, within a year of an attempt, about 1% of adults die by suicide (Hawton and Fagg

1988). Predicting future suicide with any degree of accuracy is, at present, beyond clinical skill. In one large-scale study using what was regarded as the best clinical predictors such as previous attempt, serious mental illness and admission to a mental hospital, it was found that the likelihood of predicting suicide over a succeeding 5-year period was as infrequent as three times out of a hundred (Pokorney 1983). However, most clinicians do not need to engage in this relatively long-term type of prediction. Instead, the challenge is to decide whether the patient is suicidal at the time of examination. Such a judgement is based on two things, the clinician thinking about it and the patient telling it. What the patient says is of paramount importance, as words make intention explicit. Because suicide can occur in virtually any psychiatric condition, no examination should be considered complete without asking about such intentions, often best introduced after establishing through careful examination a good clinical relationship, by enquiring about the future. Most patients are happy to unburden themselves of self-destructive pre-occupations to an understanding physician. If an attempt has been made or an intention has been disclosed to an informant, the matter should be discussed with the patient. Absence of supportive relationships, isolation and social stresses perceived by the patient as being beyond his or her capacity are negative prognostic signs. In particular, a change from a stable to an unsettled social situation may precipitate an attempt (Schmidtke et al. 1996).

Accurately identifying those at high risk of repeating parasuicidal behaviour in the future could have service benefits, because special provision could be made to prevent such a recurrence. As clinical predictors seem insufficient, some other approaches are being investigated. A prediction mechanism based on a computer model, designed for use by nursing and other paramedical staff, was found to discriminate between those who repeated parasuicide and those who did not do so (Corcoran et al. 1997). Presently this model is being tested and refined further in a prospective study by triage nurses working in accident and emergency departments.

8
Prevention of Clusters

8.1
School-Based Suicide Prevention Programmes

As discussed above, there has been considerable debate about school-based efforts to tackle suicidal behaviour in students. Some of them are of questionable efficacy and safety (Hazell and Lewin 1993). An alternative and perhaps more sensible approach has been to train teachers as gatekeepers to detect emotional problems and suicidal risk in students and refer them as appropriate, using available community resources, e.g. professional and voluntary mental health bodies (Canadian Task Force on Suicide 1994). An appropriate ethos might be to focus on positive aspects by learning coping skills that will be helpful in future crises.

8.2
Community-Based Programmes

Guidelines for programmes aimed at the prevention and containment of suicide clusters have been drawn up by the Center for Disease Control (1994) in Atlanta. The main aim of the guidelines, which must be generic by their very nature, is to have identifiable and prepared figures within the community to co-ordinate responses and to deal with the media so that panic, rumour and innuendo are reduced to a minimum. They stress the importance of using existing resources within the area and the necessity of adapting any strategy to cope with the local cultural climate.

9
Conclusion

Parasuicide prevention might well learn from the public health approach to heart disease where a change in diet, more exercise, reduction in weight, cessation of smoking and moderate drinking were found to be of such importance. A difference is, however, that in heart disease the focus was on the individual, whereas a public health approach to the prevention of parasuicide would have to focus more on social issues. The rates of parasuicide might well fall if the deprived received better education from the earliest age, including better education of the parent(s), especially as many are young, inexperienced and deprived (Kelleher et al. 1999). Children, particularly males, should be retained within the educational system. Over-crowding could be alleviated by providing adequate housing, allowing more personal space and thereby reducing the likelihood of constant confrontation. Better employment prospects are likely to reduce parasuicide, as are reduction in the use of alcohol and avoidance of the misuse of drugs (Proudfoot et al. 1997; Neeleman and Farrell 1997). The positive outcome of reduction in the use of alcohol can be seen in the former Soviet Republics, where a

restriction on alcohol during *perestroika* coincided with a sharp decline (35%) in suicide (Wasserman and Värnik 1998).

The Finnish were one of the first people to attempt to approach suicide prevention from the point of view that treatment of suicide attempters is only the tip of the iceberg in a problem whose roots reach far down into the structure of society. The obvious conclusion from this is that in order to attempt to adequately address this problem, it is necessary to make changes to our societies in ways that may conflict with deeply and sincerely held views. Another implication of their approach is that it is essential to have accurate information about all aspects of the problem, and research is therefore important at all levels of society. Internal evaluation of the strategy (Hakanen and Upanne 1996) indicates that it was successful overall in mobilising professionals both in health and non-health service sectors nationally to take responsibility and to further develop suicide prevention in practice. In addition, they reached their central strategic aim of evolving co-operation across professionals and sectors. During the implementation phase of the project, suicide frequency fell by 9% below the initial increasing rate, but Beskow et al. (1999) can only report that the strategy *may* have contributed to this reversal. The Finnish strategy makes seven main points and it is appropriate to end two chapters devoted to the problem of suicidal behaviour with their impressive insights (Tuipale 1993).

According to the strategy of "Joint Responsibility for Suicide Prevention", the rising suicide rate can be reversed if:

- Everyone who has attempted suicide receives as effective help as possible.
- Depression is recognised and the person offered all the support he or she needs; everyone suffering from serious depression should receive appropriate and effective treatment.
- We can prevent alcohol being used as the universal solution to problems, and find better means of supporting efforts to cope.
- Mental and social supports are enhanced within the treatment of somatic illness.
- A person in a life crisis receives appropriate support from friends and relatives, and from professionals when necessary.
- The risk of youngsters becoming alienated from life can be avoided, and individuals running a risk of suicide are guaranteed the possibility of coping and improving their self-esteem.
- The cultural climate in Finland, including the education system, becomes more relaxed and permissive, and less guilt promoting, stigmatising and punitive than it tends to be at present. It needs to promote belief in life, resourcefulness, self-esteem, initiative and mutual support.

10 References

Andrews JA, Lewinsohn PM (1992) Suicidal attempts among older adolescents: prevalence and co-occurrence with psychiatric disorders. J Am Acad Child Adolesc Psychiatry 31(4): 655–662

*Appleby L (1997) Assessment of suicidal risk. Psychiatr Bull 21: 193–194

Asberg M, Thoren P, Träskman-Bendz L, Bertilsson L, Ringberger V (1976) "Serotonin depression" – a biochemical subgroup within the affective disorders? Science 191: 478–480

Asberg M, Nordstrom P, Träskman-Bendz L (1986) Cerebrospinal fluid studies in suicide: an overview. Ann NY Acad Sci 487: 243–255

Bagshaw M (1988) Suicide prevention training: lessons from the Corrections Service of Canada. Prison Service J 70: 5–6

*Barraclough B, Bunch J, Nelson B, Sainsbury P (1974) A hundred cases of suicide: clinical aspects. Br J Psychiatry 125: 355–373

Beautrais AL, Joyce PR, Mulder RT (1996) Risk factors for serious suicide attempts among youths aged 13–24 years. J Am Acad Child Adolesc Psychiatry 35: 1174–1182

*Beck AT, Lester D (1976) Components of suicidal intent in completed and attempted suicides. J Psychol 92: 35–38

Beck AT, Schuyler D, Herman I (1974a) Development of suicide intent scales. In: Beck AT, Resnick HCP, Lettieri D (eds). The prediction of suicide. Charles, Maryland, pp 45–56

Beck AT, Weissman A, Lester D, Trexler L (1974b) The measurement of pessimism: the Hopelessness Scale. J Consult Clin Psychol 42: 861–865

Beck AT, Brown G, Berchick RJ, Stewart BL, Steer RA (1990) Relationship between hopelessness and ultimate suicide: a replication with psychiatric outpatients. Am J Psychiatry 147(2): 190–195

Befrienders International (1995) Study of suicide prevention within the European Community. European Commission, London

Beskow J, Kerkhof AJFM, Kokkola A, Uutela A (1999) Suicide prevention in Finland 1986–1996. External evaluation by an international peer group. Ministry of Social Affairs and Health, Helsinki

Bille-Brahe U, Andersen K, Wasserman D, Schmidtke A, Bjerke T, Crepet P, de Leo D, Haring C, Hawton K, Kerkhof AJFM, Lönnqvist J, Michel K, Phillippe A, Querejeta I, Salander-Renberg E, Temesváry B (1996) The WHO/EURO Multicentre Study: risk of parasuicide and the comparability of the areas under study. Crisis 17(1): 32–42

Böhme K (1994) Drug and alcohol induction of suicidal behaviour. In: Kelleher MJ (ed) Divergent perspectives on suicidal behaviour – proceedings of the 5th European Symposium on Suicide. DAOL, Cork, pp 128–139

Brent DA, Perper JA, Moritz G, Baugher M, Allman C (1993) Suicide in adolescents with no apparent psychopathology. J Am Acad Child Adolesc Psychiatry 32(3): 494–501

Brent DA, Bridge J, Johnson BA, Connolly J (1996) Suicidal behaviour runs in families: a controlled family study of

adolescent suicide victims. Arch Gen Psychiatry 53: 1145–1152

Brent DA, Perper JA, Moritz G, Liotus L, Scheers J, Balach L, Roth C (1994) Familial risk factors for adolescent suicide: a case control study. Acta Psychiatr Scand 89: 52–58

Callahan J (1996) Negative effects of a school suicide postvention program – a case example. Crisis 17(3): 108–115

Canadian Task Force on Suicide (1994) Suicide in Canada: update of the Report of the Task Force on Suicide in Canada. Ministry of National Health and Welfare, Ottowa

Casey PR (1989) Personality disorder and suicide intent. Acta Psychiatr Scand 79: 290–295

Center for Disease Control (1994) Suicide contagion and the reporting of suicide: recommendations from a national workshop. Morbid Mortal Weekly Report 43(RR-6): 1–19

Corcoran P, Kelleher MJ, Keeley HS, Byrne S, Burke U, Williamson E (1997) A preliminary statistical model for identifying repeaters of parasuicide. Arch Suicide Res 3: 65–74

Crighton D, Towl G (1997) Self-inflicted deaths in prison in England and Wales: an analysis of the data for 1988–1990 and 1994–1995. In: Towl G (ed) Suicide and self-injury in prisons – research directions in the 1990's. British Psychological Society, Leicester, pp 12–20

Crookall P, McLean T (1986) Evaluation of the suicide prevention training program in the Atlantic region. Correctional Service Canada, Ottawa

Daly M, Conway M, Kelleher MJ (1986) Social determinants of self-poisoning. Br J Psychiatry 148: 406–413

*De Moore GM, Robertson AR (1996) Suicide in the 18 years after deliberate self-harm: a prospective study. Br J Psychiatry 169: 489–494

Deykin EY, Buka SL (1994) Suicidal ideation and attempts among chemically dependent adolescents. Am J Public Health 84: 634–639

Dooley E (1990) Prison suicide in England and Wales 1972–1987. Br J Psychiatry 156: 40–45

Dooley E (1994) Prisons. In: Jenkins R, Griffiths S, Wylie I et al (eds) The prevention of suicide. HMSO, London, pp 104–108

Egeland JA, Sussex JN (1985) Suicide and family loading for affective disorders. JAMA 254: 915–918

Fitzsimons MM, Kelleher MJ, Keeley HS, Corcoran P, Byrne S, Williamson E, Burke U (1997) Parasuicide and general practice: a pilot study. Ir Med J 90(5): 190–192

Fombonne E (1998) Suicidal behaviours in vulnerable adolescents. Br J Psychiatry 173: 154–159

Foster T, Gillespie K, McClelland R (1997) Mental disorders and suicide in Northern Ireland. Br J Psychiatry 171: 447–452

Garrison CZ, McKeown RE, Valois RF (1993) Aggression, substance use and suicidal behaviours in high school students. Am J Public Health 83: 179–184

Gibbons JS, Butler P, Urwin P, Gibbons JL (1978) Evaluation of a social work service for self-poisoning patients. Br J Psychiatry 113: 111–118

**Goldacre M, Seagroatt V, Hawton K (1993) Suicide after discharge from psychiatric in-patient care. Lancet 342: 283–286

Gould MS, Wallenstein S, Kleinman MH, O'Carroll P, Mercy J (1990) Suicide clusters: an examination of age-specific effects. Am J Public Health 80: 211–212

Hakanen J, Upanne M (1996) Evaluation strategy for Finland's suicide prevention project. Crisis 17(4): 167–174

Hawton K, Fagg J (1988) Suicide and other causes of death following attempted suicide. Br J Psychiatry 18: 405–418

Hawton K, Fagg J, Simkin S, Bale E, Bond A (1997) Trends in deliberate self-harm in Oxford, 1985–1995. Implications for clinical services and the prevention of suicide. Br J Psychiatry 171: 556–560

*Hawton K, Arensman E, Townsend E, Bremner S, Feldman E, Goldney R, Gunnell D, Hazell P, van Heeringen K, House A, Owens D, Sakinofsky I, Träskman-Bendz L (1998) Deliberate self harm: systematic review of efficacy of psychosocial and pharmacological treatments in preventing repetition. Br Med J 317: 441–447

Hazell P, Lewin T (1993) An evaluation of postvention following adolescent suicide. Suicide Life Threat Behav 23: 101–109

Isaacson G, Rich CL (1997) Depression, antidepressants and suicide: pharmacoepidemiological evidence for suicide prevention. In: Maris RW, Silverman MM, Cannetto SS (eds) Review of suicidology. Guildford, New York, pp 168–202

Keeley HS, Corcoran P, Hennessy AM, Lawlor M (1999) Background stressors in Irish parasuicides. Poster presented at Conference of the American Association of Suicidology, 14–18 April 1999, Houston, Texas

Keeley HS, Kelleher MJ (1998) Youth attitudes to services in Ireland. Psychiatr Bull 22(4): 257

Kelleher MJ (1998a) Youth suicide trends in the Republic of Ireland. Br J Psychiatry 173: 196–197

Kelleher MJ (1998b) Suicide in schools. In: Farrell B (ed) Issues in education: changing education, changing society. ASTI, Dublin

Kelleher MJ, Kelleher MJA, Corcoran P, Daly M, Daly F, Crowley MJ, Keeley HS (1996) Deliberate self-poisoning, unemployment, and public health. Suicide Life Threat Behav 26(4): 365–373

Kelleher MJ, Keohane B, Daly C, Keeley HS, Corcoran P, Chambers D, Williamson E (1999) Individual characteristics and long-term outcome for deliberate self-poisoners. J Ir Coll Phys Surg 28(1): 5–8

Kennedy P, Kreitman N (1973) An epidemiological survey of parasuicide ('attempted suicide') in general practice. Br J Psychiatry 123: 23–34

Kerkhof AJFM, Schmidtke A, Bille-Brahe U, de Leo D, Lönnqvist J (eds) (1994) Attempted suicide in Europe: findings from the Multicentre Study on Parasuicide by the WHO Regional Office for Europe. DSWO, Leiden, pp 289–291

Kessel N, Grossman G (1965) Suicide in alcoholics. Br Med J 2: 1671–1672

Kreitman N, Philip AE (1969) Parasuicide. Br Med J 115: 746—747

Lawlor M, Corcoran P, Chambers D (2000) Suicide attempts v. deliberate self-harm: a response. Br J Psychiatry 176: 91–96

Linehan MM, Camper P, Chiles JA, Strosahl K, Shearin E (1987) Interpersonal problem solving and parasuicide. Cogn Ther Res 11(1): 1–12

Mann JJ, Stanley M, McBride A, McEwen BS (1986) Increased serotonin and β-adrenergic receptor binding in the frontal cortices of suicide victims. Arch Gen Psychiatry 43: 954–959

*Maris RW, Berman AL, Maltsberger JT, Yufit RI (eds) (1992) Assessment and prediction of suicide. Guildford, New York

McAuliffe C (1998) Student attitudes to suicidal behaviours and coping skills. Paper presented at the 7th European Symposium on Suicide and Suicidal Behaviour, 9–12 September 1998, Gent, Belgium

McKey PW, Jones RW, Barbe RH (1993) Suicide and the school: a practical guide to suicide prevention. LRP, Pennsylvania

McLeavey BC, Daly RJ, Murray CM, O'Riordan J, Taylor M (1987) Interpersonal problem-solving deficits in self-poisoning patients. Suicide Life Threat Behav 17: 33–49

Mishara BL, Baker H, Mishara TT (1976) The frequency of suicide attempts: a retrospective approach applied to college students. Am J Psychiatry 133(7): 841–844

Morgan HG, Jones EM, Owen JH (1993) Secondary prevention of non-fatal deliberate self-harm. The green card study. Br J Psychiatry 163: 111–112

Morgan HG, Pocock H, Pottle S (1975) The urban distribution of non-fatal deliberate self-harm. Br J Psychiatry 126: 315–328

Murphy GE, Wetzel RD (1992) Multiple risk factors predict suicide in alcoholism. Arch Gen Psychiatry 49(6): 459–463

Neeleman J, Farrell M (1997) Suicide and substance misuse. Br J Psychiatry 171: 303–304

Nielsen DA, Goldman D, Virkkunen M, Tokola R, Rawlings R, Linnoila M (1994) Suicidality and 5-hyrdoxyindoleacetic acid concentration associated with a tryptophan hydroxylase polymorphism. Arch Gen Psychiatry 51: 34–38

O'Donoghue JM, Panchal JL, O'Sullivan ST, O'Shaughnessy M, O'Connor TPF, Keeley HS, Kelleher MJ (1998) A study of suicide and attempted suicide by self-immolation in an Irish psychiatric population: an increasing problem. Burns 24: 144–146

Perret-Catipovic M (1999) Suicide prevention in adolescents and young adults: the Geneva University Hospital's Program. Crisis 20: 36–40

Phillips DP (1974) The influence of suggestion on suicide: substantive and theoretical implications of the Werther effect. Am Sociol Rev 39: 340–354

**Platt S (1984) Unemployment and suicidal behaviour: a review of the literature. Soc Sci Med 19(2): 93–115

Platt S, Bille-Brahe U, Kerkhof AJFM, Schmidtke A, Bjerke T, Crepet P, de Leo D, Haring C, Lönnqvist J, Michel K, Phillippe A, Pommereau X, Querejeta I, Salander-Renberg E, Temesváry B, Wasserman D, Sampaio Faria J (1992) Parasuicide in Europe: the WHO/EURO Multicentre Study on Parasuicide I: introduction and preliminary analysis for 1989. Acta Psychiatr Scand 85(2): 97–104

**Pokorney AD (1983) Prediction of suicide in psychiatric patients – report of a prospective study. Arch Gen Psychiatry 40: 249–257

Proudfoot J, Guest D, Carson J, Dunn G, Gray J (1997) Effect of cognitive-behavioural training on job-finding among long-term unemployed people. Lancet 350: 96–100

Reid S (1998) Butterflies, fractals and psychiatry. Psychiatr Bull 22: 568–570

Ringel E (1976) The presuicidal syndrome. Suicide Life Threat Behav 6(3): 131–149

Roy A, Mazonson A, Pickar D (1984) Attempted suicide in chronic schizophrenia. Br J Psychiatry 144: 303–306

Roy A, Segal NL, Sarchiapone M (1995) Attempted suicide among living co-twins of twin suicide victims. Am J Psychiatry 152(7): 1075–1076

Royal College of Psychiatrists (1996) Confidential inquiry into homicides and suicides by mentally ill people. RCPsych, London

Schmidtke A, Bille-Brahe U, de Leo D, Kerkhof AJFM, Bjerke T, Crepet P, Haring C, Deisenhammer E, Hawton K, Lönnqvist J, Michel K, Pommereau X, Querejeta I, Phillippe A, Salander-Renberg E, Temesváry B, Wasserman D, Sampaio Faria J, Weinacker B (1994a) Rates and trends of attempted suicide in Europe, 1989–1992. In: Kerkhof AJFM, Schmidtke A, Bille-Brahe U, de Leo D, Lönnqvist J (eds) Attempted suicide in Europe: findings from the Multicentre Study on Parasuicide by the WHO Regional Office for Europe. DSWO, Leiden, pp 209–229

Schmidtke A, Bille-Brahe U, de Leo D, Kerkhof AJFM, Bjerke T, Crepet P, Haring C, Hawton K, Lönnqvist J, Michel K, Pommereau X, Querejeta I, Phillippe A, Salander-Renberg E, Temesváry B, Wasserman D, Fricke S, Weinacker B, Sampaio Faria J (1994b) Sociodemographic characteristics of suicide attempters in Europe. In: Kerkhof AJFM, Schmidtke A, Bille-Brahe U, de Leo D, Lönnqvist J (eds) Attempted suicide in Europe: findings from the Multicentre Study on Parasuicide by the WHO Regional Office for Europe. DSWO, Leiden, pp 231–241

Schmidtke A, Bille-Brahe U, de Leo D, Kerkhof AJFM, Bjerke T, Crepet P, Haring C, Hawton K, Lönnqvist J, Michel K, Pommereau X, Querejeta I, Phillippe A, Salander-Renberg E, Temesváry B, Wasserman D, Fricke S, Weinacker B, Sampaio Faria J (1996) Attempted suicide in Europe: rates, trends and sociodemographic characteristics of suicide attempters during the period 1989–1992. Results of the WHO/EURO Multicentre Study on Parasuicide. Acta Psychiatr Scand 93: 327–328

Schulsinger F, Kety SS, Rosenthal D, Wender PH (1979) A family study of suicide. In: Schou M, Stromgren E (eds) Origin, prevention and treatment of affective disorders. Academic, London, pp 277–287

Schweitzer R, Klayich M, McLean J (1995) Suicidal ideation and behaviours among university students in Australia. Aust NZ J Psychiatry 29(3): 473–479

Sellers EM, Higgins GA, Sobell MB (1992) 5-HT and alcohol abuse. Trends Pharmacol Sci 13: 69–75

Shaffer D, Gould MS, Fisher P, Trautman P, Moreau D, Kleinman M, Flory M (1996) Psychiatric diagnosis in child and adolescent suicide. Arch Gen Psychiatry 142: 1061–1064

**Shneidman ES (1996) The suicidal mind. Oxford University Press, New York, pp 3–26

Steele TD, McCann TD, Ricaurte GA (1994) 3,4-Methylenedioxymethamphetamine (MDMA): pharmacology and toxicology in animals and humans. Addiction 89: 539–551

Stengel E (1952) Enquiries into attempted suicide. Proc R Soc Med 45: 613–620

Stengel E, Cook NG (1958) Attempted suicide: its social significance and effects. Chapman and Hall, London (Maudsley monograph no 4)

Taiminen TJ, Helenius H (1994) Suicide clustering in a psychiatric hospital with a history of a suicide epidemic: a quantitative study. Am J Psychiatry 151(7): 1087–1088

Tierney J (1988) Comprehensive evaluation for suicide intervention training. Unpublished doctoral dissertation. University of Calgary, Calgary

Tuipale V (1993) Suicide can be prevented: fundamentals of a target and action strategy. National Research and Development Centre for Welfare and Health, Helsinki

Wasserman D, Värnik A (1998) Suicide prevention effects of perestroika in the former USSR: the role of alcohol restriction. Acta Psychiatr Scand 98[Suppl 394]: 1–4

Wender PH, Kety SS, Rosenthal D, Schulsinger F, Ortmann J, Lunde I (1986) Psychiatric disorders in the biological and adoptive families of adopted individuals with affective disorders. Arch Gen Psychiatry 43: 923–929

Whelan C (1993) The role of social support in mediating the psychological consequences of economic stress. Sociol Health Illness 15(1): 86–101

Wilkinson G, Bacon NA (1984) A clinical and epidemiological survey of parasuicide and suicide in Edinburgh schizophrenics. Psychol Med 14: 899–912

World Health Organisation (1992) The tenth revision of the International Classification of Diseases – chapter XX: external causes of morbidity and mortality. WHO, Geneva

World Health Organisation Regional Office for Europe (1985) Targets for health for all. WHO Regional Office for Europe, Copenhagen (European health for all series no 1)

CHAPTER 11

H. Saß

Personality Disorders

Translator: A. Rodón

1 Introduction

The fascination and risks inherent in psychiatric pathology lie in the complex nature of psychiatric disorders, which is characterized by an intricate interplay of somatic functions, learning processes, acquired attitudes, and situation-specific influences. This is particularly evident with accentuated personality constitutions which conceptionally, nosologically, and diagnostically transcend and touch on various disciplines: the broad range between the healthy condition and a pathological development, between successful adaptation and dissocial development, and between constitutional temper and character variants and psychiatric illness. Moreover, the field dealing with deviant personalities is susceptible to misinterpretation in terms of anthropology, sociopsychology, and criminological policy: instead of as a physician, the psychiatrist can be perceived as an instrument to implement law and order tool. This is particularly important when dealing with dissocial behavior, which is why, in *Der Mann ohne Eigenschaften* ("The Man Without Qualities"), Musil warned our profession against becoming the backup angel of justice.

It was with the statement "Do We Need Personality Disorders?" that Tölle (1986) opened the relevant chapter in the previous, third edition of *Contemporary Psychiatry* (in German). With critical reserve toward the classification systems emerging in the Anglo-American countries at the time, he went on to outline the topic according to the German-speaking tradition, e.g., following Kretschmer and Kurt Schneider. Regarding nosological criteria, Tölle distinguished three groups of abnormal personalities, of which one resembled the psychoses (hyperthymic, paranoid, and schizoid forms), one was equivalent to character neuroses (hysterical, anancastic, depressive, in part schizoid, but also sensory disorders), and the last consisted of sociopathic or antisocial personalities. As the chapter by Hoffmann (1986) in the same book shows, despite attempts at integration using concepts from psychological and character neuroses, an incompatibility remained between the theories of neurosis and psychopathy which arose almost simultaneously nearly 100 years ago.

Personality disorders have now become well accepted and moved into the foreground of psychiatry and psychotherapy, both in science and in the doctor's office. In particular, the introduction of separate personality registration on the so-called axis II using DSM-III (APA 1980) and the subsequent editions DSM-III-R (APA 1987) and DSM-IV (APA 1994) stepped up research and publication worldwide, leading to the inception of professional journals and scientific societies devoted solely to the field of personality disorders. Today, the flood of theoretical concepts and therapeutic methods is immense and almost impossible to organize. Due to the heterogeneous nature of the subject matter, the outline which follows shows redundancies and hazy spots which at present can be avoided neither by empirically determinable criteria nor by conceptional and definitional classifications.

2 Definitions of Terms

The term "personality" is one of the most dubious words in our language. On the hand, it refers to an ethical, pedagogic value judgment, as when we say that someone has a strong personality or is a person of character. All people are persons, to be sure, but not everyone has a strong personality, which, according to Binder in the first edition of *Contemporary Psychiatry* (1964), is distinguished by a conscious, sustained exercise of thought and will. The concept of personality includes the terms temper and character, where temper refers more to the basic vital traits, while character encompasses long-range attitudes.

Using the terminology of psychiatry and psychotherapy, we will now define personality as the sum of all psychological characteristics and behaviors which lend a human being his or her unique, distinct individuality, these all being contained in aspects of perception, thought, emotion, will, and interrelationship with others. On the other hand, a personality disorder occurs if, due to degree and/or a particular combination of psychopathologically relevant features in any of these areas, there is considerable suffering and/or lasting impairment of social adaptativeness (Saß 1987).

Psychiatry and psychology have different conceptional views regarding classification and description in this field. Psychiatric thinking aims at distinguishing personality disorders according to nosological, typological, and characterological aspects. The various personality disorders are to be grouped in units distinct both from each other as well as from what is seen as normal. Methods used in depth psychology and developmental psychology, such as the psychoanalytical theory of personality, avoid a strict classification and instead depart from the assumption that all patients will show a specific disorder pattern if the biographically relevant events and influences are correlated with the development of the self.

This led to the interpersonal and sociodynamic methods used in clinical psychology, which place

greater weight on the interpersonal and interactive part of personality development, and also of personality disorders, as well as the biosocial learning theories, which take biological, intrapsychic, and environmentally specific social influences on personality disorders as their starting point (Millon 1969). Such models complement our more psychopathological concepts in interesting ways; Fiedler (1998), for example, approached personality disorders from a cognitive-behavioral direction and regarded them as personality-caused interaction disorders, and Schmitz (1999) employed a dimensional concept of dysfunctional personality styles conforming to psychoeducative aspects.

3 History of the Concepts of Abnormal Personality

3.1 French Schools of Thought

With his concept of a *manie sans délire*, Pinel (1809) pioneered scientific treatment of abnormal personality and was the first to delimit it as a nosological unit. The decisive characteristic feature was an impairment of the affective functions without a concomitant impairment of the intellect. One of his patients showed clear emotional instability and dissocial tendencies, two features that are also central to modern diagnosis. Pinel considered poor upbringing and a perverted, unrestrainable disposition as possible causes, thus giving rise to the nature versus nurture controversy, which remains current to this day. Esquirol's (1938) monomanias also show many features which are now seen as personality disorders, particularly alterations of will and emotions but not of intelligence. Unfortunately, the concept was abused insofar as perfectly circumscribed behavioral disorders came to be seen as syndromes, such as pyromania, cleptomania, erotomania, or homicidal monomania, a problem which remains rooted in modern classification systems and has led to forensically relevant misunderstandings, as in the assessment of culpability.

Particularly successful was the theory of degenerations as pathological deviances from the normal due to harmful environmental influences and heredity, severity progressing from one generation to the next and eventually leading to extinction (Morel 1857). The concept presented by Magnan and Legrain (1895) of a regular succession of certain syndromes, above all the *dégénérés supérieurs*, who showed affective peculiarities but normal intelligence, would be of significance for the future study of psychopathology. Magnan postulated a discordant interaction between the cerebrospinal centers which upset the psychic balance of the degenerate and caused a special sensitiveness. Ideas of an imbalance and a fragility due to the nervous system kept recurring in theories on abnormal personality until Dupré's *Doctrine des Constitutions* (Dupré 1925), in which he proposed the concept of a *deséquilibration mentale* in the sense of congenital psychopathic degeneration. These ideas were later combined with those from social Darwinism in many countries, where they gave rise to misguided notions about "unworthy life," but which in Germany led to the gravest consequences (Binding and Hoche 1920).

3.2 Anglo-American Traditions

In 1812, the American physician Benjamin Rush described patients with antisocial/dissocial behavior without intellectual losses, a condition which he termed "moral alienation of the mind." Prichard's (1835) concept of "moral insanity" had a similar impact in England. The increasing emphasis on individuals who showed amoral, socially harmful behavior and delinquency led to the question of whether these people were responsible for their actions. This collided with attempts at integrating "moral insanity" into a psychiatric nosology with a legalistic orientation. Later, the concept presented by Henderson (1939) of the unadapted or aggressive psychopath found its way into the Anglo-American notion of a personality disorder determined mostly by antisocial traits. The British Mental Health Act currently includes the term "psychopathic disorder" in the sense of abnormally aggressive and irresponsible behavior, a definition that has been under fire for decades (Lewis 1974).

Cleckley's monograph *The Mask of Sanity*, originally published in 1941 (Cleckley 1976), greatly influenced the American notion of psychopathy and led the way for the now important concept of the antisocial psychopath (Hare 1970, 1991). The concept of abnormal personality was also influenced by the pioneers of psychoanalytical character research, Meyer (1903) and Freud (1908). In a groundbreaking paper, Alexander (1928) distinguished between primarily self-syntonic psychopaths, whose neurotic acting out disturbs mainly those around them, and secondarily self-syntonic neurotics, who are usually their own victim.

3.3 Schools of Thought in the German-Speaking Countries

Koch (1891–1893) first employed the term "psychopathic inferiorities" for a "psychic border region" (inferiority here is closely connected with the concepts

of degeneration and must be seen somatopathologically rather than as a value judgment). He did distinguish between innate and acquired psychopathy, but classified most forms as innate psychic impairments. Beginning in 1883 and throughout the various editions of his textbook, Kraepelin developed the concept of psychopathic condition in the sense of today's understanding of abnormal personality. The concept of psychopathic personality was mentioned for the first time in the seventh edition (Kraepelin 1903/1904), where it appeared in the context of dissociality. Constitution typologies followed later, including the sketches by Kretschmer (1921) as well as other systematic typologies, e.g., the layer theory by Kahn (1928), Schultz (1928), and Homburger (1929) and the reaction typologies by Kretschmer (1921) and Ewald (1924).

As a result of Kurt Schneider's classic monograph *Die psychopathischen Persönlichkeiten* ("The Psychopathic Personalities"), originally published in 1923 (Schneider 1950a), the systematic typologies lost importance and a descriptive, analytical approach without a sociological perspective was favored. Kurt Schneider defined abnormal personalities as variations or deviations from a vaguely imagined but not precisely definable average and proceeded to point out the psychopathic personalities as such, whose abnormality causes suffering to themselves or to society. He distinguished ten types of psychopathic personality: hyperthymics, depressives, those suffering from insecurity (subdivided into anxious and obsessive–compulsive types), fanatics, show-offs, those showing mood instability, explosiveness, lack of feeling, avolition, and the asthenics. His teachings have had a decisive influence on all later concepts of personality disorder.

4 The Example of the Borderline Syndrome

As "a border region or no-man's-land" (Saß and Koehler 1983), the borderline symptom still resembles its mostly vague and poorly defined historical forerunners from which the path has led into the relative clarity of the currently used classification systems for personality disorders. In a more general form, Hughes (1884) first spoke of a "borderland" regarding unusual conditions. "Borderline" was first used as a distinct nosological term by the psychoanalyst Stern in 1938, who applied it to patients showing both psychotic and neurotic characteristics. In its eventful development leading up to today's borderline personality disorder, four theoretical currents are evident (Herpertz and Saß 1999a).

1. The first casts the borderline disorder as a subschizophrenic disorder. Kraepelin (1903/1904) already described an intermediate between pathological mental conditions and individual eccentricities, which he called a border region with undeveloped cases, *formes frustes*, of dementia simplex. Similarly, Bleuler's diagnostic category of latent schizophrenic cases included cases of peculiar, odd, eccentric individuals (E. Bleuler 1911). The Bleuler-based concepted presented by Hoch and Polatin (1949) of pseudoneurotic schizophrenia was to have serious consequences, also for psychoanalysis.
2. The second current characterizes borderline disorder as an emotionally unstable "subaffective" disorder bordering affective illness. Such a syndrome was first described by Bonet (1684) as *folie maniaco-mélancolique*. Beginning with the fifth edition of his textbook, Kraepelin described "constitutional irritations" within the psychopathic conditions. Following in the footsteps of Kraepelin, contemporary authors interpreted the borderline disorder as cyclothymic temper (Akiskal and Akiskal 1992). Kurt Schneider (1950a) included the hardening of these personalities with the emotionally unstable and explosive psychopaths.
3. The third current characterizes borderline disorder as disturbed impulse control (Clarkin et al. 1993; Herpertz and Saß 1997). This goes back to the old illness concepts of an impulsive disorder of the will (Berrios and Gili 1995). Forerunners were the *fureur sans délire* of the Swiss psychiatrist Matthey (1816) and the monomanias of Esquirol (1838), which Kraepelin grouped under "impulsive insanity." In the typology of Kurt Schneider (1950a), impulsive personality traits also fall into the category of emotionally unstable and explosive psychopaths (Herpertz and Saß 1997).
4. The fourth current, which at the moment is receiving much attention, characterizes borderline disorder as a post-traumatic stress disorder as first described in DSM-III following observations on Vietnam veterans. Up to one third of all patients meeting the DSM-III criteria for borderline disorder also showed signs of a post-traumatic stress disorder (Coryll and Zimmerman 1989). However, multiple traumas are also seen in other psychiatric illnesses (Zanarini et al. 1997), and some patients meet the criteria of borderline disorder without reporting any severe traumatic events from their childhood.

Taken together, it is evident that the rather large category of borderline syndromes evolved from a number of different forerunners and served as a diagnostic wastebasket for difficult patients. Often, it

was not used to designate any specific personality disorder, but, incorrectly, also in cases where psychosis was suspected, i.e., to conceal the physician's own diagnostic uncertainty. The decisive breakthrough needed to bring order and clarity was not to be made until the paper by Spitzer et al. (1979), who introduced the dichotomous division into emotionally unstable or borderline personality disorder on the one hand and borderline schizophrenia or schizotypal personality disorder on the other (Saß and Koehler 1983). In current diagnostic research, the borderline personality disorder has become one of the empirically best documented conditions and now serves as an example for the entire group of personality disorders.

5 Personality Disorders in the Modern Diagnostic Systems

Today's concept of personality disorder has abandoned the old controversy of neurosis (which is seen as being more acquired) versus psychopathy (which is regarded more as inherited). In the relatively assumption-free, descriptive classification systems, "personality disorder" now functions as a neutral collective term for all dysfunctional personality variants. Essentially, the modern definitions of personality disorder which appear in ICD-10 and DSM-IV go back to Kurt Schneider, since a personality disorder is diagnosed when a personality is inflexible and maladapted and this leads to clinically relevant suffering or impairment in the social, professional, or other important functions.

According to the DSM-IV and ICD-10 research criteria, diagnosis follows an operational description with a formulation of explicit inclusion and exclusion criteria in a polythetic algorithm. There is some overlap, however, since many people meet the criteria of several personality disorders. This overlap of different disorder categories is termed comorbidity, even though "joint occurrence" would be preferable, particularly with personality disorders of unclear morbidity. The so-called comorbidity often represents a methodological artifact, since identical criteria appear in the list of features of several personality disorders. The question also arises of whether the feature descriptions and the cutoff (threshold) values indicated in the criteria allow a reliable and valid assignation (Saß 1986; Herpertz et al. 1994). Furthermore, in the case of concomitant axis I disorders, a diagnosis of a comorbid personality disorder – which should only be made by means of peculiarities in the patient's medical history which existed prior to the disorder in question – would have to be uncertain (see Chap. 20, Vol. 3, Part 1; Chap. 24, Vol. 3, Part 2).

A further problem is posed by the persisting differences in terminology and concept between the classification systems, as the comparative differential typology in Fig. 1 shows. Thus, the schizotypal and narcissistic personality disorders of DSM-IV do not appear in ICD-10; instead, ICD-10 groups schizotypal disorders under schizophrenia, and the criteria for the

Kraepelin, Kretschmer K. Schneider, ICD-9	ICD-10	DSM-IV	
Fanatic Schizoid ∅	Paranoid Schizoid ∅	Paranoid Schizoid Schizotypal	A
Explosive, emotionally impoverished Moody Needing recognition ∅	Dissocial Emotionally unstable { - Borderline type / - Impulsive type } Histrionic ∅	Antisocial Borderline Histrionic Narcissistic	B
Insecure Weak-willed Obsessive-compulsive	Insecure Dependent Anancastic	Insecure Dependent Obsessive-compulsive (Passive-aggressive)	C
Depressive Asthenic Hyperthymic Cyclothymic	∅ ∅ ∅	(Depressive) ∅ ∅ ∅	+

Fig. 1. Personality disorders in the classical typology and in the modern classification systems. ∅ suggested research criteria; + "subaffective" forms

narcissistic personality disorder are found under those for the paranoid type. On the other hand, ICD-10 includes an emotionally unstable personality disorder which is divided into an impulsive type and a borderline type. In DSM-IV, the latter appears as an independent form of the borderline personality disorder. Moreover, the diagnostic threshold values also differ, with the result that prevalence rates are higher when applying ICD-10. These discrepancies between the systems make it difficult to compare studies and complicate routine diagnosis.

By now, the great variety of criteria in the modern classification systems has made operationalized diagnosis so complex as to make systematic classification a major undertaking. Moreover, diagnosis of personality is met by several relatively specific hindrances (Bohus et al. 1999). Thus, criteria for the diagnosis of personality disorders deal with subtle phenomena which must be seen from a larger chronological perspective. There can be considerable discrepancies between a patient's self-assessment and the opinion of significant others or visitors due to self-syntonia or self-dystonia. For this reason, a patient's self-assessment is of little diagnostic help, and the physician must often rely on information from other people. Thus, when using the *Aachener Merkmalsliste zur Erfassung von Persönlichkeitsstörungen* (AMPS; Aachen List of Features for Assessing Personality Disorders), we relied on quite diverse information, including a diagnostic interview with the patient, the case history, behavior observation, and a chronology of the course of illness from the patient's file (Saß and Mende 1990). For the difficult task of assessing premorbid personality disorders by analyzing case histories, special biographical inventories are recommended (von Zerssen 1994a). Table 1 shows the standardized examination procedures for assessing personality disorders which are currently available in the German-speaking countries.

6 The Five-Factor Personality Model and Its Relation to Personality Disorders

The advantages of the category-based or typological approaches currently used in psychiatry are their simple conceptualization, communicability, their wide distribution, and a general human tendency to perceive and decide in terms of categories (Saß 1986). Unfortunately, these advantages are offset by serious shortcomings. The validity of a category-based or typological model is compromised by frequent multiple diagnoses which are a result of the polythetic set of criteria. Many of the cutoff points seem arbitrary, and features falling only a little below this limit are ignored, thus leading to a relatively low interrater reliability in the diagnosis of personality disorders.

Since many psychologists dealing with personality disorders are of the opinion that a medical category-based diagnostic system cannot do justice to behavioral disorders, the concept of personality disorders is still not used in many areas of clinical psychology. Instead, such psychologists attempt to represent the various personality disorders with a dimensional personality trait model. Individuals are not classified using a category-based or typological system, but instead the various disorder features are assigned a position according to intensity in a multiaxial dimensional space. For the transition from healthy to disturbed personality, they favor a continuity model. The forerunners of dimensional models in psychiatry originate from, among others, C.G. Jung (1921), who introduced the introversion/extroversion dimension, and Kretschmer (1921), who described the dimensions of schizothymia and cyclothymia. Eysenck's (1952) influential dimensional personality model uses factor analysis to reduce the number of possible personality features to the three dimensions extroversion (sociability, activity), neuroticism (emotional instability), and psychoticism (impulsiveness, aggressiveness). Cloninger et al. (1994) also made important contributions to empirical personality research with his biopsychological personality theory, on the basis of which he developed four temper dimensions (harm avoidance, sensation seeking, reward dependence, perseverance) and three character dimensions (self-directedness, cooperativeness, and self-transcendence).

Fiske (1949) introduced a five-factor personality model, known in the English-speaking world as the Big Five (Costa and McCrae 1990). According to this model, there are five robust personality factors, i.e. neuroticism, extroversion, openness to new experiences, sociability, and conscientiousness. Von Zerssen (1994b) also includes piety as a sixth dimension and defines the sociability factor by its contrary pole, aggressiveness. The disadvantages of category-based operationalization are mostly avoided by a dimensional model, and empirical comparison of the two approaches suggests that the dimensional model is superior, at least as far as research is concerned (Rutter 1987; Widiger and Costa 1994).

Varying opinions of the dimensionality of personality disorders, as well as of the relationship between personality dimensions and personality disorders, also depend on which statistical tests are employed (Pukrop et al. 1998). Correlations are calculated, usually indirectly, using factor analysis assuming orthogonal dimensions factors, which are assigned on both a clinical and an intuitive basis. Recently, facet theory has been suggested to analyze the correlation between personality disorder and personality dimensions by

Table 1. Examination procedures for assessing personality disorders (after Bohus et al. 1999)

Type of examination	Instrument	Diagnosis systems	Source of data
All personality disorders			
Self-assessment	Personality Diagnostic Questionnaire (PDQ-R)	DSM-III-R/DSM-IV	P
	Screening Test for Co-morbid Personality Disorders (STCPD)	DSM-III-R	P
	Computerized DSM-III-R Personality Disorder Questionnaire (CDPDQ)	DSM-III-R	P
	The Temperament and Character Inventory (TCI)	–	P
	Dimensional Assessment of Personality Pathology Disorder–Basic Questionnaire (DAPP-BQ)	–	P
Checklists	Internationale Diagnosechecklisten –Persönlichkeitsstörungen (IDCL-P; International Diagnosis Checklists–Personality Disorders)	ICD-10/DSM-IV	P, I, R, H
	Aachener Merkmalsliste zur Erfassung von Persönlichkeitsstörungen (AMPS; Aachen List of Features for Assessing Personality Disorders)	ICD-10/DSM-IV	P, I, R, H
Interview	Strukturiertes Klinisches Interview für DSM-IV–Persönlichkeitsstörungen (SKID-II; Structured Clinical Interview for DSM-IV–Personality Disorders)	DSM-IV	P, R
	Standardized Assessment of Personality (SAP)	ICD-10/DSM-III-R	I
	International Personality Disorder Examination (IPDE)	ICD-10/DSM-IV	P, R, I
Individual personality disorders			
Self-assessment	Borderline Syndrome Index (BSI)	DSM-III-R	P
	Narcissism Trait Scale (NTS)	DSM-III	P
	Borderline-Persönlichkeits-Inventar (BPI; Borderline Personality Inventory)	–	P
Interview	Schedules for Interviewing Borderlines (SIB)	DSM-III	P
	Diagnostic Interview for Borderline (DIB-R)	–	P, R
	Diagnostic Interview for Narcissism (DIN)	–	P, R

P, patient; I, informant; R, rater; H, history of disease.

means of a similarities matrix between an a priori established semantic frame, the definitions system, and an empirical data structure, whose contents correspond to this system of definitions. Using this facet theory model, we empirically tested the structural similarity correlations between personality dimensions and personality disorders, confirming the continuity model and finding a meaningful relation between personality disorders and personality dimensions (Saß et al. 1995; Herpertz et al. 1997b).

7 Epidemiology, Course of Illness, and Prognosis

According to German and American studies, 3%–10% of the general population meet the diagnostic criteria for a personality disorder (Reich et al. 1989; Zimmerman and Coryll 1990; Maier et al. 1992). Compared to values seen in earlier personality diagnosis, these numbers are rather high. However, simply meeting

these criteria need not imply that the individuals are so dysfunctional and impaired as to require treatment.

Personality disorders occur more often in urban than in rural populations and tend to be more frequent in socially weaker populations. The overall sex ratio for all personality disorders is the same here, although some of them are seen more frequently in women, e.g., dependent and borderline personality disorders, while others are found more commonly in men, e.g., antisocial personality disorder (Merikangas and Weissman 1986; Weissman 1993). While these differences in frequency could indicate real gender differences, the confounding influence of social stereotypes and gender roles should not be overlooked.

Prevalence rates are much higher among nonselected psychiatric patients. After the first classification systems appeared, clinical studies found surprisingly high frequencies for personality disorders of 50%–80%, while more recent studies reported average prevalence rates of 40%–60% (Mellsop et al. 1982; Saß and Mende 1990; Oldham et al. 1992; Herpertz et al. 1994). Random samples in forensic psychiatry yielded prevalence rates of up to 80% (Dilling et al. 1984; Merikangas and Weissman 1986; Saß 1986). In a wide-scale international World Health Organization (WHO) study (Loranger et al. 1994), 39.5% of 716 psychiatric patients examined showed at least one personality disorder according to ICD-10, where individual prevalence rates fell between 15.2% (anxious personality disorder) and 1.8% (schizoid personality disorder).

The idea that personality disorders are permanent was always a part of the concept of abnormal personality and is also found in the ICD-10 and DSM-IV diagnostic guidelines. However, many authors criticize the postulate of temporal stability of personality disorders and stress the importance of situational factors. Instead of continuity and contextual independence, many cases probably exhibit a long, slowly changing course (Loranger et al. 1991). Recent overviews of empirical studies on the stability and changeability under pharmacological or psychosocial therapy partly confirm and partly refute earlier assumptions, considerable changes being observed especially with depressive forms (Sanislow and McGlashan 1998; Grilo and McGlashan 1999).

Clinical experience has shown that increasing age and decreasing vitality tend to attenuate hardened personality traits. This applies especially to those personality traits which seriously impair social functioning, such as inconstancy, dissocial behavior, autodestructiveness, and impulsiveness. Other traits can harden with advancing age, above all obstinacy and rigidity. Despite all their fundamental stability, personality disorders show more flexibility and adaptability than was earlier supposed. The degree of behavioral oddities depends on the particular situation and the demands to be met during the period of life in question, and developmental challenges can lead to a crisis. Thus, an anxious personality in early adulthood, faced with establishing autonomy, making decisions, and finding a meaning in life, will suffer acutely from its deficient social competence, a dependent or depressive personality will most likely suffer when having to cope with a loss or separation, while a histrionic personality will have great difficulties compensating age-related losses in looks and vitality.

Generally speaking, the risk of suicide is three times higher for individuals with a personality disorder than for the general population, with borderline, narcissistic, and antisocial personality disorders showing the highest numbers. These groups also show the highest degree of psychosocial impairment, with deviant actions, decreased capacity for work, and deficient skills at establishing dependable interpersonal relations. The latter is otherwise seen only in paranoids and schizoid personality disorder, with their severe tendency for social isolation. Still, catamnestic examinations over relevant periods of 10–30 years have shown sufficient social integration in two thirds of the patients even in these subgroups of severely disturbed, emotionally unstable personalities. The grave social dysfunction seen in borderline personality disorder and the repeated autodestructive actions, especially early in life, often lead to prolonged inpatient stays. Individuals with an anxious, dependent, or obsessive-compulsive personality disorder will confide in the family doctor, and occasionally in a specialist, but only with the onset of additional symptoms which warrant inpatient hospital treatment, such as anxious, dependent, or obsessive–compulsive personality disorders require. Individuals with other personality disorders, e.g., schizoid and paranoid forms, are unlikely to seek help by themselves.

The prognosis will depend on the particular type of personality disorder, any eventual comorbidity, and the degree of severity. Further prognostic factors are psychostructural maturity and the level of psychological and social functioning. In a catamnesis of 539 inpatients in the Federal Republic of Germany, Tölle (1966) found that about a third each managed their own life in a favorable, compromising, or unfavorable way, respectively. Histrionic, obsessive–compulsive, dependent, and insecure personality disorders had a good prognosis; despite an often difficult and drawn-out course of therapy, success has also been reported for borderline and narcissistic personality disorders. Prognosis is less favorable for schizoid, paranoid, and, notably, antisocial personality disorders. Prognostically favorable characteristics are motivation, trust in others, flexibility, and insight into one's own role in difficulties with interpersonal contact. Cases are com-

plicated by concomitant illness, especially addiction and affective disorders. Thus, the mortality rate for patients with personality disorders and substance abuse is three times that of patients with a single personality disorder (Bohus et al. 1999).

8 Basics of Therapy

The therapy of personality disorders employs psychotherapy, therapy with psychopharmaceuticals, and sociotherapeutic treatment. An important development in recent years has been the integrated, interdisciplinary approach (Saß and Herpertz 1999). Generally, the treatment of personality disorders requires time, as it involves the gradual reshaping of deep-seated characteristics pertaining to life, health, and social behavior. Crisis management becomes necessary if these personality traits generate problems or social conflict. It may be helpful if the patient's partner (spouse or family) can be involved. The choice of therapy will depend on the particular personality disorder and its degree of severity, as well as any concomitant psychiatric illness. The physician must always make a careful psychopathological and physical examination to detect any complicating somatic or psychiatric illness and to assess the need for medication. Psychotherapy relies mainly on supportive talks and behavioral techniques, to a lesser extent also on depth psychology and analytical therapy. In recent years, cognitive methods have gained much attention in the treatment of personality disorders, with differential indication and differential psychotherapy having become increasingly elaborated (Fiedler 1999). As far as the patient is concerned, this involves concrete, prototypic interactive disorders. Psychological and sociotherapeutic strategies can be used as complementing measures, e.g., to improve the patient's social situation or to initiate rehabilitation. Regardless of the method chosen, the establishment of a reliable, regular, and confidential therapeutic relationship is vital, as it provides the basis for one of the most important general therapeutic factors, which is the possibility of corrective emotional experiences. In addition to individual therapy, group therapy is increasingly being used now, but the patients must be compatible (Schmitz 1999; Rennenberg and Fydrich 1999).

The aim of pharmacological therapy, which is gaining in importance, is to influence the biological disposition toward dysfunctional behavior and perception. The theory can be traced back to Kraepelin and Kretschmer, who held that abnormal personality features are incompletely formed, weakened forms, or *formes frustes* of the two main groups of endogenous illnesses. This is why the psychopharmaceuticals used to treat schizophrenic and affective psychoses are also employed in the therapy of (probably) related personality disorders in order to reduce the neurochemical vulnerability to affective and cognitive dysfunctions. Pharmacological treatment is not an alternative, but merely a measure intended to support psychotherapy, although in some cases psychotherapy is only possible under the influence of stabilizing medication (Saß et al. 1998; Kapfhammer 1999; Kapfhammer and Rothenhäusler 1999). There is danger inasmuch as psychopharmaceuticals can lead to unrealistic expectations, chronic complaints about side effects, or parasuicidal actions. The possibility of supportive medication should be explained to the patient prior to the beginning of therapy. Possible detrimental effects which might carry over into the doctor–patient relationship also need to be considered. Sedatives should be approached with care, as their prolonged use can interfere with a patient's quality of life and, particularly, the ability to work. Utmost care must be taken to avoid dependence, and, except for acute crisis management, the use of benzodiazepines is particularly inadvisable for the treatment of personality disorders, which by definition have a long course of illness.

9 Individual Personality Disorders

The following section is based on the two main international classification systems. Due to the international preeminence of Anglo-American research, DSM-IV plays the leading role, but as the official diagnosis encoding system used in the German-speaking countries, ICD-10 will also be of great practical relevance as of January 1, 2000. A comment will be added wherever there is more than a slight divergence between the two systems. Further viewpoints regarding individual personality disorders will be dealt with in their own sections, which are also detailed here (e.g., Beck et al. 1993; Schmitz et al. 1996; Fiedler 1998; Bohus et al. 1999; Saß and Herpertz 1999). The section devoted to each personality disorder is divided into the historical development of the concept, a description of the clinical picture, diagnosis and differential diagnosis, prevalence, etiopathogenetic aspects, and therapy.

Since the first publication of DSM-III (APA 1980), a personality disorder has been assigned to one of three superordered clusters. Cluster A consists of personality disorders characterized by odd, eccentric behavior, i.e. the paranoid, schizoid, and schizotypal personality disorders. Cluster B consists of personality disorders characterized by emotional, dramatic, or moody

behavior, i.e., the antisocial, borderline, histrionic, and narcissistic personality disorders. Cluster C consists of personality disorders which can, in Kurt Schneider's terms, be described as anxious, fearful, or asthenic, i.e., the avoidant-insecure, dependent, and obsessive-compulsive personality disorders.

However, this cluster method, which was retained from DSM-III through to DSM-III-R and DSM-IV, has not yet been well validated empirically and is based mainly on descriptive similarities. On the whole, research by Saß et al. (1995) on a cohort of psychiatric patients with personality disorders using hierarchical cluster analysis confirmed the DSM grouping, as did research by other authors (e.g., Morey 1988), although the obsessive-compulsive personality disorder seemed to fit Cluster A better. This would seem plausible characterologically, as obsessive-compulsive subjects are characterized by rigidity, energy, perseverance and by emotional distance, similar to the other disorders in Cluster A.

9.1 Cluster A (Odd and Eccentric Personality Disorders)

9.1.1 Paranoid Personality Disorder

History

The term "paranoid personality disorder" describes a proximity to delusional (paranoid) disorders as well as to paranoid schizophrenia, but manifestations usually remain below the threshold of psychotic disorders. Paranoid features in the sense of odd personality features are mentioned in Kraepelin's textbook as "pseudo-grouch" (*Pseudoquerulant*) and "cantankerous," and Kretschmer (1921) referred to them as "expansive personalities." Kurt Schneider (1950a) mentioned the active and unrestrained nature of the fanatic and belligerent person, but also the quiet, eccentric, unrealistic, and purely fantastic forms in subdued fanatics.

Description

According to DSM-IV, individuals with a paranoid personality disorder show a pattern of distrust and suspicion, reading malevolence into the motives of others. This affects their perception not only of their friends' and their partner's loyalty and credibility, but also harmless comments and everyday behavior of people whom they come into contact with. People with a paranoid personality disorder quickly react to perceived insults with excessive force, counterattacks, and/or long-lasting enmity. They are especially sensitive about failure and supposed discrimination and are easily insulted, emotionally rigid, persevering, and belligerent, while at the same time appearing humorless and emotionally blunt. Typically, they attach exaggerated value to certain personal or imagined complexes, which can causing them to quarrel with others. Mistakes are usually blamed on others.

Typical Thinking

"I can't trust anyone," "they're trying to manipulate and exploit me," "they want to humiliate or annoy me" (Beck et al. 1993).

Differential Diagnosis

According to DSM-IV, the paranoid personality disorder is distinguished from the delusional disorder with paranoid delusion by the clearly delimited, persistent, and unrealistic delusions shown by the latter. The paranoid personality disorder, on the other hand, often shows a very real vicious circle of self-fulfilling prophecies, since the persistent distrust easily leads to information really being withheld or denied, which in the patient's mind confirms the suspicions. Occasionally, and especially during stressful times, a patient with a paranoid personality disorder may experience a loss of reality control for brief periods lasting from minutes to hours similar to that seen in psychotic episodes.

One of the special forms described by German authors is the (retiree) grouch. Sometimes, complex personalities – belligerent persons with a secret wound, a "thorn in the side" – develop a "fighting paranoia" (Kretschmer 1921). These "fighting paranoids" can be forensically important not only when they pester and insult courts and authorities, but also when they tend toward acts of violence, and extreme cases may need to be committed to an institution. Ideational fanatics follow an overvalued idea which they rank above everything else and may, depending on the times and circumstances, attain great religious or political significance or commit warlike terrorist acts and even induce others to commit (mass) suicide. In their biosocial personality theory, Millon and Davis (1996) distinguished five types of paranoid personality disorders: paranoid-narcissistic, paranoid-antisocial, paranoid-compulsive, paranoid-aggressive, and passive-aggressive and the decompensated form, the latter with schizophrenic tendencies.

Prevalence

According to studies employing DSM-IV, 0.5%–2.5% of the total population, 2%–10% of psychiatric outpatients, and 10%–30% of psychiatric inpatients showed paranoid personality disorders. Of a group of 291 outpatients with personality disorders, 22% showed a paranoid personality disorder according to DSM-III-R (Morey 1988). Using the International Personality Disorder Examination (IPDE), Loranger et al. (1994) found 2.4%, a clearly lower ratio. Regarding comorbidity with other personality disorders, the commonest are narcissistic personality disorder (2%–75%), borderline personality disorder (0%–100%), insecure personality

disorder (8%–86%), and passive-aggressive personality disorder (17%–53%) (Bernstein et al. 1993).

Etiopathogenesis
Seen psychoanalytically, delusion with projection plays a prominent role, at least with pronounced paranoid symptoms. The patient wards off conflict with his or her own aggressive impulses by transferring them onto others and perceiving them as persecution. Seen from a cognitive perspective, these paranoid patterns of behavior are an attempt at dealing with feelings of shame and humiliation. Paranoid individuals usually see themselves as being inadequate, incomplete, and useless, and so others are blamed for problems and difficulties in order to counteract these unpleasant feelings (Beck et al. 1993). Genetic findings are etiologically important, as the prevalence of paranoid personality disorder is higher among relatives of chronic schizophrenic patients and of patients with a delusional paranoid-type disorder (Kendler et al. 1984).

Therapy
See Sect. 9.1.4.

9.1.2 Schizoid Personality Disorder

History
Observing relatives of schizophrenic patients, E. Bleuler (1911) noticed that, even though they themselves had no psychotic illness, they did show certain behavioral peculiarities such as social isolation or a peculiar way of communicating and, like Kraepelin (see Sect. 4), suspected a transitional form with fluid borders between schizoid character and latent schizophrenia, with its typical thought disorders and social dysfunction. This led Kretschmer (1921) to hypothesize a continuum between premorbid personality features and schizophrenic illness, i.e., a transition from schizothymic temper to schizoid type and to psychosis. Bostroem (1926), on the other hand, disagreed with Kretschmer's hypothesis of a continuously rising degree of severity along a uniform schizoid dimension up to schizophrenia and held that, in this case, the development of specifically schizophrenic symptoms should be possible to predict from the oddities of an individual's character.

Description
The chief characteristics of the schizoid personality disorder are indifference and an inability to develop close emotional attachments to others, seclusiveness, and a reduced capacity for emotional experience and expression. Individuals with a schizoid personality disorder are shy, reserved, and seemingly indifferent to both praise and criticism. Bleuler vividly described schizoid persons as rigid, impenetrable, dry, and cold, as odd cranks, reserved eccentrics, cool and inwardly highly sophisticated aristocratic types, or scatterbrained, indifferent, and emotionally blunt odd fellows. Descriptions in Kraepelin's textbook are similar: he also noticed that dementia praecox patients often were particularly quiet and withdrawn as children.

Typical Thinking
"I feel better when I'm alone," "relationships confuse me," "I don't care what others think about me" (Beck et al. 1993).

Prevalence
The schizoid personality disorder seldom appears in a clinical setting, and prevalence lies well below 1%. According to field studies in the United States (Kalus et al. 1993), prevalence in the general population is 0.5%–1.5%. Regarding familial clustering, there is a higher prevalence of schizoid personality disorder among relatives of schizophrenic patients or those with schizotypal personality disorder.

Differential Diagnosis
The most difficult distinction to be made is that from mild forms of autism and Asperger's disorder, which also show reduced social contact, but tend to be noticed much sooner (see also Chap. 7, Vol. 3). The latter also show stereotyped or ritual behavior, motor mannerisms, and a narrowing of interests. Otherwise, schizoid personality disorder can also show brief psychotic episodes lasting a few minutes to a few hours when under stress, whereas in delusional disorders and schizophrenia these phenomena are permanent. In order to be able to make a diagnosis of schizoid personality disorder, the personality abnormalities must be present before and after the psychotic episode.

Regarding the other personality disorders, there is some overlap of symptoms with the schizotypal and the paranoid personality disorder, such as social isolation and affective narrowing, but the schizoid form lacks the cognitive losses and distorted perception of the schizotypal disorder, nor does it show the increased distrust and paranoid ideas typical of the paranoid personality disorder. It is important to distinguish it from the avoidant-insecure personality disorder, which also shows social isolation, but which in this case is due to social fear overriding a wish to engage in social contact. Schizoid individuals have little desire for social intimacy and therefore prefer to withdraw from it.

Etiopathogenesis
Regarding correlations to the schizophrenic spectrum, more psychological abnormalities were found around schizophrenia, but no significant increase in schizoid or schizotypal personality disorders. Nor is there an obvious familial clustering of personality types among relatives of schizophrenic patients which would point

to a homogenous schizoid dimension, but rather a quite heterogeneous set of abnormalities. In large-scale, albeit retrospective studies, Huber et al. (1979) and M. Bleuler (1972) found no clear premorbid schizoid personality features in two thirds and one half of the patients, respectively. Prospective high-risk studies also failed to show any significantly pronounced schizoid traits in children who later developed an illness of the schizophrenic spectrum compared to children who remained psychiatrically unremarkable (Parnas et al. 1990). On the other hand, von Zerssen (1994a) also noticed definite schizoid traits in patients showing personality disorders combined with high neuroticism. Generally speaking, although a survey across the schizophrenic spectrum revealed no close ties between schizophrenia-like premorbid personality traits and the severe forms in the sense of a specific schizothymic-schizoid-schizophrenic dimension, a clustering of general characteristics indicating mental susceptibility to disorders could be observed.

Seen psychoanalytically, schizoid individuals have particular difficulties expressing fear and enmity, which may be due to a troubled early mother–child relationship (Fairbairn 1940). Likewise, Arieti (1955) postulates that schizoid personalities develop a primarily defensive structure as a consequence of early problems, the patients being so sensitive to rejection as to repress the value of a relationship. Behavioral explanations also stress problems in dealing with feelings in the context of interpersonal relationships (Turkat 1990). The biosocial learning theory of Millon and Davis (1996) holds that pampering and protective parents teach their children to avoid stress. Lacking interpersonal experience, they are ill-equipped for handling conflicts and sticking them out, which further favors withdrawal and isolation.

Therapy

See Sect. 9.1.4.

9.1.3 Schizotypal Personality Disorder

History

The relationship of schizotypal personality disorder to schizophrenia is similar to that listed for the schizoid personality disorder. We would also like to refer to the history of the borderline syndrome, which has always included some "subschizophrenic" syndromes (see Sect. 9.2.2), as well as the schizophrenic spectrum and the borderline schizophrenia described in Danish family and adoption studies (Parnas et al. 1990).

The concept of "schizotypy," to which research by Hoch and Polatin (1949) on pseudoneurotic schizophrenia also contributed, was developed by Rado (1953) in the sense of a "schizophrenic genotype." In DSM-I (APA 1952), the schizoid personality disorder still included the now separate schizotypal and insecure personality disorder. Millon (1981) distinguished between schizoid-schizotypal and insecure-schizotypal personality variants. Spitzer et al. (1979) finally clarified the issue with their operational criteria for schizotypal personality disorder.

Description

The schizotypal personality disorder is characterized by even stronger deficits in establishing interpersonal contact than seen in the schizoid disorder. Furthermore, it shows peculiarities of perception, thinking, and behavior, disorders of attentiveness, selective attentiveness, and filtering of stimuli, cognitive gliding, and increased association via unusual selection and rating of information. Regarding language, there are unclear or strange expressions and incorrect use of words, though not to the point of associative loosening and incoherence. These individuals are loners, uncomfortable in the company of others, and do not have a good grasp of the usual means of communication such as eye contact, body language, etc. In addition, they show a pronounced fear of social situations and actively avoid them. People with this disorder often develop magic thinking, belief in the occult, and superstitions. They also tend to show odd behavior, as in their choice of dress and movement.

Typical Thinking

"I'm extremely uncomfortable when strangers talk to me," "when other people are talking, they probably have known each other for a long time and don't want me around," "I don't belong there" (Beck et al. 1993).

Differential Diagnosis

Due to close symptomatological and pathogenetic correlations to the schizophrenic spectrum, ICD-10 groups "schizotypal disorder" under schizophrenic and delusional disorders, and its symptoms overlap with those described for schizophrenia simplex in ICD-10. When distinguishing schizotypal personality disorder from social phobia in a differential diagnosis, it is important that individuals suffering from social phobia gradually relax when in contact with others, whereas those with a schizotypal personality disorder become increasingly uneasy. The schizotypal personality disorder is distinguished from schizophrenia, delusional disorder, or affective disorder with psychotic features by its lack of hallucinations and delusions. When distortions of thinking and perception occur, such as erroneous interpretation of coincidences, of comments by others, or of situations, individuals with a schizotypal personality disorder do not rigidly cling to them as do those with delusional disorders. However, similar to the paranoid and the

schizoid personality disorders, there can be brief psychotic episodes, especially when under stress. On the whole, the most difficult distinction to make is that from schizophrenia simplex or the residual type, and it should be borne in mind that it is not always possible to exclude a past schizophrenic episode with acute symptoms.

A differential diagnosis is particularly difficult in children showing social isolation, eccentric behavior, and unusual language. When making a differential diagnosis, it is important to consider autism or Asperger's disorder, as well as specific speech and communication disorders. Regarding other personality disorders, there are close similarities to the paranoid and schizoid personality disorder and further to the avoidant-insecure, narcissistic, and, above all, borderline personality disorder, which can also show psychosis-like symptoms, but is characterized by a completely different pattern of affective reactions and shows a different social behavior, social withdrawal, and avoidance of close friendships occurring only rarely.

Prevalence

According to DSM-IV, the disorder appears in 3% of the general population. There is a familial clustering inasmuch as schizotypal personality disorder tends to be commoner in first-degree relatives of schizophrenia patients, and a slight clustering of schizophrenia or other psychotic disorders has been observed in relatives of those with schizotypal personality disorder. However, the course of schizotypal personality disorder as defined in DSM-IV only rarely shows a transition to schizophrenia or another psychotic disorder.

Etiopathogenesis

Since the Danish high-risk studies on twins and adopted children, the schizotypal personality disorder is considered to be closely related to disorders at the core of the schizophrenic spectrum (Kendler et al. 1984). Family studies on schizotypal index subjects yielded corresponding findings (Battaglia et al. 1995). The hypothesis of schizotypal personality disorder as the expression of a genetically determined susceptibility to schizophrenia is supported by schizophrenia research employing biological markers, such as disorders of eye movement when following an object, disorders of attentiveness and information processing, and some abnormal social behavior.

Continuing Rado's (1953) search for a genetic component of schizophrenia, Meehl (1990) postulated an "integrative neural defect" which he called "schizotaxia" and which he saw as the only inherited factor. Individuals showing this trait were designated schizotypal or schizotypic, a term distinct both symptomatologically and theoretically from the purely descriptive DSM-IV term. Meehl believed schizotaxia to arise out of four central behavioral traits of schizotypal patients: cognitive gliding as the mildest form of the schizophrenic thought disorders; aversion toward interpersonal relationships and social fear due to a conviction of not being lovable; anhedonia with pronounced, general, and persistent impairment of the capacity for joy; and, lastly, a distinct ambivalence. The much-employed schizotypy questionnaire by Chapman et al. (1976), which is based on Meehl's concept of schizotaxia, encompasses 97 items for the three dimensions physical anhedonia, distorted perception, and magic thinking.

9.1.4 Therapy of the Cluster A Personality Disorders

Establishing an open, trusting doctor–patient relationship is of the utmost importance in psychotherapy. This relation can then be extended to relationships with others. In doing so, the considerable fear of closeness and commitment – which is also the reason why those with schizoid or schizotypal personality disorder in particular seldom seek out treatment by themselves – should never be forgotten. More usually, it is either because other people near them have taken the initiative or because there are further, strong fears, depression, or social difficulties which cause them to seek help. The distrust, the readiness to take offense, and the latent aggressiveness of patients with paranoid personality disorder require the therapist to have a great deal of patience, flexibility, and empathy. When treating patients with schizoid personality disorder, the main difficulties in establishing a therapeutic relationship are due to the fearful detachment, the inability to make contact, and the stunted affect.

Paranoid personality disorder should be treated using behavioral therapy. In view of the distrust and touchiness, care should be exercised when making references to the patient's background or psychodynamic interpretations. It is important to bear in mind the patient's tendency for erroneous interpretation of situations, the friend/enemy mentality, and specific styles of acting and interacting. Cognitive therapy requires considerable time before trust is established and the pathological thinking can be approached and dealt with (Beck et al. 1993). It is further important to improve the patient's psychosocial skills and structure his or her immediate social environment.

When treating schizoid personality disorder, the main goals are improving the patient's capacity for perception and for experiencing life more fully. This can be done using sensory, body awareness, and emotional refinement exercises. Behavior therapy should aim at training social skills and dealing with fear in social situations.

A similar approach is used for the schizotypal personality disorder, although patients tend to show cognitive distortions in thinking and in judgment, especially in interpersonal emotional situations. A combination of cognitive and behavior therapy is recommended for detecting and identifying automatic thinking and its underlying assumptions (Beck et al. 1993). The personality awareness of these patients is characterized by the belief that they are largely responsible for situations. Appropriate behavior can be shown in role-playing (acting out) and practiced in exercises involving social situations.

There are no strictly controlled studies on the pharmacological treatment of paranoid and schizoid personality disorders (Saß et al. 1998). Due to the similarities to schizophrenia shown by the schizotypal personality disorder, neuroleptics can be employed, especially with stronger schizophrenia-like symptoms, particularly since the recent atypical drugs cause fewer and less severe side effects. Concomitant depression, fear, agitation, or exacerbation of paranoid tendencies can be dealt with using symptom-oriented psychopharmacological therapy in all three disorder types. Anhedonia and lacking motivation have often been treated with tricyclic antidepressants and monoamine oxidase (MAO) inhibitors, but the new selective serotonin reuptake inhibitors (SSRI) should also be considered, although there is the danger of excessive dynamic activation and hence of causing a psychotic episode.

9.2 Cluster B (Emotionally Unstable Personality Disorders)

9.2.1 Antisocial Personality Disorder

History

The meshing of the concepts of abnormal personality and social deviance was treated in detail in the section on the history of ideas (see Sect. 3). The socially deviant personality forms are now given different diagnostic criteria, i.e., antisocial personality disorder (DSM-IV) and dissocial personality disorder (ICD-10), also called "psychopathy" by Hare (1970, 1991). The differentiation of personality disorders from pure dissociality without additional psychopathological peculiarities is of importance, especially in forensic psychiatry. This requires a distinction into more pathic and more dissocial abnormal personality variants (Saß 1987), thus yielding the following differentiation:

1. Personality disorders occur in individuals who suffer from their psychopathological peculiarities and/or whose social life is impaired by these peculiarities. Their symptoms resemble those of psychiatric patients in the strict sense.
2. Moreover, some of these individuals show a potential for conflict, as their social behavior is marked by deviance and delinquency and is evidently related to psychopathological abnormalities. Due to the close correlation between social deviance and psychopathological abnormalities, the designations antisocial personality disorder (DSM-IV) and dissocial personality disorder (ICD-10) would seem justified.
3. Some individuals show a clear and persistent disposition for deviant and delinquent behavior without psychopathologically relevant abnormalities throughout their life. This criminologically important core group corresponds to the "psychopathy" described by Hare (1970, 1991) in the strict sense: it usually shows a "dissocial character structure" (see below) and can now also be defined quite well biologically (Herpertz and Saß 1999b).

Only by means of a differentiation such as this can forensic questions of culpability, prognosis, and therapy (Saß 1987) be settled. On no account should we speak of a personality disorder when dealing with recurring social deviance and delinquency, as shown by chronic repeat offenders or professional criminals, since this diagnostic term can lead to erroneous connotations of an illness-like disorder.

Description

According to DSM-IV, the main characteristic of the antisocial personality disorder is a permanent and deep-seated tendency to infringe on and abuse the rights of others. Empirically, the "dissocial character structure" (Saß 1987) of these individuals shows the following: little introspection and self-criticism, lacking empathy, coldness, egotism, an exaggerated sense of entitlement, a paradoxical idea of adaptation, and weak or faulty social norms. Their behavior is marked by impulsiveness, unreliability, weak commitment, and an absence of guilt feelings.

Typical Thinking

"Other people are weak and deserve to be exploited," "if I want something, I'll do whatever is necessary to get it," "we live in a jungle where only the strongest survive" (Beck et al. 1993).

Differential Diagnosis

The DSM-IV definition of the antisocial personality disorder causes great difficulties when making a diagnosis, as can be seen from the rather high frequency of this diagnosis, the temporal stability of the diagnosis, insufficient consideration of symptom intensity, and the large overlap with the symptoms of substance abuse (Cunningham and Reidy 1998; Herpertz and Saß 1999b). Thus patients scoring high on Hare's (1991) psychopathology checklist also show a worse psychosocial adaptation and more attempts at

suicide in their case history (Anderson et al. 1999). Despite a continued development of the concept since DSM-III, the description in DSM-IV still lists only criminal and socially harmful behavior patterns without considering indicators of a deep-reaching personality dysfunction going beyond mere criminal acts. A further point of criticism is that insufficient attention is given to gender differences, as studies on antisocial personality disorder in women have shown that early onset of antisocial behavior in young girls shows only a weak correlation with antisocial behavior in adult women (Rutherford et al. 1999).

When making a differential diagnosis, an exclusion criterion is antisocial behavior occurring exclusively in the context of schizophrenia or a manic episode. On the other hand, an antisocial personality disorder often occurs together with conditions caused by psychotropic substances, in which case both diagnoses must be made. Regarding other personality disorders, the antisocial personality disorder is closely related to the narcissistic personality disorder, though the latter does not show tendencies for impulsiveness, aggressiveness, or cheating. There may be a partial overlap of symptoms with the histrionic, the borderline, and the paranoid personality disorder, although as a rule these do not show such lasting antisocial behavior.

Prevalence

Reports of prevalence are strongly influenced by spot checks and gender. In the general population, prevalence is about 3% for men and 1% for women. In clinical institutions it can be 3%–30%, with clearly higher values for addiction treatment centers or prisons. Studies on antisocial personality disorder (DSM-IV) in prison inmates gave a prevalence of 70%–100%, which would suggest that the category is too general (Widiger et al. 1996), whereas a study using Hare's psychopathy checklist (1991) on various forensic populations in prisons or psychiatric clinics revealed a diagnostic prevalence of 25% for psychopathy (Herpertz and Saß 1999b).

First-degree blood relatives of individuals with antisocial personality disorder show a higher prevalence of this disorder than the general population. Furthermore, there is a familial relation to somatization disorder and disorders linked to psychotropic substances, with somatization disorder being commoner in women. Studies on adopted children suggest the existence of a genetic component as well as environmental factors, since adopted children tend to resemble their biological rather than their adoptive parents regarding antisocial personality features, but the adoptive family also influences development toward antisocial behavior. The disorder can diminish in vitality with advancing age, usually from the 40th year of life onward.

Etiopathogenesis

The longitudinal study carried out by Robins (1966) on 500 difficult children showed that the best predictor for a later antisocial personality disorder is the extent of antisocial and aggressive behavior in childhood and adolescence. Rutherford et al. (1999), however, did not find general deviations from norm- and role-conforming behavior in young and adolescent girls to be reliable predictors of antisocial personality disorder in adults, but rather for antisocial traits in the strict sense of psychopathy (Hare 1970, 1991). Genetic studies on twins and adopted children suggest hereditary factors (Nigg and Goldsmith 1994). Constitutional risk factors (temper) include above-average body size, sensation seeking, and fearlessness (Raine et al. 1998). Impulsiveness also proved to be a strong predictor for an early onset of stable and pronounced delinquency (Tremblay et al. 1994). Psychophysiological risk factors for antisocial personality disorder are mainly weak electrodermal reactions in anticipation of danger or punishment or when reacting to it, indicating a low conditionability by fear and deficient avoidance behavior, while reward-related stimuli leads to a strong activation (Arnett et al. 1997).

Neurotransmitter research in the past 10 years has aimed at identifying the role of disturbed central serotonergic processes in aggressive and antisocial personalities. Regarding 5-hydroxytryptamine, which is important in behavior regulation, genetic research data indicate a relation between an "LL" genotype and impulsive-aggressive tendencies (New et al. 1998). Furthermore, central catecholaminergic activity (Coccaro et al. 1991) and testosterone (Virkkunen et al. 1994) have also been suggested to play an important role in aggressive behavior.

Therapy

Due to the pronounced aggressiveness and criminality of individuals with antisocial personality disorder, therapy is of great significance, not only for the individuals themselves, but for society at large. So far, however, results of the most varied programs have been disappointing, with similar results for the core group of psychopaths according to Hare, although for the more neurotic forms, psychodynamic and behavior therapy may offer hope. Of course, the typical behavior pattern seen in antisocial personality disorder – low frustration tolerance, impulsive behavior, short-term (immediate gain) rather than long-range thinking – is highly unfavorable for therapy. These individuals are very unreliable both in interpersonal and in therapeutic relationships. They usually do not seek psychotherapy of their own accord but on the urging or orders of others, as by court order or due to confinement in prison or a forensic psychiatric institution. Therapy is hampered by the institutional

setting and a possible tendency on the part of the patient toward manipulation and malingering, and Shapiro (1965) concluded that "psychopathic dishonesty and lying" and impulsiveness greatly limit the therapeutic possibilities in patients showing dissocial behavior.

Results of the sometimes more behaviorally, sometimes more psychodynamically oriented therapy programs for individuals with antisocial personality disorder cannot be assessed with certainty yet. Multimodal, well-structured treatment programs with cognitive and elements of behavioral therapy have tended to show positive results, whereas general group or liberal environmental therapy and unstructured psychodynamic and client-oriented sessions tend to be ineffective and may even be detrimental to prognosis (Andrews et al. 1995; McGuire et al. 1995; Müller-Isberner and Cabeza 1999).

9.2.2 Borderline Personality Disorder

History

The section on borderline syndromes (see Sect. 4) gives a description of the historical thought currents dealing with the heterogeneous "borderland." The current diagnosis of borderline personality disorder, first formulated in DSM-III (APA 1980), is based largely on the paper by Spitzer et al. (1979). Forerunners were the aggressive type of the passive-aggressive personality disorder in DSM-I (APA 1952) and the explosive/epileptoid personality disorder in DSM-II (APA 1968). Since it first appeared in DSM-III, the concept was not changed fundamentally, except that in DSM-IV (APA 1994, 1996) the original eight characteristics were complemented by the item "passing stress-caused paranoid ideas or severe dissociative symptoms."

Description

Because of its great clinical and theoretical relevance, the four most important symptom complexes of the borderline personality disorder are detailed here (Herpertz and Saß 1999a).

As experiments (Herpertz et al. 1997a) have shown, the affective instability derives from an overly sensitive reaction to low-grade but emotionally relevant stimuli, a high affective intensity, and a proneness to quick affective changes. The high affective reactivity leads to sudden, short-lived bursts of extreme mood swings. Clinically, it is further suspected that the affective excitement gradually fades away over a prolonged period of time (Linehan 1993). The affective outbursts take place under certain conditions which can almost always be elicited with the patient under behavioral and situation analysis, e.g., real or imagined experiences of abandonment and rejection. In addition, closeness is often seen as threatening. These typical situational triggers are due to an unresolved ambivalence between a need for attachment and a fear of loss of autonomy (Fiedler 1998).

Especially characteristic of borderline personality disorder is a alternating lifelong pattern of impulsively inflicting harm to oneself or to others, including suicidal actions, self-inflicted wounds, bulimic binge-and-purge attacks, periods of excessive alcohol consumption, and temper tantrums and fights (Zanarini et al. 1990). Unlike antisocial personalities, most patients try to restrain or suppress their impulses. These attempts at control, however, are crude, inflexible, and not based on long-term, stable motives and values suitable for controlling affective impulses and sudden impulses to act. This leads to unpredictable swings between a tense holding back of affective impulses on the one hand and sudden outbursts on the other (Herpertz and Saß 1997). Predominant emotions are dysphoria, anxiety, and anger, and patients with borderline personality disorder can also suffer from chronic feelings of emptiness.

Furthermore, patients with borderline personality disorder are characterized by a highly unstable image and perception of themselves, which can also include aspects of gender identity and manifest itself in form of oscillations between heterosexuality and homosexuality, sometimes also as transsexuality. A lack of orientation and plans for the future in borderline personality disorder can interfere with the patients' attempts at building an identity and cause them to drop out and change jobs frequently, but also result in indiscriminate choice of social groups or partners.

A further, important area are the dissociative or (pseudo)psychotic symptoms, whose inclusion as characteristics of borderline personality disorder and whose influence on the validity of the diagnosis remain controversial (Gunderson and Zanarini 1987; Widiger et al. 1992). Brief paranoid ideas or hallucinations have also been considered pathognomic (Zanarini et al. 1990), but other authors warn against a renewed jumbling of the concepts of borderline and schizotypal personality disorder (Serban et al. 1987).

Typical Thinking

"The world is dangerous and evil," "I'm helpless and powerless," "I'm unacceptable by nature," "I can't control my feelings and impulses" (Beck et al. 1993).

Differential Diagnosis

Contrary to earlier definitions, in DSM-IV the borderline personality disorder does not lie on a spectrum with affective and schizophrenic psychoses, although affective disorders often occur together and brief paranoid and dissociative episodes can occur under stress. The genetic link between the major affective disorders and borderline personality disorder postu-

lated earlier has again been challenged (Baron et al. 1985; Gunderson 1994; Torgersen 1994).

When making a differential diagnosis, affective disorders are of great importance, and borderline personality disorder can often resemble a depressive episode. These are not so much long-lasting, autonomous mood swings as are seen in affective disorders, but rather more intense emotional reactions to stimuli (Cowdry et al. 1991). Regarding the other personality disorders in the emotionally unstable cluster, there is considerable overlap with the histrionic personality, which also shows a craving for attention, manipulative behavior, and sudden mood swings, but fewer self-destructive acts, ended relationships, and chronic feelings of emptiness and loneliness.

Earlier, a narrower view was taken of correlations to the schizotypal personality disorder and to schizophrenia. A detailed exploration will allow a differential diagnosis in most patients. Findings that suggest borderline personality disorder include the brief duration of psychosis-like symptoms, their occurrence in an affectively highly charged context, and their connection to individuals close to the patient with whom a conflict exists. Sensory illusions are usually pseudohallucinations, which are distinct from true hallucinations since the patient recognizes them as abnormal, but still feels unable to control them. Far commoner than (pseudo)psychotic symptoms are dissociative experiences, such as childish and dissociative amnesias, less commonly also depersonalization and hypnagogic states. Severe dissociative states are of great symptomatological importance in patients who have massive traumatic experiences in their history.

There are similarities with the paranoid and narcissistic personality disorders, as such patients may also react with anger to slight stimuli, but show a more stable identity. The antisocial personality disorder shows similar manipulative behavior, but tends to use it mainly to gain advantages, whereas borderline patients use it primarily to gain the affection of significant others.

In contrast to DSM-IV, the borderline type in ICD-10 corresponds to subforms of emotionally unstable personality disorders. Similar to the "excitable psychopath" of ICD-9, the borderline type was included rather late in the drafting stage. Despite much overlap with the DSM-IV borderline personality disorder, there are clear differences. Thus ICD-9 mentions no dissociative or paranoid experiences, and feelings of emptiness and fear of abandonment are mentioned only in the research criteria. Aside from the borderline type, the emotionally unstable personality disorder of ICD-10 includes the impulsive type, which is characterized by deficient impulse and affect control, high excitability, and a tendency toward violent and threatening behavior, particularly in answer to criticism or rejection. Such a differentiation into two subtypes is not without problems, since these describe different, especially gender-specific expressions of a basically impulsive and emotionally unstable personality (Herpertz and Saß 1999a).

Prevalence

A prevalence of about 2% is reported for the general population, 10% for psychiatric outpatients, 20% for psychiatric inpatients, and 30%–60% in clinical populations with personality disorders. A total of 70% of all borderline patients are female (Widiger and Weissman 1991). Frequency is five times higher among first-degree blood relatives than in the general population. A higher familial risk also exists for substance abuse, antisocial personality disorder, and affective disorders, with a close link to impaired impulse control. Especially in young adulthood, the disorder is characterized by persistent inconstancy and chronic instability, with loss of affect and impulse control, and there is a high risk of patients harming themselves, including suicide.

Etiopathogenesis

Interpersonal stressors, especially imagined or real rejection or abandonment, trigger intense basic affective reactions such as fear, anger, or despair, as could be shown in affect induction trials (Herpertz et al. 1999). Cognitive distortions such as dichotomous thinking and generalizations cause a further exacerbation of affect and affective dedifferentiation up to strong dysphoric affect, which can lead to self-injury or bulimic binging.

The central symptom of self-injury, seen particularly in female borderline patients, is often due to an attempt to reduce unbearable tensions (Herpertz and Saß 1994). These feelings lead to a quick reorientation and a suspension of body awareness disorder right up to analgesia. In many patients, the typical course of buildup and release of tension soon leads to a habit of self-injury via operative learning (Herpertz and Saß 1999a). The tension reduction model for self-injury has now become widely accepted (Favazza and Simeon 1995; Feldman 1988) and could be proven by laboratory measurements of psychophysiological parameters (Haines et al. 1995).

Psychodynamically, problems of emotion regulation in borderline patients were seen as a consequence of a disorder that had occurred early in their development, with an emphasis on the so-called primitive defense mechanisms, especially splitting. This is why borderline patients have great difficulty in integrating ambivalent or ambitendential perceptions and emotions.

The proximity to the post-traumatic stress disorder (see Sect. 4) which has been postulated for borderline

personality disorder rests on similarities between the two regarding trigger events such as sexual abuse, manhandling, or neglect in childhood, despite many methodological doubts and the currently much-discussed false memory syndrome (Ernst 1999). Frequencies of 30%–40% are reported for physical abuse in patients with borderline personality disorder, and 25%–70% for sexual abuse (Silk et al. 1995). Further parallels between borderline personality disorder and PTSD are functional abnormalities of the autonomous nervous system, generalization of fear, and difficulties in forming habits and in purging conditioned reactions.

Therapy

Aside from the basics of building a therapeutic relationship, improving psychosocial skills, structuring the psychosocial environment, and treating dysfunctional behavior patterns, special therapies have been developed for borderline personality disorder, of which the dialectic behavioral therapy (DBT) of Linehan (1993) enjoys great popularity at present. Systematic studies have yielded much empirical data for standardized treatment and for therapy control and follow-up, while therapy manuals for depth psychology have been prepared only recently (Kernberg et al. 1999). Herpertz and Saß (1999a) recommend an interdisciplinary approach to psychotherapy with an interpretation of cognitive, behavioral, and psychodynamic methods.

In DBT (Linehan 1993; Bohus et al. 1999), various dynamically hierarchized problem areas are treated systematically in separate stages. Therapy begins with the patient being informed about the treatment, giving consent, followed by motivation and analysis of the aims. The first stage of therapy consists in dealing with suicidal and parasuicidal behavior and with behavior harmful to the therapy and to the patient's quality of life, improving behavioral skills, consciously dealing with feelings, stress tolerance, and self-management. During the second stage of therapy, the post-traumatic stress disorder – which is routinely assumed to exist – is dealt with, while in the third stage the self-esteem is raised, and individual aims are developed and implemented.

According to cognition theory, borderline personality disorder usually shows rigid basic cognitive assumptions which cause a continued distortion of everyday information processing (Schmitz 1999). Dysfunctional schemes which act as self-fulfilling prophecies and thus impair self-control must be uncovered and dealt with by being placed in the proper context of the patient's biography. After dichotomous thinking has been eliminated, the continuum technique is applied: the patient is asked to give operational definitions of both extremes of a property, e.g., a person is trustworthy, as opposed to completely untrustworthy, and then to place actual people or relationships along the continuum thus defined (Beck et al. 1993).

Psychopharmacological measures can be used for acute crisis management when treating concomitant affective disorders, and long-term when seeking to improve impulse control, where SSRI are favored particularly. Recommendations for integrated pharmacological and psychotherapeutic measures are given by Kapfhammer (1999).

9.2.3 Histrionic Personality Disorder

History

The term "histrionic personality disorder" replaces the term "hysterical personality disorder" (Histrio was an actor in ancient Rome). According to Hoffmann and Holzapfel (1999), the historical forerunner of the histrionic personality disorder was the hysterical neurosis, which was always considered inconsistent and multifaceted, with conversion symptoms and dissociative phenomena in the foreground, and all descriptions have dwelled on the variety and diversity of the disorders (hysteria imitates disease). Otherwise, the term hysteria was abandoned as a category not only because of discrimination and labeling problems, but also because it could not be shown that hysterical symptoms, as originally postulated, occurred primarily in Oedipus conflicts. Kurt Schneider (1950a) described individuals with a tendency to dramatic, theatrical behavior, suggestibility, superficial and unstable affect, and a constant striving for esteem as attention-hungry psychopaths.

Description

The main characteristics of this personality disorder are a strong need for attention and recognition, suggestibility, and a tendency for affective instability and superficiality. Histrionic personalities have a sense for atmosphere, but also a tendency for dramatizing, falseness, and coquetry. They are largely unable to maintain a steadfast pursuit of goals and value orientation and are therefore inconstant, especially in relationships with other people and partners. Individuals with histrionic personality disorder seek to impress as something else and more than they are. Symptomatology and basic personality often show a dramatic bend, with a pronounced secondary hypochondria.

Typical Thinking

"If others don't like me or admire me, I'm nothing," "feelings and intuition are much more important than rational thinking and planning," "I get what I want if I dazzle or amuse others" (Beck et al. 1993).

Based on clinical observation as well as cluster and factor analyses, Millon and Davis (1996) distinguished five subtypes of the histrionic personality disorder:

- The theatrical types, who fascinate the opposite sex with their seductive skill and ability to slip into completely different roles
- The hypomanic types, who enthuse others with their lively, extroverted manner, but disappoint with their unreliability and irresponsibility
- The infantile types, who show a high affective instability and oscillate between dependent-childish and defiant-aggressive behavior
- The flattering types, who devote themselves to others in order to secure their recognition and love
- The cunning types, who employ superficial charm to manipulate others

Differential Diagnosis

Most important is the overlap with other personality disorders from the emotionally unstable cluster. According to DSM-IV, individuals with borderline personality disorder likewise seek attention, show manipulative behavior and changeable emotions, but characteristically also self-injury, outbursts of anger, chronic feelings of emptiness, and unstable self-image. The histrionic and the antisocial personality disorders share a tendency for impulsivity, superficiality, sensation seeking, egotism, seduction, and manipulation, but not the pronounced antisocial behavior coupled with gain seeking. There is overlap of the histrionic personality disorder with the narcissistic personality disorder inasmuch as both crave attention and recognition, but the latter lacks the emotional changeability and dramatic tendencies.

Clinically relevant, though uncommon (and usually preceding, less often following psychotic episodes of various origins), are the sometimes drastic "pseudo-hysterical" cases showing aggravation, conversion, dramaticism, and improper behavior.

Prevalence

A prevalence of about 2%–3% is reported for the general population and of 10%–15% for female in- and outpatients undergoing therapy. Women are clearly overrepresented, though less than with somatization disorders (Hoffmann and Holzapfel 1999). The prevalence in neurological clinics lies between 3.5% (Kapfhammer 1999) and 5.25% (Lempert et al. 1990). Conversion symptoms and histrionic behavior can occur throughout an individual's life, but more frequently in young adulthood, and lessening with advancing age.

Etiopathogenesis

Psychodynamically speaking, the typical personality traits are based on a "hysterical" form of resolving conflicts or of defense, the preferred mechanisms being repression and denial, which also cause the amnesias and perception disorders. There is affective shift and projection of unacceptable drives and impulses. According to Hoffmann and Holzapfel (1999), of etiopathogenetic significance are the increased tendency for identification and traits such as empathy, sensitivity, and theatrical skills, which can be positive, but also allow the individual to mimic a series of conditions. According to Nemiah (1995), dissociation – which Janet already considered to be closely related to psychasthenic and hysterical disorders – plays a central role, where hysterical mechanisms of repression, denial and hyperemotionality, identification and identity shift can favor a "consciousness-impairing" effect.

According to the biosocial learning theory (Millon and Davis 1996), constitutionally fixed traits of extroverted and lively children and the interactions in the early psychosocial environment play a role, e.g., frequently changing significant others with a strong but short emotional contact, leading to a style of communication geared toward affect and effect. Cognitive behavioral therapy places an emphasis mainly on special cognitive affect-assessing processes. According to Beck et al. (1993), upbringing during infancy is important in shaping the histrionic personality, fearful and threatening situations being mastered by outward appearance and skillful role playing. In relationships, the individuals seeks to adapt perfectly to the expectations of others, yet the role played lacks self-confidence and inner conviction and remains highly dependent on a social echo.

Therapy

Psychoanalysis sought to resolve the classical Oedipus conflict, but also attempted to improve a deficient self-image (Mentzos 1982). By interpreting transfer and countertransfer processes, the therapeutic relationship was supposed to deal not only with the Oedipus conflict, but also with desires for attention and care. Today, more importance is given to modification by behavioral therapy of cognitions and affects and their interactions in the typical histrionic behavior and interaction pattern. Patients tend to fashion a rather poor identity of themselves, often with regressive changes toward childhood, helplessness, misery, and the need for support.

Further aspects to be dealt with are the global, impressionistic thinking, the craving for attention, and the dependence on recognition. The physician–patient relationship is in constant jeopardy by the patient's dramaticism, which can extend into the erotic realm. In the long run, there can be great problems when it becomes necessary to adapt to and accept aging and loss of attractiveness.

9.2.4 Narcissistic Personality Disorder

History

In psychoanalysis, narcissism was at first closely connected to an early or pre-Oedipal disorder in the development of the ego. Hartmann (1972), Kohut (1977), and Kernberg (1976) later developed the object-relation theories further, which eventually led up to the concept of the narcissistic personality disorder, which was included in DSM-III in 1980 for the first time. Subsequent studies have not proven very reliable (Gunderson et al. 1991), and the disorder has not yet been officially introduced into ICD-10, appearing only in the appendix. However, standardized examination tools have improved reliability over the past few years (Cooper and Ronningstam 1992), and the narcissistic personality disorder continues to be listed in DSM-IV.

Description

According to DSM-IV, individuals with a narcissistic personality disorder have a basically fragile self-esteem and so on the one hand revel in fantasy and grandiosity, feelings of superiority, and disdain of others, and on the other are easily offended, thin-skinned, and highly susceptible to the opinions of others. Patients with narcissistic personality disorder tend to involuntarily exploit others, thinking that their own qualities and abilities entitle them to special treatment. There is an exaggerated sense of entitlement, lack of empathy, and a strong need for recognition and admiration. Further symptoms are feelings of emptiness and futility, weak identity, and distrust. With a clearly increased self-awareness and egotism, social discomfort and fear of negative opinions predominate. A particular problem is the tendency toward depressive crises and resolute suicidality following an imagined insult.

Typical Thinking

"No-one has the right to criticize me," "since I'm superior to others, I'm entitled to special treatment and privileges," "other people should be happy to satisfy my needs" (Beck et al. 1993).

Differential Diagnosis

It must first be decided whether the feelings of grandiosity are a trait of the personality or whether they are due to a manic or hypomanic episode or perhaps to substance abuse. Criticizing or offending persons with narcissistic personality disorder can trigger depressive moods of the severity of a major depression. If all the criteria are met, both diagnoses are to be made.

The areas of overlap with other personality disorders are important, especially those in Cluster B. The main feature distinguishing the narcissistic from the histrionic, antisocial, and borderline personality disorders, whose style also includes coquetry, coldness, and social claims, respectively, is the narcissistic feeling of grandiosity. Further differences are the relative stability of identity in the narcissistic personality disorder and the absence to a large extent of self-injury, impulsiveness, and fear of abandonment, which are typical of the borderline personality disorder. The antisocial and the narcissistic personality disorders are similar in their coldness, slickness, superficiality, exploitation of others, and their lack of empathy, but the narcissistic personality disorder usually does not show precipitated actions, aggressivity, or deception.

Prevalence

According to DSM-IV, the narcissistic personality disorder has a prevalence of 2%–16% in clinical populations and below 1% in the general population. It is seen predominantly in men, with about 50%–75% of patients being male according to DSM-IV. Features can appear especially in adolescence, but without developing into a full narcissistic personality disorder later. Similar to the histrionic personality disorder, individuals with narcissistic personality disorder often have considerable difficulties coping with aging and the physical, mental, and professional losses associated with it.

Etiopathogenesis

Kernberg (1970) saw the typical narcissistic grandiosity and exploitation as a consequence of an "oral anger" arising from an early, pre-Oedipal disorder of psychosexual libido development. Emotional deprivation was hypothesized as the cause, with the grandiosity and entitlement being an exaggerated compensation and protection for a damaged self-esteem. Kohut (1977) saw narcissism in a more positive light and considered it a stage of libido in the course of a normal development, with pathological narcissism only arising with arrested development, e.g., following traumatic experiences, disappointment by the mother, or insufficiently met emotional needs. On the other hand, there is no certain evidence for these psychodynamic models of the origin of narcissistic disorder, as might be given by proof of neglect during certain stages of development in infancy and childhood.

According to the biosocial learning theory (Millon and Davis 1996), the development of narcissistic characteristics is not due to neglect by the mother, but rather to parental overvaluing and rewarding grandiose behavior, thus leading to an inflated identity and self-esteem in the child. The cognitive theory (Beck et al. 1993) suggests dysfunctional ideas about the self, the world, and the future, which would lead to distorted assumptions about one's own importance. These individuals see themselves as extraordinary and hence entitled to concentrate solely on satisfying their

own needs. Such feelings of grandiosity can be reinforced by a corresponding echo from the environment.

Therapy

The character traits of the narcissistic personality disorder cause considerable problems interacting with others, which also complicates therapy. Due to their sense of entitlement and low empathy, narcissistic personalities are not very popular, either with those around them or with therapists. Psychotherapy is initially best carried out singly due to the social touchiness and low frustration tolerance of these patients, where the therapist must gradually find out how much confrontation coupled with support the patient can take. Any threat to the patient's self-esteem can cause exacerbations and setbacks, which may result in discontinuation of therapy, but particularly in serious suicidality.

The first step in cognitive-behavioral therapy is establishing a therapeutic relationship. During this time, the therapist's authority may be challenged by the patient, with competition, downrating, and depreciation being common. The main goals of the therapy should be directed at the major components of the narcissistic disorder, i.e., the grandiosity, exaggerated sensitivity to other's opinions, and lack of empathy (Beck et al. 1993). One difficulty in the therapeutic dialogue lies in the dialectic thinking of narcissistic individuals with their strict either–or categories, the identity oscillating between great superiority and total worthlessness. It is important to slowly relativize the circular interpersonal consequences of the patient's self-centered behavior (Fiedler 1998).

9.3 Cluster C (Anxious, Avoidant Personality Disorders)

9.3.1 Avoidant-Insecure Personality Disorder

History

The term "insecure personality" was applied by Millon (1969) to individuals who, out of fear and distrust, actively avoid others, while in the schizoid personality disorder a similar pattern of social deficiency is due more to a passive remaining at a distance out of disinterest and lack of social initiative. In the German tradition, Millon was preceded by the descriptions given by Kretschmer (1921) of a "sensitive character," especially the "insecure psychopath" of Kurt Schneider (1950a), which is related to the depressive type. While the designation "avoidant personality disorder" in the DSM system particularly refers to the avoidant behavior, the German translation underlined Kurt Schneider's aspect of insecurity (Saß et al. 1996). The ICD-10 term "anxious (avoidant) personality disorder" covers both aspects, the insecurity and the avoidance, but the criteria further list the need for affection and being accepted by others.

Description

The avoidant-insecure personality disorder is characterized by a great fear of rejection and the constant effort to avoid unpleasant feelings and situations where these could arise. Despite their strong wish for affection, individuals with avoidant-insecure personality disorder avoid social relations; they are insecure, shy, tense, and anxious. Their feelings of inferiority in social contact lead to a severe restriction of their social skills and roles.

Kurt Schneider (1950a) described how, out of insecurity, people with low self-confidence frantically try to appear overly secure or conspicuous, but still have a guilty conscience and always seek the fault for everything in themselves. The excessive self-criticism of the ethically overscrupulous person is seen in the sensitive type described by Kretschmer (1921), which can even develop into paranoia. According to Kurt Schneider, the feelings of guilt and insufficiency and the overwrought conscientiousness and cleanliness of insecure and sensitive individuals can lead to compulsions, as are characteristic of anancastic psychopaths, and may sometimes be accompanied by sudden attacks of anxiety and dizziness or palpitations.

Typical Thinking

"I should avoid situations where I call attention to myself and try to be inconspicuous," "the unpleasant feelings will get worse and get out of control," "I'd never live it down if people uncovered my insecurity" (Beck et al. 1993).

Differential Diagnosis

This personality disorder shows the strongest similarity to social phobia, an example of how difficult it can be to distinguish between axis I and axis II disorders, particularly since therapy is identical for both. The overlap is ultimately due to the general criteria for personality disorder, i.e., contrary to social phobia, the avoidant-insecure personality traits must be permanent and deep-seated and must manifest themselves in different situations. In differentiating it from social phobia, Millon and Davis (1996) pointed out that, in the avoidant-insecure personality disorder, the ambivalence of socially aversive behavior is the central aspect, as well as the low self-esteem and the desire for social contact. Social phobia patients, on the other hand, show more a pattern of avoidant behavior and a strong reaction to anxiety-triggering stimuli.

When making a differential diagnosis, it is important to keep in mind the other personality disorders located in the asthenic cluster. Common features are feelings of insufficiency, hypersensitivity toward crit-

icism, and a need for recognition, but in the avoidant-insecure personality disorder the main concern is avoiding humiliation and rejection, while the dependent personality disorder is characterized mainly by its need for being cared for.

It is unclear both theoretically and empirically how the avoidant-insecure and the schizoid personality disorders are related. There are contending notions: on the one hand, it is held that the schizoid personality disorder is a variant of a more general avoidant-insecure personality disorder (Reich and Noyes 1986) or that both are variants of one and the same personality disorder (Livesley and West 1986), and on the other hand that they are two perfectly distinct personality disorders (Millon 1981; Trull et al. 1987). Patients with schizoid or schizotypal personality disorder also show social isolation, but seem to be content with this and even prefer it, whereas those with avoidant-insecure personality disorder do have a need for social contact and suffer from their loneliness.

Prevalence

Comparing various studies, Widiger (1992) found a prevalence of avoidant-insecure personality disorder (DSM-III) of 5%–35% in psychiatric populations and of about 1% in the general population. Social phobia (generalized type), dependent personality disorder, and affective illnesses play a role in comorbidity.

The disorder often begins in infancy or in childhood, with shyness, isolation, and fear of strangers and new situations. Social relations marked by shyness and insecurity can also appear in adolescence and young adulthood. With increasing life experience, these traits sometimes decrease with age.

Etiopathogenesis

Both Siever and Davis (1991) and Millon and Davis (1996) postulate a genetic component of the avoidant-insecure personality disorder. According to the biosocial learning theory, the disorder is based on a combination of a biologically determined increased readiness for fear and special social experiences, such as a perception of insufficient fear regulation and poor social support when faced with demands (Millon and Davis 1996). Kurt Schneider (1950a) saw a proximity to affective disorders for forms with a pronounced depressive component.

Interpersonal etiology models can be traced back to Horney (1945), who considered an original feeling of anxiety an expression of helplessness in the face of a parental upbringing style, and to Sullivan (1953), who pointed to difficulties in regulating self-esteem. Avoidant behavior and social isolation would thus be consequences of a protection mechanism against increased interpersonal fear in a conflict between seeking closeness on the one hand and fearing it on the other. According to cognition theory (Beck et al. 1993), insecure patients have deep-seated negative views of themselves which originate in childhood and are due to rejection and criticism on the part of significant others. This leads to social avoidance behavior as a defense mechanism against being exposed by others. Thoughts are avoided which are conducive to dysphoria, discomfort, fear, and sadness.

Therapy

The establishment of a therapeutic relationship is complicated by the patients' low self-esteem and fear of being disappointed, rejected, and criticized. Individual therapy is recommended, using empathy, caring support, and recognition to help the patient gradually build confidence and develop a certain tolerance for social situations. However, behavioral therapy approaches in group therapy of insecure personality disorder have also been tried (Renneberg and Fydrich 1999). In patients presenting affective distemper, psychopharmaceutics may be used (Kapfhammer and Rothenhäusler 1999), but benzodiazepines, although popular because of their anxiolytic action, should be strictly avoided.

9.2.3 Dependent Personality Disorder

History

Descriptions of individuals showing dependence, helplessness, lack of willpower, and a need to lean on others already appear in nineteenth-century case reports of abnormal personalities, yet without defining a distinct type. Echoes of the modern personality disorder can be found in Kurt Schneider's (1950a) weak-willed, unstable, insecure, and asthenic psychopaths.

DSM-I (APA 1952) described the dependent personality as a passive-dependent subtype of the passive-aggressive disorder, but this form no longer appeared in DSM-II (APA 1968). It was not until Millon's biosocial learning theory (1969) that passive-dependent behavior was recognized as one of eight basic personality types and was later included in the DSM-III (APA 1980). This was due to empirical examinations which, using factor analysis, identified a dependence factor with elements of passivity, submissiveness, and low self-confidence (Livesley et al. 1990).

In the ICD system, dependent features were earlier included in the "asthenic personality disorder," and the first version of ICD-10 (1991) still included a "dependent (asthenic) personality disorder." It was not until the second ICD-10 edition (WHO 1992), in agreement with DSM-III-R, that the term "dependent personality disorder" was listed as a category of its own. Conceptionally, it is closely related to "neurasthenia," which appears in ICD-10 but not in DSM-IV.

Description

The dependent personality disorder is characterized by an overpowering feeling of not being able to conduct one's own life. With a self-image of helplessness and weakness, patients seek support in all situations, from others and especially from their partner. They with a dependent personality disorder are hardly willing to assume responsibility for themselves. In a relationship, these patients experience a constant fear of loss and abandonment, coupled with efforts to adapt and to yield. They further show a cognitive distortion known as catastrophizing, i.e., a fearful and exaggerated estimate of the worst possible consequences should the relationship end. Important complications, which often are the reason for seeking therapeutic help, include depressive mood, as frequently occurs in the face of a possible or actual loss of a relationship, and anxiety disorders resulting from similar situations. Furthermore, patients, especially women, often show somatic complaints such as conversion symptoms, hypochondriac fears, and somatization syndromes. If alcohol and other psychotropic substances are taken for relief, there is the danger of addiction due to these patients' tendency toward passive conflict avoidance. In their personal relations, people with dependent personality disorder are adaptable, friendly, and submissive and may show considerable social skills in imparting feelings of importance, strength, and superiority to partners whose support they seek.

Typical Thinking

"If left to myself, I'd be helpless," "I can't take decisions for myself," "I shouldn't do anything that could hurt those who support and help me" (Beck et al. 1993).

Differential Diagnosis

Even when the concept was still being developed, it was already a constant source of criticism that characteristics of dependent individuals are shaped by certain social roles and apply especially to women (Kaplan 1983). The criteria describe a submissive form of dependence which tends to be associated with women, while dominating forms of dependence, such as occur more commonly in men, are less often diagnosed as disturbed or pathological, especially by male therapists. In dominating dependent behavior, others are instructed to do the work or make the decisions. A diagnosis of dependent personality disorder is not warranted if the dependent behavior rests on cultural values and role expectations.

In a differential diagnosis, it is important to keep in mind that dependence can occur as a consequence of affective disorders, panic disorder, or agoraphobia. Distinctive characteristics are the early onset, the chronic course of illness, and a behavior pattern which is not restricted to episodes of one of these disorders, but which can be observed throughout the entire development of the personality.

Dependent behavior can also occur in personality disorders such as the histrionic one, but such patients are distinctly dramatizing, demanding, self-centered, dazzling, and superficial, while dependent persons are docile and modest. Feelings of insufficiency, hypersensitivity toward criticism, and a need for protection and affection also occur in avoidant-insecure personality disorder, but without the readiness of dependent individuals to entrust themselves to others and to rely on them. Insecure individuals, on the other hand, tend to withdraw and avoid contact. Patients with depression and anxiety disorder who in addition show a personality disorder often exhibit a combination of dependent and insecure features (Mavissakalian and Hamman 1988).

Prevalence

The prevalence of dependent personality disorder in the general population is estimated at 1%–2%, and in clinical populations at about 10%. In psychiatric clinics, this is one of the most frequently diagnosed personality disorders. On the average, the frequency in clinical groups is about 20% (Morey 1988; Blashfield and Davis 1993).

Etiopathogenesis

Psychodynamically, the cause of dependent features was sought in a maternal behavior marked by lavish affection and excessive concern, especially early in development (Freud 1908; Abraham 1925). The fixation in the oral phase during psychological development, with a desire for pleasure and being pampered, was thought to favor a dependent behavior with poor initiative and an insistence on being cared for by others (Levy 1966). Winnicott (1965) saw the manner in which dependent subjects establish relationships as a consequence of failed identity development with the formation of a false or lost "self," leading to a helpless-insecure need for support. According to Esman (1986), there are also latent inimical feelings for the partners, with the exaggerated friendliness and submissiveness being seen as a means of maintaining the relationship. Bonding theory assumes an anxiety-laden behavior which can be traced back to unpleasant experiences in earlier relationships, with a constant fear of loss (Bowlby 1977).

Seen from a cognitive-behavioral perspective (Beck et al. 1993), dependent patients fear independence and self-reliance, have only developed a low level of self-confidence, and see themselves as helpless and incapable of life. People who underestimate their own abilities and only develop and train them marginally are increasingly ready to and in need of securing the

help of others. Due to this lack of autonomy, thinking oscillates between total helplessness and dependence and total independence and isolation.

The biosocial learning theory of Millon and Davis (1996) and Bornstein (1993) links psychodynamic and cognitive-behavioral explanations. Dysfunctional learning originates from authoritarian or overprotective parents, discouraging independence and self-reliance and instead favoring the development of a personality which needs stronger partners for protection and security. The constant reinforcement cements the fear of making one's own decisions, the increasing insecurity prevents the development of interests and their implementation, and the fear of making decisions further aggravates the cycle of dependence. This also favors the development of an often comorbid anxiety disorder and depressive moods.

Therapy aims at improving self-confidence and independence. The compliance, friendliness, and reliability of dependent patients are very helpful in establishing a therapeutic relationship. However, in view of this dependent behavior, it is important to avoid creating a codependence between patient and therapist (Benjamin 1993). At the same time, the positive traits of a dependent personality, such as good social integration and considerable professional and academic achievement, must be protected (Bornstein 1993). This can make for a high sensitivity and great stability in mutually supporting partnerships. Cognitive and behavioral therapy (Beck et al. 1993) aim at improving decision-making ability and foster patients' confidence in their own skills, a clearer perception of their own interests and possibilities for their implementation, and a gradual detachment from protectors. Role-playing and assertiveness training are helpful, as is including the partner in order to relativize dependence-encouraging behavior. Aside from individual therapy, which these patients prefer, group therapy is also indicated, both to learn from the behavior of others and to reduce dependence on the therapist. Use of the newly learned skills in society is encouraged when, after conclusion of therapy, the patient attends follow-up sessions at increasing intervals to support a gradual detachment from the therapist and to bolster confidence in the (difficult) task of leading an independent life.

9.3.3 Obsessive–Compulsive Personality Disorder

History

Obsessive–compulsive abnormalities of personality were described by Esquirol (1838), who spoke of a *délire partiel* within a *monomanie instinctive*. Disturbed emotions and impulses were central to Morel and Dagonet's view of obsessive-compulsive phenomena (Berrios 1995), while in Germany Griesinger first presented a case history with obsessive–compulsive phenomena, with Westphal describing the underlying ideations in 1877.

Kurt Schneider (1950a) classed these disorders with anancastic or obsessive–compulsive psychopaths as a subcategory of insecure psychopaths. The obsessions/compulsions stem from deep-seated feelings of insufficiency and guilt. These individuals live in constant fear of neglecting something or causing some damage, and the particular ideas are firmly rooted in the individual's wishes, values, and biography.

Since Freud's fundamental book *Charakter und Analerotik* ("Character and Anal Eroticism") (Freud 1908), psychoanalysis has seen a close connection or transition between the "anal character" and the obsessive–compulsive neurosis with well-developed symptoms. However, even psychopathologically, obsessive–compulsive traits were considered a premorbid character structure whose symptoms can progress in times of stress and lead to a full-blown illness (Weitbrecht 1963).

Description

The main characteristics of the obsessive–compulsive personality disorder are conscientiousness, perfectionism, rigidity, solidity, and conformism, which can be overvalued to the point of adversely affecting professional productivity and interpersonal relationships. These subjects show severity, seriousness, and rigidity both with themselves and with others. This stiff, morally demanding and rule-minded behavior is stubbornly held to and forced on those close to the patient. The person's own feelings and those of others, humor, and joy are all seen as suspicious and threatening. Human beings, life, and the world are seen in a negative light, and the only meaningful things in life are seen as being work, toil, and fulfilling duties.

Basically, the qualities of obsessive–compulsive personalities – such as love of order, perseverance, punctuality, and thriftiness – are socially desirable, but can be negative when exaggerated. This is the case when anxiety, scruples, weakness in making decisions, painstaking exactness and meticulousness, and inability to distinguish what is important from what is not interfere with social functioning. Exaggerated caution due to an inability to make decisions and a fear of responsibilities likewise impairs social functioning.

Typical Thinking

"If I don't observe my principles 100%, I'll drown in chaos," "the world's a dirty mess, I'm the only one who's clean and organized," "I must keep my feelings under complete control" (Beck et al. 1993).

Differential Diagnosis

The most important differential diagnosis concerns the differentiation from the obsessive-compulsive disorder, which, in contrast to the above-mentioned psychoanalytical and psychiatric traditions, plays in important role in modern classification systems. Factor analysis has allowed obsessive-compulsive symptoms to be distinguished from the obsessive-compulsive personality, even though they may overlap (Coursey 1984). Although inconsistencies remain (Zaworka and Hand 1981), a spectrum covering the obsessive-compulsive disorder and the obsessive-compulsive personality disorder has not yet been confirmed empirically. Both disorders show a tendency toward rituals, rigid behavior, an inability to distinguish what is important from what is not, and an impaired emotional expression. However, obsessive-compulsive disorder patients need not always show an obsessive-compulsive personality, but just as often exhibit insecurity and fear of criticism or punishment. In obsessive-compulsive personality disorder, the patients do not see their own thought and behavior patterns as particularly absurd, self-dystonic, and painful. They manifest themselves along a broad spectrum of situations and are more generalized than the circumscribed symptomatology of the obsessive-compulsive disorder. There are indications that, during times of great stress, some symptoms of the obsessive-compulsive personality "fall apart" and lead to an obsessive-compulsive disorder similar to that described by Kurt Schneider (1950a).

Of great importance are the interactions between the obsessive-compulsive personality disorder and depression. The negative philosophy of life and the misery associated with it can lead to distemper under stress. On the one hand, obsessive-compulsive personality traits can intensify during depression or appear as disturbing for the first time, e.g., in form of depressive insecurity, anxiety, and difficulties in making decisions; on the other hand, obsessive-compulsive behavior can lead to difficulties, and hence to reactive depression, where obsessive-compulsive personality traits and depressive symptoms are closely interwoven.

The obsessive-compulsive personality disorder overlaps with Tellenbach's "melancholic type" found in about 50% of all endogenous depression patients (Tellenbach 1983). Patients of this type are usually conscientious, dutiful, and perfectionist. In relationships, they show a need for harmony and adaptability, thus showing both obsessive-compulsive and dependent features. Inherent obsessive-compulsive tendencies can also appear in the prodromal phase of schizophrenic syndromes in an attempt at halting psychotic disintegration (Lang 1981).

Regarding other personality disorders, the narcissistic personality disorder also shows a tendency toward perfection and an annoyance at others not being able to do anything right, but narcissistic individuals lack the self-criticism that is common and at times exaggerated in obsessive-compulsive subjects. Like obsessive-compulsive patients, the schizoid personality disorder also shows an evident formalism, but in the obsessive-compulsive personality disorder this follows from overvaluing work duties and rules, whereas the schizoid personality disorder usually shows a decreased capability for intimacy and for experiencing feelings.

Prevalence

The prevalence in the general population is estimated at 2%, while among psychiatric inpatients obsessive-compulsive personality disorder shows a prevalence of 6%–9%.

Etiopathogenesis

Lately, the relation of obsessive-compulsive disorders to an exaggerated toilet training which was started too early (during the "anal phase") and failed, as is hypothesized in psychoanalysis, has lost ground. Instead, Shapiro's (1981) three elements in the psychodynamic structure of obsessive-compulsive personality disorder are now cited (Hoffman and Holzapfel 1999): first, the emotional independence of individuals with an obsessive-compulsive personality constitutes an affective I-don't-need-anyone self-sufficiency, in sharp contrast to hysteric subjects; second, obsessive-compulsives avoid independent decisions so as not to make any mistakes; and third, they have a feeling of being driven, of constantly having an imaginary foreman looking over their shoulder and finding fault with everything they do. More recent psychoanalytic hypotheses postulate less an anal-aggressive drive, but rather interpersonal tension between obedience (another's decision) and autonomy (one's own decision).

Taking an interpersonal perspective, Sullivan (1953) described an underlying insecurity in interactions in obsessive-compulsive personality disorder which causes the individual to be extremely meticulous and careful in fulfilling general rules and norms. The cognitive perspective points out the formalistic thinking and the narrowing to conventions, rules, schedules, and hierarchical structures. Beck et al. (1993) described as typical the tendency of seeing things as good/bad or right/wrong, with a decreased capability for making more subtle judgments and tolerating ambivalence.

According to Millon's biosocial learning theory (1981), the tendency of obsessive-compulsive personalities to conform stems from an excessively controlled upbringing. Attempts at independence or at behavior outside of that prescribed are punished; the child cannot develop its own identity and internalizes the

strict norms of its parents and other significant individuals.

Therapy

Psychoanalysis now plays a minor role. Instead, obsessive-compulsive personality disorder is now treated using behavioral therapy. First, an allowance must be made for the patient's needs for order, reliability, and regularity. Learning relaxation techniques and autogenous and other training can also be useful. Due to their rigid behavior and their striving for control and perfection, obsessive-compulsive patients can be quite a problem in group therapy, which is why individual therapy may be indicated at first. The patients must try to put earlier learning experiences and their lifelong striving for perfection in perspective. According to Beck et al. (1993), a series of basic cognitive assumptions can be described which may help to modify the obsessive-compulsive restrictions of behavior and emotions. These include the notions that no mistakes are allowed, that mistakes mean failure, that mistakes bring about criticism, that patients must completely control themselves and their surroundings, that every details counts, and that without rules and rituals everything is lost. Further cognitive distortions of obsessive-compulsive personality disorder which must be addressed are the dichotomous thinking, exaggeration, and catastrophizing while overvaluing one's own imperfections and flaws, the inability to concentrate only on what is relevant, and the tendency to always think in terms of "should" and "must," i.e., putting moral norms and expectations before one's own ideas and wishes.

There are interesting psychopharmacological possibilities as therapeutic aides, particularly SSRI antidepressants (Kapfhammer and Rothenhäusler 1999).

9.3.4 "Subaffective" Realm

An important and special area closely related to Cluster C are the transitional forms between the affective accentuations of personality and affective illness (Saß et al. 1993; Herpertz et al. 1998). In DSM-IV and ICD-10, all relevant changes of mood and drive are encoded as belonging to the realm of affective disorders, e.g., dysthymic and cyclothymic. Nevertheless, in the criteria of these DSM-IV axis I categories, the chronic form of symptoms is highlighted, which fits more a trait than a state character. Where a lasting course of these deviations is observed, it would therefore make sense to diagnose them as personality disorders, as has now been taken into account in DSM-IV with the reintroduction of the category of depressive personality disorder in the appendix.

These types of personality abnormalities which could be termed "subaffective" have a long history in the German tradition (Herpertz et al. 1998). Kraepelin (1903/1904) first described constitutional distempers and constitutional restlessness as independent psychopathic conditions, later however as favorable dispositions or attenuated forms of depression, mania, agitation, and cyclothymia. Kretschmer (1921) justified notions of a continuum of mental illnesses and in the affective realm described fluid borders between cyclothymic character, cycloid personality, and manic-depressive psychosis. Kurt Schneider (1950a) included lasting changes of mood and drive which had not yet reached proportions of an affective psychosis in his typology of psychopathic personalities as depressive and hyperthymic forms. The asthenic psychopath, who is characterized by low vitality and psychological and somatic weakness, also shows some overlap with the permanently depressive personality type and further with the modern somatization disorder.

In modern American diagnostic research, Akiskal and Akiskal (1992) have described a broad bipolar spectrum with episodic and chronic subaffective forms of varying degrees of severity. Patients with a cyclothymic temper are moody and impulsive and show unpredictable behavior without meeting the criteria of a bipolar manic-depressive disorder, which etiologically would seem to be endoreactive behavior, as Weitbrecht (1963) also postulated earlier for "endoreactive dysthymia." The hyperthymic temper, whose forerunners were the hyperthymic psychopath of Kurt Schneider (1950a) and the manic type of von Zerssen (1988), characterizes patients with a tendency toward euphoria, who are enterprising, and who have a positive mood and self-image up to intermittent hypomanic episodes. Due to their high activity and efficiency, these individuals often are very successful professionally and require psychiatric help or psychotherapy only when symptoms are greatly exacerbated. Subjects with a depressive temper show a permanently depressed subaffective constitution, which corresponds to the optional depressive personality disorder in DSM-IV.

Despite its clinical plausibility, the concept of independent subaffective personality disorders has once again been strongly challenged by recent empiric research (Pepper et al. 1995; Rihmer 1999). Black and Sheline (1997) found that scores for personality disorders improved in depressive patients following psychopharmacological therapy. Empirically, the depressive personality disorder shows overlap especially with dysthymia and was linked to a poor prognosis for course of illness of affective disorders (Klein and Shih 1998). Similar to Angst (1988), Akiskal now also sees the subaffective personality disorders as mild, chronic, fluctuating subsyndromal forms of a primary affective illness with a stronger biological basis (Akiskal et al. 1997).

10 Outlook

Summarizing the current knowledge on personality disorders, a considerable heterogeneity can be seen both theoretically and empirically. There is no consensus either on the relationship between normal personality, accentuated personality, and full-blown personality disorder, nor on whether a more dimensional or a more prototype- or category-oriented recording method is best. There is controversy concerning the circumstances under which these conditions arise; in addition, knowledge on comorbidity and course of illness is not perfectly consistent, and therapy employs a variety of methods. In Germany, the psychiatric view of personality disorders meets with considerable criticism from clinical and differential psychology and – with some noteworthy exceptions (Fiedler 1998; Schmitz et al. 1996) – may even be ignored as a nuisance.

What does the future hold? Substantial developments are most likely to come from a combination of modern diagnostics and pathogenetic studies. Relevant questions currently concern the relationship between category-based and dimensional descriptions of personality, a hierarchization of the classification structure of disorder characteristics, overlap or distinction between normal and abnormal personality, differentiation of trait versus state characteristics, and not least the importance of genetic and nongenetic influences on the development of personality. A critical observation of the personality factors developed so far and their relation to the personality disorders reveals only poor agreement (Dyce and O'Connor 1998), but recent data point the way (Jang et al. 1996; Clark et al. 1997; Livesley et al. 1998). The hierarchically organized personality model appears well-founded, consisting of four broader higher-order traits, i.e., emotional dysregulation, dissocial behavior, inhibition, and obsessive–compulsiveness, plus 18 more specific lower-order traits. This four-factor solution does not, however, include the dimension "openness to experiences" which figures in many other personality models (Widiger 1988; Livesley et al. 1998).

One objection raised by Livesley et al. (1998) to the model is that individuals simultaneously showing schizophrenic or affective disorders are excluded. This marks a departure from many other studies and from the principles of DSM-IV, where the independent assignment of axis II diagnosis and axis I disorder was always considered very important. Thus it is likely that personality disorders which could show a relation to axis I disorders are underrepresented, which would warrant further research on clinical groups.

The Livesley factors are stable across clinical and nonclinical groups as well as pairs of twins, but otherwise support a general dimensional model, in which personality disorders do not differ qualitatively from normal personality variants, but are merely seen as maladapted extreme variants of general personality features. A systematic interpretation of category-based diagnosis of personality disorders and of the dimensional assessment of personality features might therefore be recommended, something which cannot be reconciled in a model of normal personality.

The current problems of personality research were essentially already presented by Kurt Schneider (1950a). *Klinische Pathologie* (1950b, "Clinical Pathology") advises caution, a recommendation which remains valid, and it therefore seems fitting to cite it at the end of this chapter: Personality disorders look like diagnoses, but this is really an unjustified analogy since individuals and personalities cannot be labeled diagnostically with the same precision as illnesses and their psychological consequences can. If anything, characteristics which stand out clearly can be described and highlighted, yet without having found any similarities to symptoms of illnesses. This emphasis follows certain perspectives, such as the patient's subjective condition, his or her perception of existence and life, and the conditions for development or specific difficulties in a certain sociocultural environment. In addition to the characteristics highlighted, the same individual has myriads of others, and it is easily overlooked that labeling a certain personality disorder may give the false impression that it refers to a person's whole psychological aspect, or at least the essential part.

11 References

Abraham K (1925) Psychoanalytische Studien zur Charakterbildung. In: Abraham K (1982) Gesammelte Schriften, vol 2. Fischer, Frankfurt am Main, pp 103–160

Akiskal HS, Akiskal K (1992) Cyclothymic, hyperthymic and depressive temperaments as subaffective variants of mood disorders. In: Tasman A, Riba MB (eds) Review of psychiatry, vol 11. American Psychiatric Press, Washington DC, pp 43–62

Akiskal HS, Judd LL, Gilin JC, Lemmi H (1997) Subthreshold depressions: clinical and polysomnographic validation of dysthymic, residual and masked forms. J Affect Disord 45: 53–63

Alexander F (1928) Der neurotische Charakter. Seine Stellung in der Psychopathologie und in der Literatur. Int Z Psychoanal 14: 26–44

Andersen HS, Sestoft D, Lillebaed T, Mortensen EL, Kramp P (1999) Psychopathy and psychopathological profiles in prisoners on remand. Acta Psychiatr Scand 99: 33–39

Andrews DA (1995) The psychology of criminal conduct and effective treatment. In: McGuire J (ed) What works: reducing reoffending. Wiley, Chichester, pp 35–62

*Angst J (1998) Dysthymia and personality. Eur Psychiatry 188–197

APA (1952) Diagnostic and statistical manual of mental disorders, 1st edn. American Psychiatric Press, Washington DC

APA (1968) Diagnostic and statistical manual of mental disorders, 2nd edn. American Psychiatric Press, Washington DC

APA (1980) Diagnostic and statistical manual of mental disorders, 3rd edn (DSM-III). American Psychiatric Press, Washington DC

APA (1987) Diagnostic and statistical manual of mental disorders, 3rd rev edn (DSM-III-R). American Psychiatric Press, Washington DC

APA (1994) Diagnostic and statistical manual of mental disorders, 4th edn (DSM-IV). American Psychiatric Press, Washington DC

Arieti S (1955) Interpretation of schizophrenia. Brunner, New York

Arnett PA (1997) Autonomic responsitivity in psychopaths: a critical review and theoretical proposal. Clin Psychol Rev 17: 903–936

Baron M, Gruen R, Asnis L, Lord S (1985) Familial transmission of schizotypal and borderline personality disorders. Am J Psychiatry 142: 927–934

Battaglia M, Bernardeschi L, Franchini L, Bellodi L, Smeraldi E (1995) A family study of schizotypal disorder. Schizophr Bull 21: 1

*Beck AT, Freeman A, Pretzer J et al (1993) Kognitive Therapie bei Persönlichkeitsstörungen. Psychologie Verlags Union, Weinheim

Benjamin LS (1993) Interpersonal diagnosis and treatment of DSM personality disorders. Guilford, New York

Bernstein DP, Useda D, Siever LJ (1993) Paranoid personality disorder: review of the literature and recommendations for DSM-IV. J Pers Disord 7: 53–62

Berrios G (1995) Obsessive-compulsive disorder. In: Berrios G, Porter R (eds) A history of clinical psychiatry. Athlone, London, pp 573–592

Berrios G, Gili M (1995) Abulia and impulsiveness revisited: a conceptual history. Acta Psychiatr Scand 92: 161–167

Binder H (1964) Die psychopathischen Dauerzustände und die abnormen seelischen Reaktionen und Entwicklungen. In: Kisker KP, Meyer JE, Müller M, Strömgren E (eds) Psychiatrie der Gegenwart, vol 2. Springer, Berlin Heidelberg New York, pp 180–202

Binding K, Hoche A (1920) Die Freigabe der Vernichtung lebensunwerten Lebens. Ihr Maß und ihre Form. Meiner, Leipzig

Black KJ, Sheline YI (1997) Personality disorder scores improve with effective pharmacotherapy of depression. J Affect Disord 43: 11–18

Blashfield R, Davis RT (1993) Dependent and histrionic personality disorders. In: Sutker PB, Adams HE (eds) Comprehensive handbook of psychopathology, 2nd edn. Plenum, New York, pp 395–409

Bleuler E (1911) Dementia praecox oder Gruppe der Schizophrenien. Deuticke, Leipzig

Bleuler M (1972) Die schizophrenen Geistesstörungen im Lichte langjähriger Kranken- und Familiengeschichten. Thieme, Stuttgart

Bohus M, Stieglitz RD, Fiedler P, Berger M (1999) Persönlichkeitsstörungen. In: Berger M (ed) Psychiatrie und Psychotherapie. Urban and Schwarzenberg, Munich, pp 771–846

Bonet T (1684) Sepulchretum. Paris

*Bornstein RF (1993) The dependent personality. Guilford, New York

Bostroem A (1926) Zur Frage des Schizoids. Arch Psychiatry 77: 32–60

Bowlby J (1977) The making and breaking of addictional bonds. Br J Psychiatry 130: 201–210

Chapman LJ, Chapman JP, Raulin ML (1976) Scales for physical and social anhedonia. J Abnorm Psychol 85: 374–382

Clark LA, Livesley WJ, Morey L (1997) Personality disorder assessment: the challenge of construct validity. J Pers Disord 11(3): 205–231

Clarkin JF, Hull JW, Hurt SW (1993) Factor structure of borderline personality disorder criteria. J Pers Disord 7: 137–43

*Cleckley H (1976) The mask of sanity: an attempt to clarify some issues about the so-called psychopathic personality, 5th edn. Mosby, St Louis

Cloninger CR, Pryzbeck TR, Svrakic DM et al (1994) The Temperament and Character Inventory (TCI): a guide to its development and use. Washington University Center for Psychobiology of Personality, St Louis/MO

Coccaro EF, Lawrence T, Trestman R, Gabriel S, Klar HM, Siever LJ (1991) Growth hormone responses to intravenous clonidine challenge correlates with behavioral irritability in psychiatric patients and in healthy volunteers. Psychiatry Res 39: 129–139

Cooper AM, Ronningstam E (1992) Narcisstic personality disorder. In: Tasman A, Riba MB (eds) Review of psychiatry, vol 11. American Psychiatric Press, Washington DC, pp 80–97

Coryll W, Zimmerman M (1989) DSM-III personality disorder diagnoses in a nonpatient sample. Arch Gen Psychiatry 46: 682–689

*Costa PT, McCrae R (1990) Personality disorders and the five-factor model of personality. J Pers Disord 4: 362–371

Coursey D (1984) The dynamics of obsessive-compulsive disorder. In: Insel TR (ed) New findings in obsessive-compulsive disorder. American Psychiatric Association, Washington DC, pp 104–121

Cowdry RW, Gardner DL, O'Leary KM, Leibenluft E, Rubinow DR (1991) Mood variability: a study of four groups. Am J Psychiatry 148: 1505–1511

Cunningham MD, Reidy TJ (1998) Antisocial personality disorder and psychopathy: diagnostic dilemmas in classifying patterns of antisocial behavior in sentencing evaluations. Behav Sci Law 16: 333–351

Dilling H, Weyerer S, Castell R (1984) Psychische Erkrankungen in der Bevölkerung. Enke, Stuttgart

Dupré E (1925) La doctrine des constitutions. In: Pathologie de l'imagination et de l'émotivité. Payot, Paris

Dyce JA, O'Connor BP (1998) Personality disorders and the Five-Factor Model: a test of facet-level predictions. J Pers Disord 12(1): 31–45

Ernst C (1999) Missbrauch vergisst man nicht. Psycho 25: 613–617

Esman AH (1986) Dependent and passive-aggressive personality disorders. In: Cooper AM, Frances AJ, Sacks MH (eds) The personality disorders and neuroses. Basic Books, New York

Esquirol E (1838) Des maladies mentales, 2 vols. Baillière, Paris
Ewald G (1924) Temperament und Charakter. Springer, Berlin
Eysenck HJ (1952) The scientific study of personality. Routledge Kegan Paul, London
Fairbairn WRD (1940) Schizoid factors in the personality. In: Fairbairn WRD (ed) Psychoanalytic studies of the personality. Tavistock, London
Favazza AR, Simeon D (1995) Self-mutilation. In: Hollander E, Stein D (eds) Impulsivity and aggression. Wiley, Chichester, pp 185–200
Feldman MD (1988) The challenge of self-mutiliation: a review. Compr Psychiatry 3: 252–269
*Fiedler P (1998) Persönlichkeitsstörungen, 4th edn. Psychologie Verlags Union/Beltz, Weinheim
Fiedler P (1999) Differentielle Indikation und differentielle Psychotherapie bei Persönlichkeitsstörungen. In: Saß H, Herpertz S (eds) Psychotherapie von Persönlichkeitsstörungen. Thieme, Stuttgart, pp 63–73
Fiske DW (1949) Consistency of factorial structures of personality ratings from different sources. J Abnorm Soc Psychol 44: 329–344
Freud S (1908) Charakter und Analerotik. Fischer, Frankfurt am Main
Grilo CM, McGlashan TH (1999) Stability and course of personality disorders: Curr Opin Psychiatry 12(2): 157–162
Gunderson JG (1994) Building structure for the borderline construct. Acta Psychiatr Scand 89(379): 12–18
Gunderson JG, Zanarini MC (1987) Current overview of the borderline diagnosis. J Clin Psychiatry 48[Suppl]: 5–11
Gunderson JG, Ronningstam E, Smith LE (1991) Narcissistic personality disorder: a review of data on DSM-III-R descriptions. J Pers Disord 5: 167–177
Haines J, Williams CL, Brain KL, Wilson GV (1995) The psychophysiology of self-mutilation. J Abnorm Psychol 104: 471–89
*Hare RD (1970) Psychopathy: theory and research. Wiley, New York
Hare RD (1991) Manual for the Hare Psychopathy Checklist-Revised. Multi Health Systems, Toronto
Hartmann H (1972) Ich-Psychologie. Studie zur psychoanalytischen Theorie. Klett, Stuttgart
Henderson D (1939) Psychopathic states. Norton, New York
Herpertz S, Saß H (1994) Offene Selbstschädigung. Nervenarzt 65: 296–306
Herpertz S, Saß H (1997) Impulsivität und Impulskontrolle – Zur psychologischen und psychopathologischen Konzeptualisierung. Nervenarzt 68: 171–183
Herpertz S, Saß H (1999a) Die Borderline-Persönlichkeitsstörung in der historischen und aktuellen psychiatrischen Klassifikation. In: Kernberg OF, Dulz B, Sachsse U (eds) Handbuch der Borderline-Störungen. Schattauer, Stuttgart
Herpertz S, Saß H (1999b) Personality disorders and the law, with a German perspective. Curr Opin Psychiatry 12: 689–693
Herpertz S, Steinmeyer EM, Saß H (1994) "Patterns of comorbidity" among DSM-III-R and ICD-10 personality disorders as observed with a new inventory for the assessment of personality disorders. Eur Arch Psychiatry Clin Neurosci 244: 161–169
Herpertz S, Gretzer A, Steinmeyer EM, Muehlbauer V, Schuerkens A, Sass H (1997a) Affective instability and impulsivity in personality disorder: Results of an experimental study. J Affect Disord 44: 31–37
Herpertz S, Steinmeyer EM, Pukrop R, Woschnik M, Saß H (1997b) Persönlichkeit und Persönlichkeitsstörungen: Eine facettentheoretische Analyse der Ähnlichkeitsbeziehungen. Z Klin Psychol 26: 109–117
*Herpertz S, Steinmeyer EM, Saß H (1998) On the conceptualization of subaffective personality disorders. Eur Psychiatry 13: 9–17
Herpertz S, Kunert H, Schwenger UB, Saß H (1999) Affective response in borderline personality disorder – a psychophysiological approach. Am J Psychiatry 156(10): 1550–1556
Hoch P, Polatin P (1949) Pseudoneurotic forms of schizophrenia. Psychiatry Q 23: 248–276
Hoffmann SO (1986) Psychoneurosen und Charakterneurosen. In: Kisker KP, Lauter H, Meyer J, Müller C, Strömgren E (eds) Psychiatrie der Gegenwart, vol 1, 3rd edn. Springer, Berlin Heidelberg New York, pp 29–62
Hoffmann SO, Holzapfel G (1999) Neurosenlehre, psychotherapeutische und psychosomatische Medizin. Schattauer, Stuttgart
Homburger A (1929) Versuch einer Typologie der psychopathischen Konstitution. Nervenarzt 2: 134–136
Horney K (1945) Our inner conflicts. Norton, New York
Huber G, Gross G, Schüttler R (1979) Schizophrenie. Verlaufs- und sozialpsychiatrische Langzeituntersuchungen an den 1945 bis 1959 in Bonn hospitalisierten schizophrenen Kranken. Monographien aus dem Gesamtgebiet der Psychiatrie, vol 21. Springer, Berlin Heidelberg New York
Hughes C (1884) Borderland psychiatric records – pro-dromal symptoms of psychical impairment. Alien Neurol 5: 85–91
Jang KL, Livesley WJ, Vernon PA, Jackson DN (1996) Heritability of personality disorder traits: a twin study. Acta Psychiatr Scand 94: 438–444
Jung CG (1921) Psychologische Typen. Rascher, Zurich
Kahn E (1928) Die psychopathischen Persönlichkeiten. In: Bumke O (ed) Handbuch der Geisteskrankheiten, special part I, vol 4. Springer, Berlin, pp 227–487
Kalus O, Bernstein DP, Siever LJ (1993) Schizoid personality disorder: a review of current status and implications for DSM-IV. J Pers Disord 7: 43–53
Kapfhammer HP (1999) Integrative Therapieansätze bei Borderline Persönlichkeitsstörungen In: Saß H, Herpertz S (eds) Psychotherapie von Persönlichkeitsstörungen. Thieme, Stuttgart, pp 98–115
Kapfhammer HP, Rothenhäusler HB (1999) Zur Psychopharmakotherapie von Persönlichkeitsstörungen des Clusters C. In: Saß H, Herpertz S (eds) Psychotherapie von Persönlichkeitsstörungen. Thieme, Stuttgart, pp 171–180
Kaplan M (1983) A woman's view on DSM-III. Am Psychol 38: 786–792
Kendler KS, Masterson CC, Ungaro R, Davis KL (1984) A family history study of schizophrenia-related personality disorders. Am J Psychiatry 141: 424–427
Kernberg OF (1970) Factors in the treatment of narcissistic personality disorder. J Am Psychoanal Assoc 18: 51–58
Kernberg OF (1976) Object relation theory and clinical psychoanalysis. Aronson, New York
Kernberg OF, Dulz B, Sachsse U (eds) (1999) Handbuch der Borderlinestörungen. Schattauer, Stuttgart
Klein DN, Shih JS (1998) Depressive personality: associations with DSM-III-R mood and personality disorders and negative and positive affectivity, 30-month stability, and prediction of course of axis I depressive disorders. J Abnorm Psychol 107(2): 319–327

Koch JLA (1891–1893) Die psychopathischen Minderwertigkeiten. Maier, Ravensburg
Kohut H (1977) The restoration of the self. International Universities Press, New York
Kraepelin E (1903/1904) Psychiatrie. Ein Lehrbuch für Studierende und Ärzte, 7th edn. Barth, Leipzig
*Kretschmer E (1921) Körperbau und Charakter. Springer, Berlin
Lang H (1981) Zur Frage des Zusammenhangs zwischen Zwang und Schizophrenie. Nervenarzt 52: 643–648
Lempert T, Dieterich M, Huppert D, Brandt T (1990) Psychogenic disorders in neurology: frequency and clinical spectrum. Acta Neurol Scand 82: 335–40
Levy D (1966) Maternal overprotection. Norton, New York
Lewis A (1974) Psychopathic personality: a most elusive category. Psychol Med 4: 133–140
*Linehan MM (1993) Cognitive-behavioral treatment of borderline-personality disorder. Guilford, New York London
Livesley W, West M (1986) The DSM-III distinction between schizoid and avoidant personality disorders. Can J Psychiatry 31: 59–62
Livesley WJ, Schroeder ML, Jackson DN (1990) Dependent personality disorder and attachment problems. J Pers Disord 4: 232–240
*Livesley W, Jang KL, Vernon PA (1998) Phenotypic and genetic structure of traits delineating personality disorder. Arch Gen Psychiatry 55: 941–948
Loranger AW, Lenzenweger MF, Gartner AF et al (1991) Trait-state artifacts and the diagnoses of personality disorders. Arch Gen Psychiatry 48: 720–727
*Loranger AW, Sartorius N, Andreoli A et al (1994) The International Personality Disorders Examination. Arch Gen Psychiatry 51: 215–224
Magnan M, Legrain M (1895) Les dégénérés (état mental et syndromes épisodiques). Rueff, Paris
Maier W, Lichtermann D, Klingler T, Heun R, Hallmayer J (1992) Prevalences of personality disorders (DSM-III-R) in the community. J Pers Disord 6(3): 187–196
Matthey A (1816) Nouvelles Recherches sur les maladies de l'esprit précédées de considé. Paschoud, Paris
Mavissakalian M, Hamman MS (1988) Correlates of DSM-III personality disorder and agoraphobia. Compr Psychiatry 29: 535–544
McGuire J (ed) (1995) What works: reducing reoffending. Wiley, Chichester
Meehl PE (1990) Toward an integrated theory of schizotaxia, schizotypy and schizophrenia. J Pers Disord 4: 1–99
Mellsop GW, Varghese F, Stephens J, Hicks A (1982) The reliability of axis II of DSM-III. Am J Psychiatry 139: 1360–1361
Mentzos S (1982) Neurotische Konfliktverarbeitung. Einführung in die psychoanalytische Neurosenlehre unter Berücksichtigung neuer Perspektiven. Fischer, Frankfurt am Main
Merikangas KR, Weissman MM (1986) Epidemiology of DSM-III Axis II personality disorders. In: Frances AJ, Hales RE (eds) American Psychiatric Association annual review, vol 5. American Psychiatric Press, Washington DC, pp 258–278
Meyer A (1903) An attempt at analysis of the neurotic constitution. Am J Psychol 14: 354–367
Millon T (1969) Modern psychopathology: a biosocial approach to maladaptive learning and functioning. Saunders, Philadelphia
Millon T (1981) Disorders of personality: DSM-III, axis II. Wiley, New York
*Millon T, Davis RD (1996) Disorders of personality. DSM-IV and beyond, 2nd edn. Wiley, New York
Morel BA (1857) Traité des dégénérescences physiques, intellectuelles et morales de l'espèce humaine et des causes qui produisent ces variétés maladives. Baillière, Paris
Morey LC (1988) Personality disorders in DSM-III and DSM-III-R: convergence, coverage, and internal consistency. Am J Psychiatry 145: 573–577
Müller-Isberner R, Cabeza S (1999) Kriminalpräventive Ansätze bei persönlichkeitsgestörten Rechtsbrechern. Persönlichkeitsstörungen 4 (special issue): 91–96
Nemiah JC (1985) Dissociative disorders. In: Kaplan HI, Sadock BJ (eds) Comprehensive textbook of psychiatry, 4th edn. Williams and Wilkins, Baltimore, pp 942–957
New AS, Gelernter J, Yovell Y et al (1998) Tryptophan hydroxylase genotype is associated with impulsive-aggression measures: a preliminary study. Am J Med Genet 81: 13–17
Nigg JT, Goldsmith HH (1994) Genetics of personality disorders: perspectives from personality and psychopathology research. Psychol Bull 115: 346–380
Oldham JM, Skodol AE, Kellmann D, Hyler SE, Rosnick L, Davies M (1992) Diagnosis of DSM-III-R personality disorders by two structured interviews: patterns of comorbidity. Am J Psychiatry 149: 213–220
Parnas J, Schulsinger F, Mednick SA (1990) The Copenhagen high-risk study: major psychopathological and etiological findings. In: Straube ER, Hahlweg K (eds) Schizophrenia. Concepts, vulnerability and intervention. Springer, Berlin Heidelberg New York, pp 45–56
Pepper CM, Klein DN, Anderson RL, Riso LP, Ouimette PC, Lizardi H (1995) DSM-III-R axis II comorbidity in dysthymia and major depression. Am J Psychiatry 152: 239–247
Pinel P (1809) Traité médico-philosophique sur l'aliénation mentale, 2nd edn. Brosson, Paris
Prichard JC (1835) A treatise on insanity and other disorders affecting the mind. Sherwood Gilbert Piper, London
Pukrop R, Herpertz S, Saß H, Steinmeyer EM (1998) Personality and personality disorders. A facettheoretical analysis of the similarity relationships. J Pers Disord 12(3): 226–246
Rado S (1953) Dynamics and classification of disordered behavior. Am J Psychiatry 110: 406–416
Raine A, Reynolds C, Venables PH, Mednick SA, Farrington DP (1998) Fearlessness, stimulation-seeking and large body size at age of 3 years as early predispositions to childhood aggression at age of 11 years. Arch Gen Psychiatry 55: 745–751
Reich JH, Noyes R (1986) Historical comment on DSM-III schizoid and avoidant personality disorders. Am J Psychiatry 143: 1062
Reich JH, Yates WR, Nduaguba M (1989) Prevalence of DSM-III personality disorders in the community. Soc Psychiatry Psychiatr Epidemiol 24: 12–16
Renneberg B, Fydrich T (1999) Verhaltenstherapeutische Therapieansätze in der Gruppenbehandlung der selbstunsicheren Persönlichkeitsstörung. In: Saß H, Herpertz S (eds) Psychotherapie von Persönlichkeitsstörungen. Thieme, Stuttgart, pp 159–170
Rihmer Z (1999) Dysthymic disorder: implications for diagnoses and treatment. Curr Opin Psychiatry 12(1): 69–75
*Robins LN (1966) Deviant children grown up: a sociological and psychiatric study of sociopathic personality. Williams and Wilkins, Baltimore

Rush B (1812) Medical inquiries and observations upon the diseases of the mind. Richardson, Philadelphia

Rutherford MJ, Alterman AI, Cacciola JS, McKay JR (1999) Gender differences in the relationship of antisocial personality disorder criteria to psychopathy checklist-revised scores. J Pers Disord 12: 69–76

*Rutter M (1987) Temperament, personality and personality disorders. Br J Psychiatry 150: 443–458

Sanislow CA, McGlashan TH (1998) Treatment outcome of personality disorders. Can J Psychiatry 43: 237–250

Saß H (1986) Zur Klassifikation der Persönlichkeitsstörungen. Nervenarzt 57: 193–203

Saß H (1987) Psychopathie-Soziopathie-Dissozialität. Zur Differentialtypologie der Persönlichkeitsstörungen. Springer, Berlin Heidelberg New York

Saß H, Herpertz S (1995) Personality disorders. In: Berrios G, Porter R (eds) A history of clinical psychiatry. Athlone, London, pp 633–645

Saß H, Herpertz S (1999) Psychotherapie von Persönlichkeitsstörungen. Thieme, Stuttgart

Saß H, Koehler K (1983) Borderline-Syndrome: Grenzgebiet oder Niemandsland? Nervenarzt 54: 221–230

Saß H, Mende M (1990) Zur Erfassung von Persönlichkeitsstörungen mit einer integrierten Merkmalsliste gemäß DSM-III-R und ICD-10 bei stationär behandelten psychiatrischen Patienten. In: Baumann U, Fähndrich E, Stieglitz RD, Woggon B (eds) Veränderungsmessung in Psychiatrie und klinischer Psychologie. Profil, Munich, pp 195–206

Saß H, Herpertz S, Steinmeyer EM (1993) Subaffective personality disorders. Int Clin Psychopharmacology 8[Suppl 1]: 39–46

Saß H, Steinmeyer EM, Herpertz S, Ebel H, Herpertz S (1995) Untersuchungen zur Kategorisierung und Dimensionierung von Persönlichkeitsstörungen. Z Klin Psychol 24: 239–251

Saß H, Houben I, Herpertz S, Steinmeyer EM (1996) Kategorialer versus dimensionaler Ansatz in der Diagnostik von Persönlichkeitsstörungen. In: Schmitz B, Fydrich T, Limbacher K (eds) Persönlichkeitsstörungen: Diagnostik und Psychotherapie. Psychologie Verlags Union, Weinheim, pp 42–55

Saß H, Herpertz S, Schürkens A (1998) Medikamentöse Therapie von Persönlichkeitsstörungen. In: Gaebel W (ed) Neuroleptika bei nichtpsychotischen Störungen. Springer, Berlin Heidelberg New York, pp 194–207

Schmitz B (1999) Kognitive Verhaltenstherapie bei Patienten mit Persönlichkeitsstörungen: Behandlungsansätze und Psychoedukation. In: Saß H, Herpertz S (eds) Psychotherapie von Persönlichkeitsstörungen. Thieme, Stuttgart, pp 25–47

Schmitz B, Fydrich T, Limbacher K (eds) (1996) Persönlichkeitsstörungen: Diagnostik und Psychotherapie. Psychologie Verlags Union, Weinheim

*Schneider K (1950a) Die psychopathischen Persönlichkeiten, 9th edn. Thieme, Leipzig

*Schneider K (1950b) Klinische Psychopathologie, 5th edn. Thieme, Stuttgart

Schultz JH (1928) Die konstitutionelle Nervosität. In: Bumke O (ed) Handbuch der Geisteskrankheiten, vol 5. Springer, Berlin, pp 28–111

Serban G, Conte HR, Plutchik R (1987) Borderline and schizotypal personality disorders: mutually exclusive or overlapping? J Pers Assess 51: 15–22

Shapiro D (1965) Neurotic styles. Basic Books, New York

Shapiro D (1981) Autonomy and the rigid character. Basic Books, New York

Siever LJ, Davis KL (1991) A psychobiological perspective on the personality disorders. Am J Psychiatry 148: 1647–1658

Silk KR, Lee S, Hill EM, Lohr NE (1995) Borderline personality disorder symptoms and severity of sexual abuse. Am J Psychiatry 152: 1059–1064

*Spitzer RL, Endicott J, Gibbons M (1979) Crossing the border into borderline personality and borderline schizophrenia. Arch Gen Psychiatry 36: 17–24

Stern A (1938) Psychoanalytic investigation of and therapy in the borderline-group of neurosis. Psychoanal Q 7: 467–489

Sullivan HS (1953) The interpersonal theory of psychiatry. Norton, New York

*Tellenbach H (1983) Melancholie, 4th edn. Springer, Berlin Heidelberg New York

Tölle R (1966) Katamnestische Untersuchungen zur Biographie abnormer Persönlichkeiten. Springer, Berlin Göttingen Heidelberg

Tölle R (1986) Persönlichkeitsstörungen. In: Kisker KP, Lauter H, Meyer J, Müller C, Strömgren E (eds) Psychiatrie der Gegenwart, vol 1. Springer, Berlin Heidelberg New York, pp 151–180

Torgersen S (1994) Genetics in borderline conditions. Acta Psychiatr Scand 89(379): 19–25

Trull TJ, Widiger TA, Frances A (1987) Covariations for criteria sets for avoidant, schizoid and dependent personality disorders. Am J Psychiatry 144: 767–771

Turkat ID (1990) The personality disorders. A psychological approach to clinical management. Pergamon, New York

Virkkunen M, Rawlings R, Tokola R et al (1994) CSF biochemistries, glucose metabolism, and diurnal activity rhythms in alcoholic, violent offenders, fire setters, and healthy volunteers. Arch Gen Psychiatry 51: 20–27

von Zerssen D (1988) Der Typus manicus als Gegenstück zum Typus melancholicus in der prämorbiden Persönlichkeitsstruktur affektpsychotischer Patienten. In: Janzarik W (ed) Persönlichkeit und Psychose. Enke, Stuttgart, pp 150–171

von Zerssen D von (1994a) Prämorbide Persönlichkeit. In: Stieglitz RD, Baumann U (eds) Psychodiagnostik psychischer Störungen. Enke, Stuttgart, pp 216–229

von Zerssen D (1994b) Persönlichkeitszüge als Vulnerabilitätsindikatoren. Probleme ihrer Erfassung. Fortschr Neurol Psychiatr 62: 1–13

Weissman MM (1993) The epidemiology of personality disorders: a 1990 update. J Pers Disord 7[Suppl]: 44–62

*Weitbrecht HJ (1963) Psychiatrie im Grundriß. Springer, Berlin Göttingen Heidelberg

Westphal K (1877) Über Zwangsvorstellungen. Arch Psychiatr Nervenkrankheiten 8: 734–750

WHO (1992) International classification of diseases, 10th edn (ICD-10). Chap. V(F). World Health Organization, Geneva

Widiger TA (1992) Generalized social phobia versus avoidant personality disorder: a commentary on three studies. J Abnorm Psychology 101: 340–343

Widiger TA (1998) Four out of five ain't bad. Arch Gen Psychiatry 55: 865–866

Widiger TA, Costa PT (1994) Personality and personality disorders. J Abnorm Psychology 103: 78–91

Widiger TA, Weissman MM (1991) Epidemiology of borderline personality disorder. Hosp Commun Psychiatry 42: 1015–1021

Widiger TA, Miele GM, Tilly SM (1992) Alternative perspectives on the diagnoses of borderline personality disorder. In: Clarkin JF, Marziali E, Munroe-Blum H (eds) Borderline

personality disorder – clinical and empirical perspectives. Guilford, New York, pp 89–115

Widiger TA, Cadoret R, Hare R et al (1996) DSM-IV antisocial personality disorder field trial. J Abnorm Psychology 105: 3–16

Winnicott DW (1965) Ego distortion in terms of true and false self. In: Winnicott DW (ed) The maturational processes and the faciliating environment. International Universities Press, New York, pp 140–152

Zanarini MC, Gunderson JG, Frankenburg FR, Chauncey DL (1990) Discriminating borderline personality disorder from other axis II disorders. Am J Psychiatry 147: 161–167

Zanarini MC, Williams AA, Lewis RE et al (1997) Reported pathological childhood experiences associated with the development of borderline personality disorder. Am J Psychiatry 154: 1101–1106

Zaworka W, Hand I (1981) Die "Anankastische Persönlichkeit" – Fakt oder Fiktion? Experimentelle Diagnostik der Zwangsneurose. Z Diff Diagn Psychol 2: 31–54

Zimmerman M, Coryll W (1990) Diagnosing personality disorders in the community. A comparison of self-report and interview measures. Arch Gen Psychiatry 47: 527–531

CHAPTER

12

K.A. Halmi

Eating Disorders

1 Definition

The two major eating disorders are anorexia nervosa and bulimia nervosa. Substantial research and documentation are available for these disorders, which should be regarded as syndromes because they are not specific diseases with a common cause, common course, and common pathology.

Anorexia nervosa is a disorder characterized by weight loss, behavior directed towards losing weight, preoccupation with body weight and food, peculiar manners of handling food, intense fear of gaining weight, disturbance of body conceptualization, and amenorrhea.

Bulimia nervosa is a disorder in which the behavior of bulimia or binge eating is a predominant behavior. These individuals engage in some sort of compensatory behavior to counteract the potential weight gain from calories ingested during bingeing. They are also overly concerned about their physical appearance.

1.1 Diagnostic Criteria

The specific criteria for anorexia nervosa according to the *Diagnostic and Statistical Manual IV* (DSM-IV) are given in Appendix A (American Psychiatric Association 1994). Criterion A provides a guideline for determining the threshold which is considered "underweight," since there is no specific amount of weight loss associated with the other symptoms that constitute anorexia nervosa.

Anorectics demonstrate their intense fear of gaining weight by their intense preoccupation with thoughts of food and irrational worry about gaining weight. Criterion C describes the complexity of anorectics' disturbance of body conceptualization. Denial of illness is the hallmark symptom of this disorder. The last criterion is amenorrhea. There are two subtypes of anorexia nervosa: the bingeing/purging subtype and the restricting subtype. The latter is especially common in those who develop the illness under the age of 12.

The criteria for bulimia nervosa are more arbitrary and ambiguous compared with those for anorexia nervosa. There is no consensus on what constitutes a binge and how frequently bingeing must be in order to warrant a diagnosis of this disorder. Bulimic patients use a variety of behaviors to prevent weight gain; the most common is self-induced vomiting. Patients tend to estimate their worth in terms of body shape and weight. There are two subtypes of bulimia nervosa: the purging subtype and the nonpurging subtype. Appendix B gives the clinical criteria for bulimia nervosa.

1.2 Clinical Assessment

Current height and weight measurements are necessary for both anorexia nervosa and bulimia nervosa patients. Since there is no specific amount of weight loss associated with the other symptoms that constitute anorexia nervosa, criterion A of the diagnostic criteria merely provides a guideline for determining the threshold below which a person is considered "underweight." To determine a normal weight range for the patient, the clinician should consider the individual's body build and weight history. For children up to the age of 18, pediatric growth charts should be used. It should be noted that some children may not have weight loss, but still weigh less than their expected weight because they have failed to make weight gains while growing in height.

Many anorectic individuals deny their symptoms and are not motivated for treatment. Because of this, it is often necessary to interview family members or close friends to obtain an idea of their actual behavior. For example, they often deny an intense fear of gaining weight, but their behavior shows this fear by their intense preoccupation with thoughts of food and irrational worry about fatness. For example, they frequently look in mirrors to make sure they are thin and incessantly express concern about looking fat. They will take a great deal of time cutting up food into small pieces and rearranging food on their plates in order to eat less.

Anorectics' disturbance of body image is imbedded in a basic disturbed conceptualization of themselves. Anorectics tend to judge their self-worth predominantly by body shape and weight. Although some feel globally overweight, others realize they are thin but think certain parts of their body are too fat, particularly the abdomen, buttocks, and thighs. An overwhelming feeling of inadequacy and ineffectiveness is a core symptom of all anorectics. Their success at losing weight is an impressive achievement and boosts their self-confidence. This confidence and feeling of being in control of one area of their life augments the ability to deny the seriousness of the medical complications of their malnourished condition and thus enforces their belief that thinness defines their self-worth. Obsessive-compulsive behaviors often develop or become worse as the anorexia nervosa proceeds in severity. An obsession with cleanliness, an increase in cleaning activities, and compulsive studying are commonly observed in these patients. Perfectionistic traits are also common.

It is important to obtain a history of menstruation. In some adolescents who have never menstruated, the amenorrhea present is primary. Eating disorder patients, especially anorectics, are extremely unreliable about their menstrual histories. Many state they are

menstruating regularly because they do not wish to have the diagnosis of anorexia nervosa and they do not wish to be treated. Often, when these patients are hospitalized and observed carefully, it is obvious that they are not menstruating. In these patients, plasma levels of estrogen are low, as are levels of follicle-stimulating hormone (FSH) and luteinizing hormone (LH). This is caused by diminished secretion of gonadotropin-releasing hormone (GnRH) from the hypothalamus.

Many adolescent anorectics have delayed psychosocial sexual development, and adults often have a markedly decreased interest in sex with the onset of anorexia nervosa.

Most of the physiological and metabolic changes in anorexia nervosa are secondary to the starvation state or purging behavior. These changes revert to normal with nutritional rehabilitation and the cessation of purging behavior. In anorexia nervosa patients and bulimia nervosa patients who engage in self-induced vomiting or abuse laxatives and diuretics, hypokalemic alkalosis may develop. These patients often have elevated serum bicarbonate, hypochloremia, and hypokalemia. Patients with electrolyte disturbances have physical symptoms of weakness, lethargy, and at times cardiac arrhythmias. The latter condition may result in sudden cardiac arrest. Elevation of serum enzymes reflects fatty degeneration of the liver and is present in both the emaciated anorectic phase and during refeeding. Elevated serum cholesterol levels tend to occur more frequently in younger patients and return to normal with weight gain. Carotenemia is often observed in malnourished anorectic patients. Emaciation can also produce abnormalities in hematopoiesis, such as leukopenia and relative lymphocytosis.

Bulimia nervosa patients can have severe erosion of the enamel of their teeth, pathologic pulp exposures, loss of integrity of the dental arches, diminished masticatory ability, and an obvious unaesthetic appearance of their teeth. Parotid gland enlargement is associated with elevated serum amylase levels in bulimics who binge and vomit. The serum amylase level can be used to follow reduction of vomiting in purging patients who deny their vomiting episodes. Acute dilatation of the stomach is a rare emergency condition for patients who binge. Esophageal tears can also occur in the process of self-induced vomiting. This is usually accompanied by shock. Severe abdominal pain in the bulimia patient should alert the physician to a diagnosis of gastric dilatation and the need for nasal gastric suction, X-rays, and surgical consultation.

In bulimia nervosa and anorexia nervosa patients who vomit, ipecac may be used to induce vomiting. Cardiac failure can be caused by a cardiomyopathy from ipecac intoxication. This is a medical emergency that usually results in death. Symptoms of precardial pain, dyspnea, and general muscle weakness associated with hypotension, tachycardia, and electrocardiogram abnormalities should alert the physician to possible ipecac intoxication.

Patients who binge and purge, whether they be anorexia nervosa or bulimia nervosa patients, have a high comorbidity with major depression, alcohol and substance abuse, and other impulsive behaviors such as suicide attempts, self-mutilation, and stealing. Eating disorder patients may abuse amphetamines to decrease their appetite and lose weight. In summary, a good clinical assessment for an eating disorder patient must include specific inquiries into all of the core eating disorder symptomatology described above, as well as a comprehensive general psychiatric history because of the high comorbidity of these conditions with affective disorder, anxiety disorders, and substance abuse disorders. A serum electrolyte profile with amylase and liver enzyme levels is also essential. For eating disorder patients who have been ill for longer than 3 years, a measurement of bone density is important because of the high prevalence of decreased bone density in chronic dieting patients. This leads eventually to an increased incidence of fractures. It is important to remember that anorexia nervosa patients do not come into assessment and treatment willingly and that it is often necessary to obtain information from others in their environment (Halmi 1994).

1.3 Epidemiology

Conducting methodologically correct and meaningful studies of incidence and prevalence can be a serious problem. With a few exceptions, most of these studies on eating disorders have been conducted on restricted populations in very restricted areas of a country. Thus the true knowledge of incidence and prevalence of anorexia nervosa and bulimia nervosa within various countries is probably never known accurately. However, the studies mentioned below can give us an idea, which in some cases may be a close approximation to reality.

There has been a consistent increase in the incidence of anorexia nervosa in industrialized countries during the past three decades (Hoek 1993). For example, the most recent incidence study conducted in northeastern Scotland revealed that, between 1965 and 1991, there was almost a sixfold increase in the incidence of anorexia nervosa (from 3/100,000 to 17/100,000 cases). There are very few studies of the incidence of bulimia nervosa. In a large representative sample of the Dutch population in Holland, Hoek (1991) reported the incidence of bulimia nervosa as 9.9/100,000 per year during the period 1985–1986 and 11.4 during the period 1986–1989.

Prevalence studies of eating disorders are more abundant and easier to conduct. More recently, the prevalence of anorexia nervosa among young females in community populations has been estimated in a number of survey studies in countries such as England, Sweden, and Scotland. An average prevalence of anorexia nervosa in these countries, using strict diagnostic criteria, was 0.28% of young females (Hoek 1993).

Over fifty prevalence studies of bulimia nervosa conducted between 1981 and 1989 were reviewed by Fairburn and Beglin (1990). Those studies in which surveys and interviews were combined had a remarkably consistent prevalence rate of 1% for bulimia nervosa in adolescent and young adult women.

All studies show that eating disorders are rare among males. In clinical samples, the male to female ratio for eating disorders lies consistently between 1:10 and 1:20 (Hoek 1991).

Social and cultural factors are significant in influencing the prevalence and nature of eating disorders and dieting. The effects of the Western ideal of a slender body type have been particularly powerful. Individuals and groups who are exposed to this ideal seem to be at risk for developing an eating disorder (Crago et al. 1996).

Cross-cultural comparisons of anorexia nervosa and bulimia nervosa are fraught with the problems of correct translation of ideas and semantics. For example, in one study of Chinese patients, Lee et al. (1993) found they denied a fat phobia. However, 3 years later Lee et al. (1996) found that the majority of female Chinese undergraduate students reported feeling fat in their lower body. The authors conclude that a typically "Western" pattern of body dissatisfaction has overshadowed the traditional Chinese notions of female beauty based on the face. They further hypothesize that, in the context of a rapidly urbanizing Chinese society, this will predispose more females to weight control behavior and eating disorders. The major factors of self imposed weight loss or bingeing and purging behavior are of course present across all cultures and constitute the major criteria for the diagnosis of anorexia nervosa and bulimia nervosa. However, there may be subtle differences in the expression of these weight concerns across cultures.

1.4 Differential Diagnosis

For anorexia nervosa patients, it is important to ascertain that the patient has no medical illness that can account for weight loss. In rare circumstances, a patient may have both anorexia nervosa and a medical illness contributing to the weight loss. In such a situation, the diagnosis of anorexia nervosa is made by the positive criteria for anorexia nervosa, and both the underlying medical condition and the anorexia nervosa are diagnosed and treated as such.

Weight loss, peculiar eating behavior, and vomiting can occur in several psychiatric disorders. Weight loss frequently occurs in depressive disorders, which have several features in common with anorexia nervosa, e.g. feeling depressed, crying spells, sleep disturbance, obsessive ruminations, and occasional suicidal thoughts. However, a patient with a depressive disorder usually has a decreased appetite, whereas an anorectic patient denies the existence of an appetite. It is only in the severe stages of anorexia nervosa that the patient actually has a decreased appetite. The hyperactivity seen in anorexia nervosa differs from the agitated activity seen in depressive disorders in that anorectics' activity is planned and ritualistic, e.g. exercising programs, jogging, and cycling. The preoccupation with the calorie content of food, collecting recipes, and preparing gourmet feasts, which is typical of the anorectic patient, is not present in the patient with a depressive disorder. Unlike anorexia nervosa patients, those with depression do not have a fear of becoming fat or a disturbance in their body image.

Weight fluctuations, vomiting, and peculiar food handling may occur in somatization disorder. Generally, the weight loss in somatization disorder is not as severe as in anorexia nervosa, nor does the patient in the former disorder express a fear of becoming overweight. In addition, amenorrhea for longer than 3 months is unusual in somatization disorder.

Delusions about food in schizophrenia are seldom concerned with the calorie content of food. A patient with schizophrenia is not preoccupied with a fear of becoming obese and does not have the hyperactivity seen in an anorectic patient.

Chronic medical illnesses frequently associated with weight loss are Crohn's disease, hyperthyroidism, Addison's disease, and diabetes mellitus.

On rare occasions, a central nervous system tumor may be associated with bulimic behaviors. Overeating episodes also occur in the Klüver-Bucy syndrome, which consists of visual agnosia, compulsive licking and biting, an inability to ignore any stimulus, and hypersexuality. Another uncommon syndrome associated with hyperphagia is the Kleine-Levin syndrome, which is characterized by periodic hypersomnia lasting for several weeks.

2 Etiology and Risk Factors

A multidimensional model for conceptualizing eating disorders emphasizes the interaction of biological,

psychological, and cultural factors. Within each of these areas, research has identified factors that predispose, precipitate, or maintain the eating disorder.

2.1 Genetic Factors

Genetic factors include a genetic vulnerability or predisposition that could become manifest under adverse conditions, such as inappropriate dieting or emotional stress. The genetic predisposition could be a particular personality type, a tendency to susceptibility to psychiatric instability (especially affective or anxiety disorders), or a hypothalamic dysfunction.

In a series of 67 twin probands, Treasure and Holland (1989) found that the concordance for restricting anorexia nervosa was markedly higher for monozygotic (66%) than for dizygotic twins (0%). In a large population-based twin registry study, Kendler et al. (1991) showed that concordance for bulimia nervosa was significantly higher in monozygotic than in dizygotic twin pairs. In this study, rates of bulimia nervosa were found to be higher in individuals born more recently, giving a time cohort effect. In the first large uncontrolled study of siblings of anorectics, Theander (1970) found in a sample of 94 anorectic patients a risk of 6.6% for female siblings to be diagnosed with anorexia nervosa. In a large case-controlled design with rigorous methods of diagnosis, Strober et al. (1990) found a familial aggregation of anorexia nervosa and bulimia nervosa in the first-degree relatives of anorexia nervosa probands.

2.2 Neurobiological Factors

Most of the neuroendocrine changes in the eating disorders are directly due to dieting behaviors, weight loss, and other behaviors related to reducing caloric intake. These changes revert to normal with resumption of normal eating behavior and nutritional rehabilitation. A vulnerability for destabilization of some of the endocrine and metabolic mechanisms affecting eating behavior may influence the development of the full-blown eating disorder under stresses such as severe dieting. Increased corticotropin-releasing hormone (CRH) secretion in underweight anorectic patients and in normal-weight dieting individuals returns to normal with weight restoration. However, CRH is a potent anorectic hormone and may have a role in maintaining anorectic behaviors and initiating a relapse.

It is well established that the pituitary cells producing LH and FSH are understimulated in patients with anorexia nervosa because of hyposecretion of GnRH by the hypothalamus. Dysfunctions may be present in the neurotransmitter systems that influence GnRH release. There is considerable variability in the function of the hypothalamic–pituitary–ovarian axis in bulimia nervosa women. This is probably due to the fact that there is considerable difference in patterns of severe dieting and fasting in these patients. Some anorectic women remain amenorrheic for long periods of time (several years) after weight restoration. Studies have shown these women to be more psychologically disturbed compared to those that have a rapid resumption of menses.

It is well established that central serotonin pathways modulate feeding. Administration of serotonin antagonists produces increased food intake and subsequent weight gain; conversely, serotonin agonists decrease food intake. Kaye et al. (1991) found elevated cerebral spinal fluid (CSF) 5-hydroxyindoleacetic acid (5-HIAA) levels in anorectic patients with long-term recovery compared with control subjects. Restricting-type anorexia nervosa patients are especially rigid, inhibited, ritualistic, and perfectionistic, and these are personality traits present in obsessive–compulsive personality disorder and to some degree in obsessive–compulsive disorder. It has long been speculated that serotonin system abnormalities are present in patients with this disorder, especially since their symptoms decrease somewhat with treatment using serotonergic antagonists. In the past year, three groups (Collier et al. 1997; Enoch et al. 1998; Sorbi et al. 1998) found that the -1438A 5-HT_{2a} allele was more common in anorectics than in controls. In two other studies, Campbell et al. (1997) and Hinney et al. (1997) did not find this to be true. An abnormality of serotonin functioning in anorectic patients may contribute to the development and maintenance of anorexia nervosa.

Both serotonin agonists and antagonists have been found useful as an adjunct treatment of anorexia nervosa. Cyproheptadine, a serotonin antagonist, has been shown to facilitate weight gain in anorectic patients. This drug most likely acts on hypothalamic appetite centers to decrease satiety and increase food intake. Clomipramine and fluoxetine (serotonin reuptake inhibitors) have been useful in preventing weight relapse in anorexia nervosa and may specifically target the characteristic obsessive–compulsive behaviors that are seen with food and weight control.

There is an association of low levels of CSF 5-HIAA with impulsive, suicidal, and aggressive behavior. Animal studies have shown that impaired serotonergic function leads to overeating and obesity. The bingeing and purging behaviors of bulimic patients are suggestive of impulse control and satiety-regulation problems. Several studies have reported impaired serotonergic function in bulimic patients. Specifically, bulimic patients have lower CSF 5-HIAA levels than

control subjects and blunted prolactin responses to meta-chlorophenyl piperazine (m-CPP) and to the serotonin agonist fenfluramine (Levitan et al. 1997; Jimerson et al. 1997). There seems to be enough evidence of serotonergic dysfunction in anorexia nervosa and bulimia nervosa patients to consider vulnerability in the serotonergic neurotransmitter system as a risk factor for the development of an eating disorder.

Norepinephrine facilitates eating in animals and has reduced functioning in both underweight and weight-recovered anorexia nervosa patients. This finding suggests that brain norepinephrine abnormalities may be a risk factor for the development of or relapse in anorexia nervosa.

Animal studies have shown a role for dopamine and opioids in modulating eating behavior and pleasure–reward responses to food. Anorexia nervosa patients have a reduced response of prolactin to chlorpromazine challenge and a growth hormone L-dopa challenge. This suggests that patients with anorexia nervosa have a defect in the negative feedback system in dopamine synthesis, such as impairment at the postsynaptic receptor site. Bulimic patients without a history of anorexia nervosa appear to have lower CFS levels of homovanillic acid and a less vigorous dopamine response to a clonidine challenge than bulimic patients with a history of anorexia nervosa. These findings suggest that abnormalities in the dopaminergic pathways could lead to decreased satisfaction after eating, which in turn may facilitate binge-eating behavior. The abnormal pleasure–reward response to food seen in both anorexia and bulimia nervosa patients could be related to dopamine and its role in self-stimulating or addictive behaviors.

Neuropeptides have a role in regulating satiety, appetite, mood, and neuroendocrine functions and thus may contribute to the pathophysiology of eating disorders. Opioid agonists increase and opioid antagonists decrease food intake in animals and humans. The opioid antagonist naloxone was found to decrease bingeing in bulimic patients, but the antagonist naltrexone decreased bingeing in bulimia patients only at doses high enough to produce hepatotoxicity. The opioids are linked with pleasure–reward systems, and it is therefore possible that bulimic patients are obtaining some relief from anxiety by bingeing–purging behavior, which may activate the opioid system.

Cholecystokinin (CCK) is secreted by the gastrointestinal system in response to food intake and provides signals to the brain via the vagus nerve. CCK both in the central nervous system and periphery promotes satiety. Both pre- and postprandial plasma CCK concentrations are normal in anorectic women. In some bulimics, there was an impaired plasma CCK response after a meal. It is likely that reduced CCK response is an epiphenomenon resulting from bingeing and purging behavior, since abstinence from these behaviors results in a normal CCK response.

Neuropeptide Y (NPY) and peptide YY (PYY) are found in the central nervous system and are potent endogenous stimulants to feeding behavior. Underweight anorectic women have elevated CSF NPY levels, which reduce to normal with weight rehabilitation. CSF NPY levels are normal in bulimic patients; however, CSF PYY levels are elevated after abstinence from bingeing and purging has occurred. These findings are somewhat puzzling. The elevated NPY levels in underweight anorectics are obviously not effective in getting the anorectic woman to eat. It may be possible that the elevated PYY levels in abstinent bulimics may be a factor contributing to relapse.

Vasopressin and oxytocin are neuropeptides from the hypothalamus; the former controls water-free clearance by the kidney and the latter promotes uterine contraction during parturition and milk production during the postpartum period. Oxytocin also appears to disrupt memory consolidation and retrieval, whereas vasopressin promotes the consolidation of learning. Oxytocin also inhibits vasopressin-induced release of adrenocorticotropic hormone (ACTH) from the anterior pituitary gland. There are abnormalities in vasopressin secretion during hypertonic saline infusion in underweight and weight-restored anorectic patients and in bulimic patients. CSF vasopressin levels are high in underweight anorectic patients, and oxytocin levels are low. It is possible that a low level of oxytocin and a high level of vasopressin might work in concert to enhance the retention of distorted thinking and contribute to the anorectic patient's obsessional concerns about food. All of the above neuroendocrine findings and data are presented in a review article by Halmi (1995).

When the product leptin of the OB gene was identified, there was a rush to measure leptin levels in anorectic patients, since elevated levels of leptin were thought to cause weight loss. However, numerous studies have shown that leptin levels in anorectic patients correlate significantly with body mass index (BMI) and amount of adipose tissue. Thus emaciated anorectics have extremely low levels of leptin, which increase as patients gain weight (Eckert et al. 1998).

Because anorexia nervosa predominately starts during puberty (there is a bimodal peak for age onset at ages 14–15 and age 18), hypotheses have been generated associating endocrine changes at puberty with the cause of anorexia nervosa. Frisch (1985) hypothesized that a critical mass of fat tissue is necessary for normal gonadotropin effects and that an increase in fat tissue mass is associated with a weight increase at the time of menarche. Crisp (1984) developed the hypothesis that anorexia nervosa reflects an attempt to adapt to and

cope with maturational problems through the mechanism of avoidance of biological maturity. Crisp states that the reversal of the pubertal process is a psychological regression permitting avoidance of the self at a normal body weight. He noted that females who have an early menarche or who are mildly overweight are at a greater risk for developing eating disorders. Early menarche is usually associated with increased weight gain and change in body shape and can produce self-consciousness based on the obvious differences these young women show from their peers. This in turn can produce sensitivity to body image and an emotional vulnerability that makes them susceptible for dieting behaviors and the development of eating disorders.

2.3 Psychological Factors

Research on personality traits and psychiatric disorders in eating disorder patients is typically carried out after the eating disorder has developed. In some cases, studies have been done after patients have recovered from their eating disorder, but it has been difficult to determine whether personality features were a cause or a consequence of the eating disorder. Various studies of affective disorder in eating disorder patients have shown a lifetime prevalence rate of major depression that varies from 25% to as high as 80%. The evidence is less clear as to whether depression precedes and/or predisposes an individual to the onset of an eating disorder. At present, it can be said that major depression may precede, coincide with, or follow the onset of the eating disorder (Devlin et al. 1990).

High rates of anxiety disorders (36%–65%) have also been documented with anorexia and bulimia nervosa. Two studies found a lifetime prevalence of 65% (Halmi et al. 1991) and 60% (Toner et al. 1988) for anxiety disorders among eating disorder patients. Lifetime rates of an obsessive-compulsive disorder have varied from 26% to 44% in anorexia nervosa patients (Halmi et al. 1991; Toner et al. 1988). In a study of recovered anorexia nervosa patients, both Casper (1990) and Strober (1980) found that, compared to controls and to their sisters, anorectics with long-term recovery had greater obsessive-compulsive-like personality disorder traits. When assessed after the disorder, 23%–80% of anorectic patients received a diagnosis of personality disorder (Vitousek and Manke 1994). Restricting anorectics tend to receive cluster C (anxious-fearful) diagnoses (e.g. avoidant, dependent). By contrast, anorectic bulimics have an equal likelihood of receiving a cluster B (dramatic-erratic) personality diagnosis (e.g. borderline) or a cluster C one.

Diagnostic analyses have produced quite heterogeneous personality profiles of bulimia nervosa patients. It appears that no fundamental personality trait or traits have been identified as risk factors for the development of bulimia. Studies have shown the prevalence of personality disorders in bulimic patients have ranged from 21% to 77%, with diagnoses occurring most often in the dramatic-erratic cluster (cluster B). The rates of borderline personality disorder (BPD), in particular, have ranged from 2% to 47% (Wonderlich 1995). Most of the above psychological factors appeared to put individuals at risk for an eating disorder by increasing the likelihood that they will diet. Thus dieting itself is a major precipitating factor in the process of developing and maintaining an eating disorder.

It has been impossible to examine whether family conditions play an important role as risk factors in the development of eating disorders. Studies have been made of family interactions after patients have developed the eating disorders. In bulimic families, for example, studies have uncovered deficits such as lack of parental affection, negative, hostile, and disengaged interactions within the family, parental impulsivity, and family alcoholism and obesity (Strober and Humphrey 1987). Bulimic patients rate their families as conflictive, badly organized, noncohesive, and lacking in nuturance and caring (Wonderlich and Mitchell 1992). Anorectics perceive their families as stable, nonconflictive, cohesive, and with no lack of nurturance (Vandereycken et al. 1989). These patterns of interaction in the families of eating disorder patients are currently only descriptive and not explanatory.

Stressful life events may also be a risk factor for developing an eating disorder. There is controversy regarding the role of sexual abuse as a risk factor in the initiation of an eating disorder. Studies investigating the relationship between sexual abuse and eating disorders have produced highly discrepant results (Conners and Morse 1993). Overall, results suggest that approximately 30% of eating disorder patients have been sexually abused in childhood, and this figure is comparable to rates found in normal populations. Findings indicate that childhood psychological abuse in multiple forms increases the likelihood of lifetime comorbid axis I disorders and personality pathology among bulimic patients. A surprisingly low rate of sexual abuse has been reported among restricting-type anorectics relative to either bulimic anorectics or to normal-weight bulimics (Waller et al. 1993).

3 Course of Illness

The long-term studies of eating disorders are based on patient samples and do not include people who have eating disorders but do not get treatment. Therefore,

any generalization about the course of eating disorders can only be applied to patients who have presented for treatment. Long-term follow-up research indicates that the majority of patients with anorexia nervosa have ongoing problems with the illness (Eckert et al. 1995; Halmi et al. 1991; Theander 1985; Hsu and Crisp 1979). About one fourth recover from the disorder, one fourth stay chronically ill with no improvement, and about one half have partial improvement. Mortality rates at 10 years are 6.6%, and at 30 years are 18%–20% after presentation for treatment.

Since bulimia nervosa was first identified as a separate entity in 1979, there have been few long-term follow-up studies of this disorder. The majority of studies have had a 6-month to 2-years post-treatment follow-up. Keel and Mitchell (1997) reviewed 88 articles on the naturalistic and treatment follow-up studies of bulimia patients. Mortality rates for bulimia nervosa were estimated at between 0% and 3%. At a follow-up period of between 5 and 10 years, about 50% fully recover, while 20% continue to meet criteria for bulimia nervosa. Relapse poses a serious threat for bulimics, as about one third of recovered bulimics relapse within 4 years after treatment. About 20% of bulimics seem to sustain an unremitting bulimic illness.

3.1 Predictors of Outcome

In a 10-year follow-up study of 76 patients with severe anorexia nervosa, Eckert et al. (1995) established that those patients with partial recovery were more likely to have had bulimic symptoms at presentation for treatment with continuing bulimic symptoms through to the follow-up time 10 years later. In contrast, with the fully recovered patients, those with partial recovery expressed a greater preoccupation with weight and shape as well as being less well adjusted in their social, educational/vocational, and sexual functioning. Patients with an earlier age of onset recovered from the disorder more quickly. Some studies have found that a long duration of illness and very low weight at presentation for treatment are predictors of a bad outcome, and others have not found this to be the case. Most studies have found purging behavior to be predictive of a bad outcome.

In their review of research studies of bulimia nervosa, Keel and Mitchell (1997) concluded that personality disorders marked by problems with impulse control were associated with a worse prognosis in patients with bulimia nervosa. However, larger patient samples and longer-term follow-ups of bulimia nervosa patients are necessary to more accurately answer the questions regarding predictors of outcome in this disorder.

3.2 Standardized Assessments

The most reliable and accurate manner of making a correct diagnosis of anorexia nervosa or bulimia nervosa is by clinical interview using DSM-IV criteria. The structured clinical interview for DSM-IV axis I disorders and for personality disorders is a well-studied and frequently used semi-structured interview that is particularly helpful in establishing comorbid psychiatric disorders in eating disorder patients (First et al. 1995).

Two scales are especially useful for the documentation and measurement of the severity of core eating disorder symptomatology. The Eating Disorders Examination (EDE) is an interview for assessing current eating disorder symptoms and contains four subscales: restraint, eating concerns, shape concern, and weight concern. It also measures the frequency of binge eating and compensatory behaviors. The validity and reliability of the EDE have been well documented (Fairburn and Cooper 1984), and the instrument has been used extensively in treatment outcome studies of bulimia nervosa. The Yale-Brown-Cornell Eating Disorder Scale (YBC-EDS) is an eight-item scale assessing the severity of preoccupations and rituals common in the eating disorders. The YBC-EDS also has a set of six provisional items for assessing motivation for change. The eight-item scale and the motivational items have shown excellent reliability and convergent validity. The YBC-EDS is an easy-to-administer interview which is able to assess the severity of illness associated with the individual's unique eating disorder symptomatology (Mazure et al. 1994).

A pervasive feeling of inadequacy is common in both anorectic and bulimic patients. The Rosenberg Self-Esteem Questionnaire (RSEQ) is a ten-item self-report questionnaire with established validity and reliability and has been found in several studies to be a predictor of outcome for anorexia nervosa patients (Rosenberg 1979).

The Social Adjustment Scale (SAS) has a self-report version which assesses social functioning in several areas, including work, family relationships, social and leisure time, and economic functioning. It is widely used as a measure of general social functioning and has also been found to be a good predictor of outcome in eating disorder treatment studies (Weissman and Bothwell 1976).

Weight and height measurements are necessary for all eating disorder patients. From these measures it is possible to calculate the BMI (kg/m^2), i.e. weight in kilograms divided by height squared measured in meters. The BMI is widely used because it is a "standard score" of body mass. It is possible to set a

normal weight range across various countries and cultures by using this standard score as a measurement. Other medical assessments are discussed in the section on clinical assessments.

4 Treatment

The severity of illness will determine the intensity of treatment required for the eating disorder patient. Treatment levels range from a specialized eating disorder inpatient unit to a partial hospitalization or day program to outpatient care, depending on the weight, medical status, and other psychiatric comorbidity of the patient. Because most inpatients require intensive medical management and/or monitoring for suicidal and impulsive behaviors, controlled studies examining various treatment modalities are difficult to conduct and there have thus been very few of them. Medical management usually requires weight restoration, nutritional rehabilitation, rehydration and correction of serum electrolytes, and daily monitoring of weight, food and calorie intake, and urine output. Patients must be closely monitored for vomiting and drug abuse behavior.

Since patients with anorexia nervosa are resistance to and disinterested in treatment, there are very few outpatient controlled treatment studies. Open studies have indicated that a multifaceted treatment approach is the most effective, and this includes medical management, psychoeducation, and individual therapy utilizing both cognitive and behavior therapy principles. Controlled studies have shown that children under the age of 18 do better if they have family therapy. Nutritional counseling and pharmacological intervention can also be useful components in the treatment plan.

4.1 Pharmacotherapy

Medication is only a useful adjunct in the treatment of anorexia nervosa. The first drug used to treat anorectic patients was chlorpromazine. Surprisingly, there are no double-blind controlled studies to prove definitively its effectiveness in reducing core anorectic symptomatology and in inducing weight gain. Clinical experience has shown this medication to be particularly helpful in severely ill patients who are overwhelmed by constant thoughts of losing weight and have incessant behavioral rituals. Cyproheptadine in high doses (up to 28 mg/day) can facilitate weight gain in restricting-type anorectics and also has an antidepressant effect (Halmi et al. 1986). Some recent open studies indicate that fluoxetine may be effective in preventing relapse in anorexia nervosa.

Over a dozen double-blind, placebo-controlled trials of antidepressants, including desipramine, imipramine, amitriptyline, nortriptyline, phenelzine, and fluoxetine, have been conducted in normal-weight outpatients with bulimia nervosa. The dosage of antidepressant medication used was similar to that used for the treatment of depression. In all the trials, antidepressants were significantly more effective than placebo in reducing binge eating. In addition, the medication improved mood and reduced eating disorder symptoms, such as preoccupation with shape and weight. The abstinence rate from bingeing and purging in these studies was only 22%, which is disappointing. There are excellent reviews of these studies by Mitchell and deZwaan (1993) and by Walsh and Devlin (1992).

4.2 Cognitive-Behavioral Therapy

Cognitive-behavioral therapy (CBT) principles can be applied in both inpatient and outpatient settings. CBT has not been systematically studied in anorexia nervosa patients. There have been a few controlled inpatient studies using behavior therapy, which was found to be effective in inducing weight gain. While effective in the short term, there are no data supporting the clinical effectiveness of behavior therapy as a sole treatment for anorexia nervosa. For a review of behavioral techniques in the management of anorexia nervosa, the reader is referred to Halmi (1984).

There are no large sample size, controlled studies of cognitive therapy in anorexia nervosa patients. Monitoring is an essential component of CBT. Patients are taught to monitor their food intake, their feelings and emotions, their bingeing and purging behaviors, and their problems in interpersonal relationships. Cognitive restructuring is a method patients are taught in order to identify autonomic thoughts and challenge their core beliefs. Problem solving is a specific method whereby patients learn how to think through and devise strategies to cope with their food-related and/or interpersonal problems. For a comprehensive review of CBT in the treatment of eating disorders, the reader is referred to Fairburn and Wilson (1993), Gardner and Bemis (1992), and Kleifield et al. (1996).

CBT is the first-line treatment for bulimia nervosa. It has been found to be the most effective treatment in over 35 controlled studies. About 40%–50% of patients are abstinent from both bingeing and purging at the end of treatment (16–20 weeks). Improvement by a reduction in bingeing and purging occurred in 70%–95% of patients. Another 30% of those who did not

show improvement immediately after treatment did show improvement to full recovery 1 year after treatment. In patients where depression was comorbid with bulimia, CBT was also found to decrease depression. For a review of these studies, the reader is referred to Mitchell and Raymond (1992). In bulimia nervosa patients, CBT interrupts the self-maintaining cycle of bingeing and purging and alters the individual's dysfunctional cognitions and beliefs about food, weight, body image, and overall self-concept.

4.3 Family Therapy

A controlled family therapy study by Russell et al. (1987) showed that anorectic patients under the age of 18 benefited from family therapy, whereas patients over the age of 18 did worse in family therapy than in the control therapy. There are no controlled studies of the combination of individual and family therapy. In actual practice, many clinicians provide individual therapy and some sort of family counseling for managing anorexia nervosa patients. A family analysis should be done on all anorexia nervosa patients who are living with their families. On the basis of this analysis, a clinical judgement can be made as to what type of family therapy or counseling is advisable. Family therapy is not widely used in the treatment of bulimia nervosa. This is probably due to the fact that most patients with bulimia nervosa are in their 20s and living away from their family of origin. There is a consensus, however, that families of younger patients should be involved in their treatment.

4.4 Medical Management

The medical management of inpatients is discussed under the general treatment heading. For the anorexia nervosa outpatient, regular weight monitoring is necessary. If there is any indication that the patient is purging, as with the bulimia patient, monitoring of serum electrolytes is important. Measurement of serum amylase is also useful in a patient who may be vomiting. For chronically ill patients, measurement of bone density will be an important indication of nutritional status and a risk indicator for the development of fractures.

The majority of patients with anorexia nervosa and bulimia nervosa are chronically inflicted with core eating disorder symptomatology of varying degrees of severity. While pharmacotherapy, CBT, and family therapy have produced effective interventions, they have not had impressive cure rates. Unlike substance abuse disorders or chronic medical illnesses such as diabetes, the eating disorders have elicited morbid curiosity but not empathy from sources funding treatment studies. Future treatment research needs to examine the combination of various treatment modalities and the sequential application of different treatments.

Appendix A. 307.1 Anorexia Nervosa

A. Refusal to maintain body weight at or above a minimally normal weight for age and height (e.g. weight loss leading to maintenance of body weight less than 85% of that expected; or failure to make expected weight gain during period of growth, leading to body weight less than 85% of the expected).
B. Intense fear of gaining weight or becoming fat, even though underweight.
C. Disturbance in the way in which one's body weight or shape is experienced, undue influence of body weight or shape on self-evaluation, or denial of the seriousness of the current low body weight.
D. In post-menarcheal females, amenorrhea, i.e. the absence of at least three consecutive menstrual cycles. (A women is considered to have amenorrhea if her periods occur only following hormone, e.g. estrogen, administration.)

Specify type:

- Restricting type: during the current episode of Anorexia Nervosa, the person has not regularly engaged in binge eating or purging behavior (i.e. self-induced vomiting or the misuse of laxatives, diuretics, or enemas).
- Binge Eating/Purging type: during the current episode of Anorexia Nervosa, the person has regularly engaged in binge eating or purging behavior (i.e. self-induced vomiting or the misuse of laxatives, diuretics, or enemas).

Appendix B. 307.51 Bulimia Nervosa

A. Recurrent episodes of binge eating. An episode of binge eating is characterized by both of the following:
   (1) eating, in a discrete period of time (e.g. within any 2-h period), an amount of food that is definitely larger than most people would eat during a similar period of time and under similar circumstances

(2) a sense of lack of control over eating during the episode (e.g. a feeling that one cannot stop eating or control what or how much one is eating)

B. Recurrent inappropriate compensatory behaviors in order to prevent weight gain, such as self-induced vomiting; misuse of laxatives, diuretics, enemas, or other medications; fasting or excessive exercise.

C. The binge eating and inappropriate compensatory behaviors both occur, on average, at least twice a week for 3 months.

D. Self-evaluation is unduly influenced by body shape and weight.

E. The disturbance does not occur exclusively during episodes of Anorexia Nervosa.

Specify type:

- Purging type: during the current episode of Bulimia Nervosa, the person has regularly engaged in self-induced vomiting or the misuse of laxatives, diuretics, or enemas.
- Nonpurging type: during the current episode of Bulimia Nervosa, the person has used other inappropriate compensatory behaviors, such as fasting or excessive exercise, but has not regularly engaged in self-induced vomiting or the misuse of laxatives, diuretics, or enemas.

5 References

American Psychiatric Association (1994) Diagnostic and statistical manual of mental disorders (DSM-IV). American Psychiatric Association, Washington, DC

Campbell D, Sundaramurthy D, Markham A et al (1997) Lack of association between 5-HT_{2a} gene promoter polymorphism and susceptibility to anorexia nervosa. Lancet 351: 499–503

Casper RC (1990) Personality features of women with good outcome from restricting anorexia nervosa. Psychosom Med 52: 156–170

Collier DA, Arranz M, Lee T et al (1997) Association between 5-HT_{2a} gene promoter polymorphism and anorexia nervosa. Lancet 350: 412–413

Connors ME, Morse W (1993) Sexual abuse and eating disorders: a review. Int J Eat Dis 13: 1–11

Crago M, Schisslak CM, Estes LS (1996) Eating disturbances among American minority groups: a review. Int J Eat Dis 19: 239–248

Crisp AH (1984) Premorbid factors in adult disorders of weight, with particular reference to primary anorexia nervosa (weight phobia). J Psychosom Res 14: 1–22

Devlin MJ, Walsh BT, Kralg D et al (1990) Metabolic abnormalities in bulimia nervosa. Arch Gen Psychiatry 47: 144–148

*Eckert ED, Halmi KA, Marchi EP et al (1995) Ten year follow up of anorexia nervosa: clinical course and outcome. Psychol Med 25: 143–156

Eckert E, Pomeroy C, Raymond N et al (1998) Leptin in anorexia nervosa. J Clin Endocrinol Metab 83: 791–795

Enoch M, Kaye W, Rotondo A et al (1998) 5-HT_{2a} promoter polymorphism-1438 G/A, anorexia nervosa, and obsessive compulsive disorder. Lancet 351: 1785–1786

Fairburn CG, Beglin SJ (1990) Studies of the epidemiology of bulimia nervosa. Am J Psychiatry 147: 401–408

Fairburn CG, Cooper PJ (1984) The clinical features of bulimia nervosa. Br J Psychiatry 144: 238–246

*Fairburn CG, Wilson JT (1993) Cognitive-behaviorial therapy for binge eating and bulimia nervosa: a comprehensive treatment manual. In: Fairburn CG, Wilson JT (eds) Binge eating: nature, assessment and treatment. Guilford, New York, pp 361–404

First M, Spitzer R, Gibbon M et al (1995) The structured clinical interview for DSM-IV axis I disorders and personality disorders. Psychiatric Institute, New York

Frisch S (1985) Fatness, menarche and female fertility. Persp Biol Med 28: 611–633

Gardner DM, Bemis KM (1982) A cognitive-behavioral approach to anorexia nervosa. Cogn Ther Res 6: 1223–1250

Halmi KA (1984) Behaviorial management of anorexia nervosa. In: Garner D, Garfinkel P (eds) Treatment of anorexia nervosa and bulimia. Guilford, New York, pp 147–60

**Halmi KA (1999) Eating disorders. In: Hales KE, Yudofsky SC, Talbott J (eds) Textbook of psychiatry, 3rd edn. American Psychiatric Press, Washington, DC, pp 983–1002

Halmi KA (1995) Basic biological overview of eating disorders. In: Bloom FE, Kupfor DJ (eds) Psychopharmocology; the fourth generation of progress. Raven, New York, pp 1609–1616

Halmi KA, Eckerts E, Ladu T et al (1986) Anorexia nervosa: treatment efficacy of cyproheptadine and amitriptyline. Arch Gen Psychiatry 43: 177–181

Halmi KA, Eckert E, Marchi EP et al (1991) Comorbidity of psychiatric diagnoses in anorexia nervosa. Arch Gen Psychiatry 48: 712–718

Hinney A, Ziegler A, Nothen M et al (1997) 5-HT_{2a} receptor gene polymorphisms, anorexia nervosa and obesity. Lancet 351: 1324–1325

Hoek HW (1991) The incidence and prevalence of anorexia nervosa and bulimia nervosa in primary care. Psychol Med 21: 455–460

*Hoek HW (1993) Review of the epidemiological studies of eating disorders. Int Rev Psychiatry 5: 61–74

Hsu H, Crisp A (1979) Outcome of anorexia nervosa. Lancet 1: 61–65

Jimerson DC, Wolfe BE, Metzger ED et al (1997) Decreased seratonin function in bulimia nervosa. Arch Gen Psychiatry 54: 529–534

Kaye WH, George DT, Gwirtsman H et al (1991) Altered serotonin activity in anorexia nervosa after long-term weight restoration. Arch Gen Psychiatry 48: 556–562

**Keel P, Mitchell JE (1997) Outcome in bulimia nervosa. Am J Psychiatry 154: 313–321

Kendler KS, MacLean C, Neal M et al (1991) The genetic epidemiology of bulimia nervosa. Am J Psychiatry 148: 1627–1637

**Kleifield EI, Wagner S, Halmi KA (1996) Cognitve-behaviorial treatment of anorexia nervosa. Psychiatr Clin North Am 19: 715–737

Lee S, Ho TP, Hsu LKG (1993) Fat phobic and non fat phobic anorexia nervosa: a comparative study of 70 Chinese patients in Hong Kong. Psychol Med 23: 999–1017

Lee S, Leung T, Lee AM et al (1996) Body dissatisfaction among Chinese undergraduates and its implications for eating disorders in Hong Kong. Int J Eat Dis 20: 77–84

Levitan RD, Kaplan AS, Joffe RT (1997) Hormonal and subjective responses to intravenous meta-chlorophenyl piperazine in bulimia nervosa. Arch Gen Psychiatry 54: 521–587

Mazure CM, Halmi KA, Sunday SR et al (1994) Yale-Brown-Cornell Eating Disorder Scale: development, use, reliability and validity. J Psychiatr Res 28: 425–445

Mitchell JE, Raymond NC (1992) Cognitive-behavioral therapy and the treatment of bulimia nervosa. In: Halmi KA (ed) The psychobiology and treatment of anorexia nervosa and bulimia nervosa. American Psychiatric Press, Washington, DC

Mitchell JE, deZwaan M (1993) Pharmacological treatments of binge eating. In: Fairburn CG, Wilson JT (eds) Binge eating: nature, assessment and treatment. Guilford, New York

Rosenberg M (1979) Conceiving the self. Basic, New York

Russell GFM, Szmuckler GI, Dare C et al (1987) An evaluation of family therapy in anorexia nervosa and bulimia nervosa. Arch Gen Psychiatry 44: 1047–1056

Sorbi S, Nacmias B, Jedde A (1998) 5-HT_{2a} promoter polymorphism in anorexia nervosa. Lancet 351: 1785

Strober M (1980) Personality and symptomatological features in young, nonchronic anorexia nervosa patients. J Psychcosom Res 24: 353–359

Strober M, Humphrey LL (1987) Familial contributions to the etiology and course of anorexia and bulimia. J Consult Clin Psychol 55: 654–659

Strober M, Lampert C, Morrell W et al (1990) A controlled family study of anorexia nervosa: evidence of familial aggregation and lack of shared transmission with affective disorders. Int J Eat Dis 9: 239–253

Theander S (1970) Anorexia nervosa: a psychiatric investigation of 94 female cases. Acta Psychiatr Scand Suppl 214: 1–94

Theander S (1985) Outcome and prognosis in anorexia nervosa and bulimia: some results of previous investigations compared with those with a Swedish long-term study. J Psychiatr Res 19: 493–504

Toner BB, Garfinkel PE, Garner DM (1988) Affective and anxiety disorders in long-term follow-up of anorexia nervosa. Int J Psychiatr Med 18: 357–364

Treasure J, Holland AJ (1989) Genetic vulnerability to eating disorders: evidence from twin and family studies. In: Remschmidt H, Schmidt MH (eds) Child and youth psychiatry: European perspectives. Hogrefe and Hubert, New York, pp 59–68

Vandereycken W, Koge L, Vanderlinden J (1989) The family approach to eating disorders: assessment and treatment of anorexia nervosa and bulimia. PMA, New York

Vitousek K, Manke F (1994) Personality variables and disorders in anorexia nervosa and bulimia nervosa. J Abnorm Psychol 103: 137–147

Waller G, Halek C, Crisp AH (1993) Sexual abuse as a factor in anorexia nervosa. Evidence from two separate case series. J Psychosom Res 37: 873–879

Walsh BT, Devlin MJ (1992) The pharmacological treatment of eating disorders. Psychiatr Clin North Am 3: 149–160

Weissman J, Bothwell S (1976) Assessment of social adjustment by patient self report. Arch Gen Psychiatry 33: 1111–1115

Wonderlich S (1995) Personality and eating disorders. In: Brownell NKD, Fairburn CD (eds) Eating disorders and obesity: a comprehensive handbook. Guilford, New York, pp 171–176

Wonderlich SA, Mitchell JE (1992) Eating disorders and personality disorders. In: Yager J, Gwirtsman HE, Edelstein CK (eds) Special problems in managing eating disorders. American Psychiatric Press, Washington, DC, pp 51–86

CHAPTER 13

G. Kockott

Sexual Disorders

Translator: E. Taub

1 Introduction

Sexuality is a highly complex area of human behavior in which biological, psychological, and sociological factors all play a role simultaneously; the final result is a single, fused, unitary phenomenon that is, by nature, not exclusively biological, psychological, or sociological (Kinsey et al. 1948). The interaction of mental and biological factors is well known: according to the International Classification of Mental and Behavioural Disorders (ICD-10), sexual responsiveness is seen as a psychosomatic process. An example of this are varying social attitudes toward female sexuality, which range, even today, from acceptance in our culture to severe repression in some other cultures. It may be assumed that these differences in attitude cause differences in the sexual behavior of women in different societies.

In the late 1950s and early 1960s, Germany underwent a process of so-called sexual liberalization, which, fortunately, led to greater openness with regard to sexuality, but also to new norms, including norms of sexual performance, that created problems for many people. In recent years, the Western world has seen an increased prevalence of disorders involving diminished sexual pleasure, both in women and in men. Perhaps the cause is to be sought, at least in part, in the current sexual "oversaturation" by the mass media. An increase in disorders of sexual pleasure, specifically in women, has also been interpreted sociologically, in the light of the women's liberation movement, as a sign of women's resistance to continued male domination (Schmidt 1993).

The decision regarding which sexual behavior or behaviors to designate as deviant is very strongly influenced by sociocultural norms, which are highly divergent among different ethnic groups. Transvestitism is a paraphilia in Western culture, but transvestites are highly respected in other cultures, e.g. among some of the indigenous peoples of Southeast Asia. Among the Leptschas in India, sexual relationships are allowed between adult men and girls young enough to be legally protected from sexual contact in Western societies (i.e. below the so-called age of consent).

These sociocultural influences are taken into consideration in the international classification schemes. In the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV) of the American Psychiatric Association (APA), it is recommended that clinical evaluation take account of the ethnic, cultural, and religious affiliation of the patient, as well as his or her social background, as these factors may influence sexual desire and the individual's expectations and attitudes regarding forms of sexual behavior (APA 1994).

The international diagnostic schemes, ICD and DSM, classify sexual disorders into three major types: functional sexual disorders, disorders of sexual preference (ICD-10) or paraphilias (DSM-IV), and disorders of sexual identity. These three types of disorder are discussed in the same order in both schemes.

Homosexuality, once regarded as a disorder in both schemes, is no longer listed as such in the DSM, from its third edition onward, or in the ICD, starting with its tenth edition. Homosexuality is now regarded as an alternative lifestyle that cannot be considered a disorder. It is therefore not discussed in this chapter on sexual disorders.

2 Functional Sexual Disorders

2.1 Definition and Classification

A functional sexual disorder is defined in DSM-IV as a disturbance in the course of the sexual response cycle or pain arising in the context of sexual intercourse (APA 1994). The term "sexual dysfunction" is used to designate those disorders for which a predominantly or exclusively physical cause is assumed, while impairments that are to be regarded as being of mental origin may be designated as functional sexual disorders (Sigusch 1996). The latter are defined more precisely as those impairments of sexual behavior and of the sexual experience, as well as of modes of physiological response, that make it difficult or impossible for both partners to have a satisfying sexual interaction, even though the necessary organic substrate is present, and there is no fixation on unusual sexual objects.

The causes of functional sexual disorders, particularly in older men, may range along a continuum from the exclusively physical to the exclusively mental, with a large intervening region of mixed or uncertain etiology. The diagnosis is frequently no more specific than the overarching category of functional sexual disorder.

The transition between healthy and impaired sexuality is a gradual one, because pathological abnormalities may appear in a given individual with varying frequency. There is also some variation in the limits of individual tolerance: one person may perceive a certain type of sexual experience as normal and unimpaired, while another may not. Two criteria therefore have to be fulfilled before a disorder can be said to be present:

1. A specific functional impairment exists.
2. The impairment leads to significant distress or interpersonal difficulties.

From the practical viewpoint of the therapist, it has proved useful to describe functional sexual disorders more specifically according to their form and content. In terms of content, disorders are classified by the phase of the sexual interaction in which they appear (Table 1), with reference to the sexual response cycle described by Masters and Johnson (1967). Formal descriptive criteria include the frequency of the problem (e.g. invariably present or only occasionally so), the circumstances and conditions of its appearance, and its duration and severity.

A diagnostic scheme for sexual problems such as that given in Table 1 has the advantage of providing precise and therapeutically relevant syndrome descriptions. At the same time, it serves as a set of guidelines for the evaluation of the patient.

A few formal characteristics of functional sexual disorders deserve special emphasis, because they provide clues to the diagnosis:

- *Primary versus secondary:* A primary disorder is one that has existed since the onset of sexual activity; a secondary disorder begins after an initial, asymptomatic phase. Secondary disorders sometimes have easily discovered causes.
- *Situation-dependent versus situation-independent:* Situation-dependent disorders occur only during certain sexual activities, e.g. only during attempted coitus, but not during masturbation. Situation-dependence is a very strong indication of a disorder of mental origin. Situation-independent disorders are those that appear during sexual activity of any kind, in which case a physical cause is more likely.
- *Partner-dependent versus partner-independent:* Partner-dependent disorders, i.e. those that occur only with a specific partner, are a very strong indication of difficulties with this partner alone.

2.2 Specific Disorders

Patients with *disorders of libido* have varying degrees of loss of sexual desire and loss of sexual interest, which may reach the extent of a marked aversion, leading them to avoid any situation with sexual overtones.

Table 1. Functional sexual disorders in the different phases of the sexual interaction (with ICD-10 and DSM-IV codes for disorders of nonphysical origin)

Phases	Disorders in men	Disorders in women
1. Appetite	Persistent and significant reduction of sexual desire (F52.0; 302.71)	
	Sexual aversion, disgust, anxieties (F52.1; 302.79)	
2. Arousal	Erectile disorders: strength and duration of erection inadequate for satisfactory sexual intercourse (F52.2; 302.72)	Disorders of arousal: intensity and duration of arousal inadequate for satisfactory sexual intercourse (F52.2; 302.72)
		Vaginismus: penile intromission impossible, or possible only with pain, because of paroxysmal constriction of the vaginal introitus (F52.5; 306.51)
	Painful sexual intercourse (dyspareunia): pain in the genital area during or immediately after coitus (F52.6; 302,76)	
3. Orgasm	Premature ejaculation: ejaculation before insertion of the penis into the vagina, upon insertion, or immediately afterward (F52.4; 302.75)	Difficulties with orgasm: orgasm never, or only rarely, achieved (F52.3; 302.73)
	Absent ejaculation: lack of ejaculation and anorgasmia, despite full erection and intensive stimulation (F52.3; 302.74)	
	Ejaculation without orgasm: ejaculation without pleasure and feeling of orgasm (F52.11; 302.70)	
4. Relaxation	Post-orgasmic mood abnormalities: irritability, inner unrest, sleep disturbances, depression, crying spells, abnormal sensations in the genital area, etc.	

These disorders are to be distinguished from various situations in which sexual interest may be reduced without any disorder being present. For example, sexual desire is physiologically lower in women in late pregnancy and in the post-partum period. Deliberate sexual abstinence, which the individual may impose on himself or herself for a wide variety of reasons, also does not count as a sexual disorder.

Disorders of libido affect women more often than men. In general, sexual disinterest has become a much more common diagnosis in recent years, as mentioned above in the Introduction.

In *disorders of male sexual arousal*, the penile erection either does not develop at all or is not strong enough or does not last long enough for satisfactory coitus to be achieved. Very often, penile tumescence is entirely adequate during foreplay, but diminishes markedly at the moment of attempted penetration; this is a strong indication of a psychological cause. If the erection is weak or of variable strength during so-called petting, or if it never occurs, and if this problem is situation-independent, i.e. also occurs during masturbation and involuntary morning erections, then a physical cause should be suspected. The opposite of this form of erectile disorder is priapism, an intense and long-lasting erection that may become painful. Priapism is now seen more commonly than it used to be as a side effect of a newly developed method of treatment, so-called intracavernous injection therapy (see Sect. 2.4.1).

In *disorders of female sexual arousal*, the so-called lubrication-swelling response normally occurring during sexual stimulation develops inadequately or does not occur at all. There are considerable individual differences in this response, but its complete absence is usually accompanied by a feeling of disappointment. Many women develop the usual physical responses to sexual stimulation but feel no subjective arousal. This is most likely to be interpreted as a form of defense against sexual experience. Isolated disorders of arousal are rare in women. They are often combined with disorders of libido and orgasm.

Vaginismus is an involuntary, reflexive contraction of the muscles of the pelvic floor and the outer third of the vagina which occurs during attempted coitus. Intromission of the penis usually becomes impossible. The contraction may be of variable intensity. In severe cases, even the insertion of a tampon or a gynecological examination is impossible. Most women with this problem are able to achieve normal orgasms through other forms of stimulation.

Pain during intercourse (*dyspareunia* or *algopareunia*) occurs very rarely in men. Pain is mainly felt in the glans penis. If no physical cause, such as phimosis, can be found, then this symptom is usually due to hypersensitivity of the glans or to an intense fear of the glans being touched.

Women suffer more commonly from pain during intercourse. This problem is usually due to a physical (gynecological) cause. The particular characteristics of the painful symptom are the key to the gynecological differential diagnosis. Dyspareunia may be the result of another sexual disorder, e.g. the absence of lubrication in the female under conditions of diminished or absent libido, or a disorder of arousal, may make the sexual act painful.

The various *disorders of ejaculation* may be divided into problems of timing and disorders in which the process of ejaculation itself is disturbed. Problems of the former sort should be classified as disorders of the male orgasm, because the timing of the experience of orgasm is abnormal, but the process of ejaculation itself is not. Such problems include the following:

- *Premature ejaculation (ejaculatio praecox, orgasmus praecox):* There is a gradual transition from sexual intercourse that is experienced as rapid, but not abnormal, to the disorder known as premature ejaculation. Premature ejaculation (or premature orgasm) in the male is best defined as a disorder in which the patient has little or no control over the temporal course of the process of ejaculation; he can no longer determine the moment at which ejaculation occurs.
- *Highly delayed or absent orgasm (male anorgasmia).* This symptom is rare; it may be an expression of psychological inhibition, or it may be caused by various medications, particularly psychoactive drugs, and alcohol. A delayed or occasionally absent ejaculation is a not uncommon and quasi-normal occurrence in advanced age.

Retrograde ejaculation (ejaculation into the bladder) and *ejaculation without orgasm* are due to disturbances of the process of ejaculation. These problems are rare and usually of exclusively physical origin, and we will therefore not discuss them any further in this chapter.

At most half of all women always or almost always reach orgasm during sexual intercourse. The degree of distress caused by the absence of orgasm is highly variable. Often, the occasional or even frequent absence of orgasm is not experienced as a problem.

We draw a distinction between *total* and *coital anorgasmia.* Women with total anorgasmia have never in their lives attained orgasm during any kind of sexual activity. They may describe varying degrees of sexual arousal without orgasm, but a feeling of dissatisfaction remains. These patients seem mostly to suffer from various anxieties connected to the experience of sexuality.

Women with coital anorgasmia may experience orgasm more or less regularly during masturbation, for example, but not during coitus. Despite their basic ability to reach orgasm, these women seem to suffer greater distress because of their problem than do women with other sexual problems. Orgasm during sexual intercourse is seen as the single right way to obtain sexual satisfaction. Fear of a type of loss of control often seems to play a role, i.e. the fear of "losing face" during the experience of orgasm. This may explain why the difficulty arises only in the presence of a partner.

Post-orgasmic mood disturbances are characterized by irritability, inner unrest, and lack of release of tension after sexual contact and are usually an expression of inadequate sexual satisfaction. The cause may lie in rejection of the partner or in a not yet recognized or accepted homosexual orientation.

We have seen that there is a marked difference between the functional sexual disorders of men and women. While such disturbances may be isolated problems in men, they are much more rarely so in women: the manifestation of a disorder is often the consequence, or sometimes the cause, of other problems. Only vaginismus seems to occur in isolation.

2.3 Causes

Published data on the relative frequency of organic sexual disorders and functional sexual disorders of mental origin are highly variable, especially with regard to male functional disorders. This is both because the patient samples in published studies are often not comparable and because sexual disorders may be of multifactorial origin.

Physical and mental causes often interact with each other in a "bundle of causes," in which case an either–or diagnostic decision cannot be reached. Thus the present author (Kockott 1981) found considerable anxiety regarding inadequate sexual performance among male diabetics with erectile dysfunction. The mental component added considerably to the sexual problems caused by diabetes. A group of investigators in France (Buvat et al. 1983) treated 23 patients who had erectile dysfunction and abnormal findings on pelvic arteriography, either conservatively with psychotherapy alone, or with arteriodilators, or with both. After 6 months of treatment, mild to marked improvement was found in all groups, including the group treated with psychotherapy alone. This example underscores the suspicion, often expressed in recent years, that a demonstrated organic pathology may not necessarily represent the single, determinative etiology of a sexual disorder.

An accurate establishment of the patient's past history and present symptoms by thorough history-taking remains the most important step in the differential diagnostic assessment of male sexual disorders, even now that new diagnostic techniques have been devised, such as the intracavernous injection test with Doppler sonography (see Sect. 2.4.1). A predominantly or exclusively mental cause is likely if the symptoms arose suddenly and then persisted in a situation-dependent manner. If it can be ascertained that problems with the patient's partner or other negative life events occurred at the same time as the onset of the disorder, then a psychogenic origin is still more likely.

The following rules of thumb may be kept in mind: Functional sexual disorders of younger men are usually largely or exclusively of mental origin, while physical causes are more likely to play a contributory role with increasing age and, in advanced age, may be the predominant etiology. Functional sexual disorders in women, as far as we are now aware, are usually of mental origin regardless of the age of the patient, except for dyspareunia. Interviewing the partner may be of great help in establishing the diagnosis. Partners can often recognize contributing mental factors more easily than the patients themselves.

2.3.1 Physical Causes

Sexual disorders occurring in connection with a physical illness may be caused by the illness itself, its sequelae, and/or the measures taken to treat it. The possibility of mental factors making a further contribution should always be considered (Schover and Jensen 1988).

The most common physical causes are vascular disorders (particularly in the male), neurogenic disorders, consequences of operations in the genital area, toxic influences (drugs, alcohol) and (less commonly) side effects of medications, and (very rarely) endocrine disorders.

Sexual disorders with physical causes, unlike psychogenic disorders, have specific manifestations depending on their cause: vascular disorders in the male predominantly lead to erectile dysfunction, while the rarer endocrine disorders lead to diminished libido. Local problems in women most often lead to dyspareunia.

Approximately 50% of men with diabetes mellitus have erectile dysfunction. It is debated whether diabetic women are also more liable to sexual disorders due to diabetes and its consequences. In most cases, diabetes-related erectile dysfunction is always present and is not associated with any major loss of libido. It appears independently of the severity or duration of diabetes. Its cause has not been unequivocally

identified, although there are extensive data to suggest that it is caused by diabetes-induced vasculopathy. The latter cannot be influenced by treatment of the diabetes itself. Acute diabetic crises may lead to the rare, transient erectile disturbances that are associated with a loss of libido; these can be alleviated by control of the blood sugar level. Erectile dysfunction may also be a psychogenic phenomenon in diabetic patients, a form of "self-fulfilling prophecy" that comes true because of the patient's strong expectation that it will happen.

In older men, arterial disease, hypertension, and cigarette smoking are risk factors for the development of erectile dysfunction of vascular origin. After a myocardial infarction, bypass operation, or mild stroke, 40%–70% of patients reduce their sexual activity and report diminished sexual interest, delayed orgasms, and erectile difficulties. Sexual problems are found particularly frequently in patients who feel very ill and are anxious about their prognosis. It may therefore be assumed that these patients' sexual interest is diminished both because of their general state of physical health and because of the depressive-anxious mental state produced by the experience of illness.

A number of neurological illnesses cause sexual disorders, usually in the form of disorders of arousal. These include disseminated encephalomyelitis, polyneuropathies of various causes, traumatic injuries of the spinal cord, and temporal lobe damage (alteration of sexual interest). In such cases, the neurological manifestations are usually much more prominent clinically than the sexual dysfunction.

Operations in the genital region may impair sexuality either by their direct physical effects or through the effects they may have on body image and body perception.

Transvesical prostatectomy often leads to retrograde ejaculation into the urinary bladder, while the experience of orgasm usually remains unimpaired. After retropubic prostatectomy, erectile disturbances may appear to a degree of severity corresponding to the radicality of the required surgery. They are, however, rarer than is often feared: approximately half of the patients in one study still had partial erections after the operation, and one quarter were sometimes able to achieve coitus (Schorsch and Spengler 1981).

Hysterectomy has no direct effect on sexuality. It may indirectly benefit the patient's sexual experience if she had still had residual fears of unwanted pregnancy before the operation. Understandably, it may affect sexuality negatively if the patient had still wanted to have (further) children. We have occasionally observed a severe depressive reaction, with impairment of the sexual experience, after hysterectomy, because the patients thought of themselves as no longer being complete women.

Mastectomy has a highly negative effect in approximately 50% of women in their feminine body experience. The great advances of plastic surgery in developing techniques of breast conservation and reconstruction will bring considerable relief to women undergoing mastectomy.

Approximately 50% of chronically alcoholic men have sexual difficulties of various kinds. The frequency of such difficulties in alcoholic women has not yet been investigated to the same degree of precision. Sexual dysfunction in such cases may have several causes. Alcohol reduces the serum testosterone level directly, alcohol-induced liver damage indirectly affects sex hormone production, alcohol-induced vascular changes may cause erectile dysfunction, and the personality changes brought about by alcohol abuse may create partnership problems that impair sexuality. The psychosocial effects of chronic alcoholism lead to further tension with the partner and family and may eventually cause the partnership to fail.

Medications, particularly psychoactive medications of various kinds, sedatives, anticonvulsants, and antihypertensive agents, cause sexual problems in many patients. There is hardly any medication for which no single case of an undesired, negative effect on sexuality has been described. For the great majority of these medications, no dose-dependence of the related sexual dysfunction has been established. It follows that other factors must play at least a contributing role. These factors may lie in the illness under treatment or in other areas yet unknown. The origin of sexual problems always involves an interaction of the medical treatment of the illness with the illness itself: 17% of untreated male hypertensives and 25% of treated male hypertensives suffer from erectile disturbances (a significant difference). Both of these percentages are significantly higher than the corresponding figure for the normal population (Bulpitt et al. 1976). The undesired effects of psychoactive medications are discussed in Sect. 3.

2.3.2 Mental Causes

Sexual disorders of mental origin, in contrast to those of physical origin, have no specific relationship to their causes: all currently known psychological causes have been found in all forms of sexual disorder. Furthermore, such disorders cannot be explained as the result of a single problem alone. A bundle of causes is usually at work, even more often than in disorders of physical origin, involving an interplay of personality traits, life experiences, provoking factors, and the individual dynamics of the symptom.

The causes of functional sexual disorders are mainly to be found in the following four areas: (1) intrapsy-

chic anxieties, (2) partnership problems, (3) learning deficits and lack of sexual experience, and (4) the self-fulfilling prophecy of performance anxiety.

From the psychoanalytic and psychodynamic viewpoint, psychogenic sexual symptoms are the result of a conflict between anxiety-producing drive impulses and the defenses against them. The development of sexual symptoms serves to stabilize mental equilibrium. Sexual symptoms are accepted as a means of avoiding the anxiety-laden "deeper" conflict. Arentewicz and Schmidt (1993) list, as the major types of anxiety, those related to drives, relationships, conscience, and sexual identity. By the latter, they are referring to the patient's uncertainty as to his or her own sexual role.

Partnership problems and sexual difficulties are closely related to each other. Each of these two types of problem frequently causes the other and can thereby reinforce itself. A sexual problem inevitably affects the patient's partner, while partnership problems often affect the sexual sphere, as it is an important part of communication within the partnership; such problems may, indeed, be displaced exclusively onto the sexual area. It may then be very difficult to discover the connection between this sexual disorder and the "covert partnership problem."

Arentewicz and Schmidt (1993) have described a number of typical partner-dynamic processes of this type:

- *Delegation:* The "normal" partner has an unconscious personal interest in the functional disorder of the affected partner. He or she may need the disorder to neutralize his or her own problems or to feel superior.
- *Arrangement:* The functional sexual disorder may be an arrangement between the partners for their mutual benefit. The motive may be a common defense against sexual anxiety, e.g. when the male partner of a woman with vaginismus has an erectile disorder.
- *Turning against the partner:* The functional sexual disorder is used as a weapon against the partner in a conflict over domination.
- *Ambivalence management:* Sexuality may become an important means of regulating the correct balance in a conflict between intimacy and distance. Excessive distance from the partner may be remedied by the intimacy of sexual contact, while, in case of excessive intimacy, greater distance can be achieved through a sexual disorder.

These partner-dynamic processes are usually first recognized only when the sexual problem has become the focus of psychotherapeutic treatment.

The expression "learning deficits and lack of sexual experience" refers to a lack of information about the forms of sexuality that are currently generally practiced and experienced. Traditional, but outmoded, ideas of female and male sexuality may contrast with the partner's expectations. A person's own lack of sexual experience may also cause problems.

Practitioners of the different schools of psychotherapy all ascribe a major role to anxiety in both the development and the maintenance of functional sexual disorders. Masters and Johnson (1970) emphasized the importance of performance anxiety. It is assumed that anxieties of this type prevent sexual arousal and inhibit the autonomic nervous system to such an extent that physiological arousal becomes impossible. This assumption is not entirely undisputed, as the evidence that anxiety is the most important factor in the maintenance of functional sexual disorders is derived mainly from clinical experience rather than from empirical data.

On the basis of his investigations, Barlow (1986) was able to show that men with normal and disturbed sexuality differ from each other in several respects:

1. Anxiety inhibits sexual arousal in sexually disturbed men, but often increases it in men without sexual disorders.
2. The demand for sexual performance increases sexual arousal in normal men, but it is a distraction and an impediment for men with sexual disorders.
3. Patients with sexual disorders often have negative feelings in situations involving sexual contact, while individuals with an undisturbed sexual life mostly report positive emotions.
4. In comparison to sexually undisturbed men, sexually disturbed men tend to underestimate their degree of sexual arousal.

From these empirical findings, Barlow derived a working explanatory model for sexual disorders of mental origin: according to this model, a process of cognitive distraction, interacting with anxiety, is mainly responsible for the appearance of sexual disorders.

Langer and Hartmann (1992) developed a clinically applicable model involving four levels of causation. On the level of dispositional factors, they posited psychophysiological vulnerability and ego-structural deficits. On the next highest level, they posited intrapsychic and partnership-related etiological factors. On the level of pathophysiological mechanisms, they posited psychodynamic processes, such as anxiety defense, and psychophysiological causes, such as elevated sympathetic tone, as well as cognitive elements, such as excessive self-observation and control. On a fourth level, they posited chronifying factors including, most importantly, specific anxieties of the types discussed above.

These various clinical experiences and empirical findings can be systematized and incorporated in

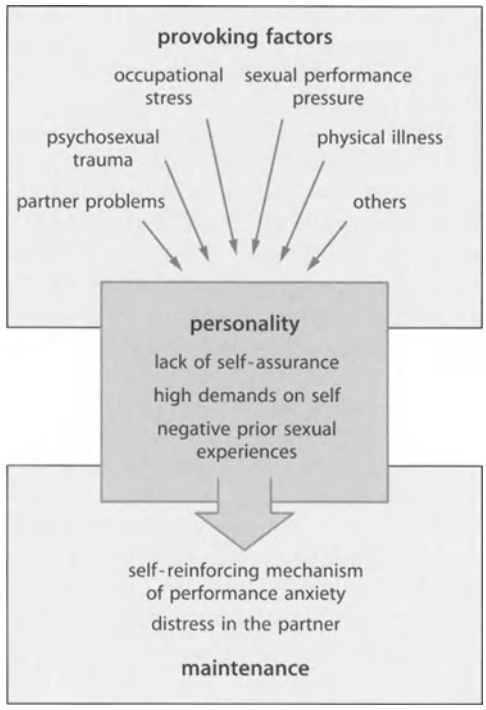


Fig. 1. Generation and maintenance of functional sexual disorders

a new theoretical viewpoint (Fig. 1). For ease of discussion, we will distinguish between the factors that provoke the disorder and those that maintain it. The connecting element is the patient's personality.

A single negative experience does not generate a sexual disorder in most cases, although an accumulation of unfavorable experiences in different areas of sexuality may do so. The provoking factors are thus not mutually exclusive, but rather additive or synergistic. The fact that one person may develop a sexual disorder after such negative experiences, while another may not, seems to be the result largely of personality variables and partly also of the partner's reaction and the patient's individual learning history. Research has yielded few firm findings with respect to the role of personality. A lack of self-assurance, low self-esteem, and a high emphasis on performance all have an unfavorable effect. As in Beck's model for the origin of depression (Beck et al. 1981), it may be assumed that life events result in a disturbance of sexuality primarily by reactivating earlier negative sexual experiences and the emotions connected to them.

Anxieties of expectation and performance always play a central role in the maintenance of functional sexual disorders, as does excessive self-observation (unless the sexual problem exclusively expresses a partnership problem). It is assumed, as a working hypothesis, that performance anxiety tends to increase by a so-called self-reinforcing mechanism.

The behavior-theoretic model is as follows: erotic situations involve the unfolding of a long chain of behaviors, which ends positively, under normal conditions, with a feeling of satisfaction and release of tension (Fig. 2).

When sexual behavior is disturbed (Fig. 3), the first, perhaps random occurrence of a functional sexual disturbance leads to performance anxiety. This anxiety prevents further normal function, which, in turn, heightens performance anxiety and leads to excessive self-observation, resulting in renewed failure.

Functional sexual disorders become firmly established through a mechanism of self-reinforcement, making them independent of the original causes that

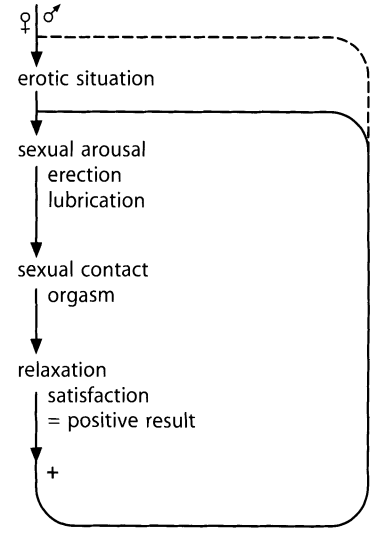


Fig. 2. Behavioral sequence of normal sexual behavior

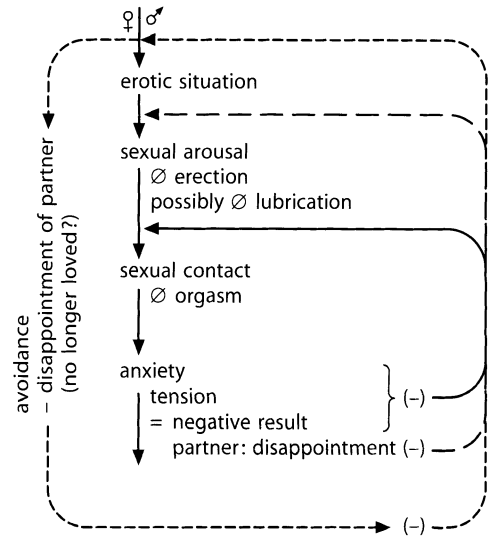


Fig. 3. Behavioral sequence of disturbed sexual behavior

provoked them; performance anxiety leads to their continued, functionally autonomous existence. Both the partner and the patient experience the disturbed sexual behavior as a disappointment. Because this disappointment further increases the patient's fear of failure, he or she begins to avoid sexuality as a means of fending off the undesirable situation. Avoidance brings the patient short-term relief, but the partner registers the patient's retreat from sexuality and relates it to him- or herself, perhaps concluding that he or she is no longer loved. Thus sexual problems open the way to partnership problems.

The foregoing is admittedly an oversimplified model of the complex mental causes of sexual disorders, but it has the great advantage of possessing explanatory power, while remaining easy for the patient to understand. This model can be used to explain the therapeutic process to affected couples and to make it clear why it is important to include the partner in the treatment.

2.4 Treatment

Not all sexual disorders require specific psychotherapy. A number of sexual problems come about merely because of lack of knowledge, absence of explanation, or a mistaken attitude toward sexuality that can be corrected by sexual counseling. Buddeberg (1996) provides an extensive survey of this topic that is relevant to clinical practice.

2.4.1 Physical and Pharmacological Treatment

When a functional sexual disorder has a predominantly physical origin, the therapeutic method of first choice is the treatment of the underlying physical illness. Frequently, however, there may be irreversible damage that will not respond to such treatment. In such cases, treatment must rely upon sexual counseling or on auxiliary pharmacological or physical measures, as will be discussed briefly in what follows.

The administration of sex hormones is appropriate only in the treatment of the rare endocrinological illnesses that are accompanied by documented androgen deficiency.

The new medication sildenafil (Viagra) has received a great deal of attention recently and will be discussed here in some detail. Sildenafil, a selective inhibitor of phosphodiesterase (PDE), has been on the market in the United States since March 1998 and in Europe since the autumn of 1998. Controlled studies have shown it to be remarkably effective in the treatment of erectile disorders of diverse etiologies.

We now know that, when the penis is in the resting state, the smooth muscle of its corpora cavernosa and arterioles is contracted, and the blood flow through the corpora cavernosa is minimal. The smooth muscle relaxes upon sexual stimulation, the flow of blood through the arterioles into the corpora cavernosa increases, and the draining veins are compressed between the corpora cavernosa and the tunica albuginea and thereby effectively sealed off. An erection results. Smooth muscle relaxation is mediated mainly by nitric oxide (NO), which is released by neurons and endothelial cells during sexual stimulation. NO, in turn, stimulates the production of cyclic guanosine monophosphate (cGMP), by activation of guanylate cyclase. In the human corpus cavernosum, cGMP is degraded mainly through the action of PDE-5. This is the point at which sildenafil acts. Sildenafil selectively inhibits PDE-5 and thereby slows the degradation of cGMP, leading to a prolongation and reinforcement of the relaxing effect of cGMP on smooth muscle and thus to longer-lasting and stronger erections. In accordance with this mechanism, sexual stimulation is a prerequisite for the effectiveness of sildenafil. Sexual stimulation initiates the development of an erection through the NO–cGMP mechanism, and the erection is then prolonged and strengthened by the slowing effect of sildenafil on the degradation of cGMP.

An improvement of erectile function is almost the only effect of sildenafil. The refractory period (i.e. the time after orgasm during which a new erection cannot occur) may be shortened; many men report more rapid ejaculations. The libido is in all likelihood not directly affected (elevated), but is certainly indirectly increased when a patient with erectile dysfunction has regained the ability to have an erection by using sildenafil.

Since its introduction, sildenafil has been tested in a number of controlled studies in different patient populations, including a total of more than 3000 men aged 19–87 years. Its effectiveness in patients with traumatic spinal cord injury and erectile dysfunction of mental origin is between 70%–80%, and in patients with severe diabetic disturbances and erectile dysfunction after prostatectomy, approximately 50%; the lower success rate in the latter groups is probably due to simultaneously existing nerve damage.

Sildenafil seems to be relatively safe to use. A few deaths have occurred in controlled studies, but the death rates were comparably low in the drug and placebo groups (0.45 versus 0.57 deaths per 100 patient-years) and were in fact lower than the expected death rate in this age-group in the general population (1.7 to 2.5 deaths per 100 proband-years), and the physicians conducting the studies found no connection between the deaths and the medication. Further deaths during treatment with sildenafil have been reported in the United States since its approval for use there. It is

not yet clear, however, whether these deaths were actually directly related to the use of sildenafil. A precise analysis of this matter is still urgently needed at present. The most recent studies show that sildenafil has no unfavorable hemodynamic effects in patients with coronary heart disease.

The major indications for sildenafil at present are erectile disturbances of physical origin, e.g. disturbances caused by diabetes and hypertension and in the aftermath of prostate surgery or traumatic spinal cord injury; severe erectile disturbance of uncertain origin in older men may also be an indication. Sildenafil is also effective against so-called psychogenic erectile dysfunction. It is theoretically conceivable that sildenafil helps men with erectile dysfunction of mental origin overcome their performance anxiety, which is often severe (see above). Sildenafil may thus be an important adjuvant treatment, i.e. a component of somatopsychotherapy. No data have been reported to date, however, regarding how treatment with sildenafil should be incorporated into psychotherapy or whether, when, and how the medication should be discontinued, etc. There are certainly other situations in which, despite an obvious mental cause of the disturbance, treatment with sildenafil is the only possible therapy, e.g. when there are partnership problems because of a psychotic illness in the spouse, but both partners strongly desire to preserve their marriage.

Sildenafil is, in all likelihood, not suitable for the treatment of disorders of libido (e.g. as a result of endocrinological or other illnesses, or the sedating effects of other medications), as our current knowledge indicates that this substance has no centrally stimulating effect.

Simultaneous treatment with sildenafil and nitrate-containing medications is absolutely contraindicated, as the mild hypotensive effect of sildenafil is considerably potentiated by nitrates, and dangerous hypotensive crises may result.

The following undesired side effects of sildenafil were described in controlled studies: flushing, headache, dyspepsia, occasional rhinitis or stuffed nose, and visual disturbances in the form of a bluish tint in the peripheral visual field. All of these side effects were said to be transient and of mild to moderate severity and remitted without specific treatment. The rate of termination of therapy because of side effects was almost equally low in treatment with sildenafil (2.5%) and placebo (2.3%). Alcohol mildly increases the hypotensive effect of sildenafil.

The effect of sildenafil is dose-dependent in the range between 25 mg and 100 mg, while increasing the dose to 200 mg or more does not increase the therapeutic effect. The effect sets in upon stimulation approximately 30 min to 4 h after the medication is taken. The manufacturer recommends that the drug be taken no more than once a day because of the possibility of increased side effects. As with all other modes of therapy for sexual disorders, it is strongly recommended that sildenafil be given in combination with sexual counseling, in which the patient's partner is included.

Nothing is known at present regarding possible problems associated with long-term use. Only a few long-term observations, involving treatment lasting up to 4 years, are now available. Physical dependence has not yet been observed. It is conceivable, but not yet demonstrated, that men might become psychologically dependent upon sildenafil.

The use of sildenafil alongside the psychotherapeutic treatment of psychogenic erectile disorders has not yet been tested systematically, and the views of psychotherapists on combination therapy of this type are highly divergent. Sildenafil might be useful for both diagnostic and therapeutic purposes if its use helps men with neurotic or partner-related conflicts to regain erections. Further difficulties and symptoms that are themselves due to the erectile dysfunction would then recede into the background. The neurotic or partner-related conflict would become clearer and thus easier to treat – or it would, perhaps, decline in importance once it became obvious that it did not, in fact, play the decisive role in causing the erectile dysfunction.

Our impression thus far is that the wave of prescriptions of sildenafil seen in the United States after its approval there is not being duplicated in Europe. Apparently, in the United States as in Europe, many men tried sildenafil for a time because of its status as a fashionable drug, and some will presumably continue to experiment with it in future. It remains to be seen whether sildenafil will have a lasting influence on human sexual behavior. We doubt it.

The World Health Organization considers sexual problems to be an impairment of an individual's health when he or she has a functional sexual disorder that causes suffering; such a disorder deserves to be considered an illness. This is the case for patients with the erectile disorders discussed above, for which sildenafil is a highly effective treatment. It follows that health insurance schemes should be required to assume the cost of such treatment. The indications for treatment should be established in a responsible manner by specialist physicians.

Controlled studies have shown that yohimbine hydrochloride is more effective than placebo in the treatment of disorders of libido in men and of erectile dysfunction. This selective, competitive α_2-adrenergic receptor blocker is thought to act centrally. Other so-called aphrodisiacs currently on the market are probably no more effective than placebo.

A number of antidepressants are said to possess a sexually stimulating activity in addition to their antide-

pressant effect. These claims are based on case reports. Such an effect was claimed for trazodone, for example, but could not be substantiated by a recently published controlled study. Reports of adverse side effects in the sexual area are far more common (see Sect. 3).

Vibrators are used in andrology as an aid in the obtaining of an ejaculate for diagnostic testing. A number of authors report the use of vibrators to treat total male anorgasmia, with occasional success. Their use as a therapeutic aid in women in the framework of so-called sex therapy is more common, but controversial among psychotherapists.

The injection of vasoactive substances into the corpus cavernosum of the penis induces an erection, which may ensue more or less rapidly and completely depending on the intactness of the corpus cavernosum and the blood vessels supplying it. This technique has therefore been in use for some time in urology as a diagnostic test for vascular abnormalities, in combination with Doppler sonography (intracavernous injection test). More recently, this technique has been put to use in the treatment of erectile dysfunction (intracavernous injection therapy).

The use of papaverine in the initial application of this technique often led to priapism, as did combination therapy with papaverine and phentolamine (though less commonly). Priapism must be alleviated within a few hours with a pharmacological antidote (vasocontrictor medication) or by surgical intervention to prevent permanent physical injury to the corpus cavernosum. Prostaglandin is currently used; this has resulted in hardly any cases of permanent erection, but complaints of local pain after the injection have become noticeably more frequent. Follow-up studies in recent years have documented high rates of discontinuation of treatment (as high as 60%), partly because of a regaining of spontaneous erections, but more commonly because of dissatisfaction with the mode of administration and disapproval by the partner. Thus informing and obtaining the consent of the partner are a necessary prerequisite for the initiation of this form of treatment.

This treatment is indicated in men with irreversible erectile dysfunction that cannot be treated successfully by other means. Its application is probably appropriate in combination with psychotherapy in men with erectile disturbances of mixed or uncertain origin in middle and advanced age. This mode of treatment will likely become much less common because of the success of treatment with sildenafil.

2.4.2 Surgical Methods

Penile prosthetic surgery has been practiced for more than 20 years. Three types of prostheses are commonly used. Silicone prostheses consist of two stiff rods that are inserted into the corpora cavernosa, thus producing a permanent erection. A second type of prosthesis, with flexible rods, leads to a type of permanent erection of intermediate stiffness. A third type consists of fluid-containing rods that can be hydraulically filled and emptied by means of a reservoir in the abdominal wall, so that an erection can be produced, and terminated, at will. The follow-up studies of patient satisfaction with this method of treatment that have been published to date were probably not carried out in representative groups of patients. Penile prostheses are currently considered a treatment of last resort for patients with a very strong desire for restoration of erectile function. The operation leads to permanent destruction of the corpus cavernosum. Prosthetic surgery has waned in popularity with the development of intracavernous injection therapy and, more recently, the introduction of sildenafil.

Surgical procedures involving the major pelvic arteries have often been reported to improve erectile function markedly as a positive side effect. The effectiveness of surgical procedures on vessels exclusively supplying the penis (usually the creation of an anastomosis) is disputed.

2.4.3 Treatment of Functional Sexual Disorders of Predominantly Mental Origin

The type of psychotherapy that should be used to treat a functional sexual disorder depends on the problem at hand. It may be necessary, in some cases, to use partner therapy, in which the sex-therapy aspect initially plays no role, or to use psychotherapy directed toward a personality disorder.

In a considerable number of patients, the mutual reinforcement of sexual and partnership problems is so strong that they can no longer be separated from each other. In such cases, sex-therapy and partner-therapy techniques are used in combination (Zimmer 1985).

The method of Masters and Johnson (1970) has proved effective for patients whose sexual problems are their most prominent symptoms. Treatment consists of a sequence of behavioral exercises that are to be carried out by the couple in between sessions with the therapist. The couple's experience with the exercises is discussed and evaluated in each session. In addition, attention is devoted to all other problems affecting the sexual area. Normal sexual behavior is gradually reconstituted in a series of intermediate steps of increasing difficulty, while other, symptom-specific treatments are concurrently applied. To treat erectile disorders, Masters and Johnson recommended the so-called "teasing method" (repeated, very brief insertion

of the penis into the vagina); to treat premature ejaculation, they recommended the "squeeze technique" (reduction of the urge to ejaculate by digital pressure on the penis); and, to treat vaginismus, they recommended the insertion of the patient's own finger or the insertion of Hegar dilators.

Treatment in the scheme of Masters and Johnson is always only one component of therapy of the couple as a whole – the "patient" is the couple. Sexual symptoms must be understood and treated on different levels. According to Schmidt (1996), the most important aspects to be considered include couple dynamics, the individual life histories of each partner, and the perspective of so-called symptom gain, as well as socio-psychological aspects (e.g. the striving for women's emancipation) that may be more or less important in individual cases.

Psychotherapy of functional sexual disorders with the inclusion of exercises, as described by Masters and Johnson, has been shown in several studies to be significantly more effective than psychotherapy without exercises.

Numerous further techniques have been proposed since 1970, and some have been scientifically studied, but the fundamental principles of therapy have remained the same:

- *Reduction of treating personnel and number of sessions:* It has been demonstrated in a number of studies that outpatient treatment in one to two weekly sessions with a single therapist is just as effective as quasi-inpatient treatment by a team of two therapists, as originally recommended by Masters and Johnson.
- *Group therapy:* The treatment of sexual problems in groups of couples with similar or different sexual disorders is just as effective as the treatment of couples separately.
- *Techniques for improving the sexual experience:* Exercises to promote physical self-awareness, practice with mechanical stimulation (e.g. the use of a vibrator when ejaculation is delayed), and the introduction of sexual fantasies have all been tested, with success.
- *Treatment of primary anorgasmia:* The technique of LoPiccolo and Lobitz (1972) has been found useful in the treatment of women with primary anorgasmia. In a program of nine sequential stages, women systematically learn to reduce their feelings of anxiety and inhibition with respect to their own bodies, to generate new, positive feelings and certain sexual capabilities, and to reach orgasm through self-stimulation.
- *Treatment of patients without partners:* Women without partners have been treated very successfully with the technique of LoPiccolo and Lobitz. No method for the treatment of men without partners has yet been adequately empirically tested. The most commonly used method at present is that suggested by Zilbergeld (1978), which mainly employs elements of self-confidence training to promote the ability to make contact with women, as well as a type of systematic desensitization that the patient carries out during masturbation to diminish sexual performance anxiety.

Therapists of different orientations have incorporated the techniques of Masters and Johnson into their own, individual theoretical frameworks and thereby developed variations of sex therapy that have been applied successfully in practice. Kaplan (1981) introduced psychodynamic and partner-dynamic aspects to sex therapy, as did the Hamburg group (Arentewicz and Schmidt 1993). Behavioral therapists use the individual techniques of Masters and Johnson selectively in individual treatment plans, according to the results of behavioral analysis. The model developed by Annon (1974, 1975) is an example of this kind of integration. In recent years, increasing attention has been paid to elements of cognitive behavioral therapy (Hawton et al. 1989).

Only a few studies have been performed on prognostic criteria for sex therapy. Thorough investigation of a large sample in England (154 couples; Hawton and Catalan 1986) revealed that a successful outcome of sex therapy is associated with the quality of the relationship between the partners, the strength of their motivation, and the extent of progress achieved in the first three therapy sessions. These criteria were basically confirmed in a later study of men with erectile dysfunction (Hawton et al. 1992).

The model of treatment devised by Masters and Johnson and its further development over the last 30 years have proved to be a major advance in psychotherapy. The present trend is toward combining sex therapy (also called "couples therapy") with the recently developed physical techniques (particularly intracavernous injection and oral medication), especially in older patients with sexual dysfunction of unclear or mixed etiology.

3 Sexuality and Psychosis

Psychotic illnesses alter patients' sexual behavior and experience. Even today, although the literature on this subject is rich in case reports and theories, data from controlled studies are scant.

Psychoses with sexual content are most frequently found among schizophrenics. The most common

thought disorders with sexual content include psychotic jealousy and delusions of being forced to engage in sexual behavior by others. These and similar delusions are probably present in 5%–20% of schizophrenics.

The most common perceptual disorders involve unpleasant sensations in the genital area, which are described by patients as burning, itching, hot, cold, or even as resembling an electric shock. The frequency of such sensations in the populations studied has been reported as being 30%–40%.

Sexual thought disorders and perceptual disorders are usually interpreted as a sign of regression accompanying a disturbance of sexual identity in psychotic patients, as a communication disorder, as an expression of sexual guilt feelings, or as a sign of anxiety connected to relationships.

Psychosis-related abnormalities of sexual behavior rarely take the form of very unusual sexual practices. They are primarily autoerotically oriented and are related, in their content, to sexual thought disorders and perceptual disorders.

Sexually deviant behavior occurs in psychotic patients as in nonpsychotic individuals. The reported data indicate, however, that sexual deviancy among psychotic patients is no more common than average. Psychiatric patients seem to be no more likely to be punished for sexual offenses than other individuals. There is also no reason to suppose that homosexual desires or behavior are any more common in psychiatric patients than in the normal population.

Schizophrenic patients seem to have less frequent sexual contacts than normal individuals and other psychiatric patients, both during acute episodes of illness and in the nonacute stage. The frequency of functional sexual disorders is estimated at 18%–60%, even after the resolution of an acute symptomatic episode.

The literature on the sexual activity of patients with affective psychoses and the frequency of functional impairments in this group is not extensive. The one finding that can be accepted as established is that the frequency of erections (including nocturnal erections) in depressive patients is lower during a depressive phase than at other times.

There is a surprising degree of variation in the frequency of sexual dysfunction and abnormalities of sexual activity reported in the literature. This is not just because of differences in study methods and in the definition of sexual problems, but also because of further factors affecting sexuality that these studies did not take always into account, including the patient's personality, premorbid sexuality, partnership situation, and psychopharmacologic treatment. Little is known about the effects of these additional factors. It does seem certain, however, that the sexual development of psychotic patients is delayed, as compared to that of normal individuals, even before the onset of acute manifestations of disease. This is particularly true of patients with schizophrenia (Nestoros et al. 1980).

Publications concerning the effect of psychopharmacological treatment on sexual experience have become more numerous only in recent years, but still largely concern single cases or small groups of patients. Controlled studies on this subject are rare, and, even today, the influence of psychopharmacological treatment on female sexuality is hardly discussed in the literature. Thus currently available data are inadequate to provide a representative picture of the sexual changes induced by psychoactive medications.

Three separate double-blind studies on small groups of patients demonstrated that neuroleptic agents diminish libido, delay ejaculation, and counteract erection (each effect was shown in a single study). If we consider the results of all of the uncontrolled case reports on this subject, the following rough picture emerges: phenothiazines may diminish libido, delay ejaculation (particularly thioridazine), and cause erectile dysfunction. In general, however, such changes are infrequently mentioned. There are a few reported cases of erectile dysfunction, delayed ejaculation, and diminished libido associated with butyrophenone treatment. Sexual disorders resulting from the use of other neuroleptics are only occasionally mentioned. The real frequency of sexual dysfunction is very likely higher than these published reports suggest. There are as yet no data concerning the so-called atypical neuroleptics.

The data on sexual side effects of the antidepressants are very heterogeneous. Well-controlled studies are rare but seem to show that functional sexual disorders are more common under treatment with antidepressants than under placebo treatment. According to Baier and Philipp (1994), the findings published to date suggest that different antidepressants may cause different, specific sexual problems depending on their particular pharmacological properties.

Sexual disorders seem to arise with increased frequency in association with the use of selective serotonin reuptake inhibitors; impairments of orgasm and ejaculation are the problems most commonly seen. In contrast, selective monoamine oxidase inhibitors and nefazodone seem to have much less of an effect on sexuality.

Our own study (Kockott and Pfeiffer 1996) of 158 psychiatric outpatients taking maintenance medication (100 schizophrenic and 58 depressed patients), and without evidence of acute or subacute disease manifestations, revealed that functional sexual disorders are common in psychotic patients. The patient needs to be asked about sexuality to discover such problems, because the patient will rarely report them spontane-

ously. Their frequency (47%) is much higher than in the normal population (approx. 10%) or than in a control group of outpatients treated for allergies (13.3%).

Disturbances of libido were the most commonly reported problem and were usually of multifactorial origin, according to expert clinical judgment. In the schizophrenic patients, an effect of the illness itself could be demonstrated. The sexually disturbed patients were somewhat "sicker" than others, although they were, on the whole, in very good mental condition. Psychosis seems to disturb sexuality considerably, as problems were present even in this significantly improved disease state. It was also shown that neuroleptic medication impairs sexuality, but its effect cannot be very large, because no significant association was found between the appearance of sexual disorders and the type, potency, or dose of neuroleptic medication given. Thus, in this study, the sexuality of schizophrenic patients was more impaired by the disease itself than by neuroleptic treatment.

4 Paraphilias

The terms "paraphilia," "sexual deviation," "perversion," "sexual variation," and "disorder of sexual preference" are used almost synonymously in the literature. Psychoanalytically oriented sexologists prefer the term "perversion" because of its prominence in the psychoanalytic theory of perversion; the term "disorder of sexual preference" was first introduced in ICD-10; and the term "paraphilia" has become usual in the English-speaking world through its use in the diagnostic manual (DSM-IV) of the American Psychiatric Association (APA).

On the level of behavior, a paraphilia or sexual deviation is best defined operationally as a sexual urge for an unusual sexual object or an unusual type of sexual stimulation.

4.1 Classification and Clinical Manifestations

Human sexuality is highly variable, both in the intensity of the experience and in its behavioral forms. As already mentioned in the Introduction, it is viewed differently in different cultures. Even within a single culture, views of sexuality may change rapidly. In the second (German) edition of this text, forms of sexual deviation were listed that are barely or not at all recognized today, such as saliromania, the urge to spray women with a fluid that the patient carries with him. In addition, sexual deviations are often given new names. Zoophilic behavior is called *Sodomie* (sodomy) in German-speaking countries today, while the same term in English also refers to homo- or heterosexual anal intercourse. The development of new technology leads to the appearance of new forms of deviation, e.g. erotophonia (the placement of obscene telephone calls to women who do not desire them), which, of course, appeared only after the invention of the telephone.

Decisions regarding what counts as a paraphilia or sexual deviation are thus highly dependent on both culture and time. The currently used international classification schemes for sexual deviations are valid only at the present time and are not fully applicable in non-Western cultures.

Gradual transitions exist from normal sexual fantasies and activities to sexual deviation. A paraphilia is present only when unusual sexual fantasies and/or activities are clearly predominant or exclusively present. According to Schorsch (1985), four levels of intensity can be distinguished:

1. A deviant impulse appears only once, or sporadically, in connection with a current conflict or a particular life crisis.
2. A deviant reaction becomes a repeated pattern of conflict resolution, without determining the patient's sexual orientation.
3. A stable, deviant orientation develops. Sexuality without deviant content cannot be experienced intensely or cannot be experienced at all (so-called fixation).
4. The stable, deviant orientation assumes a progressive course. Giese (1962) referred to this condition as *sexuelle Süchtigkeit* (sexual addiction) and described its characteristic manifestations as follows:
   a) An addiction to sensuality, in which specific stimuli take on a signal character
   b) Increasing frequency of sexually deviant behavior, with decreasing satisfaction
   c) A tendency toward anonymity and promiscuity
   d) The development of deviant fantasies and practices
   e) An "addiction-like experience" (*süchtiges Erleben*)

A fixed sexual deviation (levels 3 and 4) is characterized by the following:

- There is stereotypic, ritualized sexual behavior. The same form of sexual behavior is acted out again and again, and sexual satisfaction can only be obtained in this way.
- The partner becomes an object. The individual needs and sexual expression of the partner are of secondary importance and are accepted only when

they meet the expectations of the deviant. When they do not, they are perceived as disturbing. The partner is expected to play a specific role and is not allowed to be him- or herself. Shared experience remains entirely within the framework constructed by the deviant.
- The physical and mental satisfaction of orgasm is achievable only under the highly specific conditions characterizing the deviation, and not by ordinary coitus, which is thought of as a poor substitute for the deviant behavior.

Sexual deviations are not clearly delimited entities associated with particular abnormalities of personality, as was formerly assumed. They often appear, not in isolation, but in combination with other sexual deviations. For these reasons, there can be no generally applicable treatment program for exhibitionists or pedophiles, for example; the treatment must be individually tailored to the paraphilic patient.

The most common forms of sexual deviation are listed, with their usual international definitions, in Table 2.

4.2 Causes and Theories of Pathogenesis

The appearance of a sexual deviation in combination with a somatic or psychiatric illness is rare, and the nature of the relationship between the disorders is often unclear. An etiological relationship may be apparent in cases of mental debility or dementia. The exhibitionistic behavior of a feeble-minded patient may be a clumsy type of sexual advance that seems appropriate to the patient because he or she has never learned to do things differently. In demented patients, modes of sexual behavior may arise that are designated as deviant only because they do not conform to our current norms and conceptions of morality. In such cases, organic brain disease has impaired the patient's ability to control his or her own behavior in accordance with social convention.

An association of temporal lobe disorders with abnormal sexual behavior has been reported in a number of case studies.

There are a number of theories concerning the etiology and pathogenesis of sexual deviations, almost all of which are highly speculative.

The psychoanalytic theory of perversion has been repeatedly reformulated. In the recent literature (Morgenthaler 1974; Stoller 1979), the reparative aspect of sexual deviations is emphasized. Schorsch (1980) views the paraphilias as resulting from a severe problem of masculinity that is supposedly neutralized by the generation of the deviant symptom.

According to learning theory, on the other hand, sexual excitement in response to unusual stimuli comes about by way of classical and operant conditioning, both through sexual contacts and through fantasies

Table 2. Sexual deviations (paraphilias) and their classification in ICD-10 and DSM-IV

Sexual deviation	Characteristics	Classification	
		ICD-10	DSM-IV
	Recurrent, intense sexual urges, activities, and/or fantasies, present for at least 6 months, and involving:		
Exhibitionism	Exposure of the genitalia to unsuspecting strangers	F65.2	302.4
Fetishism	The use of inanimate objects (e.g., women's underwear)	F65.0	302.81
Pedophilia	Sexual activity with a prepubescent child or children, usually aged 13 or younger	F65.4	302.2
Transvestitism (transvestite fetishism)	The wearing of female clothing by a heterosexual man	F65.1	302.3
Voyeurism	The observation of unsuspecting individuals who are naked, in the act of undressing, or engaged in sexual activity	F65.3	302.82
Frotteurism	Touching and rubbing up against people who do not desire such behavior	F65.8	302.89
Sexual masochism	Real, not simulated acts of being humiliated, beaten, chained, or otherwise made to suffer	F65.5	302.83
Sexual sadism	Real, not simulated acts in which the mental or physical suffering (including humiliation) of the victim is sexually arousing for the affected person	F65.5	302.84
Zoophilia	Sexual activity with animals	F65.8	302.9
Erotophonia	Obscene telephone calls to unsuspecting or unwilling individuals	F65.8	302.9

during masturbation. In the view of Laws and Marshall (1990), human sexual behavior, including deviant behavior, develops according to the principles of "prepared learning," as laid down by Seligman (1970, 1971), i.e. under the influence of evolutionary, biological factors. This is thought to explain why not just any sexually neutral activity that happens to take place simultaneously with sexual arousal can become transformed into an independent source of sexual arousal.

In the theory of Money (1986), human sexuality develops through an interplay of biological and mental factors that influence the developmental process at particular times. This triad – biological factors, mental factors, and their taking effect during critical periods – is said to be decisive for the creation of sexual identity, sexual partner orientation, and the so-called "love-map," i.e. the sexual-erotic conceptual world of the individual. Money considers sexual deviancy to be a mechanism through which a sexual desire perceived as sinful is transformed into a "permitted" sexual pleasure.

4.3 Treatment

Paraphilias are not necessarily to be considered diseases that always require treatment. Deviants indeed suffer from their being different from others, but often less from the deviation itself than from the contempt and rejection that they perceive, or are actually subjected to, when their inclination becomes known. Thus counseling is often needed. Counseling may be the only therapeutic intervention required; the goal in such cases is to find an arrangement with which the affected individual and significant others can live.

Psychotherapy is the most important technique used to treat sexual deviations. It may be appropriate to combine psychotherapy with pharmacotherapy. Psychotherapy can often begin only after a reduction of libido by medication. Cyproterone acetate (CPA), a competitive testosterone inhibitor, is used for this purpose in Europe and Canada. Medroxyprogesterone acetate (MPA) is used with comparable effects in the United States, where CPA lacks government approval.

The clinical effect of CPA has been evaluated in many publications, but in only a single double-blind study. The findings may be summarized as follows: a daily dose of 100–200 mg leads to a diminution or elimination of sexual appetite after a period of 1–3 weeks. In addition, erectile function diminishes. Three weeks after the initiation of therapy, a reduction of the quantity of the ejaculate and a delay of ejaculation are usually reported. When high doses are given, a complete inability to obtain an erection or to ejaculate may result. Sexual fantasies and dreams become less frequent only in the ensuing months. CPA significantly lowers the testosterone level. The diminution of libido by CPA depends on the age of the patient and on the dose. Alcohol partially counteracts the effect. The response to CPA is highly variable among patients, but highly consistent in each individual patient.

Approximately one third of patients complain of side effects, such as fatigue, adynamia, and general reduction of performance, that appear in the second and third weeks of treatment and later resolve. A change of orientation of the sexual drive is never observed and is not to be expected by the therapist. When the medication is discontinued, the libido, erectile function, and ability to ejaculate become normal again within a few weeks. CPA is available in oral form or as an injectable solution, while MPA is available only in oral form.

The psychotherapy of patients with paraphilias is accompanied by a number of problems. Patients are often in a difficult social situation, and their motivation to undergo therapy may be highly ambivalent. Therapists assuming the treatment of a group of patients looked down upon by society may, at first, judge them not much differently. They must work to achieve a therapeutic posture. The therapist is also under pressure to succeed, even more so if the sexual deviation involves behavior injurious to others. The therapist may therefore become overcautious and institute modes of treatment that are either too rigid or too aggressive and are thus counterproductive to the goals of therapy. Because of these special difficulties, close psychotherapeutic supervision is necessary.

Research on this subject is in a highly fragmentary state. Controlled studies are rare; only a few are available in which the results of therapy are evaluated. The findings of such studies and other reports of clinical experience imply that the psychotherapy of sexual deviations is best performed according to the following guidelines:

- Psychotherapy must have a clear structure, and the boundaries of therapeutic behavior must be clearly set.
- Psychotherapy is often of a supportive nature at first; social problems must be addressed.
- Sexually deviant patients are highly variable in their suitability for therapy. The goals of treatment, forms of approach to the patient, and degree of depth must therefore be individualized. Thus the psychotherapist must not confine his or her approach to that of a single theoretical school. Integrative treatment programs that take psychodynamic and behavior-therapy concepts into account, and do not neglect pharmacological possibilities, seem to be most successful.

- A "cure" is only rarely achieved, but the patient can, at least, learn to keep his or her deviation under control and to develop a more suitable form of interpersonal communication, thereby improving quality of life.

The behavior-therapeutic approach, which has been applied by psychotherapists to a highly variable extent in the treatment of paraphilia, has four major emphases, which are independent of the type of sexual deviation (Kockott 1996). The different building blocks of behavioral therapy are used as needed to construct an individual treatment program for each patient:

1. *Covert sensitization:* The patient is instructed to think of the deviant behavior as vividly as possible. Once the mental image is clear, the patient is then told to change it suddenly and think of a particularly unpleasant experience, such as that of being surprised by a member of the family during the deviant behavior.
2. *Self-control methods:* The patient learns to distract himself or herself from deviant behaviors and to develop well-rehearsed alternative behaviors that are incompatible with the deviant behavior. For example, an exhibitionist might approach the woman to whom he desired to expose himself and, instead, ask her the time.
3. *Masturbatory saturation:* The patient masturbates to orgasm with normal sexual fantasies and then for a further time with deviant fantasies, until this activity becomes boring or unpleasant. This form of treatment is based on the learning-theory notion that normal sexual fantasies can be positively reinforced by the experience of orgasm, while deviant fantasies are extinguished, because they are no longer reinforced.
4. *Stimulus control methods:* The patient learns to recognize the situations in which the deviant behavior often occurs (e.g. unstructured free time, proximity to playgrounds) and changes his or her behavior to stay out of such situations as much as possible (by structuring free time, avoiding playgrounds, etc.).

All of the different aspects of the method of Masters and Johnson (1970) for the establishment of normal sexual behavior (see Sect. 2.4.3) can be used, depending on the individual circumstances, to develop or improve normal, nondeviant sexual behavior in sexually deviant patients. So-called "orgasmic reconditioning" is a further method. The patient is instructed to masturbate with his or her deviant fantasies at first and to switch to normal fantasies shortly before orgasm, making this switch earlier and earlier each time. The reinforcing effect of orgasm is then expected to reinforce, i.e. increase, normal masturbatory fantasies and extinguish, i.e. decrease, deviant fantasies. Although this method is often used, its effectiveness is disputed.

The usual methods of improving social skills, communication, and problem-solving behavior are applied (Margraf 1996), and the new modes of behavior that must be acquired are practiced by role-playing. For paraphilic men, stress is placed on improving behavior toward women in the social context and in the partnership, but also on improving the patient's view of himself.

Attention is increasingly being paid to relapse prevention. Several American treatment centers have taken up the relapse prevention model originally developed for alcoholism and adapted it for use in paraphilic patients. The patient is taught a number of coping strategies with which he can prevent a threatened relapse. He might, for example, carry an "emergency card" listing instructions for how to behave in an emergency and the telephone number of the treatment center.

4.4 Sexual Delinquency

Sexual delinquency is an illegal violation of the sexual autonomy of an individual, regardless of whether it is associated with a sexually deviant practice. Most sexual offenders are not paraphilic, but rather highly aggressive within the framework of normal sexuality (sexual assault, rape). Gradual transitions from nonaggressive deviancy to aggressive sexual delinquency do exist, but sadistically motivated sexual delinquency (involving a paraphilia) is rare.

Sexual delinquency creates two groups of individuals in need of treatment, the offenders and their victims. The experience of being assaulted or raped is always a trauma with major mental consequences, which have been termed post-traumatic stress reactions. Treatment is provided as in other forms of severe mental stress. Two behavior-therapy techniques are used in the treatment of women who have been raped: "prolonged exposure" and "stress-inoculation training" (SIT). Prolonged exposure is a less intense form of "flooding," while SIT is a package of treatments including muscle relaxation with controlled breathing, *Gedankenstop* ("thought stop") training, a "guided self-dialogue" in which incorrect perceptions are dealt with ("I am guilty of becoming a victim"), and "stress inoculation" in which coping strategies are developed and practiced by role-playing. A controlled study (Foa et al. 1991) showed that SIT had better results than prolonged exposure. For ethical reasons,

treatment with prolonged exposure is very rarely advisable in this patient group.

The psychotherapy of sexual offenders is particularly liable to the difficulties mentioned above in connection with the psychotherapy of paraphilic patients. The question of responsibility for possible further offenses also weighs heavily upon the therapist, who has the duty to maintain patient confidentiality, but must also find an acceptable way to prevent a relapse when the acute danger of one arises. Therapeutic "emergency measures" for use in crisis situations must be worked out with the patient in advance. If the patient is incapable of taking such measures to protect others, he must agree to protective measures on the part of the responsible government authorities.

The same general considerations apply to the psychotherapy of sexual delinquency as to that of paraphilia. Behavior-therapy methods to reduce or control sexually deviant behavior are generally not very successful in this patient group. A change of the patient's cognitive attitudes, however, is very important. Sexually aggressive offenders have extremely distorted ideas of their victims' experiences ("Women want to be raped," "Children enjoy having sex with me") and tend to use such ideas as a means of denying, or at least reducing, their responsibility for their own actions. To eliminate these cognitive distortions, the offender should be made aware of them and brought to understand the actual nature of the victim's experience, and the distortions should be changed by means of practiced role-playing. Cognitive restructuring sounds like a highly plausible therapeutic concept, but it is very difficult in practice. The effectiveness of these recommended modes of treatment has not yet been demonstrated.

Reports of primarily behavior-therapeutic treatment for sex offenders that have been published to date are mainly from Canada, the United States, and Germany. In the United States and Canada, inpatient treatment has been provided in prisons and forensic clinics for years. In Germany, there are only a few small units, in prisons and disciplinary institutions, in which sex offenders are treated. Intensive efforts are now underway to develop treatment programs that begin during the offender's incarceration and continue as outpatient therapy, because the danger of a relapse is especially great shortly after the offender is released.

There have been only a few properly controlled studies on the treatment of sexual delinquency (e.g. Schorsch et al. 1985; Marshall and Barbaree 1990). These studies do reveal, however, that sexual delinquency is treatable and that treatment reduces the rate of relapse. The success of treatment is mainly a function of the severity of the patient's psychopathology.

5 Gender Identity Disorders

The term "gender identity" denotes a person's feeling of being a man or a woman. Individuals with gender identity disorders suffer from a mismatch of their gender identity and their biological sex. The most important clinical variety of gender identity disorder is transsexualism, the persistent and complete mental identification with the opposite sex. Transsexualism will be discussed in detail below. There are also gender identity disorders occurring in childhood or adolescence or in the framework of other mental disorders, e.g. in schizophrenia.

Gender identity disorders have been described under a variety of terms, and connected with various different disease entities, since the early nineteenth century. The term "transsexualism" was first used by Hirschfeld in 1923, but entered common parlance only in the 1960s. A chapter on gender identity disorders, among which transsexualism is listed, first appeared in ICD-10. The term "transsexualism" was present in DSM-III-R, but was replaced in DSM-IV by the synonymous term "persistent gender identity disorder."

The major manifestations of transsexualism may be summarized as follows:

- *Constant and complete mental identification with the opposite sex:* transsexuals often describe this as a feeling of living in the wrong body.
- *Constant and intense discomfort with the specific features of one's own biological sex* (breasts, vagina, and menstruation in women; penis, testes, and facial hair in men): these features are totally rejected, and attempts are made to hide them by means of corsets or loose clothing or even to remove them physically; self-castration has been described.

Further symptoms resulting from these major manifestations include the following:

- The urgent *desire for sex change.* Transsexuals seek full legal and social recognition as members of the sex to which they feel they belong and actively seek sex-altering measures (hormonal treatment and surgery) as well as name changes and changes of legal status.
- The wearing of clothing associated with the opposite sex (*cross-dressing*) begins in childhood in some cases, but usually during puberty, when the mismatch between biological sex and mental gender identity becomes very obvious to the patient. Transsexuals strive, often successfully, to adopt the facial expressions and body language of the opposite sex and to work in occupations typically associated with the opposite sex.

The sexual partner orientation of male-to-female transsexuals is usually toward heterosexual men; in accordance with their perception of themselves as women, they do not view this orientation as homosexual. A few male-to-female transsexuals are sexually oriented toward female partners and perceive themselves as "lesbian." Female-to-male transsexuals seem to be almost invariably oriented toward female partners. There are also asexual individuals of both sexes.

The reported figures on the frequency of transsexualism are variable. The prevalence of the male-to-female variety is probably approximately 1:30,000, and that of the female-to-male variety approximately 1:100,000. The difference may have a sociocultural explanation. The female gender role is more flexible, at least in Western society; thus female-to-male transsexuals may find compromise solutions more easily than biologically male patients.

A responsibly performed differential diagnosis is of the utmost importance because, in severe cases of transsexualism, irreversible therapeutic measures are considered. The differential diagnosis must include the following disorders:

- *Homosexuality*, not accepted by the patient, accompanied by transvestitism: the homosexual, unlike the transsexual, basically accepts his or her biological body.
- *Fetishistic transvestitism:* A transvestite is sexually aroused by cross-dressing and does not identify permanently with the opposite sex.
- *Psychosis:* Transsexual symptoms in the framework of acute psychotic manifestations. These are only transient, and the other manifestations of psychosis are accessible to exploration by the therapist.
- *Early personality disorders*, particularly borderline pathologies, in which transsexual symptoms may appear transiently.
- *Conflicts of adolescence* with manifestations resembling transsexualism; these conflicts, too, are transient.

The essential criterion for the diagnosis of transsexualism is the complete and, above all, persistent mental identification with the opposite sex.

There is as yet no convincing explanation of the etiology of the transsexual syndrome. In children born of pregnancies supported with hormones of the opposite sex, behavioral alterations are found, including increased tomboyishness (in girls) or diminished aggressiveness (in boys), but never a disorder of sexual identity. Several authors have described electroencephalography (EEG) abnormalities in transsexuals, particularly in the frontotemporal region, but other authors have been unable to replicate these findings. Sociogenic and psychogenic etiologies are, to date, equally untenable. Sigusch (1996), who favors a psychogenic etiology, criticizes the "nosomorphic" view of transsexualism and compares it to homosexuality, which is no longer regarded as a disease state. He argues in favor of an analogous "de-pathologization" of transsexualism.

Practically all attempts to treat the transsexual syndrome with psychotherapy have failed. In the few reported cases of successful eradication of transsexualism by psychotherapy, the diagnosis of transsexualism was usually questionable.

In most cases, varying degrees of adaptation to the desired sex seem to be the most promising way to provide rehabilitation. The therapeutic procedure that has now gained international acceptance involves a series of therapeutic measures, provided over a period of years, that may culminate in a sex-change operation (Petersen and Dickey 1995). Therapeutic guidelines (so-called standards of care) for this procedure have been developed in several countries, including the United States and Germany (Becker et al. 1997).

The first step involves caring for and observing the patient for at least 1 year. The therapist tries to determine whether the transsexual desire remains stable and whether the patient can mentally accommodate a sex change. It has been found important to tell transsexuals repeatedly that the decisive therapeutic step is not the operation itself, but rather the preparation for it, which takes several years, during which the patient's ability to actually exchange gender roles is continually tested. The therapist also needs this period of time to check and consolidate the diagnosis.

The second step is the so-called everyday life test. The patient lives in the desired gender role fully, i.e. all day long, for at least 1 year. The patient must contend with the reactions provoked by the change in the social environment and determine whether the gender change has been successful.

If the patient has acquired a large measure of security in the new gender role, this may be further supported in a third step, the administration of hormones of the opposite sex. This treatment should be continued for at least 6 months before a sex-change operation. It enables the patient to experience the postoperative state before the (irreversible) operation is actually performed. Male-to-female transsexuals experience a decrease of libido, breast growth, and redistribution of subcutaneous fat, with an increased amount on the hips; while female-to-male transsexuals experience deepening of the voice, coarsening of the skin and facial features, and, in some cases, growth of facial hair.

If the first three steps have been successful, the fourth step is the sex-change operation itself. For male-to-female transsexuals, penectomy, castration, and the construction of a neovagina are usually performed. In individual cases, the construction of breasts by plastic

surgery may be necessary. For female-to-male transsexuals, bilateral mastectomy, oophorectomy, and (in some cases) hysterectomy are performed. The results of phalloplasty have been unsatisfactory to date; at present, so-called clitoral mobilization is usually performed instead.

After the operation, further treatment of the transsexual patient is still required, as both medical and mental difficulties may arise.

Pfäfflin and Junge (1992) summarized the results of a large number of follow-up studies of transsexuals who underwent sex-change operations. They concluded that the form of treatment consisting of the entire process of sex transformation is, indeed, effective. From the patient's viewpoint, treatment results in "alleviation of suffering and increased subjective satisfaction" in many different areas of life (partnership, sexuality, career, etc.).

From their review of the literature, Pfäfflin and Junge (1992) conclude that the following are positive prognostic factors for the further life course of transsexuals:

- Continued contact with a treatment facility
- Performance of the everyday life test
- Performance of hormonal treatment
- Counseling, or psychiatric or supportive psychotherapeutic treatment
- The sex-change operation itself and its quality
- Legal recognition of the sex change through a change of name and personal status

In Germany, the so-called Transsexual Law (*Transsexuellen-Gesetz*, TSG) was officially proclaimed in the Federal Legal Register[1] on 10 September 1980. The TSG provides that transsexuals may obtain a change of name and personal status through the responsible court authority. Before the name may be changed, it must be established, in two independently performed medicolegal evaluations, that the following requirements are met:

- The petitioner considers himself or herself as belonging not to the sex in which he or she was born, but to the opposite sex.
- He or she has felt compelled to live in accordance with this belief for at least 3 years.
- According to the current understanding of medical science, the petitioner's feeling of belonging to the new gender role will, to a high degree of probability, not change in future.

For a change of personal status, it is further required that the patient is not married and "permanently incapable of reproduction, and has undergone a surgical procedure to change his or her external sexual characteristics, through which a clear resemblance to the appearance of the opposite sex has been achieved." Because of the far-reaching nature of the legal decisions that are made on the basis of medicolegal evaluations, the evaluations themselves must be based on a very thorough examination of the petitioner. Guidelines for these evaluations have been developed recently (Becker et al. 1997).

Germany was the second country (after Sweden) to introduce specific legislation dealing with transsexuals. Since then, similar legal provisions have been made in other European countries, namely Italy, the Netherlands, and Turkey. In other countries, no special legislation was required, because the already existing laws were sufficient. The deciding bodies are either registry offices and comparable institutions (as in Austria, Denmark, Norway, some U.S. states, and Asia) or the courts (as in Belgium, Spain, France, and Poland).

[1] *Bundesgesetzblatt* 1980, Part 1, pp. 1654–1658.

6 References

Annon JS (1974) The behavioral treatment of sexual problems, vol 1. Enabling System, Honolulu

Annon JS (1975) The behavioral treatment of sexual problems, vol 2. Enabling System, Honolulu

APA (1994) Diagnostic and statistical manual of mental disorders, 4th edn (DSM-IV). American Psychiatric Association, Washington, DC

**Arentewicz G, Schmidt G (eds) (1993) Sexuell gestörte Beziehungen, 3rd edn. Enke, Stuttgart

Baier D, Philipp M (1994) Die Beeinflussung sexueller Funktionen durch Antidepressiva. Fortschr Neurol Psychiatr 62: 14–21

Barlow DH (1986) Causes of sexual dysfunction: the role of anxiety and cognitive interference. J Consult Clin Psychol 54(2): 140–148

Beck AT, Rush AJ, Shaw BF, Emery G (1981) Kognitive Therapie der Depression. Urban and Schwarzenberg, Munich

Becker S, Bosinski HAG, Clement U et al (1997) Standards der Behandlung und Begutachtung von Transsexuellen. Z Sexualforsch 10: 147–156, Sexuologie 2: 130–138

**Buddeberg C (1996) Sexualberatung, 3rd edn. Enke, Stuttgart

Bulpitt CJ, Dollery CT, Came S (1976) Change in symptoms of hypertensive patients after referral to hospital clinic. Br Heart J 38: 121–128

Buvat J, Dehaene L, Lemaire A, Buvat-Herbaut T (1983) Arteriell bedingte erektile Impotenz. Sexualmedizin 12: 248–251

Foa EB, Rothbaum BO, Riggs DS, Murdock TB (1991) Treatment of posttraumatic stress disorder in rape victims: a comparison between cognitive-behavioral procedures and counseling. J Consult Clin Psychol 59: 715–723

Giese H (1962) Psychopathologie der Sexualität. Enke, Stuttgart

Hawton K, Catalan J (1986) Prognostic factors in sex therapy. Behav Res Ther 24: 377–385

*Hawton K, Salkovskis PM, Kirk J, Clark CM (eds) (1989) Cognitive behavior therapy for psychiatric problems. A practical guide. Oxford Medical, Oxford

Hawton K, Catalan J, Fagg J (1992) Sex therapy for erectile dysfunction. Characteristics of couples, treatment outcome, and prognostic factors. Arch Sex Behav 21: 161–176

Hirschfeld M (1923) Die intersexuelle Konstitution. Jahrbuch Sex Zwischenstufen 23: 3–27

Kaplan HS (1981) The new sex therapy. Brunner and Mazel, New York

Kinsey AQW, Pomeroy B, Martin CE (1948) Sexual behavior in the human male. Saunders, Philadelphia

Kockott G (1981) Die sexuellen Funktionsstörungen des Mannes. Enke, Stuttgart

**Kockott G (1996) Sexuelle Störungen. In: Margraf J (ed) Lehrbuch der Verhaltenstherapie. Springer, Berlin Heidelberg New York, pp 295–318

Kockott G, Pfeiffer W (1996) Sexual disorders in nonacute psychiatric outpatients. Compr Psychiatry 37: 56–61

Langer D, Hartmann U (1992) Psychosomatik der Impotenz. Enke, Stuttgart

Laws DR, Marshall WL (1990) A conditioning theory of the etiology and maintenance of deviant sexual preference and behavior. In: Marshall WL, Laws DR, Barbaree HE (eds) Handbook of sexual assault: issues, theories and treatment of the offender. Plenum, New York, pp 209–229

LoPiccolo J, Lobitz WC (1972) The role of masturbation in the treatment of orgasmic dysfunction. Arch Sex Behav 2: 163–171

Margraf J (ed) (1996) Lehrbuch der Verhaltenstherapie. Springer, Berlin Heidelberg New York

Marshall WL, Barbaree HE (1990) Outcome of comprehensive cognitive-behavioral treatment programs. In: Marshall WL, Laws DR, Barbaree HE (eds) Handbook of sexual assault: issues, theories and treatment of the offender. New York, Plenum, pp 363–385

*Masters WM, Johnson VE (1967) Die sexuelle Reaktion. Akademische Verlagsgesellschaft, Frankfurt am Main

*Masters WM, Johnson VE (1970) Human sexual inadequacy. Little Brown, Boston

Money J (1986) Lovemaps. Irvington, New York

Morgenthaler R (1974) Die Stellung der Perversionen in Metapsychologie und Technik. Psyche 28: 1077–1098

Nestoros JA, Lehmann HE, Ban TA (1980) Neuroleptic drugs and sexual function in schizophrenia. Mod Probl Pharmacopsych 15: 111–130

Petersen ME, Dickey R (1995) Surgical sex reassignment: a comparative survey of international centers. Arch Sex Behav 24: 135–156

Pfäfflin F, Junge A (1992) Geschlechtstumwandlung. Schattauer, Stuttgart

Schmidt G (1993) Tendenzen und Entwicklungen. In: Arentewicz G, Schmidt G (eds) Sexuell gestörte Beziehungen, 3rd edn. Enke, Stuttgart, pp 1–12

Schmidt G (1996) Paartherapie bei sexuellen Funktionsstörungen. In: Sigusch V (ed) Sexuelle Störungen und ihre Behandlung. Thieme, Stuttgart, pp 180–199

Schorsch E (1980) Sexuelle Perversionen. Ideologie, Klinik, Kritik. In: Sigusch V (ed) Therapie sexueller Störungen. Thieme, Stuttgart, pp 119–158

Schorsch E (1985) Sexuelle Perversionen. Med Mensch Ges 10: 253–260

Schorsch E, Spengler A (1981) Zur psychischen und sexuellen Situation von Patienten nach Genitaloperationen. (Forschungsbericht des Sonderforschungsbereichs 1215 der Deutschen Forschungsgemeinschaft, Teilprojekt B6, Hamburg)

Schorsch E, Galedary G, Haag A, Hauch M, Lohse H (1985) Perversion als Straftat. Springer, Berlin Heidelberg New York

Schover LR, Jensen SB (1988) Sexuality and chronical illness. Guilford, New York

Seligman MEP (1970) On the generality of the laws of learning. Psychol Rev 77: 406–418

Seligman MEP (1971) Phobias and preparedness. Behav Ther 2: 307–320

**Sigusch V (ed) (1996) Sexuelle Störungen und ihre Behandlung. Thieme, Stuttgart

Stoller RJ (1979) Perversion, die erotische Form von Haß. Rowohlt, Hamburg

Zilbergeld B (1978) A guide to sexual fulfillment. Little Brown, Boston

Zimmer D (1985) Sexualität und Partnerschaft. Urban and Schwarzenberg, Munich

CHAPTER

14

W.B. Mendelson

Sleep Disorders

This work was partially supported by NIH grants 1K07 HL03640 and 1R01DA10682-01A2.

1 Introduction

There are a number of ways to classify sleep disorders, and indeed it may be some small comfort to the psychiatric community that our brethren in sleep research struggle with nosologic issues similar to those encountered when describing psychopathology. The International Classification of Sleep Disorders includes "dyssomnias" (intrinsic and extrinsic sleep disorders as well as those resulting from circadian rhythm disturbances), parasomnias (abnormal behaviors during sleep), and sleep disorders associated with mental, neurologic, or medical illness (ASDA 1997). DSM-IV describes primary sleep disorders, sleep disorders related to another mental disorder, and other sleep disorders, which include those associated with medical conditions or substance abuse (American Psychiatric Association 1994). This chapter is organized primarily on the basis of etiology and focuses on disturbances that are of most relevance to psychiatric practice; those wishing to see a broader spectrum or to obtain more detail are referred to major texts in the field (Mendelson 1987; Thorpy 1990; Poceta and Mitler 1998; Kryger et al. 2000). A detailed presentation of sleep disturbances in children, which is beyond the scope of this chapter, may be found in a recent compendium (Dahl 1996), and reviews have described the most recent findings in narcolepsy (Pelayo and Guilleminault 1998) and in sleep disturbance due to medications (Mendelson and Caruso 1998).

Let us begin with a reminder about a concept important to understanding sleep disorders in general. Since the birth of modern sleep research with the discovery of rapid eye movement (REM) sleep in the 1950s, there has been growing recognition that there are three states of consciousness: waking, REM sleep, and non-REM sleep. Each of these has a unique physiology, and indeed the physiology of REM is as different from the rest of sleep (non-REM) as the latter is from waking. Thus the regulatory mechanisms for thermal or ventilatory homeostasis found in major textbooks should be understood to be referring to processes that function when the individual is awake. It is almost as if a medical bookshelf should contain texts side by side: one for the physiology of wakefulness, and one (or even two) for physiology during sleep. Thus, while in a waking person exposure to increased ambient temperature will result in perspiration and increased ventilatory rate, and reduction in arterial oxygen saturation will cause an increase in minute ventilation, such responses are seen only very slightly if at all during REM sleep (Sullivan 1980; Parmeggiani 1980). This phenomenon explains a basic principle we will be dealing with in this chapter – that there are a number of illnesses based on pathophysiology which are manifest only during sleep. When awake, the pathophysiology is not evident, although some consequences, such as daytime sleepiness, may be. One of the best examples is found in sleep-disturbed respiration, with which we shall begin. After discussing this and another intrinsic disorder of sleep – periodic limb movement (PLM) disorder – we will describe delayed sleep phase syndrome, as an example of a circadian rhythm disorder, and then sleep disturbances due to causes extrinsic to sleep, including drugs and medical illness. We will conclude with comments about sleep in psychiatric illness, particularly depression, and two forms of insomnia which at this point are understood primarily in neuropsychological terms: psychophysiologic insomnia and sleep state misperception.

2 Sleep-Disordered Breathing

The historic roots of the understanding of sleep apnea lie in the late nineteenth century. Among those who first described sleep-disturbed breathing is S. Weir Mitchell, a neurologist important in the history of psychiatry for his rest cure for neurasthenia and for his advocacy of integrating psychiatry into the mainstream of medical research. In 1890, Mitchell described patients whose respiration was normal while awake, but who developed respiratory failure during sleep and would awaken with a sense of suffocation (Mendelson 1987); other contemporaries related upper airway obstruction to sleep disturbance and daytime somnolence (Lamacq 1887; Wells 1898). In more contemporary terms, a patient with uncomplicated sleep apnea is a person whose respiration when awake is fundamentally normal, but when sleeping develops full (apnea) or partial (hypopnea) periods of cessation of breathing. Depending on the definition used, the threshold for diagnosis is considered to be five or ten disordered breathing events (apneas and hypopneas) per hour of sleep (referred to as the apnea/hypopnea index). Most patients are unaware of having these events, although a bed partner may describe them. Typical complaints include daytime sleepiness, a sense that sleep is not restful, decreased libido, sexual dysfunction (Mendelson et al. 1990), and memory difficulties (Thorpy 1990; ASDA 1997). Some patients may report awakening with a sense of being unable to breathe, although reports of this nature should be evaluated in terms of a broad differential diagnosis including panic attacks, nocturnal asthmatic episodes, and nocturnal laryngospasm. The disordered breathing events may be

primarily obstructive (in which there is functional closure of the upper airway), central (representing a decrease in central respiratory drive), or mixed type. Obstructive sleep apnea (OSA) occurs more often in males than females by a ratio of perhaps 2:1. It tends to be an illness of middle life, although it can occur at any age, even in children, in whom there is a similar frequency in males and females. Associated features in adults include a history of snoring, obesity, and systemic hypertension. Patients with OSA have a higher rate of automobile accidents, presumably due to sleepiness, and those with an apnea/hypopnea index greater than 20 per hour have a somewhat higher mortality rate. Central sleep apnea can occur at any age, but is most common in the elderly. It, too, is more common in men than women, although after menopause this difference is reduced. A history of snoring is less common. The classical teaching is that OSA patients tend to complain of excessive sleepiness while central sleep apnea patients tend to complain of insomnia, although later studies suggest that this difference is less clear than was originally suggested (Mendelson 1995c).

As mentioned above, complaints of memory difficulties are common in OSA patients, and indeed cognitive deficits have been observed in various studies (Bedard et al. 1993). Whether treatment successfully reverses these deficits is not yet clear. One study has suggested that treated patients have improvement in long-term free recall or explicit memory (Mendelson et al. 1993), while others have shown persistent deficits in various measures even after continuous positive airway pressure (CPAP) therapy (Bedard et al. 1993).

There is evidence indicating that 20% of OSA patients currently have some form of depressive disorder using research diagnostic criteria (Reynolds et al. 1984), and the 5-year prevalence of major depressive disorder has been estimated to be 58% (Mosko et al. 1989). Some preliminary data suggest that patients with past or current treatment for depression have a significantly greater apnea/hypopnea index and total number of disordered breathing events and tend ($p < 0.06$) to have a lower minimum oxygen saturation than those with negative histories (Mendelson 1992). A later series of 199 OSA patients examined scores on the Minnesota Multiphasic Personality Inventory (MMPI) (Aikens et al. 1998). The depression scale was the most prominently abnormal scale, significantly elevated in 27% of patients, followed by hypochondriasis (22%) and hysteria (21%). The total number of MMPI elevations had a significant negative correlation with total sleep time, sleep efficiency, minimum oxygen saturation in non-REM sleep, and mean oxygen saturation in both REM and non-REM sleep. The depression scale itself did not significantly correlate with measures of sleep-disturbed respiration, although it was significantly related to total MMPI elevations. In summary, a range of symptomatology, including somatic concerns, emotional overreactivity, and other features, are often seen in OSA patients.

The original treatment for OSA was chronic tracheotomy, which, though effective, involves a major surgical procedure, requires chronic care, and is subject to various complications over time. Subsequently, uvulopalatopharyngoplasty (UPPP), which removes redundant tissue from the airway, was developed. Its advantages are that, when successful, it is a one-time procedure as opposed to other forms of treatment which require ongoing therapy; its disadvantage is that the success rate (which is highly dependent on the criteria employed) is in the range of 60%–70%, and it can also occasionally lead to complications, such as regurgitation of liquids or alterations in voice. Some centers perform specialized procedures including limited anterior sagittal mandibular osteotomy with genioglossus advancement, hyoid myotomy, maxillomandibular osteotomy with hyoid myotomy, or partial glossectomy (Powell et al. 1989; Douglas 1997). A relatively new procedure, the laser-assisted uvuloplasty (LAUP), has the advantage of being a more benign, outpatient procedure. Its effectiveness is still under investigation, but the evidence to date suggests that it is very useful for treating snoring in the absence of significant sleep-disordered breathing, but that there are little data to indicate that it is helpful for true OSA. Mechanical techniques such as tongue-retaining devices are also used, although their success rate is under investigation.

The most widely used treatment is CPAP, in which room air is introduced into the nose via a nasal mask or cushioned cannulae. It is very effective in the majority of patients. Its disadvantages are that it is a chronic treatment generally given for years and that there is often decreased compliance over time. Medical treatments at this point have been less successful. Among those that are used are protriptyline (which may increase tone of the genioglossus as well as reducing REM sleep) and medroxyprogesterone. Buspirone and fluoxetine have been used experimentally. (For all of the preceding compounds, these represent "off label" uses which are not specifically indicated in the American *Physician's Desk Reference.*) In clinical practice, this author's view is that the various medical approaches available at this time are best reserved for relatively mild cases or special situations such as for patients who are noncompliant or who refuse surgery or CPAP.

Ancillary treatments include weight loss, training to avoid sleeping on one's back, and avoidance of alcohol and other drugs which are respiratory depressants. Thus it is important to consider the possibility of sleep

apnea before administering many sedative/hypnotics to patients who may complain of poor sleep and, if features suggestive of the disorder (e.g. obesity, hypertension) are present, to obtain the services of a sleep laboratory. Drugs which are of particular concern for respiratory depression are the barbiturates. The older, long-acting benzodiazepines such as flurazepam may also exacerbate sleep-disturbed breathing; the newer, short-acting agents appear to be more benign in this regard.

In passing, it should be mentioned that one of Freud's more controversial colleagues was a Berlin otolaryngologist named Wilhelm Fliess, who believed that a number of psychiatric disorders, including depression, might be helped by operating on the nose. Although this idea has been held up to ridicule in some circles, in view of the possible association of sleep-disordered breathing with depressive symptomatology, with possible reversal after CPAP therapy (Millman et al. 1989), it may be that Fliess had more insight than he was credited for.

3 Periodic Limb Movement Disorder

Like sleep-disordered breathing, PLM is a disorder occurring primarily when the patient is asleep, with no specific manifestations in an awake patient. In this sense, it is referred to as a state-dependent motor disorder. PLM can be present in patients who complain of either insomnia or excessive sleepiness; in one case series of approximately 1700 patients in a sleep disorder center, it accounted for 13% and 6% of such cases, respectively (Mendelson 1997b). This is primarily a disorder of middle and older age and occurs equally in males and females. Typically, a patient has a normal neurologic examination when awake, and no evidence of seizure disorder on a waking or sleeping electroencephalogram (EEG), but when asleep has periodic movements of the limbs of which he or she is usually unaware. During a sleep study, these are manifested as movements in the anterior tibialis electromyogram (EMG), lasting 0.5–5.0 s, occurring approximately every 20–40 s. Often these are accompanied by brief evidence of arousal in the EEG (a K complex, alpha activity, or change to a lighter sleep stage). The most common diagnostic criteria is that there be at least five per hour. Clinically, these episodes are manifested as extensions of the big toe with dorsiflexion of the ankle, sometimes accompanied by knee flexion. In this sense, they are more reminiscent of a Babinski movement than a myoclonic jerk. The patient is usually unaware of the movements per se, but has a sense that his or her sleep is restless and unrefreshing. The bed partner, however, may be very aware of the movements, and hence it is often useful to interview him or her as well during a clinic visit.

The association of PLM with the subjective experience of poor sleep is not well understood, and indeed perhaps 6% of medical center staff with no sleep complaint meet the polygraphic diagnostic criteria (Kales et al. 1982). A later study of a clinical population complaining of either insomnia or excessive sleepiness found no correlation between the PLM arousal index (number of PLM associated with a brief EEG arousal, per hour of sleep) and subjective or objective measures of daytime sleepiness or a report of awakening refreshed or unrefreshed in the morning (Mendelson 1996). While the final answer is unclear, many clinicians believe that, although these movements can occur in people with no sleep complaints, there may be a subgroup in whom the associated arousals disturb sleep continuity to a degree that may correspond to a sense of having poor sleep.

The usual therapy for PLM has been low doses of clonazepam (Mitler et al. 1986), although compliance is often poor due to sedation; the shorter-acting temazepam (Mitler et al. 1986) and triazolam also may be used (Bonnet et al. 1990). The condition may be exacerbated by tricyclic antidepressants (Ware et al. 1984), and hence this possibility should be considered when a patient placed on these agents complains that sleep has worsened. It may occur transiently during withdrawal from barbiturates or benzodiazepines.

4 Restless Legs Syndrome

In contrast to PLM, restless legs syndrome (RLS) is a state-dependent sensory disorder. It is very different in another sense as well; while PLM is a polygraphically defined process in which the association with a subjective feeling of poor sleep is not entirely clear, RLS is primarily based on a subjective report, with minimal polygraphic evidence present. The typical history is of a dysesthesia, a "creepy crawly" or "pins-and-needles" sensation of the limbs, which occurs at rest and is relieved by movement. It is manifest as a sleep disorder insofar as patients have difficulty lying in bed to initiate sleep, but feel a need to get up and move around in order to dissipate this uncomfortable sensation. Although it can occur at any age, as with PLM it is most commonly a disorder of the middle and older years, but unlike PLM it is more common in females. It appears to have a higher frequency in patients with iron or B_{12} deficiency anemia and in patients with significant renal disease. It may also appear in pregnancy, usually after the 20th week of

gestation. Sleep studies reveal that most of these patients also have PLM (although only a small subset of PLM patients have RLS). Aside from this, there is little to be seen on polysomnography apart from increased motor movement artifact before sleep onset. It may be brought out by use of the forced immobilization test, in which a patient is immobilized on a stretcher for 60 min and sensory and motor events are recorded (Pelletier et al. 1992). The treatment of choice at this point is levodopa (Montplaisir et al. 1986), with use of bromocriptine (Von Scheele and Kempi 1990) in patients who do not respond well, although these are off-label indications for these compounds in the United States. The physician should check for, and treat, anemia. Many patients also develop anxiety or depressive symptoms which may require attention.

5 Circadian Rhythm Disorders: Delayed Sleep Phase Syndrome

Delayed sleep phase syndrome (DSPS) is a disorder in which the endogenous central nervous system (CNS) mechanisms for regulating the sleep/wake cycle operate in a stable but delayed phase relationship to the more typical times of waking and sleeping. In theory, it results from an inadequate ability of endogenous pacemakers to be reset by zeitgebers in the environment (primarily sunlight) to a typically scheduled 24-h day. A more extreme form of dysfunction, in which the pacemakers operate completely independently of the environment, such that the sleep/wake cycle tends to manifest the endogenous rhythm of approximately 25 h, is found in the much less common non-24-h sleep/wake syndrome. Patients with DSPS may present with a complaint of initial insomnia (Weitzman et al. 1981) or with difficulty with sleepiness in the morning. Onset is often in adolescence, and it is unusual to have symptoms appear for the first time after age 30. The data on the sex ratio are inconsistent, but it probably has approximately equal frequencies in men and women in the adult form. Typically a patient will describe being unable to fall asleep before 2:00 or 3:00 A.M. and then difficulty getting up before late morning. This may lead to problems with tardiness at school or work. There is often a history of use of sedative/hypnotics without much benefit, and this author has the clinical impression that many patients have been treated unsuccessfully for depression. Many patients have self-medicated with alcohol, and their sleep disturbance may be complicated by dependence.

A sleep study in DSPS usually reveals a very long sleep latency (time from the beginning of the recording until the patient first enters sleep), but then a relatively normal pattern, with minimal arousals. Thus this is a disorder of the timing, rather than the quality, of sleep. The original treatment for DSPS was chronotherapy (Czeisler et al. 1981), i.e. gradually moving the patient's time of sleep around the clock until his or her bedtime occurs at a more typical hour. In the short term, this is a fairly disruptive intervention, and compliance is often poor. A much more convenient therapy is to expose the patient to bright lights for 2 h in the morning, which phase-advances the sleep/wake rhythm to a more socially acceptable time (Rosenthal et al. 1990). In passing, it should be mentioned that there is one small case series of patients who became suicidal when given bright light therapy for seasonal affective disorders (Praschak-Rieder et al. 1997); the significance of this possible unusual association is under investigation.

6 Sleep Disturbance Due to Medications or Medical Illness

A wide range of prescription medications may result in disturbed sleep. The following may cause insomnia:

- CNS stimulants, e.g. dextroamphetamine
- β-Adrenergic agonists
- Steroids
- Calcium channel blockers
- Lovastatin
- Aminophylline
- β-Adrenergic blockers
- Thyroid preparations
- Cancer chemotherapeutic agents
- Fluoxetine

Other medications may cause daytime sleepiness, including the following:

- Sedative/hypnotics, particularly long-acting agents, e.g. flurazepam
- Phenytoin
- Disulfiram
- Tricyclic antidepressants
- Histamine-1 antagonists
- Centrally acting α-adrenoreceptor agonists

In psychiatric practice, a common dilemma is that patients may develop insomnia when placed on monoamine oxide (MAO) inhibitors or some selective serotonin reuptake inhibitors (SSRI), particularly fluoxetine. The ideal response is to lower the dose or discontinue the medication, but of course clinically this may not be possible. In that situation, low bedtime doses of trazodone may be added, although the

physician should be aware of rare instances of confusional states which may result.

It is also important to assess caffeine intake. In addition to caffeinated beverages, significant quantities are also found in a variety of nonprescription medications sold for a variety of reasons, including weight loss, headache, and menstrual discomfort. It is important to look for a present or past history of alcohol dependence. Even after a patient is "dry," sleep disturbances may persist for as long as 2 years.

Any medical disorder that causes pain or discomfort may of course disturb sleep, whether this be arthritic pain keeping an elderly patient awake or difficulty in breathing experienced by a congestive heart failure patient when lying down. In these types of patients, therapy is best directed at the underlying disorder rather than the sleep disturbance per se. Thus for the elderly arthritic patient, for example, the best remedy may be an analgesic rather than a hypnotic, and for the congestive heart failure patient a reassessment of the diuretic or cardiac medicines may be in order. The general principle, then, is to treat the underlying process whenever possible. This is true whether the fundamental difficulty is a medical problem or psychiatric disorder, as will be discussed in the next section.

7 Sleep in Depression

Sleep disturbance is reported by about 90% of patients with major affective disorder (Nelson and Charney 1980), and improvement in sleep has been interpreted by some to be one of the first signs of recovery (Mayer-Gross et al. 1960). A smaller group – perhaps 10% – complain of hypersomnolence, and these tend to be the bipolar or atypical depression patients. Sleep laboratory data indicate that total sleep and the amount of slow-wave sleep (stages 3 and 4) are reduced, while the number of awakenings during the night is increased (Mendelson 1987). Although theories put forth in the 1960s suggested a link between reductions in REM sleep and depression, the total amount of REM sleep is normal or only slightly reduced. A number of features of REM sleep are altered, however. It tends to occur earlier, often less than 60 min after sleep onset, a phenomenon known as a "short REM latency," which has been viewed by some investigators as a possible biological marker of depression. In addition, the first REM period is often longer, and characterized by more eye movements, than in controls. These changes have been interpreted in a variety of ways, and possible hypotheses have been that this reflects an alteration in the homeostatic regulation of sleep, a phase advance of a pacemaker regulating REM sleep, or increased cholinergic relative to noradrenergic activity (Mendelson 1987). Perhaps the most important point, however, is that a short REM latency is fairly sensitive, but not very specific, for depression, as it occurs in a variety of other conditions. Among these are obsessive-compulsive disorder, anorexia nervosa, drug and alcohol withdrawal, some instances of schizophrenia, divorcing women, and alcoholics with secondary depression. This has led some investigators to speculate that the short REM latency represents a marker of stress or illness severity rather than depression per se.

In practical terms, management of the sleep disturbance is best handled by aggressive treatment of the depression itself. There is little evidence to support the common practice of adding a benzodiazepine to the therapeutic regimen early in treatment. Usually a sedating tricyclic antidepressant or trazodone at bedtime provides relief during the period before the antidepressant effect is manifest. As mentioned earlier, a more difficult problem is the sleep disturbance which may result from MAO inhibitors and some serotonin reuptake inhibitors, which is best dealt with by dose adjustment or addition of a low dose of trazodone.

In passing, it should be mentioned that the sleep disturbance may not be merely an epiphenomenon of depression, but rather may in some sense be intimately related to the pathology. A variety of manipulations of sleep, including REM sleep deprivation, and partial or total sleep deprivation may effectively treat the constellation of depressive symptoms in a manner comparably effective as conventional therapies (Gillin 1983). The mechanism by which these treatments act remains to be elucidated. It is not clear, for instance, whether total sleep deprivation acts by altering circadian processes or whether it produces a nonspecific activation as opposed to a very specific improvement in depression. Although such techniques are very labor intensive and hence mostly of heuristic value, they do suggest a close relationship between disorders of sleep and affect. There are, of course, a wide variety of sleep disturbances in nonaffective psychiatric disorders, and the reader is referred to a recent thorough review (Kerkhofs 1997).

8 Chronic Insomnia

Many patients present with long-standing insomnia, but without evidence of major psychiatric or medical disorders, and do not manifest specific pathophysiology such as sleep apnea or PLM. These patients, who represent perhaps 20% of those seen in a typical sleep disorders center (Mendelson 1997b), are primarily

understood in neuropsychological terms. They are often divided into those with sleep disturbance manifest on the sleep recordings and in whom conditioned behaviors contribute to poor sleep (psychophysiologic insomnia) and those whose complaint of sleep disruption seems very disproportionate to the relatively mild changes found in their polygraphically measured sleep (sleep-state misperception insomnia). We will examine each of these in turn.

8.1 Psychophysiologic Insomnia

Of the two, psychophysiologic insomnia is by far the most common and is manifest by anxiety about not sleeping, which then in turn contributes to perpetuating the poor sleep. As in sleep-state misperception, it is more common in women than men; the typical age of onset is young adulthood, although patients are often not seen clinically until middle age. Phrased in behavioral terms, the typical patient has developed conditioned behaviors incompatible with sleep which become associated with the act of going to bed. The cues may be external (an association of the bedroom environment with a sense of distress about sleeping) or internal (a conditioned worrying about being unable to sleep). A typical story, which is not usually evident until after several visits, is that at some time in the past there was an emotionally traumatic event, such as an unhappy ending to a romantic relationship, which might be expected to lead to sleep disturbance in almost anyone. Typically, of course, a form of grief reaction occurs, and the sleep disturbance resolves along with the other signs of stress. These patients, however, seem to focus on the sleep disturbance rather than the reality issues; over time, the original trauma is sorted out, but the insomnia has taken on a life of its own. Clinically, it is characterized by difficulty going to sleep (as opposed to awakening during the night) and excessive worry about sleep. When questioned, the patient may report that he or she sleeps better when away from home (i.e. away from the stimuli associated with poor sleep) and that he or she can fall asleep when not trying (e.g. when watching television in the living room), but not when trying to sleep after going into the bedroom. The act of trying to sleep thus becomes a behavior which inhibits the ability to fall asleep.

8.2 Sleep-State Misperception

In sleep-state misperception, a patient who complains of chronic difficulty in going to sleep or awakening during the night is found to have relatively normal sleep as determined in the laboratory. The discrepancy between the patient's reported experience and the physiological measures has led to the hypothesis that this disorder results from the patient's misperception of his or her state of consciousness. An insight into this process comes from studies going back as far as the 1960s in which it was shown that many poor sleepers, when awakened by an auditory stimulus shortly after EEG-defined sleep onset, tend to report that they believe that they had been awake even though the polygraph indicated that they had been in non-REM sleep (Mendelson 1990). Later studies have indicated that, when such patients are given hypnotics such as triazolam or zolpidem, their reports of whether they had been awake or asleep become more consonant with the EEG (Mendelson 1993, 1995b); reports of having been awake or asleep in normal sleepers are not altered under these experimental conditions (Mendelson 1995a). An implication of these studies, then, is that hypnotic medications may not only increase the amounts of EEG-defined sleep, but may also alter the subjective experience of being asleep in some insomniacs.

Incidentally, it should be mentioned that there are also some patients with idiopathic insomnia, which typically begins in childhood and lasts throughout their lives. Both clinically and polygraphically, this may be mild or severe. The etiology is unknown, but presumably represents failure of sleep-inducing or -maintaining systems. When seen in childhood, there is often a history of birth complications, dyslexia, or hyperkinesis. In adults, there is often an overlay of suspiciousness, irritability, and dysphoric qualities, although there is no evidence of a higher frequency of major depressive disorder. Key elements of the history are that the sleep difficulty goes back to childhood and that the severity does not follow a temporal path parallel to periods of stress or emotional upset. In contrast, a diagnosis of sleep disturbance associated with depression, for instance, would require that there be a clear temporal pattern in which the sleep disturbance is coincident with clinical depressive symptomatology.

Although the processes which maintain the insomnia symptoms in some chronic insomnia patients are beginning to be understood (e.g. the conditioned behaviors incompatible with sleep in psychophysiologic insomnia), little is known about the underlying predisposing factors. One promising hypothesis suggests that there is a state of hyperarousal, as measured by heart rate, core temperature, and metabolic rate (Bonnet and Arand 1997).

8.3 Treatment

Both pharmacologic and nonpharmacologic treatments are available for the primary insomnias. The major sedative/hypnotics used are the benzodiazepines and the newer non-benzodiazepine γ-aminobutyric acid (GABA)A-benzodiazepine receptor agonists, including zaleplon, zolpidem and zopiclone, which have recently been reviewed (Mendelson and Caruso 1998). Unless daytime sedation is specifically desired, the shorter-acting agents are generally more appropriate. A recent review of reported adverse reactions to sedative/hypnotics during a 3-year period in a large teaching hospital indicates that most are mild and can be viewed as extensions of the therapeutic effect; the median frequency of reported adverse effects was 1 in 10,000 doses administered (Mendelson et al. 1996). In patients in whom traditional sedative/hypnotics are not effective, many clinicians use low bedtime doses of sedating tricyclic antidepressants or trazodone, although these are "off-label" indications in the United States. There has been a great deal of interest in the use of melatonin as a hypnotic; although this is an area in which there is not yet general agreement, this author's view is that there are very little data to support its use in insomnias which are not due to circadian rhythm disturbances (Mendelson 1997a).

Nonpharmacologic approaches to the primary insomnias are based on behavioral techniques (Mendelson 1987). Although they differ in their specifics, most have elements in common. These include documenting the amount of sleep, emphasizing self-management, encouraging the patient to take a more active stance, and helping the patient to see the sleep difficulty in the broader context of his or her waking life. Probably the two most widely used approaches are sleep restriction and stimulus control. Sleep restriction therapy (Spielman et al. 1987) is based on the notion that excessive time in bed helps perpetuate sleep disturbance, although the original cause may have been some other process. The technique is to make sleep more efficient by first shortening time in bed and then slowly increasing it; a 35-week follow-up study showed continued benefits as measured by total sleep, sleep efficiency, and waking time after initial sleep onset. The stimulus control technique (Bootzin and Engle-Friedman 1987), derived from operant conditioning principles, seeks to remove behaviors that are incompatible with falling asleep. Such behaviors include, for instance, having unhappy discussions with one's spouse in the bedroom, using the bed as a place to plan tomorrow's activities, or worrying about not sleeping. Thus the patient is instructed to strip the bedroom of all activities except sleeping and sex. The patient is told that if he or she is lying in bed unable to sleep, to get up and go into another room, and not return to bed until he or she feels able to sleep. After a 4-week treatment period, benefits have been reported to be maintained at 6-week follow-up (Puder et al. 1983). It should also be mentioned that medication and nonpharmacologic techniques are not necessarily incompatible. One study has indicated that patients who received a combination of behavioral therapy and triazolam were doing better in terms of total sleep and feeling rested in the morning 5 weeks after treatment compared to patients in monotherapy (Milby et al. 1993).

9 Conclusion

Sleep difficulties may be seen as symptoms of underlying processes that include intrinsic pathophysiologies of sleep, circadian rhythm disorders, medical or psychiatric illness, drugs, conditioned behaviors incompatible with sleep, or misperception of states of consciousness. For any given patient, the goal is to identify the underlying disturbance and provide a specific treatment.

10 References

Aikens JE, Caruana-Montaldo B, Vanable PA, Tadimeti L, Mendelson WB (1998) Depression and general psychopathology in obstructive sleep apnea. Sleep [Suppl] 21: 71 (abstr)

American Psychiatric Association (1994) Diagnostic and statistical manual of mental disorders, 4th edn. American Psychiatric Association, Washington

ASDA (1997) International classification of sleep disorders, revised: diagnostic and coding manual. American Sleep Disorders Association, Rochester, pp 1–401

Bedard MA, Montplaisir J, Malo J, Richer F, Rouleau I (1993) Persistant neuropsychological deficits and vigilance impairment in sleep apnea syndrome after treatment with continuous positive airway pressure (CPAP). J Clin Exp Neuropsychol 15: 330–341

Bonnet MH, Arand DL (1997) Hyperarousal and insomnia. Sleep Med Rev 1: 97–108

Bonnet MH, Dexter JR, Arand DL (1990) The effect of triazolam on arousal and respiration in central sleep apnea patients. Sleep 13: 31–41

Bootzin RR, Engle-Friedman M (1987) Sleep disturbances. In: Edelstein BA, Carstensen LL (eds) Handbook of clinical gerontology. Pergamon, New York.

Czeisler CA, Richardson GS, Coleman RM et al (1981) Chronotherapy: resetting the circadian clock of patients with delayed sleep phase insomnia. Sleep 4: 1–21

Dahl RE (1996) Sleep disorders. Child Adolesc Psychiatr Clin North Am 5: 1–767

Douglas NJ (1997) Surgical treatment for obstructive sleep apnoea. Sleep Med Rev 1: 77–86

Gillin JC (1983) The sleep therapies of depression. Prog Neuropsychopharmacol Biol Psychol 7: 351–364

Kales A, Bixler EO, Soldatos CR et al (1982) Biopsychobehavioral correlates of insomnia. I. Role of sleep apnea and nocturnal myoclonus. Psychosomatics 23: 589–600

Kerkhofs M (1997) EEG sleep in non-affective psychiatric disorders. Sleep Med Rev 1: 109–118

Kryger MH, Roth T, Dement WC (2000) Principles and practice of sleep medicine, 3rd edn. Saunders, Philadelphia, pp 1–1336

Lamacq L (1887) A propos de quelque cas de narcolepsie. Rev Med 17: 699–714

Mayer-Gross W, Slater E, Roth M (1960) Clin Psychiatry 211: 381–395

Mendelson WB (1987) Human sleep: research and clinical care. Plenum, New York, pp 96–97

Mendelson WB (1990) Insomnia: the patient and the pill. In: Bootzin RR, Kihlstrom JF, Schachter DL (eds) Sleep and cognition. American Psychological Association, Washington, pp 139–147

Mendelson WB (1992) Depression in obstructive sleep apnea. Sleep Res 21: 230

Mendelson WB (1993) Pharmacologic alteration of the perception of being awake or asleep. Sleep 16: 641–646

Mendelson WB (1995a) Effects of flurazepam and zolpidem in the perception of sleep in normal volunteers. Sleep 18: 88–91

Mendelson WB (1995b) Effects of flurazepam and zolpidem on the perception of sleep in insomniacs. Sleep 18: 92–96

Mendelson WB (1995c) The relationship of sleepiness and blood pressure to respiratory variables in obstructive sleep apnea. Chest 108: 966–972

Mendelson WB (1996) Are periodic leg movements associated with clinical sleep disturbance? Sleep 19: 219–223

Mendelson WB (1997a) Efficacy of melatonin as a hypnotic agent. J Biol Rhythms 12: 651–656

Mendelson WB (1997b) Experiences of a sleep disorders center: 1700 patients later. Cleveland Clin J Med 64: 46–51

Mendelson WB, Caruso C (1998) Pharmacology in sleep medicine. In: Poceta JS, Mitler MM (eds) Sleep disorders: diagnosis and treatment. Humana, Totowa, pp 137–160

Mendelson WB, Gujavarty K, Slintak C, Schwartz J (1990) Reported sexual dysfunction in obstructive sleep apnea patients. Ann Clin Psychiatry 2: 93–101

Mendelson WB, Maczaj M, Putnam K, Weingartner H (1993) Cognitive measures in obstructive sleep apnea patients before and after treatment. Sleep Res 22: 235 (abstr)

Mendelson WB, Thompson C, Franko T (1996) Adverse reactions to sedative/hypnotics: three years' experience. Sleep 19: 702–706

Milby JB, Williams V, Hall JN, Khuder S, McGill T, Wooten V (1993) Effectiveness of combines triazolam-behavioral therapy for primary insomnia. Am J Psychiatry 150: 1259–1260

Millman RP, Fogel BS, McNamara ME, Carlisle CC (1989) Depression as a manifestation of obstructive sleep apnea: reversal with nasal continuous positive airway pressure. J Clin Psychiatry 50: 348–351

Mitler MM, Browman CP, Menn SJ et al (1986) Nocturnal myoclonus: treatment efficacy of clonazepam and temazepam. Sleep 9: 385–392

Montplaisir J, Godbout R, Poirier G, Bedard M (1986) Restless leg syndrome and periodic movements of sleep: physiopathology and treatment with L-dopa. Clin Neuropharmacol 9: 456–463

Mosko SS, Zetin M, Glen S et al (1989) Self-reported depressive symptomatology, mood ratings, and treatment outcome in sleep disorders patients. J Clin Psychiatry 45: 51–60

Nelson JC, Charney DS (1980) Primary affective disorder criteria and the endogenous-reactive distinction. Arch Gen Psychiatry 37: 787–793

Parmeggiani PL (1980) Temperature regulation in sleep. In: Orem J, Barnes CD (eds) Physiology and sleep. Academic, New York, pp 98–145

Pelayo R, Guilleminault C (1998) Narcolepsy and excessive daytime sleepiness. In: Poceta JS, Mitler MM (eds) Sleep disorders – diagnosis and treatment. Humana, Totowa, pp 95–116

Pelletier G, Lorrain D, Montplaisir J (1992) Sensory and motor components of the restless legs syndrome. Neurology 42: 1663–1666

Poceta JS, Mitler MM (1998) Sleep disorders: diagnosis and treatment. Humana, Totowa, p 1

Powell NB, Guilleminault C, Riley RW (1989) Surgical therapy for obstructive sleep apnea. In: Kryger MH, Roth T, Dement WC (eds) Principles and practice of sleep medicine. Saunders, Philadelphia, pp 706–721

Praschak-Rieder N, Neumeister A, Hesselmann B, Willeit M, Barnas C, Kasper S (1997) Suicidal tendencies as a complication of light therapy for seasonal affective disorder: a report of three cases. J Clin Psychiatry 58: 389

Puder R, Lacks P, Bertelson AD, Storandt M (1983) Short-term stimulus control treatment of insomnia in older adults. Behav Ther 14: 424–429

Reynolds CF III, Kupfer DJ, McEachran AB, Taska LS, Sewitch DE, Coble PA (1984) Depressive psychopathology in male sleep apneics. J Clin Psychiatry 45: 287–290

Rosenthal NE, Joseph-Vanderpool JR, Levendosky AA, Johnston SH, Allen R, Kelly KA, Souetre E, Schultz PM, Starz KE (1990) Phase-shifting effects of bright morning light as treatment for delayed sleep phase syndrome. Sleep 13: 354–361

Spielman AJ, Saskin P, Thorpy MJ (1987) Treatment of chronic insomnia by restriction of time in bed. Sleep 10: 45–56

Sullivan CE (1980) Breathing in sleep. In: Orem J, Barnes CD (eds). Physiology and sleep. Academic, New York, pp 214–272

Thorpy MJ (1990) Handbook of sleep disorders. Dekker, New York, pp 1–817

Von Scheele C, Kempi V (1990) Long-term effects of dopaminergic drugs with restless legs. Arch Neurol 47: 1223–1224

Ware JC, Brown FW, Moorad PJ, Pittard JT, Murphy M, Franklin D (1984) Nocturnal myoclonus and tricyclic antidepressants. Sleep Res 13: 72 (abstr)

Weitzman ED, Czeisler CA, Coleman RM et al (1981) Delayed sleep-phase syndrome. A chronological disorder with sleep-onset insomnia. Arch Gen Psychiatry 38: 737–746

Wells W (1898) Some nervous and mental manifestations occurring in connection with nasal disease. Am J Med Sci 116: 677–682

CHAPTER

15

R.J. Hodgson, R. Budd, M. Griffiths

Compulsive Behaviours

1 Introduction

Many of the compulsive behaviours or behavioural addictions that are the focus of this chapter appear to fit the DSM-IV definition of impulse control disorders. This definition includes failure to resist an impulse, drive or temptation to perform some act that is harmful to the person or others; an increasing sense of tension or arousal before committing the act; and an experience of immediate pleasure, gratification or release when the act is initiated or completed. Compulsive behaviours such as compulsive buying or compulsive gambling also share many characteristics with drug dependence as well as with obsessive–compulsive disorders. However, in order to justify the use of the term compulsive behaviour rather than excessive, impulsive or addictive behaviour, there is a need to explain the use of the term compulsive.

It is tempting to suggest that there are compulsions that function to avoid distress, such as obsessive–compulsive disorders, and compulsions that function to achieve pleasure or gratification, such as compulsive buying or sexual compulsions. Certainly this distinction might apply to the initial stages of the development of the habit, but when a compulsion is full blown, the situation is more complicated. Although a sexual compulsion, for example, results in immediate pleasure, there are also occasions when the avoidance of preoccupation and frustration becomes more important than the achievement of gratification.

Thus a closer analysis leads to the conclusion that most compulsive behaviours have one characteristic in common, i.e. resisting an urge results in an unpleasant experience. This unpleasant experience could be called anxiety if a threat is involved, and frustration if there is a loss of an expected gratification (Table 1).

The following definition of a compulsive behaviour flows from this framework. It focuses upon the expected consequences of resisting the performance of an urge.

A compulsive behaviour is an active avoidance response. The avoidance is of intense, expected subjective experiences such as those associated with anxiety, frustration, guilt, anger or withdrawal symptoms. The other aspect of compulsion that needs to be emphasised is ambivalence. There is often a desire to carry out the compulsion and often a desire to resist.

Although this approach to compulsion applies to a wide range of behaviours such as drug use and obsessive–compulsive disorders, this chapter will focus upon the following disorders: compulsive gambling, compulsive buying, compulsive exercise, compulsive stealing, sexual compulsions and technological compulsions.

Table 1. Experiences associated with compulsive avoidance and compulsive appetites when behaviour is resisted or initiated after a long delay

	Behaviour resisted	Behaviour initiated after delay
Compulsive avoidance (e.g. OC disorder)	Anxiety	Relief of anxiety
Compulsive appetites (e.g. sexual compulsions)	Frustration	Relief of frustration

OC, obsessive–compulsive.

2 Compulsive Gambling

2.1 Definition

Compared to the other compulsive behaviours gambling, has been the subject of many academic and clinical publications, but the majority are speculative and do not provide definitive guidelines. Future epidemiological studies will be on a better footing thanks to the recent development of a reliable and valid method of identifying compulsive gamblers in community and clinical population.

The South Oaks Gambling Screen (SOGS) (Volberg and Steadman 1989) has been shown to be reliable and valid in identifying compulsive gamblers within a range of settings (Lesieur and Blume 1987). The SOG includes 20 items derived from the criteria published in the third edition of the *Diagnostic and Statistical Manual* (DSM-III), including the following:

- Gambled more than intended
- Have felt guilty about gambling
- Unable to stop gambling
- Others have criticised gambling

The 4-week test–retest reliability for SOGS is 0.71, and the internal consistency is 0.97. The correlation with a detailed clinical assessment based upon DSM-III-R is 0.94.

2.2 Prevalence

Estimates of the number of probable adult pathological gamblers vary from 0.2% to 1% in the United Kingdom (Dickerson 1974; Royal Commission 1978), 0.77% to 3.4% in the United States (Kallick et al. 1979; Culleton 1985; Sommers 1988; Volberg and Steadman 1988,

1989) and 0.25% to 1.73% in Australia (Dickerson and Hinchy 1988). These surveys have also indicated that pathological gambling is twice as common among males as it is among females, that non-whites have higher rates than whites and that those with poor education are more likely to be pathological gamblers (Lesieur and Rosenthal 1991).

As with many other excessive and dependence behaviours, availability is an important influence. For example, a study of prevalence across different states in the United States found that the number of opportunities to gamble at casinos, with slot machines, on sports and in teletheatres was associated with greater per capita incidence of Gamblers Anonymous chapters (Lester 1994). It is likely that the prevalence of excessive gambling will increase if opportunities to gamble continue to increase in many countries.

The acquisition, development and maintenance of pathological gambling is an area that is continually disputed. The exact causes and reasons for continuing gambling behaviour seem to be dependent upon the individual, but there also seem to be some general underlying factors and recurring themes. Problem gambling generally begins in adolescence and may start following a major life stress (Wolkowitz et al. 1985), e.g. the death of a parent or the birth of the first child. Such events may induce a need to escape from the problems of reality (Moran 1970). Prior to the age of 15, other predisposing factors may include serious family problems (e.g. divorce of parents), inappropriate school or parental discipline, exposure to gambling in childhood and/or adolescence (by family and/or peers) and even familiar emphasis on material symbols rather than savings (American Psychiatric Association 1980). According to Custer and Custer (1981), there are several "soft signs" of pathological gambling, including a higher than average I.Q., a lively and energetic personality, a predilection to taking risks, a lack of hobbies and interests, a low boredom threshold, episodic insomnia and "workaholic" tendencies.

As with all compulsive and dependence behaviour, it is important to recognise the heterogeneity of gambling. Firstly, there are many types of gambling, including the casino, cards, lotteries, horse and greyhound racing, bingo, gaming machines, dice, the stock market, games of skill and sports. Secondly, many types of people take many different routes into gambling and experience a range of consequences. Volberg and Steadman (1988), for example, identified two distinct groups of excessive gamblers based upon income. Lower-income gamblers were less likely to be white or male, had significantly lower levels of education and were more likely to have tried to stop gambling. They were less likely to have tried many types of gambling and tended to gamble less frequently. The higher-income group experienced more frequent criticism by others about their gambling and had argued more often with others about this behaviour. They had more frequently borrowed from banks and loan companies or chased in stocks and bonds. It is probable that different treatment interventions will be found to be appropriate for different types of people who gamble excessively in different settings with different levels of severity.

2.3 Treatment

Although in most spheres of treatment, therapists have claimed a reasonable amount of success, it is probably fair to say that gambling has been fairly resistant to therapy interventions and that the gamblers themselves may be reluctant and resistant. The thorough meta-analysis of gambling treatment outcomes carried out by Walker (1992) claimed that only about a third of gamblers (37%) are still abstinent after 2 years. A recent study of cognitive-behavioural approaches in the treatment of excessive gambling on slot machines (Echebura et al. 1996) suggests that cue exposure and response prevention should be considered, especially since this approach has proven effectiveness with other compulsions (Rachman and Hodgson 1980). This successful treatment for gambling involved gradual in vivo exposure which encouraged participants to experience the desire to gamble and learn to resist this desire. At the start of treatment, high-risk situations were deliberately avoided and access to money was controlled. These rules were then gradually relaxed as treatment progressed. Cue exposure and response prevention was superior to cognitive restructuring therapy and also to a combined cognitive-behavioural approach at the 1-year follow-up point. Cue exposure resulted in a 69% rate of successful abstinence, whereas other treatments resulted in only 38% success.

Psychodynamic "cures" are at present almost non-existent (at least in the published literature), and aversive behavioural techniques are rarely used except in conjunction with other forms of treatment. The influence of the spouse of the adult gambler in successful treatment appears to be crucial, and it would seem that the most potentially effective treatments are those which acknowledge the multivariant nature of and influences on pathological gambling (e.g. cognitive, social and physiological) and utilise more than one therapeutic approach. The only "new" (and some might say radical) approach to the treatment of pathological gamblers could be in the field of psychopharmacology, which perhaps might see the use of beta blockers to decrease arousal levels during gambling (Brown 1986)

or the administration of opiate antagonists to block euphoric beta endorphin effects (Blaszczynski et al. 1986). There are already some published case studies on the success of treating pathological gambling using pharmacological agents, for instance, including the use of lithium carbonate (Moskowitz 1980) and clomipramine (Hollander et al. 1992).

3 Compulsive Buying

3.1 Definition

In the small body of literature on the subject, the condition is also referred to as oniomania, buying mania, compulsive consumption, compulsive shopping and addictive or impulsive buying. The ubiquitous and increasing use of credit cards is paving the way for an epidemic of compulsive buying, and yet there are still only a few research studies on which to base assessment and treatment advice (McElroy et al. 1994; Christenson et al. 1994).

As with many other excessive behaviours, compulsive buying can lead to severe personal distress and family disruption (Christenson et al. 1994). The main characteristics of this disorder are an irresistible urge to buy and a feeling of immediate relief following a purchase. Attempting to resist the urge upon the realisation that buying has become a severe problem often follows the build-up of large debts and criticism from family and acquaintances as well as legal problems and feelings of guilt.

In the study by Christenson et al. (1994), most compulsive buyers described experiencing irresistible urges, uncontrollable needs or mounting tension that could only be relieved by buying. Over 90% of them frequently attempted to resist the urges to buy, but were unsuccessful 75% of the time. Typically 1–5 h elapsed between initially experiencing an urge to buy and an eventual purchase. Usually compulsive buyers do not decide in advance exactly what they intend to buy.

Based upon a detailed study of 20 compulsive buyers, McElroy et al. (1994) proposed the following preliminary operational criteria:

1. Maladaptive preoccupation with buying or shopping, or maladaptive buying or shopping impulses or behaviour, as indicated by at least one of the following:
   a) Frequent preoccupation with buying or impulses to buy that are experienced as irresistible, intrusive, and/or senseless.
   b) Frequent buying of more than can be afforded, frequent buying of items that are not needed or shopping for longer periods of time than intended.
2. The buying preoccupation, impulses or behaviour cause marked distress, are time-consuming, significantly interfere with social or occupational functioning or result in financial problems (e.g. indebtedness or bankruptcy).
3. The excessive buying or shopping behaviour does not occur exclusively during hypomania or mania.

Note that for the first criterion either excessive buying or a preoccupation with buying must be present, but not necessarily both. For the second criterion, as with obsessive-compulsive and substance misuse disorders, the preoccupation or the buying behaviour must lead to marked distress, financial difficulties or interference with adaptive social or occupational functioning.

3.2 Treatment

There are no studies of effective treatments for compulsive buying, even though there are at least two self-help books which focus on this disorder (Damon 1988; Wesson 1991). McElroy et al. (1994) asked 20 patients to evaluate their response to treatments they had received. Neither insight-oriented psychotherapy nor pharmacological interventions (thymoleptic medications) appeared to be effective in the long term. It is also interesting to note that none of the patients had received cognitive-behavioural therapy for their compulsive buying.

4 Compulsive Exercise

4.1 Definition and Prevalence

Unfortunately, the lack of adequate research means that it is impossible to estimate the prevalence of compulsive over-exercising in the general population. To date, what evidence there is on this problem has mostly taken the form of clinical case reports (e.g. de la Torre 1994; Griffiths 1997a) and, where data have been reported, it is difficult to draw any firm conclusions since excessive exercising has been defined in vastly different ways by different authors (see Davis and Fox 1993; Davis et al. 1993). While over-exercising has traditionally been defined by the frequency and/or

duration of exercising (de Coverley Veale 1987), recent research has suggested that in order to adequately delineate this problem, it may be necessary to consider exercise cognitions as well as exercise behaviour (Davis and Fox 1993; Long et al. 1993).

4.2 Association with Anorexia

The lack of evidence concerning excessive exercising is perhaps surprising given that this problem has now been acknowledged for some time (de Coverley Veale 1987). However, what research there is has tended to focus on the relationship between over-exercising and eating disorders. It is known, for example, that there are high rates of eating disorders among athletes (Johnson et al. 1990) and that excessive exercising is commonly associated with anorexia nervosa (Crisp et al. 1980) as well as with dieting and weight preoccupation in the general population (Davis and Fox 1993). With regard to the hypothesised relationship between eating disorders and excessive exercising, it has been argued by some authors that regular exercise may foster weight preoccupation, while other authors have suggested that over-exercising and weight preoccupation are symptoms of the same underlying psychological disturbance (Eisler and Le Grange 1990).

4.3 Mechanisms of Exercise Dependence

However, recent research does suggest that it may be possible to isolate a group for whom excessive exercise is a problem in its own right, rather than simply being a feature of anorexia (Davis and Fox 1993). This problem has variously been termed exercise dependence (de Coverley Veale 1987), exercise abuse (de la Torre 1994) or exercise addiction (Kagan and Squires 1985). It appears that, in line with our previously presented model of compulsive behaviour, compulsive exercising may be maintained by the experience of an aversive state (withdrawal) when regular exercise is stopped (Morris et al. 1990; de Coverley Veale 1987). It has further been hypothesised that this may be mediated by the endogenous opiates (Russell et al. 1987). Similarly, some evidence suggests that when excessive exercisers are prevented from exercising, they experience guilt, dysphoria and the compulsion to exercise harder next time in order to make up for the missed exercise session or sessions (Davis and Fox 1993; Long et al. 1993).

A recent review of the exercise addiction literature by Murphy (1994) outlined the three most significant psychophysiological explanations for exercise addiction – the thermogenic hypothesis, the catecholamine hypothesis and the endorphin hypothesis. The thermogenic hypothesis suggests that exercise increases body temperature (which reduces tonic muscle activity), thus reducing somatic anxiety (de Vries 1981; Morgan and O'Connor 1988). The catecholamine hypothesis suggests that exercise releases catecholamines, which are strongly implicated in the control of attention, mood, movement and endocrine, cardiovascular and stress responses (i.e. dopamine, adrenaline, noradrenaline) (Kety 1966). Further to this, high levels of catecholamines are thought to be associated with euphoria and positive mood state. The third hypothesis – the endorphin hypothesis – is the best known and most empirically researched. It is also the hypothesis that seems to be accepted by the general public. For instance, in a study carried out by Griffiths and Duff (1993) of lay beliefs about 17 potentially addictive behaviours, it was reported that exercise addiction was perceived to be more of a physiological addiction than a psychological one when compared with other behaviours such as gambling, television watching and sex. The endorphin hypothesis suggests that exercise produces endogenous morphines (i.e. endorphins), which leads to a reinforcing enhanced mood state. Despite the general acceptance of the endorphin hypothesis, there remains little conclusive evidence to indicate a precise mechanism of effect (Murphy 1994). Moreover, not only does the type of exercise (i.e. steady state or progressive and incremental) and the mode (e.g. running, step aerobics) need to be taken into account in the role of mood modulation, but (perhaps more importantly) the person's psychological constitution.

4.4 Treatment

There are no systematic outcome studies for excessive exercise reported in the literature. However, given the apparent similarity between this and other compulsive behaviours, we might expect that a broad cognitive-behavioural approach might be an effective treatment.

5 Kleptomania (Compulsive Stealing)

5.1 Definition

Kleptomania has been recognised as a psychological disorder since the early nineteenth century. DSM-IV

classifies kleptomania as an impulse control disorder that is characterised by the failure to resist impulses to steal along with the experience of tension prior to committing a theft, which is relieved once the theft has been committed. Thus, for repetitive theft to be classified as kleptomania, the motive of the theft must not be financial gain, but rather the avoidance of an aversive psychological state. In this regard, kleptomania has the characteristic features of a behavioural compulsion.

5.2 Prevalence

The main sources of data on the prevalence of kleptomania are studies of shoplifters who have been referred by the courts for psychiatric assessment. In their review of these studies, McElroy et al. (1991) note that these studies suggest that an average of 4% of apprehended shoplifters meet the criteria for kleptomania. However, these authors suggest that this figure may underestimate the true prevalence of kleptomania in the general population due to systematic sampling biases.

It appears that kleptomania may be more common in women than in men, just as shoplifting is more common among women (McElroy et al. 1991). While it has been argued that the higher prevalence of kleptomania among women may simply reflect cultural stereotypes that pathologise criminal behaviour in women, it is likely that this higher prevalence may be due, at least in part, to the more frequent access women have to situations in which they can steal.

5.3 Association with Other Psychological Problems

It has been noted by many authors (e.g. Gudjonsson 1987) that kleptomania is frequently associated with a range of other psychological problems including depression, eating disorders, anxiety symptoms and a range of obsessive-compulsive behaviours. Given that kleptomania has often been defined as shoplifting that is "out of character", it is not surprising that this behaviour, like other compulsive behaviours, often leads to severe personal distress and significant social and family problems.

5.4 Treatment

While no controlled outcome studies have been reported for the treatment of kleptomania, a number of single case studies of behavioural interventions have been reported (Gauthier and Pellerin 1982; Glover 1985; Kraft 1970). These have focused on the use of desensitisation, contingency management and response prevention in the treatment of kleptomania.

6 Sexual Compulsions

6.1 Definition

Problem behaviours involving sexual compulsions have variously been termed sexual addiction, hypersexuality, hyperphilia and, most commonly, compulsive sexual behaviour (Coleman 1991). While most authors acknowledge that the term compulsive sexual behaviour describes a set of different, but related, problem behaviours, there is currently no agreed classification of compulsive sexual behaviour. For example, while DSM-IV classifies these problems as psychosexual disorders not elsewhere classified, others have classified these behaviours as impulse control disorders (for a discussion of the problem of classification of compulsive sexual disorders, see Travin 1995).

Despite the problem of classification, one important distinction to make is that between the compulsive enactment of normative (non-paraphilic) and non-normative (paraphilic) sexual behaviour. While the classification of some sexual acts as paraphilic is dependent upon social and cultural norms that have changed over time, this distinction does nonetheless have clinical utility. Most notably, many paraphilic sexual compulsions may be best treated within a forensic setting.

Coleman (1992) has proposed a classification of non-paraphilic compulsive sexual behaviours that may prove helpful to the clinician:

- Compulsive cruising and multiple partners: This involves compulsively searching for new sexual partners to satisfy the individual's need for sexual intercourse.
- Compulsive fixation on an unobtainable partner: This is frequently associated with unrealistic sexual fantasies concerning the targeted partner.
- Compulsive auto-eroticism: This involves compulsive masturbation that results in occupational, social or interpersonal problems or self-injury.
- Compulsive multiple love relationships: This involves compulsively searching for the experience of new-found love, with each relationship ending following the initial romance.

- Compulsive sexuality within a relationship: This involves an insatiable and compulsive demand for sexual intercourse.

6.2 Mechanisms of Compulsive Sexual Behaviour

While Travin (1995) has noted that compulsive sexual behaviour has been variously hypothesised to be a spectrum affective disorder, an impulse control disorder or an "addiction", he suggests that it should not be defined as an obsessive-compulsive disorder as the behaviour is ego-syntonic. In other words, he argues that the individual is motivated to engage in the sexual behaviour rather than resist the behaviour, as is the case with obsessive-compulsive disorders. However, in line with our previously described model of behavioural compulsions, Coleman (1991) has noted that "compulsive sexual behaviour ... is driven by anxiety reduction mechanisms rather than by sexual desire. The obsessive thoughts and compulsive behaviours serve the function of temporarily reducing anxiety and stress, but they create a self-perpetuating cycle" (p. 37). In this way, non-paraphilic compulsive sexual behaviours may be best defined by reference to the mechanisms that motivate the behaviour.

6.3 Prevalence

While there are no studies on the prevalence of such problems, Coleman (1992) has estimated, on the basis of her clinical experience in this area, that compulsive sexual behaviours may occur in as many as 5% of the adult population over their life time.

6.4 Treatment

There are no reported treatment outcome studies of non-paraphilic compulsive sexual behaviours. There are, however, case reports of the use of lithium carbonate, selective serotonin reuptake inhibitors (SSRI) and anti-androgens in the treatment of these problems (for a review of these case reports, see Coleman 1992). In addition, it has been suggested that psychological therapy may be effective in treating compulsive sexual behaviours. In particular, given the clear similarities that have been noted between such behaviours and other behavioural compulsions (Coleman 1991), we might expect cognitive-behavioural approaches such as cue exposure and response prevention to be potentially profitable treatment procedures.

7 Technological Compulsions

As our use of technology grows and increasingly becomes a major part of our everyday leisure activities, there is also the potential for the development of technological compulsions or addictions. The term technological addictions has been used to cover computer addiction (e.g. hacking, Internet usage), video and computer games, fruit machines, pinball machines and virtual reality addiction (Griffiths 1995b). Our use of technology can lead to excessive use, unsuccessful efforts to cut down, giving up other social, occupational and recreational activities and persisting despite recurring physical, psychological, social or occupational problems. Since technological compulsions will undoubtedly increase in the future, there is a need for more systematic research of the field. Three of these compulsions will be examined in this chapter: fruit machine playing, computer game playing and Internet usage. Fruit machines are a type of gambling (slot) machine. Three reels contain differing numbers of symbols (usually representations of fruit, e.g. cherries, lemons) which spin on predetermined time cycles. "Pay-out" rates of machines vary between 70%–90%, and money is won when the payline (i.e. the middle row) shows a winning symbol combination, e.g. a row of three cherries.

7.1 Compulsive Fruit Machine Playing

Compulsions to play fruit machines involve excessive person-machine interaction and to some extent rely on the machine's structural characteristics, i.e. features which have been put into the machine to make them more tempting (e.g. light and sound effects, graphics, pseudo-skill buttons). It could be argued that it is these features which make compulsive fruit machine playing more of a technological compulsion than a gambling compulsion. It is also worth noting that compulsive fruit machine playing also tends to affect adolescents more than adults.

7.1.1 Prevalence

There have been about 20 studies examining adolescent fruit machine playing (Griffiths 1995a). Research

figures suggest that 10% of adolescents are regular fruit machine players (playing at least once a week) and that up to 6% of adolescents are probably compulsive players or have severe playing difficulties (Fisher 1993; Griffiths 1995a). All studies have reported that boys play on fruit machines more than girls and that as fruit machine playing becomes more regular, it is more likely to be a predominantly male activity. Very few female adolescent fruit machine addicts have been identified by researchers.

Like other compulsive behaviours, fruit machine compulsion causes the individual to engage in negative behaviours such as truanting in order to play the machines (Huff and Collinson 1987; Moran 1987; National Housing and Town Planning Council 1988; Leeds Polytechnic cited in Long 1989; Griffiths 1990), stealing to fund machine playing (Barham and Cormell 1987; Moran 1987; Spectrum Childrens Trust 1988; Griffiths 1990), getting into trouble with teachers and/or parents over machine playing (Moran 1987; Griffiths 1990), borrowing or using of lunch money to play the machines (National Housing and Town Planning Council 1988; J. Rands and M. Hooper, unpublished; Griffiths 1990), poor schoolwork (Moran 1987; Griffiths 1990) and in some cases aggressive behaviour (Moran 1987; Griffiths 1990).

There is also a small body of evidence (Griffiths 1995a) suggesting that there may be at least two types of compulsive fruit machine player. The first type appears to be addicted to the fruit machine itself (a "primary addiction") and plays to test his or her skill, to get social rewards and most of all for excitement, i.e. plays fruit machines because of their arousing properties. The second type appears to play machines as a form of escapism, where the machine is possibly an "electronic friend", i.e. plays them because of their tranquillising properties. This is what could be termed a "secondary addiction" in that the player uses the machines to escape the primary problem (e.g. broken home, physical disability, relationship break-up). If the primary problem is resolved, the excessive playing disappears. Such a distinction has obvious clinical usefulness.

7.1.2 Treatment

There has been little in the way of treatment approaches specifically geared towards adolescent fruit machine gamblers. However, interventions and treatments for compulsive fruit machine playing do not differ significantly from other forms of compulsion or addiction treatment. The real difference is that fruit machine addiction is often harder to spot.

7.2 Compulsive Computer Game Playing

7.2.1 Prevalence

Home computer game playing is now a popular activity among children and adolescents (Griffiths 1993), with approximately one third of children and adolescents playing computer games most days (Griffiths 1997b). However, for a minority of children and adolescents, home computer games can take up considerable time and result in compliance use. Studies by Griffiths and Hunt (1993, 1995) and Parsons (1995) reported that 7% and 9% of schoolchildren, respectively, claimed to spend at least 30 h a week playing video games. The question to ask is what longitudinal effect of an activity that takes up 30 h of leisure time a week has on the educational and social development of children and adolescents. At present we do not know the answer to this, but any child who engaged in any activity excessively every day over a number of years from a young age would have their social and/or educational development negatively affected in some way.

7.2.2 Characteristics

Like fruit machine playing, computer game playing compulsions tend to be prevalent in adolescents. More recently, research has identified behavioural signs of video game compulsions. These have included stealing money to play arcade games (Klein 1984; Keepers 1990), stealing money to buy new games cartridges (Griffiths and Hunt 1993), engaging in minor delinquent acts (Kestenbaum and Weinstein 1985), using lunch money to play (McLure and Mears 1984), truanting from school to play (Keepers 1990; Griffiths and Hunt 1993), not doing homework/getting bad marks at school (Griffiths and Hunt 1993; Phillips et al. 1995), sacrificing social activities to play (Egli and Meyers 1984; Griffiths and Hunt 1993), irritability and annoyance if unable to play (Griffiths and Hunt 1993; Rutkowska and Carlton 1994), playing longer than intended (Griffiths and Hunt 1993; Phillips et al. 1995) and an increase in self-reported levels of aggression (Griffiths and Hunt 1993).

It must be recognised that computer games for some schoolchildren can be potentially compulsive – the underlying process of which may be similar to pathological gambling. For a complete understanding of the issues involved in the development and maintenance of computer game playing and its treatment at excessive levels, considerably more input will be needed from both a clinical and educational perspective.

7.2.3 Treatment

To date there have been very few accounts of treating computer game compulsions. Kuczmierczyk et al. (1987) reported the successful treatment of an 18-year-old college student. Kuczmierczyk and his colleagues assumed that compulsive video game playing was conceptually similar to pathological gambling and used a cognitive-behavioural modification approach in their treatment. Using a combination of self-monitoring, biofeedback assisted relaxation training, in vivo exposure and response prevention, a 90% reduction of playing was observed which persisted at 6- and 12-month follow-ups. Keepers (1990) treated a 12-year-old boy who was brought for psychiatric help by his mother because he was playing video games excessively. He was placed in a residential treatment centre and given family therapy. At 6-month follow-up, no recurrence of the boy's difficulty was noted. Keepers also considered his patient's behaviour to be reminiscent of pathological gambling.

7.3 Compulsive Internet Use

7.3.1 Characteristics

A few surveys have tried to examine the concept of Internet compulsions and addiction (e.g. K. Young, unpublished; Brenner 1997). None of the surveys to date conclusively shows that compulsive Internet use exists. At best, these studies indicate that Internet addiction may be prevalent in a significant minority of individuals, but that more research using validated survey instruments and other techniques (e.g. in-depth qualitative interviews) is required.

Both Griffiths (1996, 1997c) and Young (1996) have described what they consider to be detailed case studies of Internet addicts. The people they describe appeared to suffering from a genuine compulsion that is comparable with other more accepted compulsions and addictions. Taking all the case study and survey evidence together, it can be argued that excessive usage in a majority of cases appears to be purely symptomatic, but that for what appears to be an exceedingly tiny minority, the Internet *may* be compulsive. The case study accounts of Griffiths (1997c) show that the Internet was used to counteract other deficiencies in the person's life (e.g. relationships, lack of friends, physical appearance, disability, coping) and that all of the people studies used the computer for social contact (mostly for Internet Relay Chat services).

Kimberly-Young is founding President of the On-Line Addiction Centre (http://netaddiction.com). She has put forward the following warning signs based upon her own work:

1. Losing track of time once on-line
2. Minimising to others the amount of time spent on-line
3. Anticipation of on-line usage
4. Others complaining about the amount of time spent on-line
5. Others complaining about a large phone bill related to time spent on-line
6. Daily home use of on-line services
7. High phone bills related to on-line service fees and connect time
8. Being logged on to the personal account while at work
9. Poor interpersonal relationships due to increased time spent on-line
10. Sneaking on-line when the significant other is not at home

7.3.2 Treatment

As yet there are no recognised treatments for compulsive Internet use, although at a theoretical level, cognitive-behavioural treatments might be the most effective.

8 Cognitive-Behavioural Treatment of Compulsive Behaviour

There is insufficient evidence to strongly support the effectiveness of any treatment approach in the area of compulsive behaviour (or behavioural addictions). Nevertheless, if it is possible to generalise from obsessive–compulsive behaviours and addictive behaviours, then the following treatment plan would be worth investigating:

1. Motivational interviewing
2. Identifying high-risk situations or cues
3. Developing specific coping responses for the most important cues
4. Developing a general coping strategy to be used in a variety of situations
5. Exposure to cues and rehearsal of coping skills
6. Developing social relationships

The treatment programme outlined above relies upon evidence drawn from research on the treatment of alcohol dependence, compulsive eating and obsessive–compulsive disorder. It also relies upon general cognitive-behavioural theory and evidence of effec-

tiveness. There is a need to test the effectiveness of this approach in the area of compulsive behaviours.

9 References

American Psychiatric Association (1980) Diagnostic and statistical manual of mental disorders, 3rd edn. American Psychiatric Association, Washington, DC

Barham B, Cormell M (1987) Teenage use of amusement arcades in Bognor Regis. WSIHE; Bognor Regis

Blaszczynski AP, Winter SW, McConaghy N (1986) Plasma endorphin levels in pathological gambling. Sixth National Conference on Gambling and Risk Taking (1984, Atlantic City, New Jersey). J Gambling Behav 2(1): 3–14

Brenner V (1997) Psychology of computer use. XLVII. Parameters of Internet use, abuse and addiction: the first 90 days of the Internet Usage Survey. Psychol Rep 80: 879–882

Brown RI (1986) Arousal and sensation seeking components in the general explanation of gambling and gambling addictions. Int J Addict 21(9–10): 1001–1016

Christenson GA, Faber RJ, de Zwaan M, Raymond NC, Specker SM, Ekeren MD, Mackenzie TB, Crosby RD, Crow SJ, Eckert ED, Mussell MP (1994) Compulsive buying: descriptive characteristics and psychiatric comorbidity. J Clin Psychiatry 55(1): 5–11

Coleman E (1991) Compulsive sexual behaviour: new concepts and treatments. J Psychol Hum Sex 4(2): 37–52

Coleman E (1992) Is your patient suffering from compulsive sexual behaviour? Psychiatr Ann 22(6): 320–325

Crisp AH, Hsu LKG, Harding B, Hartshorn J (1980) Clinical features of anorexia nervosa. J Psychosom Res 24: 179–191

Culleton RP (1985) A survey of pathological gamblers in the State of Ohio. Transition Planning Associates, Philadelphia

Custer RL, Custer LF (1981) Soft signs of pathological gambling. Paper presented at the 5th National Conference on Gambling, Reno, Nevada, University of Nevada, May 1981

Damon J (1988) Shopaholics: serious help for addicted spenders. HP Books, New York

Davis C, Fox J (1993) Excessive exercise and weight preoccupation in women. Addict Behav 18(2): 201–211

Davis C, Brewer H, Ratusny D (1993) Behavioural frequency and psychological commitment: necessary concepts in the study of excessive exercising. J Behav Med 16(6): 611–628

de Coverley Veale DM (1987) Exercise dependence. Br J Addict 82(7): 735–740

de la Torre J (1994) Mens sana in corpore sano, or exercise abuse? Clinical considerations. Bull Menninger Clin 59(1): 15–31

de Vries HA (1981) Tranquilizer effect of exercise: a critical review. Phys Sports Med 9(11): 47–53

Dickerson MG (1974) The effect of betting shop experience on gambling behaviour. Ph.D. dissertation, University of Birmingham

Dickerson M, Hinchy J (1988) The prevalence of excessive and pathological gambling in Australia. J Gambling Behav 4: 135–151

Echebura E, Baez C, Fernandez-Montalvo J (1994) Comparative effectiveness of different therapeutic modalities in the psychological treatment of pathological gambling: an experimental study. Anal Modicacion Conducta 20: 617–643

Egli EA, Meyers LS (1984) The role of video game playing in adolescent life: is there a reason to be concerned? Bull Psychonom Soc 22: 309–312

Eisler I, Le Grange D (1990) Excessive exercise and anorexia nervosa. Int J Eating Disord 9: 377–386

Fisher S (1993) Gambling and pathological gambling in adolescents. J Gambling Stud 9: 277–288

Gauthier J, Pellerin D (1982) Management of compulsive shoplifting through covert desensitisation. J Behav Ther Exp Psychiatry 13: 73–75

Glover JH (1985) A case of kleptomania treated by covert sensitization. Br J Clin Psychol 24(3): 213–214

Griffiths MD (1990) The acquisition, development and maintenance of fruit machine gambling in adolescents. J Gambling Stud 6: 193–204

Griffiths MD (1993) Are computer games bad for children? Psychologist Bull Br Psychol Soc 6: 401–407

Griffiths MD (1995a) Adolescent gambling. Routledge, London

Griffiths MD (1995b) Technological addictions. Clin Psychol Forum 76: 14–19

Griffiths MD (1996) Internet "addiction": an issue for clinical psychology? Clin Psychol Forum 97: 32–36

Griffiths MD (1997a) Exercise addiction: a case study. Addict Res 5: 161–168

Griffiths MD (1997b) Video games and children's behaviour. In: Charlton T, David K (eds) Elusive links: television, video games and children's behaviour. Park Published Papers, Cheltenham, pp 66–93

Griffiths MD (1997c) Technological addictions: looking to the future. Paper presented at the 105th Annual Convention of the American Psychological Association, Chicago, Illinois, August 1997

Griffiths MD, Duff J (1993) Etiologies of addictive behaviour: a survey of non-professional peoples' beliefs. Addict Res 1: 199–206

Griffiths MD, Hunt N (1993) The acquisition, development and maintenance of computer game playing in adolescence. Paper presented at the British Psychological Society London Conference, City University, December 1993

Griffiths MD, Hunt N (1995) Computer game playing in adolescence: prevalence and demographic indicators. J Community Appl Soc Psychol 5: 189–194

Gudjonsson GH (1987) The significance of depression in the mechanism of 'compulsive' shoplifting. Med Sci Law 27: 171–176

Hollander E, Frenkel M, Decaria C, Trungold S, Stein DJ (1992) Treatment of pathological gambling with clomipramine. Am J Psychiatry 149(5): 710–711

Huff G, Collinson F (1987) Young offenders, gambling and video game playing. Br J Criminol 27: 401–410

Johnson WG, Corrigan SA, Schlundt DG, Dubbert PM (1990) Dietary restraint and eating behaviour in the natural environment. Addict Behav 15(3): 285–290

Kagan DM, Squires RL (1985) Addictive aspects of physical exercise. J Sports Med Phys Fitness 25(4): 227–237

Kallick M, Suits D, Dielman T, Hybels J (1979) A survey of American gambling attitudes and behavior. Institute for Social Research, University of Michigan, Ann Arbor

Keepers GA (1990) Pathological preoccupation with video games. J Am Acad Child Adolesc Psychiatry 29: 49–50

Kestenbaum GI, Weinstein L (1985) Personality, psychopathology, and developmental issues in male adolescent video game use. J Am Acad Child Psychiatry 24: 325–337

Kety SS (1966) Catecholamines in neuropsychiatric states. Pharmacol Rev 18: 787–798

Klein MH (1984) The bite of Pac-man. J Psychohistory 11: 395–401

Kraft T (1970) A short note on forty patients treated by systematic desensitization. Behav Res Ther 8: 219–220

Kuczmierczyk AR, Walley PB, Calhoun KS (1987) Relaxation training, in vivo exposure and response-prevention in the treatment of compulsive video-game playing. Scand J Behav Ther 16: 185–190

Lesieur HR, Blume SB (1987) The South Oaks Gambling Screen (SOGS): a new instrument for the identification of pathological gamblers. Am J Psychiatry 144(9): 1184–1188

Lesieur HR, Rosenthal RJ (1991) Pathological gambling: a review of the literature. J Gambling Stud 7: 5–39

Lester D (1994) Access to gambling opportunities and compulsive gambling. Int J Addict 29(12): 1611–1616

Long J (1989) Playing the machine: amusement arcade ethics. Leisure Management 9(8): 65–66

Long CG, Smith J, Midgley M, Cassidy T (1993) Over-exercising in anorexic and normal samples: behaviour and attitudes. J Mental Health UK 2(4): 321–327

McElroy SL, Hudson JI, Pope HG, Keck PE (1991) Kleptomania: clinical characteristics and associated psychopathology. Psychol Med 21(1): 93–108

McElroy SL, Keck PE, Pope HG, Smith JMR et al (1994) Compulsive buying: a report of 20 cases. J Clin Psychiatry 55(6): 242–248

McLure RF, Mears FG (1984) Video game players: personality characteristics and demographic variables. Psychol Rep 55: 271–276

Moran E (1970) Gambling as a form of dependence. Br J Addict 64: 419–428

Moran E (1987) Gambling among schoolchildren: the impact of the fruit machine. National Council on Gambling, London

Morgan WP, O'Connor PJ (1988) Exercise and mental health. In: Dishman RK (ed) Exercise adherence: its impact on public health. Human Kinetics, Champaign, pp 91–121

Morris M, Steinberg H, Sykes EA, Salmon P (1990) Effects of temporary withdrawal from regular running. J Psychosom Res 34(5): 493–500

Moskowitz JA (1980) Lithium and lady luck; use of lithium carbonate in compulsive gambling. NY State J Med 80(5): 785–788

Murphy MH (1994) Sport and drugs and runner's high (psychophysiology). In: Kremer J, Scully D (eds) Psychology in sport. Taylor and Francis, London

National Housing and Town Planning Council (1988) Gambling machines and young people. National Housing and Town Planning Council, London

Parsons K (1995) Educational places or terminal cases. Young people and attraction of computer games. Paper presented at the British Sociological Association Annual Conference, April 1995

Phillips CA, Rolls S, Rouse A, Griffiths M (1995) Home video game playing in schoolchildren: a study of incidence and patterns of play. J Adolesc 18: 687–691

Rachman S, Hodgson RJ (1980) Obsessions and compulsions. Prentice Hall, Englewood Cliffs

Royal Commission (1978) Report of the Royal Commission on Gambling. HMSO, London

Russell JC, Epling WF, Pierce D, Amy RM, Boer DP (1987) Induction of voluntary prolonged running by rats. J Appl Physiol 63(6): 2549–2553

Rutkowska JC, Carlton T (1994) Computer games in 12–13 year olds' activities and social networks. Paper presented at the British Psychological Society Annual Conference, April 1994

Sommers I (1988) Pathological gambling: estimating prevalence and group characteristics. Int J Addict 23: 477–490

Spectrum Childrens Trust (1988) Slot machine playing by children: results of a survey in Taunton and Minehead. Spectrum Childrens Trust, London

Travin S (1995) Compulsive sexual behaviours. Department of Psychiatry, Bronx-Lebanon Hospital Center, New York, USA. Psychiatr Clin North Am 18(1): 155–169

Volberg RA, Steadman HJ (1988) Refining prevalence estimates of pathological gambling. Am J Psychiatry 145: 502–505

Volberg RA, Steadman HJ (1989) Prevalence estimates of pathological gambling in New Jersey and Maryland. Am J Psychiatry 146: 1618–1619

Walker MB (1992) The psychology of gambling. Pergamon, Oxford

Wesson C (1991) Women who shop too much: overcoming the urge to splurge. St. Martin's, London

Wolkowitz OM, Roy A, Doran AR (1985) Pathological gambling and other risk-taking pursuits. Psychiatr Clin North Am 8: 311–322

Young K (1996) Psychology of computer use. XL. Addictive use of the internet: a case that breaks the stereotype. Psychol Rep 79: 899–902

CHAPTER

16

J.B. Saunders, T. Sitharthan, P.B. Krabman

An Overview of Substance Use Disorders and Their Management

1 Introduction

The misuse of psychoactive substances is of public health and social concern throughout the world. The commonest substances used are alcohol and tobacco, followed by cannabis, sedative-hypnotic drugs, opioids, psychostimulants and hallucinogens, roughly in that order. Most drugs that are misused have the potential to induce dependence. Many have intrinsic tissue toxicity and give rise to a range of physical and neuropsychiatric complications. Illicit drug use may be complicated by adverse effects of adulterants, and the intravenous route may act as the means of transmission of viral, bacterial and other communicable diseases. Adding further to the burden are the lifestyles associated with drug use, which may involve prostitution, sexually transmitted disease, malnutrition and the legal and penal consequences of crime.

The purpose of this chapter is to provide an overview of the epidemiology of substance use, the behavioural and clinical syndromes associated with it, techniques of diagnosis, and management strategies. In the chapters that follow, detailed information will be given about the syndromes and management approaches applicable to specific substances. The overview presented here will cover the main types of psychoactive substances, with the exception of tobacco. The reason for omitting reference to tobacco is that assessment and management of smoking-related disorders, and the provision of smoking cessation therapies, is usually a function of internal medicine clinics rather than mental health or alcohol and drug services.

2 Epidemiology of Substance Use

2.1 Alcohol

Alcohol is ubiquitous in Western societies, and its consumption has increased steadily in many countries in Asia and Africa over the past 30 years. A single summary measure, the per capita intake, is used to define "high-consumption" countries such as France, Portugal and Hungary (which average 20–35 l per adult per year), "moderate-consumption" countries such as the United States, Canada, United Kingdom, Australia and Germany (7–15 l per year) and "low-consumption" countries such as the Nordic countries, Israel, Turkey, China and India (1–6 l per adult per year).

Per capita alcohol intake has declined in many developed countries over the past 10–15 years, notably in France and Italy, where decreases of around 25% have been recorded. On the other hand, intake has increased substantially in Japan, China and several South East Asian countries. The prevalence of alcohol-related problems is correlated with the overall level of consumption, but there are substantial variations related to the different patterns of use. Specifically, Mediterranean countries, where drinking is often a regular daily activity, have a high incidence of chronic physical complications such as cirrhosis of the liver. Nordic countries have an intermittent pattern of drinking; overall consumption is low, but there is a disproportionately high incidence of trauma and acute physical sequelae.

2.2 Sedative-Hypnotic Drugs

There are numerous proprietary and traditional drugs that have sedative, anti-anxiety and hypnotic properties. Barbiturates, introduced in the late 1950s, were responsible for widespread morbidity due to their high dependence potential and toxicity in overdose. They have largely been supplanted by the benzodiazepines. These drugs were considered to have relatively low dependence potential, but there is now compelling evidence that the risk even at usual therapeutic doses is considerable: 40% of individuals taking a long-acting benzodiazepine for 3 months or more develop an identifiable withdrawal syndrome when it is ceased (Petursson 1994).

2.3 Inhalants

Inhalants are commonly misused by groups that do not have access to the more usual psychoactive substances. They include pre-teenage children and people living in isolated communities. Inhalants include petrol (gasoline), paint thinners, glues and adhesives, nail varnish removers, cleaning fluids and lighter fuels. Users tend to be male adolescents, particularly those who come from impoverished or isolated backgrounds. This group tends to experience significant social and psychological problems (Smart 1986) and is prone to depression (Jacobs and Ghodse 1987). Inhalants are considered "gateway" drugs in Japan (Fukui 1994). About 34% of drug-dependent patients in a mental hospital in Japan had a problem of inhalant abuse, and approximately 20% had started using inhalants before the age of 20 years (Fukui 1994).

2.4 Cannabis

Cannabis is the most widely used illicit drug in developed countries. Its use increased during the 1960s, and although this has declined in some countries, the majority of young people would have some experience with it. Frequent use is less common, its dependence potential being relatively low, and in developed countries only about 3% of the 18- to 40-year-old population are regular daily users. Reflecting the sporadic nature of its use among the majority, physical sequelae are comparatively uncommon. More potent forms of cannabis, with higher concentrations of the most active cannabinoid, Δ-9-tetrahydrocannabinol (THC), are becoming available. Cannabis-related psychosis appears to be increasing in prevalence as a result.

2.5 Opiates

Results of population surveys in the USA and Europe indicate a lifetime prevalence of heroin use of 1%–2%. In several countries this has increased over the past decade. This percentage rises to 4%–6% among young adults in the 18- to 30-year age-group in certain cities in the USA and Southern Europe. Overall, 0.3%–0.6% of adults in North America and Europe are dependent on heroin, with the highest rates being found in the USA, Italy, Spain, Switzerland and the UK (European Monitoring Centre for Drugs and Drug Addictions 1999). In Europe, deaths from heroin use, principally from overdose, have doubled over the past 15 years. Thus, while heroin use is relatively rare in the general population, it is associated with disproportionate costs in terms of drug-related problems and deaths. It is estimated that 25% of opiate-dependent patients are likely to die within 10–20 years of active drug misuse. Causes of death include unintentional overdose, suicide, homicide, infectious diseases, and toxic reactions to the drug or adulterants.

2.6 Psychostimulants

Three groups of psychostimulants are commonly misused: (1) amphetamine and its derivatives, (2) cocaine and (3) prescribed stimulants. Amphetamines were marketed at the beginning of the twentieth century as anti-depressants and appetite suppressants. They have been used by troops in wartime, by long-distance drivers and students, but increasingly are used by economically deprived young people. Amphetamines have become one of the most popular types of illicit drug in many countries in Europe, Australia and Japan. In recent surveys in Australia, 15% of young males admitted to using amphetamines, 3% to cocaine and 10% to Ecstasy (Department of Community Services and Health 1999). Ecstasy (methylenedioxymethylamphetamine) is one of several amphetamine derivatives that have been popularised as a teenage drug. It has both hallucinogenic and stimulant properties.

Cocaine has been used by the indigenous peoples of South America for centuries. In Europe, it was widely used in the early twentieth century and was readily available in proprietary tonics and medicines until the mid-1920s. In the 1960s, there was a wave of intravenous cocaine use, and in the late 1970s the free-base form of cocaine became commonly available, especially in the USA. Its use spread from the elite to the economically deprived and ethnic minorities in inner city areas. The number of people in the USA who have ever used cocaine increased five-fold from 5 million in 1975 to 25 million in 1985, by which time 3 million were considered cocaine dependent (Kleber 1988).

Prescription stimulants are abused predominantly by young women, often in a quest to maintain a low body weight. They include methylphenidate and fenfluramine. Caffeine and related compounds are contained in many brands of compound analgesic.

2.7 Hallucinogens

Thousands of naturally occurring hallucinogens have been identified. They occur in various species of mushroom and nuts, most commonly. Synthetic drugs such as lysergic acid diethylamide (LSD) became widely used in the 1960s and were an important aspect of youth drug culture. Less widespread now, there remain pockets of LSD use in many developed countries.

2.8 Polydrug Abuse

Polydrug misuse has become increasingly common, particularly among lower socio-economic groups, and it is estimated that 5% of adults use two or more illicit drugs. Concurrent alcohol, tobacco and/or benzodiazepine use add to the adverse consequences to health. The more chaotic drug use is, the worse the prognosis.

3 Predisposing Factors

The reasons for substance use are complex and interconnected. The reinforcing properties of the drug, genetic factors, cognitive factors, faulty learning, personality characteristics, availability of the substance, peer pressure, cultural factors, media influence and religious beliefs are all important in influencing why an individual initiates substance use, and why others do not. Likewise, the pharmacological properties of the substance, personality and social characteristics of the user and the setting in which the substance is used are important in influencing continuing use. Why people decide to stop and then re-engage is also determined by a host of factors and cannot be reduced to a single reason or cause.

It is useful to conceptualise substance use as being determined by three interacting forces: (1) the pharmacological properties of the drug, (2) the characteristics of the individual and (3) the influence of the environment. This conceptualisation has analogies to the bio-psychosocial model of mental and behavioural problems.

3.1 Nature of the Drug

Substances with abuse potential have two predominant characteristics. They have pleasurable psychic effects, and they are (nearly all) rapidly acting. The abuse potential is related both to the drug's intrinsic properties and to the mode of administration. Drugs that are smoked or injected intravenously are more reinforcing than those taken by mouth (Strang et al. 1998). There is no common mechanism of action. Some substances such as alcohol and barbiturates are relatively non-specific in their actions, having general central nervous system depressant effects. Others such as opiates interact with specific receptors. One possible "final common pathway" for the pleasurable and reinforcing effects is the dopaminergic system in the mid-brain. Following repeated exposure, a state of neuroadaptation develops to many substances. This may reflect downregulation of specific receptors or changes in secondary messenger systems. Neuroadaptation leads to tolerance and, in the absence of the drug, a withdrawal syndrome.

3.2 Individual Factors

There are numerous factors which influence individual susceptibility to developing substance use disorders. There is increasing evidence that genetic factors are important, particularly in relation to alcohol use and dependence, where perhaps 50% of the variance in susceptibility is explained on genetic grounds. The mode of inheritance seems to be polygenic, but with dopamine D_2 receptor polymorphism exerting a dominant influence (Noble 1999). Alcohol dependence is genetically linked to certain personality characteristics such as impulsiveness and novelty seeking. However, there is no evidence of a predisposing "drug-dependent personality" as was once thought, nor of a "depressive spectrum disorder" predisposing to alcohol dependence. Childhood conduct disorder and attention deficit hyperactivity disorder appear to be risk factors for later substance use disorders. Physical and sexual abuse are increasingly recognised as predisposing to the more severe dependencies. Individuals with antisocial personality disorder or post-traumatic stress disorder are more likely than the general population to experience problems with substance abuse. In general, where use of a substance is uncommon in the general population, individual predisposing factors are more likely to be identified than when a substance is commonly used.

3.3 Environmental Factors

The prevalence of alcohol and drug use varies widely from country to country, an observation which is largely explained by sanctions and constraints placed on substance use, the role of the substance in society or in particular subcultures, economic opportunity and its physical availability. When a substance is illegal and/or costly, its use tends to be lower than when it is widely available and cheap, though this may be at the expense of greater social damage when it is used. The use of legal substances, such as alcohol and tobacco, is responsive to promotional campaigns. Correspondingly, when restrictions are placed on their use (e.g. by legislating on the minimal legal age, random breath testing), levels of use and associated problems decline.

4 Core Syndromes

4.1 Spectrum of Use and Misuse

Patterns of substance use range from an occasional experimental or recreational exposure to continuous heavy daily consumption. A spectrum is observed not only in levels of use, but also in the psychological and

behavioural accompaniments of substance use, and in the prevalence and nature of substance-related problems. Diagnostic systems reflect this spectrum, defining a range of substance use disorders. The historical development of diagnostic schemata has been summarised elsewhere (see e.g. Babor 1992; Schuckit 1996; Krabman and Saunders 1996). ICD-10 and DSM-IV, the two current international classification systems for psychiatric disorders, both contain a separate section for substance use disorders. They agree in the main with regard to the diagnosis of dependence, but their approach to non-dependent use, termed "abuse" in DSM-IV and "harmful use" in ICD-10, differs significantly.

4.2
Non-dependent Use

Although drug and alcohol services tend to cater for those with established dependence, it is necessary to acknowledge that, at a population level, the majority of morbidity and mortality results from non-dependent substance use. Non-dependent use infers a level or pattern of repeated substance use that does not meet the criteria for dependence yet leads to recurrent and significant substance-related problems or confers the risk of harm. It encompasses hazardous use, harmful use (an ICD-10 diagnosis), and abuse (DSM-IV). Identifying non-dependent problem users is important, as their substance use can often be influenced through interventions much less costly and more effective than the treatment of dependence (Saunders 1989; Bien et al. 1993; WHO Brief Intervention Study Group 1996).

4.3
Hazardous Use

Hazardous or "at-risk" alcohol and drug use is defined as a repetitive pattern of use that significantly increases the risk of harmful physical and psychological consequences. As such, it is significant in terms of preventive medicine and public health, although it is not included in ICD-10 of DSM-IV. It is employed by the World Health Organization (WHO) and by some national authorities in public education campaigns. For example, hazardous alcohol consumption is defined by the Australian National Health and Medical Research Council (Hawks and Pols 1992) in terms of regular daily intake (more than 40 g alcohol for men and 20 g for women) and/or the amount consumed in any one occasion (more than 60 g alcohol for men and 40 g for women). Hazardous use can be extended beyond measures of consumption to include such at-risk behaviours as sharing intravenous needles, binge use with intermittent severe intoxication and use in unsafe settings such as while driving or operating machinery.

The importance of this diagnosis is that it identifies a substantial group in the general population which is at increased risk of experiencing substance-related social and medical problems and of progressing to harmful use/abuse or dependence (Conigrave et al. 1995a). Brief interventions providing information and advice have been shown to be effective in this group in reducing hazardous consumption and adverse consequences (Bien et al. 1993; WHO Brief Intervention Study Group 1996).

4.4
Harmful Use

Harmful use is defined in ICD-10 as a pattern of substance use that is actually causing physical or psychological harm to the individual. No specific level or pattern of substance use is predicated, although it is widely assumed that it would also fulfil the criteria for hazardous use.

4.5
Substance Abuse

DSM-IV takes a different approach to the issue of non-dependent use, focusing on social and interpersonal problems as the core of the concept of substance abuse. Failure in role obligations, and recurrent legal, social or interpersonal problems are listed as diagnostic criteria. These problems are prevalent in a wide diversity of cultures (Helzer and Canino 1989), and continued use despite them does demonstrate a dysfunctional pattern of use. Abuse can thus be defined as a significant failure to conform to the prevailing social standards with regard to substance use, even though those standards may vary with culture, gender and generation.

4.6
Dependence

Dependence is the diagnostic concept at the heart of both ICD-10 and DSM-IV. It was characterised in 1976 by Edwards and Gross, who described a "dependence syndrome" in relation to alcohol use. This concept has since been applied to a wide variety of psychoactive substances (Edwards 1986) and has largely replaced earlier formulations of alcoholism (see e.g. Jellinek 1960) and drug addiction. The syndrome encompasses a series of cognitive, behavioural and physiological

changes, the central elements of which are an increasing priority given to substance use over other behaviours, a strong desire (or "craving") to take the substance, impaired control over its use, tolerance, withdrawal symptoms (in some cases) and continued use despite harmful consequences. No assumptions are made about causality. In both diagnostic systems, the presence of three (out of six or seven) criteria is required for the diagnosis of dependence. Dependence is partially correlated with, but conceptually distinct from, the level of use and severity of problems stemming from substance use.

A diagnosis of dependence can be applied to all classes of psychoactive substance. This is especially relevant as increasingly individuals are using more than one drug, either concurrently or consecutively. The same criteria apply to all substances, but some will be less relevant to some classes or, rarely, not applicable at all (e.g. withdrawal symptoms are not specified for hallucinogens).

Both ICD-10 and DSM-IV allow a diagnosis of dependence without a requirement for the presence of the physiological phenomena of tolerance, withdrawal and relief of withdrawal by substance use. This broader view is especially relevant to drugs such as cocaine, which do not have prominent physical withdrawal phenomena. Some evidence suggests that the more psychological aspects of dependence often develop prior to the onset of significant withdrawal symptoms (Chick and Duffy 1979), so such an approach may be more sensitive to detecting moderate or early dependence. However, the physical criteria cluster most strongly together in factor analytic studies, and most strongly predict outcome. Both the DSM-IV and ICD-10 include the diagnostic specifier "with or without physiological dependence" to allow a differentiation to be made on this issue. Among treatment populations, dependence with physiological criteria appears to be more severe, and to be associated with more problems and a greater likelihood of relapse, than dependence without

Clinical Features of Depenence

A preoccupation with the substance and its use is a central element of the dependence syndrome. A great deal of thought and energy is expended on obtaining and using the substance, but also conversely on efforts to control and limit its use. A proportion of users will never attempt to limit their use, but the majority will have repeated unsuccessful attempts to stop or cut down on their consumption. Many report a strong urge or sense of compulsion to use (often termed a "craving") in an uncontrolled way after an initial "priming dose". Craving may be heightened by situational cues such as the smell or sight of alcohol or the drug in question, or proximity to people or places associated with its consumption. This dimension can be assessed by asking questions about a perceived compulsion to use. It can also be measured by observing behaviour, i.e. frequent failure to control substance use or failure to keep to set limits.

As dependence develops, the individual's life becomes increasingly centred on the substance, over and above other interests and role obligations. This is termed the "salience" of substance use. A large amount of time and mental energy is spent obtaining and using the substance at the expense of activities that were once considered important. Dependent individuals will often be able to list a number of past hobbies or interests that they no longer pursue due to their focus on substance use. Evidence of the high priority given to substance use is often demonstrated by continued use despite harmful consequences. Although the dependent individual is aware that their substance use is causing serious medical, family, occupational, financial or legal problems, they continue to use, thus worsening these problems.

When tolerance develops, a greater dose of the substance is required to achieve an equivalent effect, and there is a diminished effect from a given dose. Tolerance is seen in particular with alcohol, opiates and stimulants, where dependent users may maintain near-normal function after a dose that would be fatal to the naive subject. A withdrawal state develops when the dependent individual ceases or reduces substance use, and a cluster of cognitive, behavioural and physiological symptoms are experienced that are usually the phenomenological opposite of the drug's effect in intoxication. The desire to reduce or avoid withdrawal symptoms is seen by some as a key contributor to ongoing dependent substance use. Frequently, the dependent individual will take the substance or another related substance to avoid or alleviate withdrawal symptoms. Some manage to avoid withdrawal symptoms altogether by maintaining continuous high blood levels of the substance. Tolerance and withdrawal are primarily manifestations of a neuroadaptive state; however, they are also modified by environmental factors and conditioning. Previously drug-dependent individuals can experience withdrawal phenomena on returning to an environment associated with drug use (Grant et al. 1990).

these features (Langenbucher et al. 1997; Schuckit et al. 1998).

In their initial description of the dependence syndrome, Edwards and Gross (1976) described the rapid reinstatement of a dependent pattern of use, once a previously dependent individual recommences substance use after a period of abstinence. This phenomenon has been observed in animals and in humans (Edwards 1990). Follow-up studies confirm that return to controlled use in a previously dependent individual is uncommon (Orford et al. 1976; Cohen et al. 1989). Thus, for patients who are dependent to the extent that they experience recurrent withdrawal symptoms, it is advisable to recommend a goal of abstinence.

Lay groups such as Alcoholics Anonymous (AA) and equivalent 12-step programmes for drug dependence teach that life-long abstinence is the only means of overcoming dependence and that any use will inexorably lead to a loss of control and a return to old patterns of consumption. Controversy exists over the extent to which the ubiquity of these teachings has contributed to the self-fulfilling prophecy of previously dependent users reverting to uncontrolled use after treatment. This assertion is supported by the finding that outcome is linked as strongly to beliefs about ability to control drinking as to severity of dependence prior to treatment (Heather et al. 1983).

4.7 Withdrawal State

The "withdrawal state" or withdrawal syndrome is a central element of dependence, but it is also a diagnosis in its own right in ICD-10 and DSM-IV, reflecting the fact that it is often the presenting diagnosis when people attend for treatment. The nature of the withdrawal syndrome varies according to the pharmacological class of the substance to which dependence has developed. Thus central nervous system depressants such as alcohol and sedative-hypnotics have a withdrawal syndrome characterised by tremor, agitation, hallucinations and seizures, while withdrawal from stimulant drugs involves lethargy and inertia. Withdrawal is often accompanied by psychological symptoms such as anxiety, depression and sleep disturbance.

The onset and duration of withdrawal symptoms is related to the duration of action of the substance. Withdrawal from substances with shorter duration of effect such as alcohol or opiates will usually commence within 48 h and have settled within a week. Withdrawal from substances with longer half-lives such as diazepam or methadone may not commence for several days to weeks and can continue for 6–8 weeks. Predictors of increased severity of withdrawal include sudden cessation of use, high levels of consumption prior to cessation, previous severe withdrawal symptoms, concurrent medical disorders and metabolic disturbances. Withdrawal may be complicated by seizures, delirium or depression.

5 Complications of Substance Misuse

Substance misuse may have both acute and chronic adverse effects on the user. These may be direct consequences of use, such as alcoholic liver disease or human immunodeficiency virus (HIV) infection acquired through sharing contaminated injecting equipment, or indirect effects such as the medical consequences of malnutrition due to loss of income, or depression and loneliness related to family separation. Many adverse effects of substance use occur in non-dependent users.

5.1 Alcohol

Misuse of alcohol can lead to a vast array of physical, psychological and social complications. The range of physical disorders is particularly broad, reflecting the fact that alcohol possesses much greater intrinsic tissue toxicity than most psychoactive substances. Table 1 lists some of the recognised complications of harmful alcohol use and dependence.

5.2 Sedative-Hypnotic Drugs

By contrast, use of sedative-hypnotics is associated with little physical pathology. Serious morbidity and death is usually related to intentional overdose, often in combination with alcohol or other central nervous system depressants. Chronic use may produce cognitive impairment, although this is substantially reversed following prolonged abstinence.

5.3 Inhalants

Chronic inhalant abuse carries an increased risk of cognitive impairment (Zur and Yule 1990).

Table 1. Alcohol-related problems

Nature of problem	Type	Problem
Physical	Intoxication	Stupor Ataxia Coma Respiratory depression Inhalation of vomit and asphyxiation
	Liver disease	Fatty liver Alcoholic hepatitis Cirrhosis
	Gastrointestinal diseases	Carcinoma of oropharynx Reflux oesophagitis Gastritis Acute pancreatitis Chronic pancreatitis
	Brain damage	Wernicke-Korsakoff syndrome "Frontal lobe" syndrome Cerebellar disease Subdural haematoma Hypertensive stroke
	Other neurological diseases	Optic neuropathy Peripheral neuropathy Autonomic neuropathy
	Cardiovascular diseases	Cardiomyopathy Alcohol-induced arrhythmias Hypertension
	Respiratory diseases	Aspiration pneunomia Lobar pneumonia Pulmonary tuberculosis
	Endocrine and metabolic disorders	Hypothalamic–pituitary dysfunction Testicular atrophy Hypoglycaemia Pseudo-Cushing's syndrome Gout
	Musculoskeletal disorders	Acute rhabdomyolysis Chronic myopathy Osteoporosis Avascular necrosis of hip Gout
	Trauma	Head injury Fractures of rib Fractures of long bones Multiple trauma (e.g. from road accidents)
	Fetus	Fetal alcohol syndrome Low birth weight Second-trimester abortion
Psychological	Withdrawal state	Anxiety and agitation Malaise and lethargy Tremor Confusion Disorientation Hallucinations Paranoia
	Psychiatric disorders	Anxiety states Phobic disorders Panic attacks Depression Amnesic episodes Acute alcoholic hallucinosis Chronic delusional psychosis ("alcoholic paranoia") Deterioration in moral and ethical standards Self-centredness Suicidal ideation Suicidal attempt
Social	Domestic and allied problems	Loss of friends Deterioration in marital and other significant relationships Psychological symptoms in spouse, e.g. stress anxiety, tension Domestic arguments Domestic violence Neglect of children Separation Divorce
Social	Occupational	Absenteeism Poor work performance Unexplained absences Demotion Dismissal Unemployment
	Financial problems	Loss of regular income Hardship due to money spent on alcohol Gambling debts Victim of fraud
	Legal problems	Drink-driving offences Property crime Assault Homicide Prostitution

Lead encephalopathy is a serious complication of the use of petrol that has lead added to improve combustion.

5.4 Cannabis

The major health risks of acute cannabis use are impairment in psychomotor performance and psychosis. Long-term use may produce subtle cognitive impairment of memory, attention and organisation and integration of complex information. Evidence suggests that the longer the period that cannabis is used, the more pronounced the cognitive impairment is. While subtle, these impairments may affect everyday functioning, particularly in adolescents with marginal educational aptitude and among adults in occupations that require high levels of cognitive capacity. Repeated high-level cannabis use may be followed by the development of a psychosis (Hall 1998). Other recognised adverse effects include respiratory diseases associated with smoking such as chronic bronchitis. There is a concern that cannabis use may be a risk factor for lung and oesophageal cancers. Subfertility in cannabis users has been described.

5.5 Opiates and the Complications of Injecting Use

Heroin and the other opiates possess relatively little intrinsic tissue toxicity. Most morbidity and mortality results from unintentional overdose or from the injecting mode of drug use.

Sharing injecting equipment results in rapid transmission of disease, particularly HIV, hepatitis B (HBV) and hepatitis C (HCV). Drug use also leads to high-risk behaviour and the likelihood of contracting these diseases through sexual and other contacts. Insoluble additives to the drug (such as talc and starch) may cause granulomas in the lungs and thromboembolism. Bacterial, viral and fungal infections result from the injection site being a portal for entry of micro-organisms. Much effort in recent years has been directed to reducing communicable diseases among injecting drug users.

Bacterial infections may present as local infections, such as cellulitis, skin abscesses and thrombophlebitis, or systemic ones, such as septicaemia, bacterial endocarditis, thromboembolic disease and systemic abscesses in the brain, lung spleen and liver. Osteomyelitis and septic arthritis are also well-known complications. The emergence of multi-drug-resistant tuberculosis among HIV-infected drug users is of concern.

The prevalence of HIV infection among injecting drug users shows striking geographical differences not only between countries, but also within national boundaries and even within areas in the same city. In Europe and the USA, injecting drug users make up the second most important risk group for HIV/acquired immunodeficiency syndrome (AIDS) infection, with approximately one third of AIDS patients reporting injecting drug use as a risk factor. Among injecting drug users, HIV prevalence rates range between virtually zero and 48% among different states in the USA and from less than 4% in London to more than 60% in Edinburgh. The initially high (64%) HIV prevalence rate reported in Edinburgh between 1980 and 1985 has since fallen to 15% among more recent cohorts of injecting drug users. The introduction and expansion of drug use in developing countries is often followed by outbreaks of HIV infection. In Burma, Thailand and Manipur (India), the HIV epidemic has spread rapidly among injecting drug users, with prevalence rates of 30%–90% being described (Poshyachinda 1993).

The prevalence rates for HCV in the general population vary from 0.2% to 2.6%, the highest prevalence rates being in Asia, the Middle East, Southern Europe and South America. Injecting drug use (specifically sharing contaminated injecting equipment) is the single most important risk factor for HCV. High prevalence rates of anti-HCV seropositivity (68%–96%) have been reported among injecting drug users worldwide. Transmission is rapid, and 70%–80% of injecting drug users become anti-HCV positive within 1–2 years of use (Bell et al. 1990). Evidence of past or present HBV is seen in up to 80% of injecting drug users. In most there is complete recovery, but the infection remains active in approximately 5%. Regular alcohol consumption by patients with HCV leads to more rapid progression of the liver disease (Saunders and Devereaux, in press). HCV and alcohol appear to have important synergistic effects in alcohol-related liver damage and progression to cirrhosis and hepatocellular carcinoma.

5.6 Psychostimulants

Recurrent high-level use of amphetamine, cocaine and prescribed psychostimulants may lead to acute or chronic psychosis, even in the absence of pre-morbid vulnerability. Auditory and visual hallucinations, prominent paranoid ideation, thought disorder and hypomania are characteristic features. These drugs are also vasoconstrictors, and cocaine in particular is a recognised cause of acute myocardial infarction and

ventricular arrhythmias in young individuals in the absence of underlying cardiac disease. It can also cause vasculitis, infarcts, ischaemic limbs, thrombosis, thromboembolism and stroke. Other complications of cocaine use include bowel infarction and rhabdomyolysis, which may progress to acute renal failure if not promptly diagnosed and treated. Sniffing of cocaine free base may result in ulceration and perforation of the nasal septum.

6
Principles of Diagnosis

6.1
Substance Use Disorders Frequently Underdiagnosed

Substance use disorders are commonly underdiagnosed in both general medical and psychiatric populations. This is true of in-patient and out-patient settings. A missed diagnosis may lead to insufficient or inappropriate treatment, and thus to poorer outcomes. Likely contributory factors are the subtlety of presentation of drug problems, the complex interactions with other psychiatric disorders, failure of the patient to admit to the problem due to denial or lack of insight, and a low index of suspicion on the part of clinicians.

6.2
Subtlety of Presentation

One reason for a failure to recognise problems related to substance misuse is that these problems often present in a form that mimics other conditions. This is true in physical medicine, where excessive alcohol consumption is often the underlying cause for a number of common conditions such as hypertension, obesity, hyperlipidaemia or gout. An individual in acute intoxication or withdrawal may present with nausea and vomiting, diarrhoea, cardiac arrhythmia or seizures.

In psychiatric practice, drug intoxication and withdrawal can present as delirium, psychosis, sleep disturbances, anxiety, depression or agitation. Chronic substance misuse may lead to depression, dementia, phobias, persistent paranoia and psychosis. The only pointer to substance use may be an unexplained failure of social functioning, such as occupational problems in a skilled and previously successful businessman or failure to cope at home in an elderly person. Hidden substance misuse is a common cause of unexplained treatment failure, especially the pattern of repeated sudden deteriorations on discharge from in-patient care, followed by rapid improvement with re-admission.

6.3
Dual Diagnosis

Studies of in-patient and out-patient psychiatric treatment populations show a prevalence of a co-morbid substance misuse diagnosis of 30%–50% (Drake and Wallach 1989; Siegfried 1998). Substance abuse may precipitate or induce another psychiatric disorder or may be secondary to another disorder, or both may be the result of a common genetic, psychological or social pathway (Minkoff 1994). Cause and effect may be difficult to define and are often of little relevance to practical treatment issues. What is important is that substance misuse be considered as a possible cause of acute psychiatric symptomatology even in patients with a known chronic psychiatric illness, and that substance misuse is properly assessed and treated even when it appears to be "secondary" to another illness. Dual diagnosis is associated with poorer prognosis for both the substance misuse disorder and the other psychiatric disorder.

6.4
Validity of Self-Reported Substance Use

Another factor contributing to the under-diagnosis of substance misuse problems is that the individual may not perceive their substance misuse as a problem in situations where their level of use, although hazardous, is commensurate with that of their peers. The high prevalence of denial in substance misusers leads to a lack of awareness of the negative physical, social and psychological consequences of their use and contributes to under-reporting of such problems. Under reporting is also common in settings where excess consumption is socially unacceptable (e.g. alcohol consumption in Indian women) or illegal. This contrasts with clients who have self-referred for substance abuse treatment, among whom self-report is much more valid and reliable.

6.5
Practical Relevance

The end result of under-diagnosis of substance misuse disorders is inappropriate or inadequate assessment

and treatment. Symptoms caused by substance misuse can be attributed to other illnesses and over-treated. Ongoing untreated substance misuse can lead to non-compliance, disruption of therapeutic milieus and dangerous interactions between the substance and medications (Schwartz et al. 1993).

6.6 Techniques of Diagnosis

The likelihood and accuracy of diagnosis can be improved by appropriate training of clinical staff and a high index of suspicion based on knowledge of the incidence of substance use problems and the subtlety of their presentation. A full review of substance use should be included as a routine part of initial assessment or admission procedures. Brief screening questionnaires such as CAGE (Mayfield et al. 1974), the Munich Alcoholism Test (MALT; Feuerlein et al. 1979) or Alcohol Use Disorders Identification Test (AUDIT; Saunders et al. 1993) for alcohol use and the Severity of Dependence Scale (Gossop et al. 1995) and Drug Check (Kavanagh et al. 1999) for drug use help to identify likely substance misusers for more detailed review.

Knowledge of the local spectrum of substance use is important in terms of substances used, methods of administration and the social and physical environments in which use occurs. Accurate assessment is facilitated by knowledge of the local vocabulary for describing drug types and quantities. Where substance misuse is suspected, all avenues of inquiry should be used to assess the situation.

Information gained from these inquiries should be fed back to the patient in order to reinforce the negative effects of their substance use or to challenge denial of misuse in a situation where it is strongly suspected.

Assessment should include a specific substance use history:

1. Lifetime history:
   - Age at first use of substance
   - Age of first intoxication
   - Age of first substance-related problem
   - Periods of heaviest consumption in the past
2. Current history:
   - Levels and patterns of consumption
   - Date/time/quantity of last use
3. Aspects of dependence (current and previous)
4. Physical, psychiatric, social or legal problems
5. Family history of substance use or psychiatric disorder
6. Attitudes to substance use: reasons why the patient uses the substance and the negative effects (if any) the patient attribute to his or her substance abuse
7. Perception of a substance abuse problem: what does he or she see as the solution to this problem?
8. Previous help-seeking/treatment, attempts to cut down/cease substance use, out-patient or in-patient treatment, involvement with 12-step groups.

Appropriate interview techniques will improve the validity of self-report in patients who wish to hide or minimise their problem. The topic should be approached in a casual, non-judgemental manner. The "top-high" technique involves suggesting an extremely high intake, giving "permission" for the patient to admit to a high intake ("How many bottles of beer do you drink per day? Thirty?" "No! Fifteen at the most".). Where substance use is suspected, questioning techniques can place the onus of denial on to the patient ("How much marijuana have you smoked today?", not "Do you ever smoke marijuana?"). The clinician can then move from less threatening to more threatening subject areas, i.e. from legal to illegal substance use or from family history to past to current use. It is admissible to ask about the presence of a problem before it is related to substance use ("Do you get anxious and shaky sometimes?", then "Is that worse at a particular time of day?" and then "Is that worse the morning after you have been drinking?", not "Does drinking cause you to get anxious and shaky next morning?"). Questions on dependence are important in view of its prognostic and treatment implications.

General psychiatric, medical and social history provides useful information on the presence, severity and time course of any neuropsychiatric, physical or social complications of substance misuse. A working knowledge of these complications (discussed elsewhere in this chapter) is essential, as they may often be the only pointers to the presence of a substance misuse problem. Once again, the interview techniques discussed above can improve the validity of self-report.

Corroborative history can be obtained from relatives, friends, other treating professionals, such as the local doctor, and from old notes. Such information should be considered in the light of the overall clinical picture and not regarded as a "gold standard" by which to judge self-report. Understatement of substance use by collateral sources can occur due to co-dependence, fear of retribution or fear of negative social or legal consequences if use is revealed. Overstatement can result from false assumptions based on previous behaviour or hearsay or from a desire for retribution.

Physical and mental state examination may reveal evidence of recent drug use (smell of alcohol on breath,

signs of intravenous needle use), an intoxication or withdrawal syndrome, or of the acute or chronic complications of substance misuse. Substance misuse is a common cause of a general decline in global functioning, evidenced by a poor state of dress, general health, nutrition and hygiene. Neuropsychiatric testing may reveal patterns of abnormality consistent with the known effects of substance use, such as that of alcohol-related brain damage.

Laboratory tests may provide some evidence of substance misuse. In the acute situation, the presence and amount of most substances can be detected by urine or serum tests or (in the case of alcohol) breath tests. As most are cleared rapidly from the body, however, tests will be negative if no recent use has occurred. Exceptions are THC, which is stored in fat and may be detectable in urine and serum for several weeks after cessation of marijuana use, and long-acting benzodiazepines such as diazepam, which may be detectable for several days. Abnormal values for mean corpuscular volume (MCV) and γ-glutamyltransferase (GGT), both widely available tests, may indicate excessive alcohol consumption over the preceding weeks. Both have good specificity but poor sensitivity. More recently, carbohydrate-deficient transferrin (CDT) has been identified as a sensitive and highly specific marker of excessive alcohol consumption (Conigrave et al. 1995b).

If substance misuse is suspected, observation over time during treatment can provide additional information. Admission to hospital of a substance misuser may be followed by rapid resolution of symptoms, onset of a withdrawal syndrome, drug-seeking behaviour or signs of intoxication after ward leave. Observed tolerance to benzodiazepines or opiates may indicate alcohol/sedative or opiate abuse, respectively.

7 Management

Treating alcohol and drug problems is a major challenge for mental health professionals. In general, the prognosis and response to treatment is better when the problem is identified at an early stage. Correspondingly, the prognosis is worst and response to treatment least likely when dependence has become well established, the associated lifestyle entrenched, and physical and other complications severe.

When the disorder has reached the stage of dependence, it has a chronic, relapsing course. It is not likely to respond in the long term to a single session of treatment, no matter how intensive. Relapses following treatment are the rule rather than the exception. About two thirds of the patients attending treatment programmes for addictive disorders relapse within 3 months following treatment. Similarly, half of opiate users relapse within 6 months after completing in-patient treatment (Gossop et al. 1989), and 97% relapse in 1 year (Vaillant 1988).

Interventions may be classified into those that aim to assist patients in achieving abstinence and maintaining a relapse-free lifestyle, and those that respond to the statistical likelihood of relapse by aiming for stable maintenance on a substitute drug or continued use with a minimum of harmful effects (a "harm-reduction" philosophy). Although total abstinence from all psychoactive substances may be the ideal goal, it is achievable in only a minority of substance users (Seivewright and Greenwood 1996). Hence, it is crucial for mental health and other health professionals to embrace and promote an approach which recognises that there are diverse treatment goals, relevant to different individuals and treatment settings. Interim goals may differ from long-term ones. The multiplicity of non-abstinence goals is encapsulated in the harm-reduction approach (Des Jarlais 1995; O'Hare et al. 1992). In the following section, we will briefly review various interventions for alcohol, opiates and psychostimulants. We shall focus on those appropriate for ambulatory care and post-detoxification settings.

7.1 Alcohol

No one treatment approach is effective for all individuals with alcohol problems (Institute of Medicine 1990). A major reorientation of treatment has come about with the evidence that intensive in-patient treatments are, at best, only slightly more effective than briefer out-patient treatments (Annis 1986; Keso and Salaspuro 1990). Several important factors should be considered in advising on the form of treatment. They include the severity of alcohol dependence and the severity of alcohol-related problems, availability and quality of social support, strength of coping skills, patients' readiness to change, severity of cognitive impairment and personal choice of treatment goals. However, assignment to treatment is not a precise science (Project MATCH Study Group 1997), and the availability of different treatments and the skills, experience and enthusiasm of a therapist in a particular approach are also important factors.

Reviews of the effectiveness and cost-effectiveness of treatment of alcohol problems indicate that many widely used treatments have yet to demonstrate their effectiveness convincingly and/or are very expensive to

deliver (Finney and Monahan 1996; Holder et al. 1991; Saunders 1989). Those for which there is acceptable evidence of effectiveness include brief interventions, cognitive-behaviour therapy (CBT), cue exposure treatment, pharmacotherapies such as acamprosate and naltrexone, social skills training, behavioural marital therapy and stress management training. Other techniques are beneficial but are too expensive for widespread use. Some, such as AA and other 12-step approaches, are widely practised, but are not well founded on empirical evidence.

Brief Interventions

Brief Interventions for alcohol problems are usually between one and three sessions, which vary from 5 to 30 min. Hazardous and harmful drinkers who are not dependent on alcohol generally benefit from this approach (Babor 1994; Bien et al. 1993; WHO Brief Intervention Study Group 1996). Typically, following a standardised brief assessment, e.g. AUDIT (Saunders et al. 1993), patients are provided feedback regarding excessive alcohol use and related problems and, through reflective listening, guided to weigh the pros and cons of continued use versus change. Self-monitoring and simple cognitive and behavioural tips to reduce drinking are generally preferred and reportedly used by patients (Sitharthan et al. 1996).

Cognitive-Behaviour Therapy

CBT involves functional analysis of drinking behaviour, goal setting, self-monitoring and self-appraisal of progress. Patients are trained to identify high-risk situations and apply problem-solving coping skills to avoid the problem situation or drink less in such situations. In addition, patients are encouraged to gradually engage in different problem situations of graded difficulty to enhance their self-efficacy (Miller et al. 1992; Sitharthan and Kavanagh 1990). Most importantly, patients are trained to cope with setbacks by viewing "slips" as learning experiences to prevent further relapses from occurring (Marlatt and Gordon 1985). Although CBT is generally applied to moderating drinking, similar strategies can be applied to promote abstinence. CBT can be provided in individual or group format.

Cue Exposure Treatment

Recently, there has been an interest in applying cue exposure therapy for alcohol problems. The idea of this approach is that drinking cues (e.g. sight, smell, taste) are thought to acquire potency through conditioning, and extinction would require repeated exposure to alcohol cues with response prevention. In a trial aimed at moderation drinking, non-dependent problem drinkers were provided priming doses of alcohol (i.e. patients were provided two or three standard drinks) followed by exposure to alcohol cues and then prohibited from drinking further (Sitharthan et al. 1997). This approach to treatment appears promising.

Pharmacotherapies

Pharmacological treatments have received much attention since the early 1990s. The oldest drug used in the treatment of alcohol dependence is disulfiram (Antabuse), an inhibitor of aldehyde dehydrogenase. By itself, disulfiram has only a modest effect on abstinence rates among alcohol-dependent patients (Fuller et al. 1986). However, when used as an adjunct with behavioural interventions, it has produced favourable outcomes. In combination with behavioural counselling, contingency contracting and, where possible, supervised administration, treatment compliance, abstinence rates and overall outcomes are substantially improved (Azrin et al. 1982).

There has been a great upsurge in interest in pharmacological agents to reduce craving and avert relapses. Naltrexone, an opioid receptor antagonist, which is considered to block the reinforcing effects of alcohol, has been shown to reduce relapse rates in several randomised controlled trials when compared with a placebo preparation. Results are best when naltrexone is combined with CBT (O'Malley et al. 1992; Center for Substance Abuse Treatment 1998) and are most impressive in the first 3–6 months of therapy.

Acamprosate (calcium acetyl homotaurinate) has been shown in several controlled trials to increase abstinence rates among alcohol-dependent patients (Sass et al. 1996). Acamprosate appears to complement behavioural and psychosocial treatments (Whitworth et al. 1996). Both naltrexone and acamprosate are well tolerated in alcohol-dependent individuals. Fluoxetine and imipramine may be useful adjuncts to relapse prevention programmes, particularly for dependent drinkers with co-morbid depression (Kranzler et al. 1995; McGrath et al. 1996). Combinations of pharmacotherapies may result in improved outcomes compared with single agents (Besson et al. 1998; Swift 1999).

Social Skills Training

Social skills training involves teaching patients assertion skills, including drink refusal and general communication skills. Clients who are dependent on alcohol are usually provided social skills training in a group format. Through role playing and modelling, the therapist trains the patient to respond to different difficult high-risk situations. Attention is given to not just what the patient might say in a high-risk situation, but also how it might be said effectively. Social skills training can be part of a global relapse prevention treatment programme.

Behavioural Marital Therapy

Behavioural marital therapy provides an opportunity for couples to negotiate treatment goals, report on treatment progress and improve communication and problem-solving skills. Patients may also be prescribed pharmacotherapy, and the spouse/partner learns to reinforce non-drinking and other appropriate behaviours. In addition, the couples also invest time to engage in shared rewarding activities (O'Farrell 1993). Behavioural marital therapy can be conducted with individual couples or delivered in a group format.

Stress Management Training

Stress management training is suited for patients who need to learn to relax and reduce their stress. Patients can be trained in various relaxation and meditation techniques. However, improvement is dependent on compliance to practice these techniques on a regular basis. Hence, engagement in pleasant activities may be as beneficial as relaxation training for some patients to reduce stress. In addition, cognitive strategies aimed at challenging and correcting maladaptive thinking patterns (Beck et al. 1993; Ellis et al. 1988) can also be used as part of general stress management training.

Twelve-Step Approaches

Worldwide, AA and similar 12-step movements represent the most widespread, often most accessible and sometimes the only available form of care and support for people with alcohol and drug problems. This mutual support group of "recovering alcoholics" is organisationally, doctrinally and financially independent of any government or professional body, but is incorporated into many funded and professionally sanctioned treatment programmes.

AA teaches that alcoholism is a disease and that the only possible way of recovery is through lifetime abstinence from alcohol. This recovery is to be gained through admitting helplessness and seeking assistance from a "higher power" (which is not aligned with a particular deity), considering the effects of drinking on others and where possible making amends, and assisting other alcoholics to obtain and maintain sobriety.

AA fellowships hold regular meetings at which a number of speakers will recount their own histories, describing their drinking, its negative consequences and how they came to recovery. The content of these meetings reminds the members of their own past and the need to remain sober and may powerfully challenge new attendees who relate to certain aspects of the stories of others. AA provides social support formally in the allocation of a "sponsor" to new members and informally by providing opportunities for friendship and social interaction in an alcohol-free environment.

Considering that AA was established in 1935, there has been remarkably little scientific investigation into the conceptual basis of the 12-step approach or into the outcome of treatment in comparison with other modalities. Although many individuals attribute their long-term abstinence to involvement with AA (Chappell 1993), the only comprehensive randomised controlled trial of a 12-step approach in comparison with other alcohol dependence treatments is Project MATCH (1997). Outcome at 12 months was similar in those groups receiving a 12-step approach compared with two groups receiving cognitive–behavioural or motivational enhancement therapies. The 12-step approach was slightly superior when judged on abstinence duration criteria (Project MATCH Study Group 1997).

7.2 Opiates

There are three broad approaches to the treatment of opiate dependence: psychosocial interventions, substitution therapies and opioid antagonists.

Psychosocial Interventions

Psychosocial interventions for opiate dependence are generally modelled on treatments for alcohol dependence. Preventing or reducing drug use, reducing health care utilisations, crime and legal expenditure and improving health and psychosocial well-being are the principal treatment objectives. However, compared with the alcohol treatment literature, there are few treatment outcome studies. Cognitive behaviour therapy is reported to be effective for patients who are depressed or experiencing other psychiatric symptoms (Woody et al. 1995), but recruiting patients into such therapies is often difficult. Group relapse prevention plus participation in self-help groups has been found to be more effective in reducing opiate use, unemployment and criminal activities (McAuliffe 1990).

Agonist Maintenance

Substitution therapies are particularly valuable in the management of opiate dependence. To date, the most effective form of treatment is methadone maintenance (Ball and Ross 1991). Methadone is a long-acting opioid agonist that is well absorbed after oral dosage and binds to opioid receptors. Methadone maintenance was introduced in the 1960s in an effort to provide some "pharmacological stability" and as an attempt to blunt the effect of injected heroin. It has been employed in several ways since then and in combination with various forms of counselling and psychological therapy. It results in cessation of illicit

opiate use in only a minority of persons, at least in the early months. However, the frequency of injecting and the overall amount used are reduced by some 90%, with attendant benefits in reducing needle/syringe sharing, infective complications, preoccupation with drug use and property crime. With time, life becomes more settled, relationships are more stable and employment prospects improve.

The benefits of methadone maintenance are maintained for as long patients are on treatment. However, there is a high rate of relapse into illicit opiate use, needle sharing and HIV risk behaviour when maintenance is ceased. This applies particularly when patients are coerced or expected to progressively reduce their dose. Contemporary practice emphasises the need for a more participatory approach (Caplehorn et al. 1998), with patients taking the initiative to commence a dose-reduction regime. High-dose methadone maintenance (80–100 mg per day) is more effective in retaining patients in treatment and reducing extraneous heroin use than low-dose regimes (30–40 mg per day). In a recent large-scale study, patients maintained on dosage of less than 70 mg/day were more likely to use heroin during methadone maintenance treatment (Hartel et al. 1995). Split-dosing (half of the dose in the morning and half in the evening at a second visit to the clinic) (Kidorf and Stitzer 1996) and take-homes doses reduce illicit substance use among patients on methadone maintenance.

Buprenorphine is a shorter-acting partial opioid agonist. It is at least as effective as methadone for maintenance treatment of individuals with moderate to severe dependence. It appears to be easier to withdraw off than methadone and is safer in overdose because it results in little respiratory depression. Its wider safety margin make dosing every alternative day an option, which is more convenient for patients, more cost-effective and reduces the risk of diversion of take-home doses. Levo-alpha-acetylmethadol (LAAM) is a long-acting analogue of methadone. Patients may receive their dose every 48–72 h. It appears to be an effective maintenance agent when patients have been initially stabilised.

The use of heroin for chronic heroin dependence which is resistant to other treatments has recently been evaluated in Europe. Significant reductions in illicit drug use and criminal activity and substantial improvements in psychosocial outcomes were documented. Supervised administration of heroin appears to be viable. However, much of the benefit probably derived from the intense psychosocial support provided to patients in the programme.

Antagonists

Pharmacotherapies can also be oriented towards abstinence. Naltrexone is a long-acting antagonist which is administered orally and can block opioid receptors for up to 72 h. The positively reinforcing euphoric effects of heroin are blocked, and craving for drugs is reduced substantially. Patients need to undergo detoxification and be drug-free for 7 days before starting naltrexone. Alternatively, they can be offered rapid opiate detoxification in which naltrexone is administered under sedation or anaesthetic. This precipitates a withdrawal syndrome, which is controlled by clonidine and drugs to suppress gastric secretions, and offers a relatively comfortable and rapid means of induction onto naltrexone (Brewer 1997). Naltrexone is more effective than a placebo drug in reducing heroin use, but there is a high attrition rate from treatment, particularly compared with methadone. It is best administered under direct supervision. There is a high incidence of depression in the early stages of treatment, and there is some evidence of greater effectiveness if naltrexone is used in combination with a serotonin reuptake inhibitor.

Harm Reduction Strategies

Ongoing opiate use and at-risk sexual behaviour is common despite methadone maintenance treatment or psychosocial interventions. Reducing the harm associated with these practices is an important goal. Sterile injecting equipment can be made available, along with instructions on how to inject safely and information to prevent overdosing. Similarly, condoms should be made available for patients, and they should be provided with information about safe sexual practices and communicable diseases.

There has been a considerable debate as to whether controlled availability of heroin (or other opiates) is a legitimate policy for the health care system. Given the findings that, once maintained on heroin, they show substantial improvements in their health status, stability of residence and employment status, there is substantial support for this approach. Most countries have been reluctant to endorse such programmes, but there is a growing movement to allow prescription of injectable opiates under close supervision.

7.3 Psychostimulants

There are no specific cognitive, behavioural or other psychosocial interventions of sustained effectiveness for patients who abuse or are dependent on cocaine. Cognitive-behavioural relapse prevention was more effective than standard case management, particularly for severe cocaine users (Carroll et al. 1994). Providing gift vouchers that can be exchanged for retail items has been shown to increase compliance among cocaine-dependent patients in providing clean urine. Similar

voucher exchange methods produced favourable results among injecting polydrug users (Silverman et al. 1996). A community reinforcement approach plus incentives were found to produce better outcomes than 12-step counselling. Intensive day hospital treatment is as beneficial as intensive in-patient treatment for cocaine-dependent patients (Alterman et al. 1994). Cognitive–behavioural relapse prevention approaches which include identifying high-risk situations and planned coping in such situations, along with social support, and contingency contracting should be considered as part of the treatment plan for patients who abuse or are dependent on cocaine. Antidepressants (desipramine and the selective serotonin reuptake inhibitors) ameliorate the depression of psychostimulant withdrawal and appear to improve treatment compliance in the first few weeks. However, there are no pharmacological treatments of proven effectiveness in promoting long-term abstinence from cocaine (American Psychiatric Association 1995).

7.4 Simultaneous Versus Sequential Cessation

It is not uncommon for substance misuse patients to abuse more than one drug. Many opiate and cocaine users tend to abuse alcohol and other substances. As one psychoactive substance may serve as a cue or trigger to use another, it may be necessary for patients to try to abstain from all psychoactive substances. However, simultaneous cessation from all substances, including tobacco, will be considered a very difficult task by many patients. If sequential cessation is preferred by the patient, a clearly defined contract, specifying a timetable of cessation events, should be developed. The patient should be provided with professional support through individual or group counselling to achieve sequential cessation.

If simultaneous cessation is required (by legal mandate), contingency contracting, compulsory breath and urine toxicology testing, regular attendance at intensive individual counselling, participation in group therapy and other self-help groups and, where possible, family support and supervision should be considered as part of the overall treatment plan, along with appropriate pharmacological adjuncts.

7.5 Managing Psychiatric Co-morbidity

Several epidemiological studies have found a high prevalence of psychiatric disorders among substance misusers and of substance misuse among psychiatric patients (Alaja et al. 1998; Brown et al. 1995; Hall 1996). While there are suggestions regarding case management of dual-diagnosis patients, randomised controlled trials evaluating the effectiveness of different methods of psychological and pharmacological interventions for dual-diagnosis patients are few (Sitharthan et al. 1999).

Patients with mental illness usually have better access to treatment services than patients with a co-morbid substance use disorder (Minkoff and Drake 1991). It is not uncommon for psychiatric patients with alcohol problems to feel "unwelcome" in alcohol treatment agencies (Institute of Medicine 1990).

At a practical level, there are several problems that arise in the day-to-day management of such patients. One is inadequate liaison between addiction treatment agencies and specialist psychiatric settings. Secondly, many psychiatric patients may not relate well to the demands placed on them by specialist addiction settings. Thirdly, diagnosis of co-morbid affective and anxiety disorders or schizophrenia cannot be reliably established until the patient is alcohol or drug free for 4 weeks or more, a goal which is often difficult to achieve. Liaison between addiction treatment agencies and specialist psychiatric clinics can lead to more effective detection and intervention for alcohol and drug problems (Sitharthan et al. 1999).

Providing an environment that fosters an abstinent lifestyle should be part of the clinical management of patients. In addition to providing compassionate support and care, patients should be trained in appropriate coping skills to prevent re-engagement in substance use (Kavanagh et al. 1998). Encouraging them to attend a group such as Dual Recovery Anonymous may provide peer support for such patients. Voluntary participation in self-help groups may also reduce the overall costs of providing professional treatment (Humphreys and Moos 1996).

It is not uncommon for many psychiatric patients to engage in high-risk sexual practices. Hence, educating patients about HIV transmission and teaching them behavioural skills to reduce HIV risk reduction is crucial. Behavioural skills include teaching patients to identify high-risk situations leading to substance use and unsafe sexual practices, learning problem-solving skills to negotiate safe sexual practices and assertion skills to refuse participating in high-risk sexual behaviours (Kalichman et al. 1995).

Addiction treatment services generally aim to promote total abstinence from all substances. It may not be possible for dual-diagnosis patients to achieve this goal. Treatment agencies must be sensitive to the special needs of patients with a dual diagnosis and may need to work within a harm-minimisation/reduction policy. Increased compliance and reductions in overall health care utilisation should be considered

as legitimate treatment goals. As dual-diagnosis patients frequently report low levels of satisfaction with family relationships (Dixon et al. 1995), involving family members in planned intervention must be explored.

8 Conclusions

Substance misuse and dependence continue at a high level worldwide. Reductions in the prevalence of alcohol and sedative-hypnotic misuse in developed countries have been offset by substantial increases in developing countries and in growing illicit drug markets. There is increasing diagnostic precision in the main international diagnostic systems, which reflects much research into the spectrum of substance misuse, the phenomenology of dependence and the natural history of the dependence syndrome. There is also a wider array of interventions, including more potent brief treatments for hazardous alcohol use, pharmacotherapies for alcohol dependence and substitution therapy and antagonists for the management of opiate dependence.

The last decade has witnessed a growing demand for evidence-based interventions and clinical accountability. Gordis (1996), commenting on alcohol research and practice, asserted that clinical practice that functions on the basis that "Deep in my heart I know our program works" can no longer be funded. We believe the same sentiment should be extended to other forms of substance misuse and related co-morbid disorders. Economic constraints, sophistication in implementing and evaluating treatment outcome research and the development of innovative treatment programmes will compel us to think constantly as to how best to treat our patients.

9 References

Alaja R, Seppä K, Sillanaukee P, Tieriari P, Huyse FJ, Herzog T, Malt UF, Lobo A and the European Consultative Liaison Workgroup (1998) Physical and mental comorbidity of substance use disorders in psychiatric consultations. Alcohol Clin Exp Res 22: 1820–1824

Alterman AI, O'Brien CP, McLellan AT, August DS, Snider EC, Droba M, Cornish JW, Hall CP, Raphaelson AH, Schrade FX (1994) Effectiveness and costs of inpatient versus day hospital cocaine rehabilitation. J Nerv Ment Dis 182: 157–163

American Psychiatric Association (1995) Cocaine related disorders: treatment principles and alternatives. Am J Psychiatry 152[Suppl]: 36–39

Annis H (1986) Is inpatient rehabilitation of the alcoholic cost effective? Adv Alcohol Substance Abuse 5: 175–190

Azrin NH, Sisson RW, Meyers R, Godley M (1982) Alcoholism treatment by disulfiram and community reinforcement therapy. J Behav Ther Exp Psychiatry 13: 105–112

Babor TF (1992) Substance-related problems in the context of international classificatory systems. In: Lader M, Edwards G, Drummond C (eds) The nature of alcohol and drug-related problems. Oxford University Press, Oxford, pp 83–97

Babor TF (1994) Avoiding the horrid and beastly sin of drunkenness: does dissuasion make a difference? J Consult Clin Psychol 62: 1127–1140

*Ball JC, Ross A (1991) The effectiveness of methadone maintenance treatment: patients, programs, services and outcomes. Springer, Berlin Heidelberg New York

Beck AT, Wright FD, Newman CF, Liese BS (1993) Cognitive therapy of substance abuse. Guilford, New York

Bell J, Batey RG, Farrell GC, Crewe EB, Cunningham AL, Byth K (1990) Hepatitis C virus in intravenous drug users. Med J Aus 153: 274–276

Besson J, Aeby F, Kasas A, Lehert P, Potgieter A (1998) Combined efficacy of acamprosate and disulfiram in the treatment of alcoholism: a controlled study. Alcohol Clin Exp Res 22: 573–579

Bien TH, Miller WR, Tonigen JS (1993) Brief interventions for alcohol problems: a review. Addiction 88: 315–336

Brewer C (1997) Ultra-rapid antagonist-precipitated opiate detoxification under general anaesthesia or sedation. Addict Biol 2: 291–302

Brown SA, Inaba RK, Gillin JC, Schuckit MA, Stewart MA, Irwin MR (1995) Alcoholism and affective disorder: clinical course of depressive symptoms. Am J Psychiatry 152: 45–51

Caplehorn JRM, Lumley TS, Irwig L, Saunders JB (1998) Changing attitudes and beliefs of staff working in methadone maintenance programs. Aust NZ J Pub Health 22: 505–508

Carroll KM, Rounsaville BJ, Gordon LT, Nich C, Jatlow P, Bisighini RM, Gawin FH (1994) Psychotherapy and pharmacotherapy for ambulatory cocaine abusers. Arch Gen Psychiatry 51: 177–187

Center for Substance Abuse Treatment (1998) Naltrexone and alcoholism treatment. US Department of Health and Human Services, Rockville MD

Chappell J (1993) Long term recovery from alcoholism. Psychiatr Clin North Am 16: 177–187

Chick J, Duffy JC (1979) Application to the alcohol dependence syndrome of a method of determining the sequential development of symptoms. Psychol Med 9: 313–319

Cohen SI, Lichtenstein E, Prochaska JO et al (1989) Debunking myths about self-quitting: evidence from 10 prospective studies of persons who attempt to quit smoking by themselves. Am Psychol 44: 1355–1365

Conigrave KM, Saunders JB, Reznik RB (1995a) Predictive capacity of the AUDIT questionnaire for alcohol-related harm. Addiction 90: 1479–1485

Conigrave KM, Saunders JB, Whitfield JB (1995b) Diagnostic tests for alcohol consumption. Alcohol Alcohol 30: 13–26

Department of Community Services and Health (1991) Statistics on drug abuse in Australia Canberra. Australian Government Publishing Service, Canberra

*Des Jarlais DC (1995) Harm reduction – a framework for incorporating science into drug policy. Am J Public Health 85: 10–12

Dixon L, McNary S, Lehman A (1995) Substance abuse and family relationships of persons with severe mental illness. Am J Psychiatry 152: 456–458

Drake RF, Wallach MA (1989) Substance abuse among the chronically mentally ill. Hosp Commun Psychiatry 40: 1041–1045

Edwards G (1986) The alcohol dependence syndrome: a concept as stimulus to enquiry. Br J Addict 81: 171–183

Edwards G (1990) Withdrawal symptoms and alcohol dependence: fruitful mysteries. Br J Addict 85: 447–461

Edwards G, Gross MM (1976) Alcohol dependence: provisional description of a clinical syndrome. Br Med J 1: 1058–1061

Ellis A, McInerney JF, DiGiuseppe R, Yeager RJ (1988) Rational-emotive therapy with alcoholics and substance abusers. Pergamon, New York

European Monitoring Centre for Drugs and Drug Addiction (1999) Annual report on the state of the drug problem in the European Union. Office for Official Publications of the European Communities, Luxembourg

Feuerlein W, Ringer C, Kufner H, Antons K (1979) Diagnosis of alcoholism: the Munich Alcoholism Test (MALT). Curr Alcohol 7: 137–147

Finney JF, Monahan SC (1996) The cost-effectiveness of treatment for alcoholism: a second approximation. J Stud Alcohol 57: 229–243

Fukui S (1994) Epidemiological trends of solvent abuse and dependence. J Ment Health 40: 3–11

Fuller RK, Branches L, Brightwell DR, Derman RM, Emrick CD, Iber FL, James KE, Lacoursiere RB, Lee KK, Lowenstam I, Maany I, Neiderhiser D, Nocks JJ, Shaw S (1986) Disulfiram treatment of alcoholism: a Veterans Administration Cooperative Study. JAMA 256: 1449–1455

Gordis E (1996) Alcohol research: at the cutting edge. Arch Gen Psychiatry 53: 199–201

Gossop M, Darke S, Griffiths P, Hando J, Powis B, Hall W, Strang J (1995) The Severity of Dependence Scale (SDS): psychometric properties of the SDS in English and Australian samples of heroin, cocaine and amphetamine users. Addiction 90: 607–614

Gossop M, Green L, Phillips G, Bradley L (1989) Lapse, relapse, and survival among opiates after treatment. Br J Psychiatry 154: 348–353

Grant KA, Hoffman PL, Tabakoff B (1990) Neurobiological and behavioural approaches to tolerance and dependence. In: Edwards G, Lader M (eds) The nature of drug dependence. Oxford University Press, Oxford, pp 135–165

Hall W (1996) What have population surveys revealed about substance use disorders and their comorbidity with other mental disorders? Drug Alcohol Rev 15: 157–171

Hall W (1998) Cannabis use and psychosis. Drug Alcohol Rev 17: 433–444

Hartel DM, Scheonbaum EE, Selwyn PA, Kline J, Davenny K, Klein RS, Friedland GH (1995) Heroin use during methadone maintenance treatment: the importance of methadone dose and cocaine use. Am J Public Health 85: 83–88

Hartnoll RL (1994) Opiates: prevalence and demographic factors. Addiction 89: 1377–1383

Hawks DV, Pols R (1992) Is there a safe level of daily alcohol consumption for men and women? Australian Government Publishing Service, Canberra

Heather N, Rollnick S, Winton M (1983) A comparison of objective and subjective measures of alcohol dependence as predictors of relapse following treatment. Br J Clin Psychol 22: 11–17

Helzer JE, Canino GJ (1989) The implications of cross-national research for diagnostic validity. In: Robins LN, Barrett JE (eds) The validity of psychiatric diagnosis. Raven, New York

Holder HD, Longabaugh R, Miller WR, Rubonis AV (1991) The cost-effectiveness of treatment for alcohol problems: a first approximation. J Stud Alcohol 52: 517–540

Holman CDJ, Armstrong BK, Arias LN, Martin CA, Hatton WM, Hayward LD et al (1988) The quantification of drug caused morbidity and mortality in Australia. Department of Community Services and Health, Canberra

Humphreys K, Moos RF (1996) Reduced substance-abuse-related health care costs among voluntary participants in Alcoholics Anonymous. Psychiatr Serv 47: 709–713

**Institute of Medicine (1990) Broadening the base of treatment for alcohol problems. National Academy Press, Washington

Jacobs AM, Ghodse H (1987) Delinquency and regular solvent abuse: an unfavourable combination? Soc Sci Med 24: 863–866

Jellinek EM (1960) The disease concept of alcoholism. Hillhouse, New Haven

Kalichman SC, Sikkema KJ, Kelly JA, Bulto M (1995) Use of a brief behavioural skills intervention to prevent HIV infection among chronic mentally ill adults. Psychiatr Serv 46: 275–280

Kavanagh DJ, Young R, Boyce L, Clair A, Sitharthan T, Clark D, Thompson K (1998) Substance treatment options in psychosis (STOP): a new intervention for dual diagnosis. J Mental Health 7: 135–143

Kavanagh DJ, Saunders JB, Young R, White A, Jerver L, Clair A, Wallis J (1999) Drug Check: a new screening tool for substance abuse in mental disorders. University of Queensland, Brisbane

Keso L, Salaspuro M (1990) Inpatient treatment of employed alcoholics: a randomised clinical trial of Hazelden-type and traditional treatment. Alcohol Clin Exp Res 14: 584–589

Kidorf M, Stitzer ML (1996) Contingent use of take-homes and split-dosing to reduce illicit drug use of methadone patients. Behav Ther 27: 41–51

Kleber HD (1988) Cocaine abuse: historical, epidemiological, and psychological perspectives. J Clin Psychiatry 49[Suppl]: 3–6

*Krabman P, Saunders JB (1996) Diagnostic criteria for substance misuse and dependence. In: Rommelspacher H, Schuckit MA (eds) Bailliere's clinical psychiatry: drugs of abuse. Baillière Tindall, London, pp 375–404

Kranzler HR, Burleson JA, Korner P, Del Boca FK, Bohn MJ, Brown JB, Liebowitz N (1995) Placebo-controlled trial of fluoxetine as an adjunct to relapse prevention in alcoholics. Am J Psychiatry 152: 391–397

Langenbucher J, Chung T, Morgenstern J, Labouvie E, Nathan PE, Bavly L (1997) Physiological alcohol dependence as a

"specifier" of risk for medical problems and relapse liability in DSM-IV. J Stud Alcohol 58: 341–350

Marlatt GA, Gordon JR (1985) Relapse prevention: maintenance strategies in the treatment of addictive behaviors. Guilford, New York

Mayfield D, McLeod G, Hall P (1974) The CAGE questionnaire: validation of a new alcohol screening instrument. Am J Psychiatry 131: 1121–1123

McAuliffe WE (1990) A randomized controlled trial of recovery training and self-help for opioid addicts in New England and Hong Kong. J Psychoactive Drugs 22: 197–209

McGrath PJ, Nunes EV, Stewart JW, Goldman D, Agosti V, Ocepek-Welikson K, Quitkin FM (1996) Imipramine treatment of alcoholics with primary depression. Arch Gen Psychiatry 53: 232–240

Mendenhall CL, Seeff L, Diehl AM, Ghosn SJ, French SW, Gartside PS, Rouster SD, Buskell-Bales Z, Grossman CJ, Roselle GA et al (1991) Antibodies to hepatitis B virus and hepatitis C virus in alcoholic hepatitis and cirrhosis: their prevalence and clinical relevance. The VA Cooperative Study Group. Hepatology 14: 581–589

Miller WR, Leckman AL, Delaney HD, Tinkcom M (1992) Long-term follow-up of behavioural self-control training. J Stud Alcohol 53: 249–261

Minkoff K (1994) Models for addiction treatment in psychiatric populations. Psychiatr Ann 24: 412–417

Minkoff K, Drake R (1991) Dual diagnosis of major mental illness and substance disorder. Jossey-Bass, San Francisco

Noble EP (1994) Polymorphisms of the D2 dopamine receptor gene and alcoholism and other substance use disorders. Alcohol Alcohol Suppl 2: 35–43

Noble EP (1999) The D_2 dopamine receptor gene: a review of association studies in alcoholism and phenotypes. Alcohol 16: 33–45

O'Farrell TJ (1993) Treating alcohol problems: marital and family interventions. Guilford, New York

*O'Hare PA, Newcombe R, Matthews A, Buning E, Drucker E (1992) The reduction of drug-related harm. Routledge, London

*O'Malley SS, Jaffe AJ, Chang G, Schottenfeld RS, Meyer RE, Rounsaville B (1992) Naltrexone and coping skills therapy for alcohol dependence. Arch Gen Psychiatry 49: 881–887

Orford J, Oppenheimer E, Edwards G (1976) Abstinence or control: the outcome for excessive drinkers two years after consultation. Behav Res Ther 14: 409–418

Petursson H (1994) The benzodiazepine withdrawal syndrome. Addiction 89: 1455–1459

Poshyachinda V (1993) Drug injecting and HIV infection among the population of drug abusers in Asia. Bull Narcot 45: 77–90

*Project MATCH Group (1997) Matching alcoholism treatments to client heterogeneity. Project MATCH posttreatment drinking outcomes. J Stud Alcohol 58: 7–29

Sass H, Soyka M, Mann K, Zieglgansberger W (1996) Relapse prevention by acamprosate. Results from a placebo-controlled study on alcohol dependence. Arch Gen Psychiatry 53: 673–680

Saunders JB (1989) The efficacy of treatment for drinking problems. Int Rev Psychiatry 1: 121–138

Saunders JB, Devereaux BM. Epidemiology and comparative incidence of alcohol-induced liver disease. In: Sherman D, Preedy V, Watson R (eds) Ethanol and the liver. Harwood, Reading (in press)

*Saunders JB, Aasland OG, Babor TF, De la Fuente JR, Grant M (1993) Development of the Alcohol Use Disorders Identification Test (AUDIT): WHO collaborative project on early detection of persons with harmful alcohol consumption II. Addiction 88: 791–804

Schuckit MA (1996) DSM-IV: was it worth all the fuss? In: Saunders JB, Whitfield JB (eds) The biology of alcohol problems. Elsevier, Oxford, pp 459–469

Schuckit MA, Smith TL, Daeppen J, Eng M, Li T, Hesselbrock VM, Nurnberger JI, Bucholz KK (1998) Clinical relevance of the distinction between alcohol dependence with and without a physiological component. Am J Psychiatry 155: 733–740

Schwartz LS, Lyons JS, Stulp F, Hassan T, Jacobi N, Taylor J (1993) Assessment of alcoholism among dually diagnosed psychiatric inpatients. J Subst Abuse Treat 10: 255–261

Seivewright NA, Greenwood J (1996) What is important in drug misuse treatment? Lancet 347: 373–376

Siegfried N (1998) A review of comorbidity: major mental illness and problematic substance use. Aust NZ J Psychiatry 32: 707–717

Silverman K, Higgins ST, Brooner RK, Montoya ID, Cone EJ, Schuster CR, Preston KL (1996) Sustained cocaine abstinence in methadone maintenance patients through voucher-based reinforcement therapy. Arch Gen Psychiatry 53: 409–415

Sitharthan T, Kavanagh DJ (1990) Role of self-efficacy in predicting outcomes from a programme for controlled drinking. Drug Alcohol Depend 27: 87–94

Sitharthan T, Kavanagh D, Sayer J (1996) Moderating drinking by correspondence: an evaluation of a new method of intervention. Addiction 91: 345–355

Sitharthan T, Sitharthan G, Hough M, Kavanagh D (1997) Cue exposure in moderation drinking: a comparison with cognitive-behavior therapy. J Consult Clin Psychol 65: 878–882

Sitharthan T, Singh S, Kranitis P, Currie J, Freeman P, Muragesan G, Ludowici L. Integrated drug and alcohol intervention (IDAI): development of an opportunistic intervention program to reduce alcohol and other substance use among psychiatric patients. Aust NZ J Psychiatry (in press)

Smart R (1986) Solvent use in North America: aspects of epidemiology, prevention and treatment. J Psychoactive Drugs 18: 87–96

Strang J, Bearn J, Farrell M, Finch E, Gossop M, Griffiths P, Marsden J, Wolff K (1998) Route of drug use and its implications for drug effect, risk of dependence and health consequences. Drug Alcohol Rev 17: 197–211

*Swift RM (1999) Drug therapy for alcohol dependence. N Engl J Med 340: 1482–1490

Vaillant GE (1988) What does long-term follow-up teach us about relapse and prevention of relapse in addiction? Br J Addiction 83: 1147–1157

Whitworth AB, Fischer F, Lesch OM, Nimmerrichter A, Oberbauer H, Platz T, Potgieter A, Walter H, Fleischhacker WW (1996) Comparison of acamprosate and placebo in long-term treatment of alcohol dependence. Lancet 347: 1438–1442

WHO Brief Intervention Study Group (1996) A cross-national trial of brief interventions with heavy drinkers. Am J Public Health 86: 948–955

Woody GE, McLellan AT, Luborsky L, O'Brien CP (1995) Psychotherapy in community methadone programs: a validation study. Am J Psychiatry 152: 1302–1308

Zur J, Yule W (1990) Chronic solvent abuse. I. Cognitive sequelae. Child Care Health Dev 16: 1–20

CHAPTER

17

M. Soyka

Abuse of, and Dependence on, Medically Prescribed Drugs

Translator: E. Taub

1 Problems of Classification

The diagnosis of abuse of, and dependence on, substances that are partly or mainly prescribed by physicians is associated with major problems of classification, more so than in the case of alcohol and other "classical" addictive drugs, such as opioids or cocaine. The reason is partly that these substances are often given to treat mental or psychovegetative disorders; they thus possess a therapeutic effect that other psychotropic substances do not. A further reason is that the ICD-10 and DSM-IV criteria for psychotropic substance abuse and dependence are not easy to apply to medically prescribed substances.

"Harmful use" of a substance, in the terminology of ICD-10, involves actual harm to the individual consuming it, i.e. an injury to health rather than merely negative social effects. In DSM-IV, on the other hand, negative social effects are among the criteria for the diagnosis of substance abuse. Even protracted long-term consumption of certain hypnotics and anxiolytics may not lead, in all cases, to an immediately recognizable or readily diagnosable injury to health, even when injurious use is clearly present from the clinical perspective in view of the severe associated psychosocial problems.

The diagnosis of dependence on a medically prescribed substance may also be problematic. The desire or compulsion ("craving") to consume a psychotropic substance is typically not as strong in such cases as in alcohol or opiate abuse; furthermore, many of these substances do not demonstrably impair self-control. The development of tolerance is nearly always present with long-term consumption of addictive drugs, but need not be so in cases of medication dependence: thus clinicians have long been aware of the existence of so-called low-dose benzodiazepine dependence, in which there is neither tolerance nor an escalation of dose, and many frequently abused medications, such as anabolic steroids, do not give rise to tolerance at all in the narrow sense of the term.

It may also be difficult to distinguish the physical or mental manifestations of withdrawal from those of the underlying illness ("symptom re-emergence") when medications prescribed to treat mental health problems, such as hypnotics, are discontinued or when their dosage is reduced. To illustrate the problem, an epileptic patient treated with benzodiazepines might have seizures when they are discontinued, but such seizures would not be called "withdrawal seizures." In analogous fashion, the re-emergence of anxiety or of sleep disorders after the discontinuation of hypnotics or anxiolytics is not necessarily the result of a dependence syndrome. In recent publications, terms such as "rebound insomnia" and the like have been used to describe these phenomena.

In summary, the ICD-10 and DSM-IV criteria for abuse, harmful use, and dependence on psychotropic substances are somewhat problematic for many of the classes of medications listed below, and the diagnosis must also take other clinical aspects into account (e.g. behavioral abnormalities, "doctor shopping," etc.). Moreover, a clear distinction must be drawn between primary and secondary medication dependence (as in alcohol and drug abuse).

2 Epidemiology

The available estimates of the frequency of abuse of, and dependence on, medically prescribed drugs in the general population are highly variable. It is clear from official statistics that a number of substance classes, in particular hypnotics/anxiolytics, tranquilizers, and analgesics, are among the more commonly sold medications of all kinds. Glaeske (1996) estimated that approximately 6%–8% of all prescribed medications have an intrinsic potential for abuse.

There have been major changes in prescribing behavior in recent years. First of all, 38% of all medications sold require no prescription (so-called self-medication); analgesics account for the major share of these.

Community-based studies of medication use have revealed that approximately 6%–12% of the adult population is under treatment with psychoactive medications (Cooperstock and Parnell 1982; Weyerer and Zimber 1997). Population-based studies in a number of countries have demonstrated a relationship between the use of psychoactive medications and the age and sex of the consumer (Larson et al. 1991). Such medications, particularly tranquilizers and hypnotics, are used more frequently by women, and still more so with advancing age. Most psychoactive medications are not prescribed by psychiatrists. According to a representative study by Beardsley et al. (1988), two thirds of all psychoactive medications are prescribed by general practitioners and internists, and only about 17% by psychiatrists or other specialists.

The problem of dependence on medication is common knowledge not only among clinicians, but also in the general public. Thus the rate of prescription of benzodiazepines is steadily falling (though from a high baseline), and barbiturates now play no more than a secondary role. On the other hand, the prescription of codeine-containing drugs primarily

designated as antitussives has risen astronomically in the past few years, even though these are frequently used in the illegal drug scene as so-called escape drugs. Because of their relatively short half-life (4–6 h), medications containing codeine or dihydrocodone seem to be particularly unsuitable for substitution treatment, as compared with methadone.

Keup (1993) attempted to assess the addictive potential of medications in a relatively systematic fashion by means of the so-called early warning system. The latter has been operative in Germany since 1976 and involves the determination of all substances and preparations taken and abused by a representative, random sample of individuals with a history of substance abuse, including approximately 700–900 individuals annually.

Somewhat surprisingly, it has occasionally been stated in Germany in the last few years that antidepressants and neuroleptics are medications with addictive potential. This is probably because many individuals dependent on alcohol and illegal drugs value their general psychomotor damping effect, which reduces anxiety and tension. It is highly misleading, however, when medications such as amitriptyline and doxepine are mentioned in the same breath as tramadol, morphine, levomethadone, and clomethiazol in a table of abused medications listed by frequency of prescription, as in the last Addiction Yearbook (*Jahrbuch Sucht* 1998) published by the Hauptstelle gegen die Suchtgefahren (German Central Agency Against the Dangers of Addiction).

The so-called German Federal Study has been carried out at irregular intervals of 2–5 years since 1980 by order of the Federal Ministry of Health and has provided important data on the use of psychoactive substances among adults in Germany. It was last performed in 1997 on a sample of 8020 adults aged 18–59 (Kraus and Bauernfeind 1998). With regard to the prevalence of medication use, it was found that almost twice as many women (19.5%) as men (11.5%) had taken psychoactive substances, e.g. analgesics, laxatives, and appetite suppressants, in the 4 weeks before the study. The types of medication most commonly taken included analgesics (13.5% in women vs. 8.6% in men), hypnotics (3.2% vs. 2%), tranquilizers (4.4% vs. 2%), so-called stimulants (1% vs. 0.8%), laxatives (3.1% vs. 0.9%), and appetite suppressants (1.2% vs. 0.4%). It should be kept in mind that these figures do not reflect the use of psychoactive substances by persons over 60, which, according to all currently available data, is even higher.

Although certain psychoactive medications and analgesics are commonly prescribed, as stated above, it should be pointed out explicitly that the prescribing habits revealed by statistics in no way imply the development of an addiction in every case. Even so, the number of individuals dependent on medications in Germany has been estimated to be as high as 1.4 million (Hüllinghorst 1996). Women considerably outnumber men as consumers of practically all of the medications named in this chapter, with the exception of anabolic steroids. According to a review by Kraus (1996), 60% of those studied had taken one or more medications at least once a week in the month before the study (men 12.8%, women 19.3%). A further finding is that many of the substances named (psychoactive medications in particular) are increasingly consumed with advancing age and are very commonly prescribed to residents of old-age homes and nursing homes (for a review, see Weyerer and Zimber 1997).

An important, frequently overlooked aspect of dependence on medically prescribed substances is the associated mortality. Studies by Poser et al. (1990, 1992) revealed that the mortality of patients taking psychoactive substances was 2.1 times higher than that of a control group. This figure was lower than that found in other forms of addiction. Alcoholism led to a fivefold increase in the consumption of illegal drugs and to a 17-fold increase in mortality. Abuse was much less dangerous than dependence with respect to increased mortality (elevation by a factor of 1.1, as compared to 3.8). In cases of addiction that went on to take a favorable course, mortality was essentially normalized. The reported causes of death in patients dependent only on prescribed medication are of interest. Among such patients who died, one third died by suicide, and 8% as the result of accidents or intoxications. These findings may also indicate a high degree of psychiatric comorbidity among patients dependent on prescribed medication.

3 Individual Substance Classes

3.1 Anxiolytics and Hypnotics

Benzodiazepines

The benzodiazepines have largely replaced the "classical" hypnotics and anxiolytics, such as the barbiturates, because of their low toxicity and are among the more commonly used medications of all kinds. The therapeutic effect of these substances is due to specific binding to the γ-aminobutyric acid $(GABA)_A$ receptor, and the pharmacological potency of each benzodiazepine substance is a function of its affinity for this receptor. The effect results from GABA-induced neural inhibition in various regions of the central nervous system. All benzodiazepines have hypnotic, anxiolytic, anticonvulsant, and muscle-relaxing properties,

although there is wide variation in the clinical spectrum of the individual substances. The major indications for their use in psychiatry are anxious conditions and sleep disturbances. They are also used to treat withdrawal syndromes and withdrawal psychoses, as well as epileptic seizures.

It is important to note that benzodiazepines are also used for indications other than these classical ones, often without adequate medical justification. The study by Geiselmann et al. (1989) revealed a marked degree of "therapist non-compliance" with respect to the generally applicable recommendations for the use of benzodiazepines, which are clearly often given for nonindicated purposes. In addition, they are often administered in the long term and not merely in the short term, as generally recommended, and they are sometimes given even to patients with primary substance abuse. This study shows that erroneous or inadequately justified prescribing behavior on the part of the therapist may actually favor the development of substance abuse and dependence in many cases.

The most important side effects of benzodiazepine overdose include excessive sedation, disturbance of motor coordination (ataxia), somnolence and lethargy, delirium and disorientation, amnesia, and aggravation of dementia. Higher doses primarily impair cognitive and psychomotor function.

Although the short-term administration of benzodiazepines is regarded as relatively unproblematic, there have been numerous reports of long-term administration for months or years leading to the development of dependence. Some patients may develop a marked tolerance for the drug, which is, however, not as severe as with barbiturates. Tolerance is due to adaptation of receptors in the central nervous system, rather than to pharmacokinetic changes; benzodiazepines, unlike other hypnotics, do not induce the production of metabolic enzymes in the liver.

Benzodiazepine abuse and dependence are clearly very common. The findings of a number of studies suggest that benzodiazepines with particularly short half-lives or strong anxiolytic profiles have an especially high dependence potential (see Poser and Poser 1996; Soyka et al. 1988), yet essentially all of them have a considerable potential for addiction. As remarked above, benzodiazepine dependence differs from other types of addiction in that it may develop without a progressive increase in dosage, i.e. without any major degree of tolerance. Severe mental and physical withdrawal phenomena may follow the discontinuation of benzodiazepines even in the setting of low-dose benzodiazepine dependence (Soyka et al. 1988).

Linden et al. (1998) found that patients with low-dose benzodiazepine dependence, even those undergoing long-term treatment, had a type of medication craving, as well as many other, major mental and physical symptoms. The authors interpreted these findings as showing that low-dose benzodiazepine dependence is indeed dependence in the proper sense of the term.

Low-dose benzodiazepine dependence is said to be present when the daily dose does not exceed the equivalent of approximately 30 mg diazepam. Experience has shown that benzodiazepines should be withdrawn slowly and that patients taking higher doses of medication have a higher risk of developing withdrawal psychoses and seizures.

It is important to point out that most cases of benzodiazepine abuse and dependence are secondary, i.e. that they arise in the setting of other types of dependence (e.g. alcoholism) or multisubstance abuse. Patients who abuse and are dependent on other substances should thus be considered a high-risk group for benzodiazepine dependence, even when benzodiazepines are used preventively. The elderly are a further risk group: several studies have shown that patients in homes for the elderly and nursing homes are very frequently treated with psychotropic drugs, and with benzodiazepines in particular. As mentioned above, such treatment may lead to a certain degree of cognitive impairment and confusion; moreover, the elderly often have an impaired ability to metabolize benzodiazepines. Clinical and epidemiologic studies have shown that some types of psychoactive drugs, particularly tranquilizers, are prescribed very frequently to patients living in nursing homes and homes for the elderly (Weyerer and Zimber 1997b).

More than 20 different benzodiazepines are now on the market in Germany, including short-acting (triazolam), intermediate (alprazolam, lorazepam), and long-acting preparations (diazepam, chlordiazepoxide). The abuse potential of benzodiazepines is less than that of other hypnotics (including barbiturates); however, because they have largely displaced other hypnotics and tranquilizers in psychiatric practice, benzodiazepine dependence is now the most frequent indication for withdrawal treatment. A special role is played by the medication flunitrazepam (Rohypnol), which has achieved a regrettable degree of prominence in the drug scene and is often taken by many addicts because of its strong sedative effect and strong psychotropic activity. It is also sold on the black market in greater volume than other benzodiazepines.

The clinical manifestations of benzodiazepine intoxication are essentially no different from those seen with other sedatives, hypnotics, and anxiolytics. They possess a wide therapeutic window (i.e. there is a wide range between the minimal therapeutic dose and the minimal toxic dose), and thus lethal outcomes are to be feared only in cases of mixed intoxication.

Among its mental manifestations, benzodiazepine intoxication leads to the following: affective disinhibi-

tion; emotional lability; impairment of cognitive function, judgment and critical ability; impairment of psychomotor performance, including driving ability; and various other behavioral abnormalities, most importantly profound sedation or coma. Its most prominent general medical and neurological complications are slurred speech, incoordination, ataxia, unsteady gait, respiratory depression, and sometimes hypo- or areflexia. The differential diagnosis includes alcohol and other intoxications, hypoglycemia, epilepsy, cerebral hemorrhage or infarction, myocardial infarction, and other illnesses.

No specific measures are needed to treat less severe intoxications. The administration of the short-acting benzodiazepine antagonist flumazenil (Anexate, 1 ampoule IV) is indicated less for therapeutic than for diagnostic purposes (exclusion of benzodiazepine or multisubstance intoxication). Patients suffering exclusively from benzodiazepine intoxication awaken spontaneously in a very short time.

Benzodiazepine withdrawal phenomena are extremely variable. The individual manifestations of sedative and hypnotic withdrawal are very similar to those of alcohol withdrawal, though their course is somewhat different (usually prolonged). Perceptual abnormalities are very common, though not pathognomonic. Abuse of, or dependence on, these substances is often secondary, i.e. in association with the consumption of other psychotropic substances, particularly alcohol and opioids. Primary dependence is relatively rare and, as experience has revealed, develops most frequently in patients with an underlying psychiatric illness. The clinical picture of benzodiazepine withdrawal phenomena is correspondingly variable.

In general, two types of withdrawal syndrome can be distinguished on the basis of their course: a mild, brief withdrawal syndrome ("minor withdrawal") and a severe one accompanied by psychosis ("major withdrawal"). Perceptual abnormalities are highly characteristic, though not pathognomonic, of benzodiazepine withdrawal and may be deceptively similar to symptoms of neurological disease. Withdrawal psychoses are found mainly when higher doses of benzodiazepines have been consumed. When benzodiazepines have been administered in the long term, their sudden discontinuation or excessively abrupt reduction is contraindicated. A stepwise tapering of medication ending in discontinuation has proved useful in withdrawal treatment (e.g. halving of the dose every 5 days; Soyka et al. 1988).

As discussed above, the differentiation of substance-induced disorders from the manifestations of an underlying psychiatric illness is particularly difficult in the setting of hypnotic or anxiolytic withdrawal. Withdrawal psychoses in the form of delirium and schizophreniform psychosis are comparatively rare. Epileptic seizures, in contrast, are relatively frequent in benzodiazepine withdrawal. It is known that benzodiazepines can lead to amnestic syndromes and to an impairment of psychomotor performance, including driving ability. The personality changes that they produce are difficult to classify and often have manifestations of both anxiety and depression.

Barbiturates

Barbiturates, like benzodiazepines, exert their effects by way of the $GABA_A$ receptor. They have a markedly higher abuse potential than the benzodiazepines, and lethal intoxications are common. Most barbiturates have a very long half-life. Unlike the benzodiazepines, they have no specific antidote.

Tolerance and dependence often develop very rapidly. Metabolic tolerance occurs because of the induction of hepatic enzymes.

When they have been taken for a considerable length of time, barbiturates often have a paradoxical activating and mood-brightening effect. High doses often lead to a state of intoxication. Because barbiturates have a powerful hypnotic effect and suppress rapid eye movement (REM) sleep, their withdrawal often leads to REM rebound with increased dreaming, nightmares, and sleep disturbances. Delirium, hallucinations, disorientation, agitation, and epileptic seizures are common in barbiturate withdrawal.

Newer Non-benzodiazepine Hypnotics

The newer GABAergic non-benzodiazepine hypnotics, zolpidem and zopiclon, have found wide application in recent years. There have been a few case reports of abuse and dependence with both of these substances (Bottlender et al. 1997; Thome et al. 1997), but primary dependence is exceptional. For zolpidem, in particular, many very careful investigations of its possible potential for abuse and dependence have revealed these dangers to be considerably lower than with other hypnotics, including benzodiazepines; long-term use of zolpidem is relatively rare (Keup 1998). These substances have not found their way into the drug scene to date. Nonetheless, the possibility of intoxication, abuse, and withdrawal cannot be excluded in individual cases.

Clomethiazol

Clomethiazol requires special mention in this discussion. Dependence on clomethiazol often develops in the setting of primary alcohol dependence. It is a highly effective medication in the treatment of the alcohol withdrawal syndrome and of delirium (reviewed in Soyka 1995, 1997), but, when given in the outpatient setting, it often leads to a displacement of addiction and to secondary dependence. It should thus be given in the outpatient setting only in exceptional cases, when specially indicated, and only

in the short term. Clomethiazol withdrawal is often very difficult and protracted.

Other Hypnotics

The "classical" non-barbiturate hypnotics, like the barbiturates, have now been largely, though not completely, supplanted in clinical use by the benzodiazepines. A number of single and compound preparations of this type remain available in Germany, including methaqualone, meprobamate, chloral hydrate, piperidine derivatives, and bromourea derivatives. The latter substance class not only has a considerable dependence potential, but is also associated with a specific clinical pattern of chronic intoxication. Bromism is characterized by irritability, depressive mood disturbance, sedation, purpura, excessive tearing, intestinal complaints, anorexia, headache, and disturbances of speech and coordination, as well as acneiform cutaneous abnormalities. The diagnosis may be established by determination of the serum bromine concentration.

3.2 Caffeine

Caffeine is a mild psychostimulant unrelated to any other class of psychostimulants, e.g. the amphetamines. It is present not only in coffee, tea, cola, chocolate, and cocoa, but also in numerous combination drugs, particularly analgesics. It has various pharmacological effects. In addition to its stimulating effect, there is also a very mild euphoria and an acceleration of breathing. In higher doses (250–600 mg), caffeine produces nervousness, anxiety, restlessness, difficulty falling asleep, tremor, and hyperesthesia. The public has become increasingly aware in recent years that caffeine can be used for "doping," as it improves performance in endurance sports. The psychotropic effect of caffeine is largely due to an interaction with adenosine receptors in the central nervous system (for a review, see Poser and Poser 1996), but it also influences other neurotransmitters, such as dopamine, as well as second messenger systems.

Caffeine abuse is generally seen in connection with doping or with analgesic consumption. Caffeine is only rarely abused as a single substance (as sometimes occurs, for example, in penal institutions).

Caffeine intoxication, which requires doses of over 500 mg, is also very rare. Anxiety, insomnia, mood changes, tachycardia, and cardiac dysrhythmia, as well as elevated blood pressure and gastrointestinal complaints, are the most prominent manifestations. The consumption of high doses of caffeine can lead, in very rare cases, to acute psychosis (Poser and Poser 1996) or to epileptic seizures.

Caffeine-induced panic attacks are a special phenomenon. Patients with a predisposition to such attacks should avoid consuming caffeine. Usually, however, fairly high doses of caffeine are required to induce panic attacks in patients with anxiety disorders. This phenomenon is not observed in healthy individuals.

The so-called mixed analgesics are by far the most commonly abused caffeine-containing medications. Their chronic use may lead to headache as the most prominent symptom.

3.3 Mixed Analgesics

Mixed analgesics are medications in which a peripherally acting analgesic or antipyretic is combined with one or more centrally acting substances. The centrally acting substances most commonly found in such preparations were at one time barbiturates and tranquilizers, though opioids, alkaloids, and caffeine were also used. Combination drugs, e.g. caffeine with acetylsalicylic acid, paracetamol, or propyphenazone, are among the medications more commonly used in Germany at present. Some require a prescription, while others are available over the counter. Mixed analgesic abuse is one of the more common forms of medication abuse; it is seen most often among individuals suffering from headaches, but also among those with other underlying physical illnesses.

The clinical picture of mixed analgesic abuse is highly characteristic. Its neuropsychiatric manifestations include persistent headache, fatigability, frequent or intensified migraine attacks, personality changes, avid desire for medication, and, with heavy consumption of barbiturates or tranquilizers, seizures and withdrawal delirium. The cardiovascular manifestations include hypertension, arteriosclerosis, coronary heart disease, cerebral infarction, and an increased frequency of preeclampsia. The cutaneous changes are also relatively typical, including paleness and abnormal pigmentation. These patients often seem prematurely aged.

Renal damage, specifically analgesic nephropathy, is a potential concern in chronic analgesic consumption. Gastrointestinal disturbances also occur, particularly an increased frequency of pancreatitis, as well as hematologic disturbances (anemia) and an increased frequency of osteoporosis and infertility. Headache is a relatively typical finding with consumption of mixed analgesics that contain ergot alkaloids. A dull, bilateral headache occurs when the medication is taken, and severe migraine attacks occur when it is withdrawn. Caffeine withdrawal headaches were already mentioned above. In general, patients chronically consuming

mixed analgesics tend to suffer from dull, pressure-type headaches.

The withdrawal of mixed analgesics is often difficult and protracted and usually involves the administration of sedating tricyclic antidepressants of the doxepine class. Oral metoclopramide is recommended for the treatment of nausea and vomiting, while severe migraine attacks can be treated with newer antimigraine agents of the sumatriptan type (Imigran). Further treatment depends on the clinical course. Psychotherapy, instruction in relaxation techniques, self-help groups, or biofeedback may be indicated.

3.4 Psychostimulants and Appetite Suppressants

Cocaine and most psychostimulants with similar effects are illegal drugs and are thus not discussed in this chapter. Some psychostimulants, however, do have a narrow range of clinical application. Amphetamine may be prescribed and sold as the racemic substance, but not as dextro- or levo-amphetamine. The abuse of appetite suppressants is a major problem mainly among patients with eating disorders.

The effect of amphetamines is essentially due to their indirect stimulation of the release of newly synthesized catecholamines, primarily dopamine, from presynaptic terminals. This results in behavioral stimulation, elevated psychomotor activity and, with higher doses, stereotypic behavior patterns and aggression. These effects are found primarily in adults; in children, on the other hand, amphetamine derivatives are used to treat aggressive behavior and hyperactivity. Aside from this special area of clinical application, amphetamines also play a role in the current treatment of narcolepsy and of overweight. The latter indication is highly controversial; it may fall still further out of favor because of recent developments in pharmacology (see below).

It is generally recommended that appetite suppressants such as fenfluramine be used as monotherapy only for short periods of less than 3 months. The consumption of appetite suppressants has risen exponentially in the last decade, at least in the United States (Kahn et al. 1998). Some of the newer so-called appetite suppressants, not only fenfluramine but also the serotonin reuptake inhibitors fluoxetine and sertraline, enhance serotonergic rather than dopaminergic neurotransmission. In contrast, the mechanism of action of the "classical" appetite suppressant methamphetamine is similar to that of cocaine. The consumption of amphetamine in high doses has been repeatedly reported to lead to violent outbursts, paranoid psychosis, and other psychotic behavioral abnormalities. It probably has neurotoxic effects.

Moreover, express mention should be made of the cardiotoxicity of appetite suppressants, which has been reported with increasing frequency only in the last few years. Two recently published studies (Kahn et al. 1998; Jick et al. 1998) revealed an elevated risk for cardiac valvular insufficiency in association with the consumption of fenfluramine, phentermine, and dexfenfluramine.

3.5 Anabolic Steroids and Other Substances Used for Doping

Anabolic steroids and other substances used for doping constitute a very heterogeneous substance class comprising psychostimulants and caffeine (see above), other stimulants, opioids, and, most importantly, anabolic steroids and nonsteroidal anabolic substances. The "red list" includes a large number of substances that enhance performance in competitive situations and, like anabolic steroids, lead to increased strength and muscle mass. Most of these substances are sold illegally, but there are also medically justifiable indications for their use. Anabolic steroids are used to treat lack of appetite, generalized weakness, muscle atrophy, and bone decalcification, among other conditions. The consumption of anabolic steroids does not lead to euphoria or to other obvious psychotropic effects.

The question whether a physician may prescribe anabolic substances to a bodybuilder is to be answered as follows, from the legal perspective: a physician can and may prescribe such preparations, whether or not the patient is (coincidentally) engaged in bodybuilding, but a strict medical indication must be present for their use (Schlund 1998). If the physician prescribes such a substance in the knowledge that the "patient" is not ill and wants to take it only to improve athletic performance or to increase muscle mass, then he or she is behaving unethically from the legal viewpoint and is, furthermore, violating the fundamental canons of the medical profession. From the medical perspective, careful ethical consideration is necessary before anabolic substances can be administered in the absence of a compelling medical indication.

It is of great concern that not only high-performance athletes, but also amateur athletes and bodybuilders are now taking such substances. Recent studies of amateur athletes engaged in fitness training have revealed that the consumption of anabolic steroids is surprisingly common. Boos et al. (1998) surveyed participants in commercial athletic studios and found that 24% of the men and 8% of the women were taking anabolic medications. In 94% of such cases, the substances taken were highly hepatotoxic. The source of the medications consumed was also of interest: they

were mainly obtained on the black market, but approximately 14% were prescribed by physicians.

The social framework of doping-substance abuse is somewhat different in the United States, where this phenomenon has been recognized for some time, than in Germany. The abuse of anabolic steroids in the United States is by no means confined to high-performance athletes, as studies by American sports physicians have revealed. It is thought that not only the high social standing of the successful athlete, but also the potential receipt of athletic scholarships to finance college education are important sociological factors motivating the consumption of such substances (for a review, see Boos et al. 1998). Nevertheless, a study performed by Boos's group revealed that a lower educational level is strongly associated with a higher degree of consumption.

It is not entirely clear whether the abuse of these substances leads to dependence, as postulated by a number of American investigators (e.g. Kashkin 1992). Favoring this hypothesis is the fact that 72% of the athletes abusing medications who were questioned in the study by Boos et al. (1998) wanted to continue taking them. Moreover, there seems to be a relatively high correlation between the use of anabolic steroids and the consumption of cocaine, marijuana, and other intoxicating drugs, with the exception of alcohol.

Relatively little is currently known about the possibly adverse mental effects of long-term consumption of these substances. The mechanisms leading to dependence have also been little studied to date. The desired increase in muscle mass and altered body image are probably such powerful psychological reinforcers in themselves that they can perpetuate the self-administration of anabolic steroids (Julien 1997). The changed personal appearance and the associated increase in self-esteem are a sufficient explanation in many cases.

The physical harm resulting from the use of these substances is quite extensive. There is damage to the cardiovascular system (arteriosclerosis, heart failure), the endocrine system (hypothyroidism), the immune system, the musculoskeletal system, and, relatively frequently, the liver, including the development of hepatic tumors. The sudden death of high-performance athletes in connection with doping has repeatedly come to the attention of the public. In men, anabolic steroids lead to gynecomastia, testicular atrophy, acne, and baldness; in women, they lead to masculinization, with an increase in body hair, beard growth, and other changes. (Such abnormalities were readily apparent in many high-performance female athletes in East Germany during the period of Communist rule, and the fact that they were due to anabolic steroid use has now been officially acknowledged.) Long-term mental effects may include depression, fatigue, restlessness, reduced appetite, loss of libido, and sleep disturbances. It has been stated that mood fluctuations, aggression, suicidality, and psychoses can also be induced (Kashkin 1992; Haupt 1993).

The withdrawal of anabolic steroids may be associated with depression, fatigue, restlessness, sleeplessness, and reduction of appetite and libido, possibly also headache and, in some cases, suicidality. A clearly delineated psychiatric withdrawal syndrome has not yet been described; withdrawal-associated psychoses or bipolar disorders have not been found to occur, though depression has been reported (Julien 1997).

3.6 Other Substances

Many other medically prescribed substances are subject to abuse, and many of these have no psychotropic effects. The abusive consumption of glucocorticoids has been reported in very rare cases, while the abuse of anticholinergics is somewhat more common. The latter may lead to dry mouth, fatigue, mydriasis, tachycardia, constipation, cutaneous changes, and hyperthermia. Frequent psychiatric manifestations of anticholinergic abuse include delirium with confusion, sleeplessness, excitement, and hallucinations. There have also been sporadic observations of abuse of tri- and tetracyclic antidepressants and of clonidine.

The abusive consumption of antihistamines is somewhat more common, e.g. diphenhydramine, promethazine, and doxylamine, which are available in many combined preparations (antiemetics, antitussives, and antiallergic medications). Isolated abuse or dependence is hardly ever seen with these substances. Their abuse generally causes excessive sedation; delirium occurs rarely, in cases of overdose.

The abuse of laxatives and diuretics is much more common. The latter are predominantly prescribed by physicians. Female patients with eating disorders are particularly likely to be affected. Bodybuilders and prostitutes also take diuretics to alter the results of urinalyses or to achieve a desired "fighting weight" or appearance. The desire for a better appearance with regard to muscle mass and bodily proportion is an important factor in the abuse of these substances. Laxatives, in particular, are also commonly abused by elderly patients without any underlying psychopathology. The consequences of long-term laxative consumption may include chronic constipation or diarrhea, abdominal pain and cramps, and even, in severe cases, toxic megacolon. The abusive consumption of diuretics can lead to hypokalemia.

There have been only sporadic reports of the abuse of so-called aphrodisiacs, i.e. substances thought to increase sexual desire, heighten the sexual experience,

or improve potency. This is a very heterogeneous class of substances, most of which probably do not possess their purported effects. Amyl nitrite, a powerful vasodilator, is frequently used as an aphrodisiac. The recent introduction of sildenafil (Viagra) in the treatment of erectile dysfunction (see Chap. 13, this volume) and the associated public interest make it seem likely that cases of abuse will be observed more commonly in future. It has been reported that the simultaneous use of Viagra and Ecstasy is already common in the "techno" scene. The danger is obvious.

4 Potential Future Developments

In conclusion, we will venture to make a psychiatric prognosis and predict that a negative development will take place in the coming years. It is now clear that a number of new, so-called lifestyle medications will be introduced in the near future (see *Deutsches Ärzteblatt* 1998, no. 36, B1681). *Deutsches Ärzteblatt* (German physicians' weekly) counts Viagra, Xenical, and Propecia – used to combat erectile dysfunction, overweight, and baldness, respectively – among these medications. Almost nothing is known at present about their possible psychotropic effects and other side effects or about their spectrum of clinical application. From the psychiatric viewpoint, it may be assumed that the appearance of this very heterogeneous group of substances, administered for entirely novel indications, will call forth a desire for "improved well-being" among many people. Even if these substances turn out to have no psychotropic effects, they will likely be abused with increasing frequency in the future.

Just as doping substances, such as anabolic steroids, have recently come into widespread use among certain subgroups of the population, it can be assumed that these or similar lifestyle medications will achieve prominence in the public consciousness and in the treatment of various disorders of greater or lesser severity. In addition, because (as is now clear) only a small fraction of such medications will be paid for by medical insurance, they will increasingly be taken by self-medication or via private prescription. The Internet, too, is already an avenue for the distribution of many substances. Many medications with psychotropic effects, as well as Viagra, can now be ordered electronically with minimal difficulty and in the complete absence of medical supervision.

From the psychiatric viewpoint, it would be naïve for us to shut our eyes to these developments. In terms of prevention, we must work constantly to inform patients and the public of the risks of the new, so-called lifestyle medications, in order to prevent them from being abused.

5 References

Beardsley RS, Gardocki GJ, Larson DB, Hidalgo J (1988) Prescribing of psychotropic medication by primary care physicans and psychiatrists. Arch Gen Psychiatry 45: 1117–1119

Boos C, Wulff P, Kujath P, Bruch HP (1998) Medikamentenmißbrauch beim Freizeitsportler im Fitnessbereich. Dtsch Ärzteblatt 95: B774–778

Bottlender R, Schütz C, Möller HJ, Soyka M (1997) Zolpidem dependence in a patient with former polysubstance abuse. Pharmacopsychiatry 30(3): 108–112

Cooperstock R, Parnell P (1982) Research on psychotropic drug use. A review of findings and methods. Soc Sci Med 16: 1179–1196

Geiselmann MB, Linden M, Sachs-Ericsson N (1989) Benzodiazepine prescriptions and therapist non-compliance. Euro Arch Psychiatr Neurol Sci 239: 185–187

Glaeske G (1996) Psychotrope und andere Arzneimittel mit Mißbrauchs- und Abhängigkeitspotential. In: Deutsche Hauptstelle gegen die Suchtgefahren (DHS) (ed) Jahrbuch Sucht 97. Neuland, Geesthacht, pp 32–54

Haupt HA (1993) Anabolic steroids and growth hormone. Am J Sports Med 21: 468–474

Hüllinghorst (1996) Versorgung Suchtkranker in Deutschland. In: Deutsche Hauptstelle gegen die Suchtgefahren. In: Deutsche Hauptstelle gegen die Suchtgefahren (DHS) (ed) Jahrbuch Sucht 97. Neuland, Geesthacht, pp 128–142

Jick H, Vasilakis C, Weinrauch LA, Meier CR, Jick SS, Derby LE (1998) A population-based study of appetite-suppressant drugs and the risk of cardiac-valve regurgitation. N Engl J Med 339(11): 719–724

**Julien RM (1997) Drogen und Psychopharmaka. Spektrum, Heidelberg

Kashkin KB (1992) Anabolic steroids. In: Lowinson JH, Ruiz P, Millman RB, Langrod JG (eds) Substance abuse: a comprehensive textbook, 2nd edn. Williams and Wilkins, Baltimore, pp 380–395

**Keup W (1993) Mißbrauchsmuster bei Abhängigkeit von Alkohol, Medikamenten und Drogen. Frühwarnsystem-Daten für die BRD 1976 bis 1990. Lambertus, Freiburg

Keup H (1998) Zolpidem und Zopiclon geringeres Mißbrauchspotential im Vergleich zu Benzodiazepinen. Arzneimitteltherapie 16: 246–253

Khan MA, Herzog CA, St. Peter JV et al (1998) The prevalence of cardiac valvular insufficiency assessed by transthoracic echocardiography in obese patients treated with appetite-suppressant drugs. N Engl J Med 339(11): 713–718

Kraus L (1996) Ergebnisse der Repräsentativerhebung zum Gebrauch psychoaktiver Substanzen 1995. Neuland, Geesthacht

Kraus L, Bauernfeind R (1998) Repräsentativ-Erhebung zum Gebrauch psychoaktiver Substanzen bei Erwachsenen in Deutschland 1997. Sucht 44 (special issue 1)

Larson DB, Lyons JS; Hohmann AA, Beardsley RS, Hidalgo J (1991) Psychotropics prescribed to the US elderly in the early

and mid 1980s: prescribing patterns of primary care practitioners, psychiatrists, and other physicans. Int J Geriatr Psychiatry 6: 63–70

Linden M, Bär T, Geiselmann B (1998) Patient treatment insistence and medication craving in long-term low-dosage benzodiazepine prescriptions. Psychol Med 28: 721–729

*Poser W, Poser S (1996) Medikamente – Mißbrauch und Abhängigkeit. Thieme, Stuttgart

Poser W, Poser S, Thaden A, Eva-Condemarin P, Dickmann U, Stötzer A (1990) Mortalität bei Patienten mit Arzneimittelabhängigkeit und Arzneimittelabusus. Suchtgefahren 36: 319–331

Poser W, Poser S, Eva-Condemarin P (1992) Mortalitity in patient with dependence on prescription drug. Drug Alcohol Depend 30: 49–57

Schlund GH (1998) Rechtliche Aspekte des Arzneimittelmißbrauchs. Dtsch Ärzteblatt 95: B779–781

Soyka M (1997) Alkoholismus. Wissenschaftliche Verlagsgesellschaft, Stuttgart

**Soyka M (1998) Medikamenten- und Drogenabhängigkeit. Wissenschaftliche Verlagsgesellschaft, Stuttgart

**Soyka M (1999) Alkoholabhängigkeit. Springer, Berlin Heidelberg New York

Soyka M, Steinberg R, Vollmer M (1988) Entzugsphänomene bei schrittweisem Benzodiazepinentzug. Nervenarzt 59: 744–748

Thome J, Wiesbeck GA, Becker T (1997) Zum Abhängigkeitspotential der nicht-Benzodiazepin-Hypnotika Zolpidem und Zopiclon. Nervenheilkunde 16: 575–578

Weyerer S, Zimber A (1997a) Psychopharmakagebrauch und -mißbrauch im Alter. In: Förstl H (ed) Lehrbuch der Gerontopsychiatrie. Enke, Stuttgart

Weyerer S, Zimber A (1997b) Abhängigkeit und Mißbrauch von Alkohol und Medikamenten in Alten- und Pflegeheimen. In: Watzl H, Rockstroh B (eds) Abhängigkeit und Mißbrauch von Alkohol und Drogen. Hogrefe, Göttingen, pp 159–184

CHAPTER 18

L.G. Schmidt

Alcoholism: Aetiology, Epidemiology and Diagnosis

Translator: S. Johnson

1 Aetiology

The term alcoholism is often used to encompass both alcohol misuse and alcohol dependence. It describes repeated impairment of an individual's health and social functioning due to frequent alcohol consumption. This disorder has a multifactorial aetiology. Thus a variety of factors influence whether a young person decides to become a regular drinker, and other circumstances determine whether the excessive drinking which may follow persists despite consequent health and social problems and becomes uncontrollable, so that it can generally no longer be given up.

1.1 Psychological Models

Psychological models are generally based on the experiences of people who have not developed alcoholism. They tend to report that alcohol in low concentrations can reduce the feelings of tension associated with stressful situations, such as after the mental effort involved in a difficult day at work or when coping with a new situation (tension reduction hypothesis). However, consumption of large quantities of alcohol results rather in the opposite effects, especially once blood alcohol levels begin falling. In this instance, people tend to experience an increase in muscle tension and greater feelings of unease and tension.

Excessive alcohol drinking is a behaviour which can be explained according to the established principles of learning theory. Thus the reduction in tension due to an ungratified wish and the positive subjective effects resulting from alcohol act as a reinforcer, so that the probability of further behaviour oriented towards consumption increases (operant conditioning). Classical conditioning processes also play a significant role, as previously neutral stimuli which become associated with alcohol drinking become triggers for craving for alcohol ("cue reactivity") and hence for the pattern of consumption with which they are associated. It has thus been demonstrated that even for subjects who are not dependent, expectations regarding effects associated with alcohol ("alcohol expectancies") closely correlate with actual alcohol consumption (Cooper et al. 1992).

Investigations in personality psychology have not led to identification of any cluster of characteristics which might constitute a specific "addictive personality" structure and thus be predisposing for the development of dependence (Nathan 1988). The generally accepted psychodynamic view is that alcohol weakens the demands experienced from a stern superego. Classical psychoanalytic theories assume that at least a proportion of those with alcohol problems have become fixated at the oral stage of psychic development, so that drinking alcohol makes frustrations more bearable. Alcohol is also postulated as a means by which people with narcissistic disturbances may experience feelings of power and importance which might not otherwise be accessible to them. Some empirical studies indicate that certain psychopathological factors, particularly antisocial personality disorder, are often associated with alcoholism. Raised rates of co-morbidity with anxiety disorders have been found in some epidemiological studies, but the prevalence of depressive illnesses among relatives of people with alcoholism remains in dispute (Maier 1995).

Some prospective studies have indicated that early experiences of loss have an important role in the development of psychiatric illnesses in adult life (Vaillant and Milofsky 1982). With alcoholism as with depression, loss through separation rather than through death has a particular association with later illness. This demonstrates the special importance of the quality of emotional relationships in the family. Growing up in a dysfunctional family characterised by a high level of psychological, physical or sexual violence is likely to constitute an important risk factor for the later development of alcoholism. Substance misuse by parents and norms in the peer group are influences which shape the drinking style of young people as they grow up, but which probably have less effect on the extent of associated problems related to alcohol (Berman and Noble 1993; Oyefeso 1994). Unemployment emerges as a particular risk factor for increased alcohol consumption among young people (Janlert and Hammarström 1992).

1.2 Sociocultural Models

The influence of ethnic and cultural factors in moulding the drinking customs of a society is well-known. In some cultures, stern religious prescriptions regulate alcohol use. On the whole, measures which reduce access to substances with addictive potential (age restrictions, taxes) seem to be more effective in limiting addiction problems than punitive measures (convictions and imprisonment). In the past, the so-called abstinence cultures (e.g. Islam, where there are religious sanctions against use of alcohol) have been distinguished from permissive cultures (e.g. the Mediterranean countries, where social drinking customs tend to promote a pathological pattern of continuous steady drinking and consequent inability to abstain) or ambivalent cultures (e.g. American puritanism; Scan-

dinavia, where because of a tight, socially controlled drinking style, excess typically takes the form of episodic binges with loss of control). Where alcohol use is not embedded in the rites and norms of a culture, the effects of alcohol can turn out to be very destructive, as indicated by the example of the American Indians or Australian Aborigines. When evaluating observations of this type, the potentially important role of genetic variations needs also to be remembered (e.g. in relation to availability of the enzymes which break down alcohol). Overall, socio-cultural theories have significant heuristic value, but are generally not verifiable or falsifiable through empirical data.

1.3 Biological Models

1.3.1 Predisposition to Excessive Drinking

The clustering of alcoholism in families has been recognised since the classical works of Aristotle and Plutarch and has also been established in many more recent studies. Family investigations demonstrate the power of genetic factors, indicating that the relative of an alcohol-dependent person is three or four times as likely to develop an alcohol problem as someone without such a family history. Further, twin studies regarding children of alcoholic parents show higher concordance rates for alcohol-related disorders for identical than for fraternal twins. Studies in which data from twins are used to construct models quantifying the relative influences of family and individual-specific environmental factors suggest that individual-specific environmental factors (e.g. critical life events in adult life) have about as strong an effect as genetic factors (Maier 1995).

Family environment seems above all to transmit a similar drinking style and in this way to influence the development of excessive consumption (as defined in the misuse concept) and alcohol-related problems. Its effect on whether alcohol dependence develops seems to be negligible in comparison. Finally, adoption studies have shown that the adopted-away children of alcoholic parents retain a higher risk of alcohol disorders, even if they have grown up with healthy adoptive parents. The risk is not increased much further if the adoptive parents themselves have alcoholism.

The genetic background of alcoholism is complex: it is accepted that multiple genetic factors interact with environmental factors. Thus, for example, a subtype of alcoholics with high genetic predisposition, above all men with antisocial personality characteristics and early illness onset (type II, "male-limited"), may be distinguished from a second subtype with more limited genetic loading and later onset (type I, "milieu-limited") (Cloninger 1987a). It is unlikely that a single major locus is responsible for all familial variants of alcoholism. Rather, genetic variance is likely to be explained much more through a multiplicity of susceptibility genes, whose products presumably influence the effect of alcohol in particular ways. Alcoholism research using animal models has indicated that the various components of the dependence syndrome, such as alcohol preference, sensitivity, neuroadaptation and withdrawal symptoms, may be separated from one another in terms of genetics and behavioural biology (Crabbe et al. 1994). Each behavioural characteristic is probably determined by the interaction of multiple genes (Lander and Schork 1994). Individually, these genes are probably neither necessary nor sufficient to produce the characteristic and have on their own only a limited role in shaping the characteristic ("quantitative trait loci").

Little is known about the influence of specific gene variations on illness risk. The only risk-modulating genes identified with any certainty act via alcohol dehydrogenase (ADH) and acetaldehyde dehydrogenase (ALDH) to control alcohol metabolism. Thus it has been known for some time that Asians who develop a flushing syndrome after alcohol consumption have a degree of protection from alcoholism because of an isoenzyme deficiency. The flushing syndrome primarily reflects a raised acetaldehyde level, resulting from a deficiency in mitochondrial (ALDH2). This occurs only in Asians and arises through a point mutation on chromosome 12. Thus Asians who are homozygous for this variant (ALDH2*2) have an 18 times lower risk of alcoholism than control subjects. However, this gene variation is too rare in European populations to play any significant protective role among them.

Because of the particular significance of the dopamine-driven mesolimbic–mesocortical reward and motivation system in addiction processes, researchers have recently concentrated on investigating the genes which regulate dopamine metabolism as candidate genes. Since 1990, the taqA-polymorphism ($DRD_2{}^*A_1$ allele), which lies very close to the dopamine D_2 receptor gene on chromosome 11, has been investigated in many association studies. Initial positive findings about its association with alcoholism (and other disorders such as drug dependence, Tourette's syndrome, autism) could not, however, be replicated in subsequent studies. This was attributable to confounding differences, above all ethnic differences, between patients and control groups in the original work. Thus population genetic studies have shown that the $DRD_2{}^*A_1$ allele varies ten-fold in frequency between different ethnic groups and is also combined in different ways with other polymorphic systems

(haplotypes) and genes (gene interactions). In any case, the first published haplotype investigations from the United States and Taiwan have shown no association with alcoholism (Kidd et al. 1996).

Despite these largely negative results, individual variations of a large number of neurotransmitter, neuromodulator, receptor and transporter systems may be significant for the development and maintenance of alcohol dependence. Thus my own investigations suggest that the A_9 allele of the dopamine transporter has a modifying influence on development of the withdrawal syndrome (Sander et al. 1997; Schmidt et al. 1998). The U.S. Collaborative Study on the Genetics of Alcoholism (COGA) project constitutes a hypothesis-free approach to discovering unknown gene regions which contribute to vulnerability to alcoholism and to alcoholism-correlated phenotypes such as personality type, blood and electroencephalographic (EEG) markers. In this project, the whole genomes of members of 105 "informative" families containing alcoholic probands were systematically screened with the aid of genome scanning in sections of 20 centimorgans (this corresponds to a region of 20 million base pairs, on which generally around 300 genes are coded) (Reich et al. 1998). This process gave indications of linkage with alcoholism for gene locuses on chromosomes 1 and 7, and for a locus on chromosome 4 which appears to exert a protective effect via the nearby ADH locus.

As well as a genetically linked predisposition, there are also indications that early acquired damage in the form of organic lesions or psychological experiences of trauma give rise to enduring characteristics ("traits"), which can in turn promote the development of alcohol dependence. For example, in subjects with a history of perinatal trauma who later develop alcoholism, a wide range of disturbances of functioning, emotions and behaviour have been found, often in the context of a hyperkinetic syndrome (Donovan 1986). In these cases, objective evidence can be found of disturbances of neuronal information processing, particularly in the forms of deficits in neuropsychological performance and neurophysiological abnormalities (flattening of the P300 amplitude in the evoked potentials, low α_a activity on EEG examination).

There is increasing evidence that early psychosocial stressors are associated with variations between individuals in the development of stress sensitivity and later alcohol responsiveness. High sensitivity to stress and marked damping of stress through ethanol seem to favour the development of alcohol dependence, a finding supported by high-risk paradigm research (Finn and Pihl 1987). In animal research, it has been shown that the hypothalamic–hypophyseal–adrenal system (HPA system) is an important biological substrate for individual differences in sensitivity to alcohol (and drugs) and influences individual readiness to self-administer addictive substances (Fahlke et al. 1994). This is also supported by findings that corticotrophin-releasing factor (CRF) is present in smaller concentrations in the hypothalamus, amygdala and prefrontal cortex in alcohol-preferring rats (developed through breeding) than in non-alcohol-preferring rats (Ehlers et al. 1992).

Corresponding findings from animal studies indicate abnormalities in the central serotonin system in subjects who have been subject to a social stress ("handling") immediately after birth or who have grown up with their peers rather than their mother (Higley et al. 1991). It may be that, in humans, reduced functional capacity in the central serotonin system represents a neurobiological correlate of reduced affect and impulse control. Increased susceptibility to anxiety or uncontrolled impulsivity associated with a serotonergic abnormality such as this (as opposed to testosterone-linked competitive aggressiveness) have also been considered as potential precursors to human alcoholism and elements in manifest alcohol dependence (Fils-Aime et al. 1996).

The tendency of alcohol (and other addictive drugs such as heroin, cocaine and nicotine) to modify the behaviour and motivation of organisms in a very characteristic way, so that they become oriented towards renewed self-administration ("positive reinforcement"), is explained in terms of effects on reward centres in the brain. These centres are usually activated by naturally arising stimuli such as nourishment or sexual activity, albeit not as intensively as by alcohol. A central neuronal cluster among the structures designated by Olds (1956) as pleasure centres is the mesolimbic–mesocortical dopamine system, which can be activated by alcohol and other addictive drugs (DiChiara and Imperato 1988). This projects from the ventral tegmental area (VTA) to the nucleus accumbens and beyond this to the prefrontal cortex and other limbic structures. In relation to this, Wise (1996) has developed the dopamine hypothesis of addiction, based on the psychomotor activating and motivational effects of dopamine after administration of addictive substances. According to the findings obtained by Herz (1995), processes dependent on dopamine (D_1) play a particular role in the perception of stimuli as rewarding or unpleasant. DiChiara and North (1992) distinguish processes involving psychomotor activation and orientation of the organism towards potentially rewarding, "incentive" (and conditioned) stimuli, which are all mediated by dopaminergic neurones, from the predominantly opioid mechanisms involved in the act of consumption. Thus both systems act as interfaces for conversion of motivations into behaviour. In addition, an important function in the anticipation and prediction of so-called incentive stimuli is ascribed

to certain dopaminergic neurones (Schultz 1997). Finally, this mesolimbic–mesocortical reward system is also modulated by γ-aminobutyric acid (GABA)-ergic, glutamatergic and serotonergic input.

The dopamine- and opioid-dependent interactions of alcohol with the GABAergic system in other regions of the brain (and voltage-gated calcium channels) is also of substantial clinical significance, as feelings of relaxation and calm and possibly also anxiolytic effects depend on these. Self-administration experiments with alcohol indicate that there are individual variations in the responsiveness of this system at systemic, cellular and molecular levels (DeWitt et al. 1989), which exist independently of the functioning of the stress axis (Piazza et al. 1996). For example, Wafford et al. (1991) have reported that potentiation of GABA effects of alcohol (in low doses) is dependent on the γg_{2L} subunit of the pentameter $GABA_A$-benzodiazepine (BZD)-receptor complex and that such potentiation no longer takes place with point mutations or deletions at this site. The findings by Volkow et al. (1993, 1995) indicate that people with alcohol dependence have reduced GABAergic activity in particular regions of the brain (e.g. orbitofrontal cortex, striatum, thalamus) and that this is manifest especially under stress and may correspond with excessive dopamine release in response to new stimuli. Compensation for this GABAergic deficit through alcohol use could thus lead to alcoholism.

Another target system for alcohol is the glutamatergic system, most important for its excitatory effects in the central nervous system (CNS), where it is the main system acting in opposition to the GABAergic system. This system is also characterised by great heterogeneity, e.g. in the structure and functioning of *N*-methyl-D-aspartate (NMDA)-linked ionic channels and the receptor subtypes and splicing variants associated with these, a heterogeneity which is reflected in varying degrees of alcohol sensitivity (Tsai et al. 1995).

Finally, clinical investigations show in a very general way that reduced sensitivity of neurobiological systems to alcohol is an important element in the development of alcohol dependence (Schuckit 1994). Thus sons from families with histories of alcoholism, when compared with those with no such family history, experience specific intoxicating doses of alcohol as less intensive in their effects, they respond less ataxically to motor tests ("body sway") and are less responsive to neuroendocrine tests (e.g. in terms of reduced adrenocorticotropic hormone, prolactin and cortisol release). Genetically and probably also environmentally related disturbances of neuronal information processing, recognisable through a reduced P3 amplitude in event-related potentials, further increase the risk of alcoholism (Begleiter et al. 1984). These findings are interpreted as meaning that, in a society in which alcohol drinking is widespread, appropriately predisposed individuals will experience few aversive effects from it (e.g. headaches, giddiness, nausea). Such effects are therefore less likely to prevent them from continuing to drink, and they will therefore experience higher ethanol concentrations, which are associated with greater risk of dependence and adverse consequences.

1.3.2 Development of Alcohol Dependence

The indicator of the onset of dependence is the transition from controlled to uncontrolled alcohol consumption. A requirement for this is repeated or chronic exposure of the brain of an organism to alcohol, with such a transition taking place above all when alcohol intake is determined by free choice (Wolfgramm 1995). The most important development in addiction research in recent years has been the recognition that the cerebral processes arising from positive reinforcement are progressively altered by alcohol- and drug-related sensitisation processes, providing the best explanation for the escalation of craving (Robinson and Berridge 1993).

Sensitisation to addictive substances involves a state of increased responsiveness of the dopaminergic mesolimbic reward system to the effects of addictive substances. This state is recognisable through an increased tendency to psychomotor activation (raised motility), which can be induced in particular through administration of psychostimulants such as cocaine or amphetamine (Wise 1996) and may also be triggered to a degree by alcohol (Weiss et al. 1993). It is further manifested in an increased sensitivity of this system to the motivational effects of addictive substances. A key aspect of increasing sensitisation is that euphoria-inducing effects do not increase, but the "incentive salience" of alcohol, drugs or associated stimuli is experienced and influences behaviour more and more forcefully. The dopaminergic reward system is thus the central neuronal circuit for anticipation and assessment of alcohol and drug-related cues. It awards these stimuli, especially those which have activated it, greater and greater attractiveness as consumption increases, and is thus also involved in the development of "craving". These attractive (incentive) effects of alcohol and drugs are distinguished from effects related to consumption, which are predominantly mediated by opioid pathways.

1.3.3 Persistence of Dependence and Vulnerability to Relapse

Dependence, tolerance and tendency to relapse (vulnerability to re-exposure) can be seen as neuroadaptive

sequelae to chronic alcohol exposure, arising at intersystemic, cellular and molecular levels (Koob and Bloom 1988). If we trace the dependence-inducing (i.e. anaesthetic–narcotic) effect of alcohol back to its physicochemical effects on the nerve cell membrane (fluidisation), many selective effects on particular neuronal systems, receptor subtypes, ion channels and intracytoplasmic and intranuclear processes have been identified and form the basis of addiction-specific changes in neurones (Nestler and Aghajanian 1997).

Thus, after acute consumption of addictive substances, the immediate early genes are first switched on and code for transcription factors (fos, jun), through which short-term and temporary effects on other target genes are exercised. With the transition to chronic consumption, the G-protein pattern in the nerve cell membrane changes, resulting in modification of the processes of signal transduction. Finally, formation of the so-called late (onset) Fos-related antigens (F ras) results in activation of the late (onset) genes. These alter the gene expression of the neurones in a long-term, possibly permanent, way. Thus a modulation occurs in the genomic programme, involving alteration in the work plan of the neurone cells and long-term changes in cell proliferation, task differentiation and specific functions.

These possibly irreversible changes in brain metabolism, which are associated with alteration of gene expression by addictive substances, may represent the molecular correlate of the phenomenon called addictive memory. A particular role may be played in its development by glutamatergic neurones, which develop certain connectivities through the mechanism of long-term potentiation and the close proximity of glutamatergic and opiodergic synapses in the hippocampus. When the different neuronal units are considered at an intersystem level, increasing dependence is accompanied by an increasing imbalance in facilitatory and inhibitory processes. As these changes at molecular, cellular and intersystemic levels probably do not all begin at precisely the same time, the "point of no return" or transition from controlled to dependent consumption is probably not a distinct event, but a continuous process (Coper et al. 1990).

Well-being is motivated by the experience of pleasant or euphoric states ("positive reinforcement"). Consumption of addictive substances has traditionally been regarded as maintained by a motivation to avoid or limit withdrawal symptoms ("negative reinforcement"), which tends to lead to renewed intake of the addictive substance. Typical of this are the morning withdrawal symptoms of people with alcohol dependence, which result from falling alcohol concentrations and are alleviated by consumption of alcohol on an empty stomach. Other neurobiological systems (e.g. the locus coeruleus) have also been ascribed behaviour-motivating functions in connection with this process; however, the significance of physical dependence-maintaining systems for perpetuation of addictive behaviour is not now believed to be very great.

In the context of addictive illness, the term relapse refers to renewed consumption of the addictive substance with early loss of control over this behaviour and accelerated reinstatement of the full dependence syndrome (Edwards and Gross 1976). This process requires a reward system which is already sensitised through conditioning. It begins when re-exposure to cue constellations previously closely associated with consumption of addictive substances sets off neurobiological processes which are themselves very similar to the effects of alcohol and drugs.

Thus the "craving" which induces relapse may result from a process in which substance-associated cue constellations (such as an alcohol-dependent person's visit to their regular local bar) trigger automatic processes resulting in dopamine release in a sensitised reward system. Internal processes (e.g. particular mood states) in an alcohol-dependent person may also induce such a process, as may the first mouthful of alcohol ("priming"), which is recognised as often leading to a stronger craving for drink and renewed addictive behaviour.

This means that an addict is characterised by a particular vulnerability to re-exposure, which fundamentally differentiates him or her from a healthy person and which cannot be reversed (Coper et al. 1990). Euphoric experiences are not "forgotten" either by an individual or by the relevant neurobiological structures in his or her brain, but retain a lasting influence. Knowledge is still limited about the particular molecular effects of excessive drinking on the brain maturation of young people and about the details of the psychological processes associated with this.

2 Epidemiology

2.1 Prevalence

In 1993, Germany had the highest per capita alcohol consumption of all nations, at 12.3 l of pure alcohol per annum. Surveys indicate that around 90% of the population have at some time tried alcohol, and around 60%–70% currently drink. Figures from the Deutsche Hauptstelle gegen die Suchtgefahren (the central German organisation concerned with substance misuse) suggest that around 5% of men and 2% of women are dependent on alcohol. However, this is probably an underestimate, as the method of case

definition does not correspond to modern diagnostic criteria. On the basis of such operationalised diagnostic criteria, 13% of the general population in Germany and in the United State report a lifetime history of alcohol use which is a hazard to their health (around 21% of men and 5% of women) (Bronisch and Wittchen 1992; Robins et al. 1984).

The 6-month prevalence of alcohol related disorders is around 5.1% (3.1% for conspicuous to severe forms; Fichter et al. 1996). According to studies by the World Health Organization (WHO), 6% of patients in German general practices are alcohol dependent (1-month prevalence), while a further 4% meet criteria for the diagnosis of harmful use or misuse. According to more recent general population investigations, 9.7% of young people between 14 and 24 years have a misuse and 6.2% a dependence diagnosis according to DSM-IV (Holly et al. 1997). Thus alcohol misuse and dependence are among the commonest forms of psychological disorder in the Western world (apart from nicotine dependence).

A wide range of consequences result from this. For example, foetal alcohol syndrome is the commonest cause of mental handicap, and alcoholism the most frequent cause of malnutrition states among adults. A total of 25% of all accidents at work are related to alcohol, and alcohol plays a part in around 20% of fatal road traffic accidents. While moderate alcohol consumption may have a protective effect against coronary heart disease, heavy misuse, especially when combined, as it often is, with tobacco consumption, greatly increases the risk of carcinomas of the oral cavity, pharynx, larynx and oesophagus. A total of 15% of people with alcohol disorders meet criteria for cirrhosis of the liver; and suicide, alcoholic myopathy and hypertension also contribute to an overall shortening of life expectancy of around 10–15 years.

Another form of impairment is alcoholic dementia, estimated as contributing 10% to the total number of cases of dementing illnesses. Thus, in Germany, around 30,000–40,000 people a year die of the consequences of excessive alcohol consumption, with smoking also contributing to this increased mortality. In American healthcare services, around 15% of the budget is taken up by alcohol-related disorders. The total direct and indirect economic costs (through loss of working hours, consequences of accidents, early retirement, treatment costs for related illnesses, rehabilitation measures) are likely to amount to more than DM 80 billion a year in Germany. This may be compared with state income of around DM 8 billion from taxes on alcohol (and DM 13 billion from tobacco).

The manifestations of alcohol-related disorders are extremely variable. While Jellinek (1960) still assumed alcoholism (at least the γ-form) to be a relentlessly progressive type of illness, the longitudinal studies conducted by Vaillant (1996) and the group working with Edwards (Taylor et al. 1986) indicate that individuals have greatly fluctuating symptom courses with varying degrees of severity. It has also been observed that the earlier alcohol-related problems arise ("early-onset alcoholism"), the more frequent co-morbid psychiatric disorders, especially antisocial personality disorders, are. The latter are characterised by behavioural abnormalities in family life, school and peer group and are frequently also associated with drug misuse.

In a second group of alcohol-dependent subjects ("late onset alcoholism"), medical problems are particularly prominent, while social difficulties are rather less conspicuous. Overall, however, the progression of alcoholism does have a typical course (Schuckit et al. 1993; Nelson et al. 1996); excessive drinking (end of the second decade of life) is followed by disturbances in a wide variety of areas of life, after which loss of control, development of tolerance, withdrawal manifestations and finally severe professional, social and health difficulties develop (middle to end of the thirties). Mortality is influenced by volume of drinking, health problems (above all liver diseases) and also unemployment; a good prognosis on the other hand is related to generally stressful circumstances, family environment and individual resources for actively combating the illness (Finney and Moos 1992).

2.2 Co-morbidity

Figures for co-morbidity of depression and alcoholism vary between 3% and 98% depending on survey method, diagnostic system used, method of defining illness, time criteria (e.g. primary, secondary) and time course (e.g. incidence, lifetime prevalence). They are also dependent on the context of the survey (e.g. specialist clinic, primary care, general population) and important sociodemographic variables (such as age, gender, marital status, social background, employment or presence of cognitive impairment).

Meta-analyses suggest that around 40% of people with alcohol disorders admitted to a psychiatric clinic have recognisable depressive syndromes. With regard to the sequence of lifetime history, in around 40% of these patients an alcohol problem has appeared first, followed later by depressed mood, i.e. primary alcoholism and secondary (major) depression, and in about the same number the presence of a primary depression prior to alcoholism can be established (time sequence cannot be established in 20%). Careful analysis of the sequence and severity of disorders, however, suggests that only around 5% of alcoholics have experienced an affective illness of the severity of

a major depression before the onset of excessive drinking and alcohol-related problems (Schuckit 1994).

People with alcohol disorders have a higher risk than the general population of developing another psychiatric disorder. The strongest association (in relation to lifetime diagnoses of alcohol dependence and misuse) is found for antisocial personality disorder (indicated by an odds ratio of 21.0); following this are drug misuse (7.2), mania (6.2), schizophrenia (4.0), panic disorder (2.4), dysthymia (1.8) and major depression (1.7) (Helzer and Pryzbek 1988). The current interpretation of the association found between alcohol and anxiety disorders in some studies is that generalised anxiety often precedes alcoholism, whereas panic disorder is usually observed as a sequel to it (Kushner et al. 1990).

3 Diagnosis

3.1 Diagnostic Criteria

Misuse, designated in ICD-10 as harmful use, is characterised according to clinical diagnostic guidelines by a pattern of consumption which results in damage to health. This can be a physical disorder, such as hepatitis after injecting substances, or a psychological disorder e.g. a depressive episode after massive alcohol consumption (WHO 1992). Harmful using behaviour is often criticised by others and often also has a variety of negative social consequences. Disapproval from others or from a whole society of the pattern of use or of a particular substance is not an indicator of harmful use, nor are negative social consequences such as imprisonment, job loss and marital problems (WHO 1992).

The DSM-IV criteria, in contrast, do include the social dimensions, defining alcohol misuse (abuse) as a maladaptive pattern of substance use which leads to clinically significant impairments or suffering, with at least one of the following criteria being met in the course of a 12-month time period (APA 1994):

1. Repeated substance use, which leads to failure to meet important obligations at work, school or home
2. Repeated substance use in situations in which there may be physical danger because of the consumption
3. Recurring problems with the law in connection with the substance use
4. Continuing substance use despite constant or repeated social or interpersonal problems, which are caused or reinforced by the effects of the psychotropic substance

According to ICD-10, a dependence syndrome involves a cluster of physiological, behavioural and cognitive phenomena in which the use of the substance takes on a much higher priority for a given individual than other behaviours that once had greater value. A central descriptive characteristic of the dependence syndrome is the desire (often strong, sometimes overpowering) to take psychotropic substances or medications (whether prescribed by a doctor or not), alcohol and tobacco (WHO 1992).

Accordingly, the diagnosis of dependence should be made only when at some stage in the last year, three or more of the following criteria have been met at the same time:[1]

1. A strong desire or sense of compulsion to take a psychotropic substance.
2. Difficulties in controlling substance-taking behaviour in terms of its onset, termination, or levels of use.
3. A physiological withdrawal state when substance use has ceased or been reduced, as evidenced by substance-specific withdrawal symptoms or use of the same or a closely related substance for the purpose of alleviating or avoiding withdrawal symptoms.
4. Evidence of tolerance, such that increased doses of the psychotropic substance are required in order to achieve effects originally produced by lower doses (clear examples of this are found in alcohol-dependent individuals who may take daily doses sufficient to incapacitate or kill non-tolerant users).
5. Progressive neglect of alternative pleasures or interests because of substance use; increased amount of time necessary to obtain or take the substance or to recover from its effects.
6. Persisting with alcohol use despite clear evidence of overtly harmful consequences, such as harm to the liver through excessive drinking, depressed mood following heavy substance use (or drug-related deterioration in cognitive functioning); efforts should be made to determine that the user was actually, or could be expected to be, aware of the nature and extent of the harm.

Narrowing of the personal repertoire of patterns of substance use has also been described as a characteristic feature (e.g. a tendency to drink alcoholic drinks in the same way on weekdays and weekends, regardless of social constraints that determine appropriate drinking behaviour) (WHO 1992). The seven DSM-IV

[1]The first edition of ICD-10 had six criteria, whereas in the second the two concerning withdrawal were merged and the criterion concerning stereotyped repertoire of alcohol consumption removed.

criteria for dependence are very comparable to the six ICD-10 criteria, with the additional criterion no. 5 in DSM-IV ("great deal of time spent on obtaining the substance, taking it, or recovering from its effects") subsumed under criterion no. 5 in ICD-10. A more specific coding is made in ICD-10 to describe course (abstinent, current substance use, constant or episodic substance misuse), while in DSM-IV there is also subtyping in relation to the presence or absence of physiological dependency.

3.2 Empirical Research on Diagnosis

Two lines of development may be traced in the history of diagnostic practices in alcohol-related disorders (Schmidt 1995). While Huss, Jellinek and latterly a large American specialist association have adhered to a categorical or unitary concept of alcoholism, the WHO, influenced by Edwards (Edwards and Gross 1976) have developed a biaxial (or dichotomous) classification. This biaxial classification differentiates patients with alcohol dependence from those who are not dependent but incur alcohol-related disabilities through habitually drinking excessive quantities of alcohol. The definition of misuse or harmful use in ICD-10 is rather vague (based on an intuitive global judgement without time or frequency criteria), and correspondingly the reliability of this diagnosis has also proved to be low. For example, prevalence rates for harmful use (as defined by ICD-10) are around twice as high as for misuse (at least when still applying DSM-III-R).

Longitudinal investigations have, however, also shown that for 70% of patients with a misuse diagnosis, the diagnosis remained unaltered after 4 years. A series of investigations have shown that an overall distinction can be made between two groups of excessive drinkers: those who are not dependent and can return to asymptomatic drinking, and a dependent group for whom this is no longer the case. An research question still to be fully resolved is for which patients and in what circumstances excessive drinking does eventually lead to dependence.

Studies investigating the reliability and validity of the dependence syndrome indicate that both diagnostic systems result in similar prevalence rates and that there is evidence of the same syndrome in a great variety of the countries. Investigations into construct validity (at least when based on the dependence syndrome as defined by Edwards) have shown that the symptoms load onto a common factor and correspond to clinical degree of severity.

Evidence for the unidimensional nature of the dependence syndrome is its extensive independence from measures of damaging consequences, while its universality is demonstrated by evidence of its existence for a range of different substances. The syndrome becomes more homogeneous as severity increases. Thus the high inter-correlations found in investigations in clinical groups have not been replicable among young people or in population surveys. In relation to this, it should be noted that ICD-10 and DSM-IV are "polythetic" systems, in which no single diagnostic criterion is necessary or sufficient and the diagnosis of dependence can be made when the minimum quantity of three criteria (not empirically validated) is reached.

In contrast, Edward's concept of the dependence syndrome is nomothetic, i.e. the set of criteria he proposes (narrowing of the drinking repertoire, excessive salience of drinking, increased tolerance, repeated withdrawal symptoms, drinking to relieve or avoid withdrawal symptoms, subjective awareness of a compulsion to drink, rapid reinstatement of withdrawal symptoms after abstinence) are regarded as sufficient and necessary for a diagnosis to be made, and all the phenomena described in these criteria are expected to be present to a lesser or greater extent where dependence is diagnosed. According to this concept, dependence is a unitary syndrome, whose development is facilitated or inhibited by a variety of factors, above all psychopathological ones.

3.3 Instruments

A large variety of standardised instruments are available to detect and describe alcohol-related disorders (Table 1). The screening instruments most frequently used in English-speaking countries to detect clear-cut cases of alcoholism in clinical research or epidemiology are the Michigan Alcoholism Screening Test (MAST), the MacAndrew Scale and also the CAGE questions. The MALT and the *Kurzfragebogen für Alkoholgefährdete* (KFA) are the principal instruments used in German-speaking countries. The Alcohol Use Disorders Identification Test (AUDIT), which consists of ten questions, is suitable for early recognition.

Detailed questionnaires such as the *Trierer Alkoholismus Inventar* (TAI) or the *Fragebogen zur Klassifikation des Trinkverhaltens* (FTA) are available as instruments for clinical case diagnosis. The goal of the Addiction Severity Index (ASI) and the Severity of Alcohol Dependence Questionnaire (SADQ) is to measure the severity of dependence. The *Lübecker Alkoholabhängigkeitsskala* (LAS) represents a new development in quantification of the development of alcohol dependence and the secondary characteristics of the dependence syndrome. The *Baseler Drogen- und Alkoholfragenbogen* (BDA) has been developed as a

Table 1. Screening instruments for alcoholism

Instrument	Full title	Author(s)	Items	Comments
CAGE	Detecting Alcoholism – the CAGE questionnaire	Ewing (1984)	4	Brief, for routine use
MAST	Michigan Alcoholism Screening Test	Selzer (1971)	25	For primary care
MAC	MacAndrew Alcoholism Scale	MacAndrew (1965)	49	Part of the MMPI
KFA	Kurzfragebogen für Alkoholgefährdete	Feuerlein et al (1976)	22/23	Four domains of drinking behaviour
MALT	Münchner Alkoholismus Test	Feuerlein et al (1977)	24 self- and 7 observer-rated items	Self- and observer-rating
AUDIT	Alcohol Use Disorders Identification Test	Saunders et al (1993)	10 core and 8 clinical questions	Early detection

method of delineating multiple dependence. For research purposes, ICD-10 recommends the use of "craving scales" to document desire for alcohol (e.g. *Lübecker Craving-Risiko-Rückfall-Fragebogen*, LCRR; Obsessive Compulsive Drinking Scale, OCDS) (Table 2).

Structured questionnaires have also been developed in which generally comprehensible questions have been formulated on the basis of the operationalised criteria found in modern diagnostic systems. These are indispensable not only in clinical diagnosis, but also in precise epidemiological research. For example, the Composite International Diagnostic Interview (CIDI) is a fully standardised interview, which allows diagnosis based on ICD-10 and DSM-III-R criteria (WHO 1990). The core version of the CIDI covers a comprehensive range of psychiatric disorders. However, if more detailed information is needed about time

Table 2. Instruments for clinical case diagnosis

Instrument	Full title	Authors	Items	Comments
AUI	Alcohol Use Inventory	Wanberg et al (1977)	228/147	16 areas of dependent drinking behaviour
TAI	Trierer Alkoholismus Inventar	Funke et al (1987)	90	German adaptation of the AUI; 7 subscales
FTA	Fragebogen zur Klassifikation des Trinkverhaltens	Roth (1986)	89	Profile analysis, typology based on Jellinek
ASI	Addiction Severity Index	McLellan et al (1980)	9 areas	Measurement of severity of dependence
EuropASI	Addiction Severity Index	Gsellhofer et al (1993)	9 areas	
SADQ	Severity of Alcohol Dependence Questionnaire	Stockwell et al (1979)	33	5 scales
GABS	Göttinger Abhängigkeits-Skala	Jacobi et al (1987)	22	5 subscales, German version of the SADQ
LAS	Lübecker Alkohol-Abhängigkeitsskala	John et al (1992a,b)	29	Primary and secondary characteristics of the dependence syndrome
BDA	Baseler Drogen- und Alkoholfragenbogen	Ladewig et al (1976)	59	Multiple dependences
LCRR	Lübecker Craving-Risiko-Rückfall-Fragebogen	Veltrup (1994)	VAS scale and 33 items	Measurement of "craving"
OCDS	Obsessive Compulsive Drinking Scale	Anton et al (1996)	14 and 4 VAS items	Different components of "craving"
OCDS-G	Obsessive Compulsive Drinking Scale	K. Mann and K. Ackermann (in preparation)	14 and 4 VAS items	German version of the OCDS

Table 3. Further instruments for measurement of disorders in people with alcohol problems

Instrument	Full name	System/authors	Comments
CIDI	Composite International Diagnostic Interview	ICD-10; Robins et al (1988)	Standardised diagnostic interview
CIDI-SAM	Composite International Diagnostic Interview – Substance Abuse Module	ICD-10; DSM-III-R; Cottler et al (1989)	Substance-specific questions
SKID	Strukturiertes Klinisches Interview nach DSM-III-R	DSM-III-R; Wittchen et al (1990)	Structured interview
SCAN	Schedules for Clinical Assessment in Neuropsychiatry	ICD-10; DSM-III-R; WHO (1995)	Structured interview
CIWA-Ar	Clinical Institute Withdrawal Assessment for Alcohol (revised)	Sullivan et al (1989)	Withdrawal symptoms
MAWS	Mainz Alcohol Withdrawal Scale	Banger et al (1992)	Withdrawal symptoms
AWDI	Alcohol Withdrawal Diagnostic Inventory Assessment	Sellers et al (1991)	Withdrawal symptoms
BDI	Beck Depression Inventory	Beck et al (1961)	Depressive symptoms
STAI	State-Trait-Anxiety Inventory	Laux et al (1981) based on Spielberger et al (1970)	Anxiety symptoms
TPQ	Tridimensional Personality Questionnaire	Cloninger (1987b), Cloninger et al (1993)	Personality structure (3 or 7 factors
NEO-FFI	Neurotizismus-Extraversion-Offenheit-Fünf-Faktoren-Inventar	Borkenau and Ostendorf (1993) ("big five")	Personality structure based on Costa and McCrae (1989)
FTND	Fagerström Test for Nicotine Dependence	Heatherton et al (1991)	Co-morbid nicotine dependence

course, the various consequences, the quantity and frequency of alcohol consumption (as an indicator of degree of severity) and the pattern of use of different substances in a particular population, the substance abuse module of the CIDI (CIDI-SAM) should be used as a separate instrument (see Table 3).

Biochemical laboratory tests are often also regarded as a helpful aid to diagnosis. Blood alcohol concentration may be measured as an indicator of acute intoxication, which clinically may be associated with disturbances of conscious level, cognitive functions, perception, affect, behaviour or other psychophysiological functions and reactions. Thus the first signs of motor and cognitive impairment appear with a blood alcohol level of 0.2%–0.3%; between 0.8% and 2% there are disturbances of coordination, judgement and affect; over 2% there are disturbances of memory (so-called alcoholic blackouts); and levels over 3% begin to be life-threatening. Other laboratory parameters such as γ-glutamyltransferase (GGT) and mean corpuscular volume (MCV) are often regarded as markers of chronic tissue damage due to alcohol, but their sensitivity and specificity are unsatisfactory. Carbohydrate-deficient transferrin (CDT) is superior in this respect and is suitable for detecting chronic excessive alcohol consumption and even relapse (Schmidt et al. 1997).

Quantitative measures are very useful in the treatment of the withdrawal syndrome, and various scales have been developed for this purpose (Table 3). The Beck Depression Inventory (BDI) is generally recommended as a measure of associated depressive syndrome, Spielberger's *State-Trait Angstinventar* (STAI) for anxiety and, as a measure of personality, Cloninger's Tridimensional Personality Questionnaire (TPQ) or the Neuroticism-Extraversion-Openness Five-Factor Inventory (NEO-FFI) by Costa and McCrae (1989). As well as these, a great variety of inventories of associated complaints and problems are in use. Because of the high prevalence of co-existing nicotine dependence among people with alcohol disorders (between 80% and 90%), measurement of smoking habits is increasingly required. The modern ICD-10 and DSM-IV diagnostic systems are suitable for multiaxial assessment of patients.

Finally, tendencies to lying, dissimulation or falsification (a fundamental characteristic of people's reports regarding alcohol) may limit the validity of diagnostic results. However, people with alcohol disorders have no greater overall tendency to lie in self-reports than is encountered in the non-alcoholic population. Dissimulation occurs above all in diagnostic interviews for forensic purposes, so that 50% of forensic patients are thought to conceal their alcohol problems.

Alcohol-dependent patients also dissemble regarding their problems much more frequently in the phase of initial clinical contact than directly after withdrawal or at the beginning of residential treatment. Typically, the distortion consists of alcoholic patients willingly recognising their problems, but seeing them more as cause than as consequence of drinking.

In view of the immense health and socio-political significance of alcoholism, techniques for early diagnosis and brief intervention have been recently developed. These have proved effective particularly as a means of reducing psychological and physical damage among patients with less severe and chronic alcohol problems (Schmidt 1997).

4 References

Anton RF, Moak DH, Latham PK (1996) The Obsessive Compulsive Drinking Scale. Arch Gen Psychiatry 53: 225–231

APA (1994) Diagnostic and statistical manual of mental disorders, 4th edn (DSM-IV). American Psychiatric Association, Washington, DC

Banger M, Benkert O, Röschke J, Herth T, Hebenstreit M, Philipp M, Aldenhoff JB (1992) Nimodipine in acute alcohol withdrawal state. J Psychiatr Res 26: 117–123

*Begleiter H, Porjesz B, Bihari B, Kissin B (1984) Event-related potentials in boys at risk for alcoholism. Science 225: 1493–1496

Berman SM, Noble EP (1993) Childhood antecedents of substance abuse. Curr Opin Psychiatry 6: 382–387

Borkenau P, Ostendorf F (1993) NEO-Fünf-Faktoren-Inventar (NEO-FFI) nach Costa und McCrae. Hogrefe, Göttingen

Bronisch T, Wittchen HU (1992) Lifetime and 6-month prevalence of abuse and dependence of alcohol in the Munich follow-up study. Eur Arch Psychiatry Clin Neurosci 241: 273–282

**Cloninger CR (1987a) Neurogenetic adaptive mechanisms in alcoholism. Science 236: 410–416

Cloninger CR (1987b) A systematic method for clinical description and classification of personality variants. Arch Gen Psychiatry 44: 573–588

Cloninger CR, Svrakic DM, Przybeck TR (1993) A psychobiological model of temperament and character. Arch Gen Psychiatry 50: 975–990

Coper H, Rommelspacher H, Wolfgramm J (1990) The "point of no return" as a target of experimental research on drug dependence. Drug Alcohol Depend 25: 129–134

Cooper ML, Russell M, Skinner JB, Frone MR, Mudar P (1992) Stress and alcohol use: moderating effects of gender, coping and alcohol expectancies. J Abn Psychol 101/1: 139–152

Costa PT, McCrae RR (1989) NEO-PI/FFI manual supplement. Psychological Assessment Resources, Odessa/FL

Cottler LB, Robins LN, Helzer JE (1989) The reliability of CIDI-SAM: a comprehensive substance abuse interview. Br J Addict 84: 801–814

Crabbe JC, Belknap JK, Buck KJ (1994) Genetic animal models of alcohol and drug abuse. Science 264: 1715–1723

DeWitt H, Pierri J, Johanson CE (1989) Assessing individual differences in ethanol preference using a cumulative dosing procedure. Psychopharmacology 98: 113–119

**DiChiara G, Imperato A (1988) Drugs of abuse preferentially stimulate dopamine release in the mesolimbic system of freely moving rats. Proc Natl Acad Sci USA 85: 5274–5278

DiChiara G, North RA (1992) Neurobiology of opiate abuse. Trends Pharmacol Sci 13: 185–192

Donovan JM (1986) An etiologic model of alcoholism. Am J Psychiatry 143: 1–11

**Edwards G, Gross M (1976) Alcohol dependence: provisional description of a clinical syndrome. Br Med J 1: 1058–1061

Ehlers CL, Chaplin RI, Wall TL, Lumeng L (1992) Corticotropin-releasing factor (CRF): studies on alcohol-preferring and non-preferring rats. Psychopharmacology (Berl) 106: 359–364

Ewing JA (1984) Detecting alcoholism – the CAGE questionnaire. JAMA 252: 1905–1907

Fahlke C, Engel JA, Eriksson CJP, Hard E, Söderpalm B (1994) Involvement of corticosterone in the modulation of ethanol consumption in the rat. Alcohol 11: 195–202

Feuerlein W, Küfner H, Ringer C, Antons K (1976) Kurzfragebogen für Alkoholgefährdete. Arch Psychiatr Nervenkrankh 222: 139–152

Feuerlein W, Ringer C, Küfner H, Antons K (1977) Diagnose des Alkoholismus. Der Münchner Alkoholismus-Test (MALT). Munch Med Wochenschr 119: 1275–1282

Fichter MM, Narrow WE, Roper MT et al (1996) Prevalence of mental illness in Germany and the United States. J Nerv Ment Dis 184: 598–606

Fils-Aime ML, Eckardt MJ, George DT, Brown GL, Mefford I, Linnoila M (1996) Early-onset alcoholics have lower cerebrospinal fluid 5-hydroxyindoleacetic acid levels than late-onset alcoholics. Arch Gen Psychiatry 53: 211–216

Finn PR, Pihl RO (1987) Men at high risk for alcoholism: the effect of alcohol on cardiovascular response to unavoidable shock. J Abn Psychol 96: 230–236

Finney JW, Moos RH (1992) The long-term course of treated alcoholism. II. Predictors and correlates of 10 years functioning and mortality. J Stud Alcohol 53: 142–153

Funke W, Funke J, Klein M, Scheller R (1987) Trierer Alkoholismus Inventar (TAI). Göttingen, Hogrefe

Gsellhofer B, Fahrner EM, Platt J (1993) Deutsche Version des European Addiction Severity Index (EuropASI) nach dem amerikanischen Original von T. McLellan (5th edition 1992) und der europäischen Version EuropASI von A. Kokkevi, C. Hartgers, P. Blanken et al. IFT Institut für Therapieforschung, Munich

Heatherton TF, Koslowski LT, Frecker RC, Fagerström KO (1991) The Fagerström Test for Nicotine Dependence (FTND). Revision of the Fagerström Tolerance Questionnaire. Br J Addict 86: 1119–1127

*Helzer JE, Pryzbeck (1988) The cooccurrence of alcoholism with other psychiatric disorders in the general population and its impact on treatment. J Stud Alcohol 49: 219–224

Herz A (1995) Neurobiologische Grundlagen des Suchtgeschehens. Nervenarzt 66: 3–14

Higley JD, Hasert MF, Suomi SJ, Linnoila M (1991) Nonhuman primate model of alcohol abuse: effects of early experience, personality and alcohol consumption. Proc Natl Acad Sci USA 88: 7261–7265

Holly A, Türk D, Nelson CB, Pfister H, Wittchen HU (1997) Prävalenz von Alkoholkonsum, Alkoholmißbrauch und -ab-

hängigkeit bei Jugendlichen und jungen Erwachsenen. Z Klin Psychol 26: 171–178

Jacobi C, Brand-Jacoby J, Marquardt F (1987) Die "Göttinger Abhängigkeitsskala (GABS)": ein Verfahren zur differentiellen Erfassung der Schwere der Alkoholabhängigkeit. Suchtgefahren 33: 23–36

Janlert U, Hammarström A (1992) Alcohol consumption among unemployed youths: results from a prospective study. Br J Addict 87: 703–714

Jellinek EM (1960) The disease concept of alcoholism. Hillhouse, New Haven

John U, Schnofl A, Veltrup C, Bunge S, Wetterling T, Dilling H (1992a) Sekundärmerkmale des Alkoholabhängigkeitssyndroms: Entwicklung eines diagnostischen Fragebogens. Sucht 38: 362–370

John U, Veltrup C, Schnofl A, Bunge S, Wetterling T, Dilling H (1992b) Entwicklung eines Verfahrens zur Erfassung von Ausprägungen der Alkoholabhängigkeit aufgrund von Selbstaussagen: die Lübecker Alkoholabhängigkeitsskala (LAS). Sucht 38: 291–333

Kidd KK, Pakstis AJ, Castiglione CM et al (1996) DRD2 haplotypes containing the TaqI A1 allele: implications for alcoholism research. Alcohol Clin Exp Res 20: 697–705

Koob GF, Bloom FE (1988) Cellular and molecular mechanisms of drug dependence. Science 242: 715–723

Kushner MG, Sher KJ, Beitman BD (1990) The relation between alcohol problems and anxiety disorders. Am J Psychiatry 147: 685–695

Ladewig D, Graw P, Miest C, Hobi V, Schwarz E (1976) Basler Drogen- und Alkoholfragebogen (BDA). Pharmacopsychiatry 9: 305–312

Lander ES, Schork NJ (1994) Genetic dissection of complex traits. Science 265: 2037–2048

Laux L, Schaffner P, Glanzmann P (1981) State-Trait Angstinventar (STAI). Beltz, Weinheim

MacAndrew C (1965) The differentiation of male alcoholic outpatients from nonalcoholic psychiatric outpatients by means of the MMPI. Q J Stud Alcohol 26: 238–246

Maier W (1995) Mechanismen der familiären Übertragung von Alkoholabhängigkeit und Alkoholabusus. Z Klin Psychol 24/2: 147–158

McLellan AT, Luborsky L, Woody GE, O'Brien CP (1980) An improved diagnostic evaluation instrument for substance abuse patients. J Nerv Ment Dis 168: 26–33

Nathan PE (1988) The addictive personality is the behavior of the addict. J Consul Clin Psychol 56/2: 183–188

Nelson CB, Little RJA, Heath AC, Kessler RC (1996) Patterns of DSM-III-R alcohol dependence symptom progression in general population survey. Psychol Med 26: 449–460

*Nestler AJ, Aghajanian GK (1997) Molecular and cellular basis of addiction. Science 278: 58–63

Olds J (1956) Pleasure centers in the brain. Sci Am 195: 105–106

Oyefeso A (1994) Sociocultural aspects of substance use and misuse. Curr Opin Psychiatry 7: 273–277

Piazza PV, Rougé-Pont F, Deroche V, Maccari S, Simon H, LeMoal M (1996) Glucocorticoids have state-dependent stimulant effects on the mesencephalic dopaminergic transmission. Proc Natl Acad Sci USA 93: 8716–8720

**Reich T, Edenberg HJ, Goate A et al (1998) Genome-wide search for genes affecting the risk for alcohol dependence. Am J Med Genet 81(3): 207–215

Robins LN, Helzer JE, Weissman MM, Orvaschel H, Gruenberg E, Burke JD, Regier DA (1984) Lifetime prevalence of specific psychiatric disorders in three sites. Arch Gen Psychiatry 41: 949–958

Robins LN, Wing JK, Wittchen HU et al (1988) The Composite International Diagnostic Interview. An epidemiological instrument. Arch Gen Psychiatry 45: 1969–1077

**Robinson TE, Berridge KC (1993) The neural basis of drug craving: an incentive-sensitization theory of addiction. Brain Res Rev 18: 247–291

Roth J (1986) Fragebogen zur Klassifikation des Trinkverhaltens Alkoholabhängiger. Psychodiagnostisches Zentrum der Humboldt Universität Berlin, Berlin

Sander T, Harms H, Podschus J et al (1997) Allelic association of a dopamine transporter gene polymorphism in alcohol dependence with withdrawal seizures or delirium. Biol Psychiatry 41: 229–304

Saunders JB, Aasland OG, Babor TF, de la Fuente JR, Grant M (1993) Development of the Alcohol Use Disorders Identification Test (AUDIT): WHO collaborative project on early detection of persons with harmful alcohol consumption II. Addiction 88: 791–804

Schmidt LG (1995) Diagnostische Aufgaben bei Alkoholmißbrauch und -abhängigkeit. Z Klin Psychol 24/2: 98–106

Schmidt LG (1997) Frühdiagnostik und Kurzintervention beim beginnenden Alkoholismus. Dt Aerztebl 94: A2905–2908

Schmidt LG, Schmidt K, Dufeu P, Kuhn S, Ohse A, Rommelspacher H, Müller C (1997) Superiority of Carbohydrate Deficient Transferrin (CDT) to g-Glutamyl-Transferaseγg (g-GT) in detecting relapse in alcoholism. Am J Psychiatry 154: 75–80

Schmidt LG, Harms H, Kuhn S, Rommelspacher H, Sander T (1998) Modification of alcohol withdrawal by the A9 allele of the dopamine transporter gene. Am J Psychiatryl 55: 474–478

*Schuckit MA (1994) Low level of response to alcohol as a predictor of future alcoholism. Am J Psychiatry 151: 184–189

Schuckit MA, Smith TL, Anthenelli R, Irwin M (1993) Clinical course of alcoholism in 636 male inpatients. Am J Psychiatry 150: 786–792

**Schultz W (1997) Dopamine neurons and their role in reward mechanisms. Curr Opin Neurobiol 7: 191–197

Sellers EM, Sullivan JT, Somer G, Sykora K (1991) Characterization of DSM-III-R criteria for uncomplicated alcohol withdrawal provides an empirical basis for DSM-IV. Arch Gen Psychiatry 48: 442–447

Selzer ML (1971) The Michigan Alcoholism Screening Test: the quest for a new diagnostic instrument. Am J Psychiatry 127: 1653–1658

Spielberger CD, Gorsuch RL, Lushene RE (1970) STAI. Manual for the State-Trait-Anxiety-Inventory. Consulting Psychologists, Palo Alto

Stockwell T, Hodgson R, Edwards G, Taylor CRA, Rankin H (1979) The development of a questionnaire to measure severity of alcohol dependence. Br J Addict 74: 79–87

Sullivan JT, Sykora K, Schneiderman J, Naranjo CA, Sellers EM (1989) Assessment of alcohol withdrawal: the revised Clinical Institute Withdrawal Assessment For Alcohol Scale (CIWA-Ar). Br J Addict 84: 1353–1357

Taylor C, Brown D, Duckitt A, Edwards G, Oppenheimer E, Sheehan M (1986) Alcoholism and the patterning of outcome: a multivariate analysis. Br J Addict 81: 815–823

Tsai G, Gastfriend DR, Coyle JT (1995) The glutamatergic basis of human alcoholism. Am J Psychiatry 152: 332–340

Vaillant GE (1996) A long-term follow up of male alcohol abuse. Arch Gen Psychiatry 53: 243

Vaillant GE, Milofsky E (1982) Natural history of male alcoholism. Arch Gen Psychiatry 39: 127–133

Veltrup C (1994) Erfassung des "Craving" bei Alkoholabhängigen mit Hilfe eines neuen Fragebogens (Lübecker Craving-Risiko-Rückfall-Fragebogen). Wien Klin Wochenschr 106: 75–79

Volkow ND, Wang GJ, Hitzemann R et al (1993) Decreased cerebral response to inhibitory neurotransmission in alcoholics. Am J Psychiatry 150: 417–422

Volkow ND, Wang GJ, Begleiter H et al (1995) Regional brain metabolic response to lorazepam in subjects at risk for alcoholism. Alcohol Clin Ex Res 19: 510–516

Wafford KA, Burnett DM, Leidenheimer NJ (1991) Ethanol sensitivity of the $GABA_A$ receptor expressed in xenopus oocytes requires eight amino acids contained in the g_{2L} subunit in the receptor complex. Neuron 7: 27–33

Wanberg KW, Horn JL, Foster FM (1977) A differential assessment model for alcoholism. The scales of the Alcohol Use Inventory. J Stud Alcohol 38: 512–543

Weiss F, Lorang MT, Bloom FE, Koob GF (1993) Oral alcohol self-administration stimulates dopamine release in the rat nucleus accumbens: genetic and motivational determinants. J Pharmacol Exp Ther 267: 250–258

WHO (1990) CIDI – The Composite International Diagnostic Interview. Core Version 1.0. World Health Organization, Geneva

WHO (1992) The ICD-10 classification of mental and behavioural disorders, clinical descriptions and diagnostic guidelines. World Health Organization, Geneva

WHO (1995) Schedules for Clinical Assessment in Neuropsychiatry (SCAN). World Health Organization, Geneva

Wise R (1996) Neurobiology of addiction. Curr Opin Neurobiol 6: 243–251

Wittchen HU, Zaudig M, Schramm E, Spengler P, Mombour W, Klug J, Horn R (1990) Strukturiertes Klinisches Interview für DSM III-R (SKID). Beltz, Weinheim

Wolfgramm J (1995) Abhängigkeitsentwicklung im Tiermodell. Z Klin Psychol 24/2: 107–117

CHAPTER

19

K.F. Mann

Alcoholism: Clinical Syndromes and Treatment

Translator: S. Johnson

1 Introduction

The previous chapter describes definitions and diagnostic criteria for alcohol misuse and dependence, as well as recent advances in understanding their aetiology. In this chapter, new findings regarding clinical syndromes and treatment are outlined; as in the last chapter, these represent a significant advance on previous textbook knowledge.

Alcohol dependence is now primarily understood as a disease of the brain which in certain social contexts gives rise to particular behaviour patterns. As proposed by Heimann (1992), this disorder may be regarded as a neurobiological model system which allows study of the effects of a defined toxin on various brain structures and functions, including the regenerative and restorative changes observed under conditions of abstinence. Work based on such a formulation opens a "window onto the brain", as understanding of the effects of alcohol can be used to deepen our knowledge about the structure and functioning of the brain. This therefore represents an extension of the heuristic value of the dementia model (Mann and Widmann 1995).

There have also been exciting changes in the treatment of alcohol misuse and dependence. The international trend is towards day hospital or primarily out-patient treatment. In-patient types of treatment are still offered to dependent individuals who have a poor prognosis or are severely damaged. The old division between medically oriented detoxification and psychotherapeutically oriented rehabilitation is becoming more and more blurred, and targeted psychotherapeutic interventions have come to be seen as indispensable during withdrawal (a process previously generally described as detoxification). Pharmacological interventions are now increasingly used to augment psychological relapse prevention work.

The structure of this chapter needs to accommodate these far-reaching changes in treatment practices. Treatment options are described in sequence according to the severity of misuse or dependence for which they are appropriate. Thus this survey begins with brief out-patient interventions and extends across a range of treatments to in-patient psychological treatment.

2 Clinical Syndromes

2.1 Acute Alcohol Intoxication

The following symptoms are observed as blood alcohol levels rise in simple drunkenness, depending on the extent of habituation to alcohol and development of tolerance: elevated mood, a reduction of anxiety and tension and an increase in impulses and motor activity. With increasing doses, dysphoria, irritability, dysarthria and disturbances in attention, awareness, judgement and co-ordination set in, together with drowsiness, which may progress to reduced conscious level and coma (Julien 1997).

Pathological drunkenness is a syndrome whose existence remains in dispute. It no longer features in DSM-IV, as it is not considered an adequately valid diagnosis (APA 1994). In the German literature, too, discussion of its diagnostic specificity has been predominantly critical (Athen 1986). Further, the symptoms of simple drunkenness can hardly be regarded as non-pathological. Pathological drunkenness describes a state which is not just a quantitative progression from simple drunkenness, but something really qualitatively different. Triggered even by small quantities of alcohol, in the majority of cases there are disturbances in orientation and in consciousness, progressing to sleep and amnesia (twilight state). Further striking manifestations are personality-incongruent disturbances of behaviour, often with aggressiveness and anxious, irritable mood.

2.2 Alcohol-Induced Amnesias (Blackouts)

Anterograde amnesias and gaps in memory are common disturbances in alcohol-dependent subjects, but can also occur in healthy individuals after sufficiently severe states of drunkenness. As a rule, amnesia is complete (Goodwin et al. 1969). These episodes are thought to occur particularly in the presence of rapidly rising blood alcohol levels (Ryback 1970). The pathogenesis has not as yet been elucidated (Charness 1992).

2.3 Withdrawal Syndrome

The withdrawal syndrome has the following principal characteristics: alcohol-dependent people who inten-

tionally or unintentionally end or interrupt their drinking (e.g. when admitted to hospital) may develop withdrawal symptoms within hours or a few days. These show great inter-individual variation in type and intensity and are categorised according to the systems affected (Table 1).

As a rule, the withdrawal syndrome subsides after 3–7 days, and longer courses are rare. About a third of patients need pharmacological treatment. A wide range of medications have been proposed for this purpose, including chlormethiazole, benzodiazepines, clonidine and carbemazepine.

Scholz (1982) has described a so-called protracted alcohol withdrawal syndrome. This is described as lasting several months and consisting of fluctuating levels of disturbances such as dysphoria, anxiety, sleep disorders and episodes of sweating. This syndrome has recently been discussed again more frequently, as there are at least temporal associations with increased vulnerability to relapse. Most relapses occur in the first few months after withdrawal, and craving for alcohol appears to be particularly irresistible during this period. There may be a neurobiological basis for this phenomenon, in the form of altered expression of various intracellular proteins (e.g. fos, jun) of the *N*-methyl-D-aspartate (NMDA) receptors, which are up-regulated during chronic intoxication (Buck and Harris 1991). Activation of "late-onset genes" may result in long-term changes in the gene expression of the nerve cells (see Chap. 18, Vol. 3, Part 2).

In delirium tremens, disorientation is a major element added to the symptoms already described, which are generally present in a severe form. Patients are disoriented with respect to time, place and circumstances and sometimes also person. Conscious level is not necessarily reduced. Visual hallucinations (e.g. insects, small animals) are often observed. About half of all episodes of delirium begin with a cerebral convulsion. The pathophysiology remains uncertain. Speculation that disturbances in water balance with electrolyte imbalances might be causal has not been confirmed (Flink et al. 1954; for an overview, see Lishman 1990). Delirium tremens is always a life-threatening condition, which as a rule requires intensive treatment.

Alcoholic hallucinosis is rare. Compared with individuals with alcohol dependence and no hallucinosis, patients with alcoholic hallucinosis are characterised by earlier onset and larger volumes of alcohol consumed, and sometimes also by high levels of drug use (Tsuang et al. 1994). Difficulties sometimes arise in differentiating this from delirium tremens. While various of the symptoms take a similar form (e.g. feelings of anxiety, psychomotor arousal, vivid hallucinations – predominantly auditory in the case of alcoholic hallucinosis), autonomic manifestations and disturbance of orientation are entirely absent. A particular factor differentiating schizophrenia from alcoholic hallucinosis is that passivity phenomena are very rare in the latter (Soyka 1990). The age of illness onset is earlier in schizophrenia.

Alcoholic hallucinosis is principally treated with neuroleptics, which do not need to be prescribed on a long-term basis. The prognosis of acute alcoholic hallucinosis is good. In approximately 80% of patients, symptoms abate within weeks or a few months (Glass 1989). The course is described as chronic if symptoms persist for more than 6 months (Benedetti 1952). In these cases, prognosis is unfavourable, regardless of whether neuroleptics have been prescribed.

Table 1. Withdrawal symptoms

System affected	Symptoms
Gastrointestinal	Nausea; diarrhoea
Circulatory	Tachycardia; hypertension
Autonomic	Increased sweating; sleep disturbances; moist, cool extremities
CNS	Generalised seizures (grand mal); tremor; dysarthria; ataxia; inner agitation; increased impulsiveness; anxious, dysphoric, depressed mood; hallucinations (predominantly visual); tendency to take fright; short episodes of disturbed awareness with transient dissociation

2.4 Alcohol-Related Pathological Jealousy

Alcohol-related pathological jealousy is a rare disorder, observed almost exclusively in men (Kolle 1932). It not infrequently results in criminal behaviour. Patients have an unshakeable belief that their partners are unfaithful. Onset is insidious, and it can therefore be readily differentiated from delirium tremens and alcoholic hallucinosis. The aetiological significance of the disturbances of sexual functioning which frequently occur in alcoholics has not been clearly established (Laux and Reimer 1979). Treatment is difficult, and success is limited regardless of whether psychopharmacological treatment with neuroleptics or psychotherapeutic measures are used. Achieving abstinence is important, and following this the symptoms may very slowly recede (Vauhkonen 1968).

2.5 Structural and Functional Changes in the Nervous System

The findings described in this section do not solely occur in the presence of alcohol dependence. Evidence of them may also be found in the context of alcohol consumption fitting diagnostic criteria for harmful use or misuse.

Recent work has confirmed older findings (Courville 1955) of atrophic changes in the cerebrum in about a third to half of patients (Miyakawa et al. 1977). The subcortical white matter is particularly affected. There is probably also a reduction in the number and density of neurones in the cortex (Harper et al. 1987), although final resolution of persisting doubts about this (Jensen and Pakkenberg 1993) has not been possible (for a review, see Mann and Widmann 1995).

Imaging procedures allow the reduction in volume in the subcortical and cortical regions to be measured (Jernigan et al. 1991a). These changes are partly reversible with abstinence (Carlen et al. 1978; Schroth et al. 1985), although full recovery to a normal state does not occur (Muuronen et al. 1989). These changes are correlated to some extent with deficits in cognitive performance (Mann 1992). In severe cases, disorders of memory and fine motor performance also occur (Parsons et al. 1987). The neuropsychological deficits do not correspond closely to the patterns characteristic of damage to particular brain regions. The overall picture suggests much more a global deterioration in brain functioning (the mild generalised dysfunction hypothesis of Parsons et al. 1987), which varies from a mild to a severe form. Cognitive and visuo-motor disturbances are also partly reversible with abstinence (Brandt et al. 1983; Mann 1992).

The pathophysiological basis of the atrophic changes in the brain and their partial reversal has been intensively investigated for many years. According to the "rehydration hypothesis" (Carlen et al. 1978), the chronic effects of alcohol mediated by anti-diuretic hormone (ADH) lead to dehydration of the brain with a consequent reduction in volume. It is postulated that this can then be reversed by abstinence, with re-entry of water into the brain explaining the observed increases in volume. This hypothesis has largely been refuted through experimental neuropathological investigation (Harper et al. 1988) and imaging procedures (Mann et al. 1993a,b). The alternative hypothesis of partial regeneration of neuronal material with new growth of axons and dendrites and increased synapse formation long met with considerable scepticism. It seemed incompatible with the fact that, as neurones are post-mitotic cells, regeneration by cell division cannot occur. However, this view failed to take into account the finding that subtotal nerve cell damage can occur (Ferrer et al. 1986; Harper and Corbett 1990). For such partially damaged neurones, new growth in axons and dendrites and renewed synapse formation is quite conceivable. This process has been demonstrated in animal experiments (McMullen et al. 1984; Dlugos and Pentney 1997). These findings underline the way in which alcohol dependence can act as a model for the study of neuronal plasticity.

Wernicke's encephalopathy is commoner than previously believed (3%–12% of individuals with alcohol dependence; Dufour 1993) and appears to affect women more frequently (Glenn 1993). The basis is thiamine deficiency. A genetically determined reduced affinity of the enzyme transketolase for the co-factor thiamine pyrophosphate has been described (Blass and Gibson 1977). This could explain why, despite the high prevalence of alcoholism, this disorder is not seen more frequently. The principal symptoms of the syndrome are clouding of consciousness, ataxia and disorders of the eye muscles (ophthalmoplegias, paralysis of conjugate gaze, pupillary disturbances) and nystagmus. These conspicuous manifestations are often preceded by a prodrome involving gastrointestinal disturbances and fever. In neuropathological terms, the characteristic changes are primarily in the brain regions around the third and fourth ventricles, in the aqueduct, the dorsomedial nucleus of the thalamus and the mamilliary bodies. A pattern of gliovascular damage has been described (Peiffer 1985). Haemorrhagic lesions are characteristic (Colmant 1965; Victor et al. 1971). There may also be a subclinical form of Wernicke's encephalopathy (for a review, see Lishman 1987). This is supported especially by the discrepancy between the frequency of neuropathological evidence in autopsy studies and the relative rarity of a clinical diagnosis of the disorder being made (Harper 1983).

Treatment of Wernicke's encephalopathy requires immediate administration of vitamin B_1 (50 mg thiamine i.v. and 50 mg i.m.). The very rarely encountered hazard of anaphylactic reactions should not result in avoidance of parental administration, as many alcoholics do not absorb oral thiamine (Thomson and Cook 1997). Treatment with thiamine is continued following hospital admission. The ophthalmoplegic symptoms of the illness generally recede quite quickly with such treatment, while the confusion may persist for days or weeks. Nystagmus and ataxia can continue for longer and in some cases persist indefinitely.

Korsakow's syndrome often begins with Wernicke's encephalopathy, but sometimes also with a confusional state of other origin (Victor et al. 1971). The principal symptoms are disturbances of long- and short-term memory, together with confabulation, and of concentration and orientation. There are often additional symptoms in the form of polyneuropathy, chronic eye muscle and pupillary disturbances and nystagmus. The

course is generally chronic. The mortality rate is around 15%–20%. An increase in prevalence has been reported in recent years (e.g. Ramayya and Jauhar 1997), and women are more often affected than men (Victor et al. 1971; Torvik et al. 1982). Autopsy studies indicated that in this disorder the specific neuropathological changes are found much more frequently than the corresponding clinical diagnosis is made (Harper 1983). Neuropathological and computed tomography (CT) and magnetic resonance imaging (MRI) studies indicate that cortical atrophy occurs in addition to the pattern of damage characteristic of Wernicke's encephalopathy (Victor 1990; Jernigan et al. 1991b). Lishman (1990) reported the direct neurotoxic effect of alcohol and its breakdown products, which also particularly affects the cortex and may compound the central changes resulting from thiamine deficiency.

Influenced by Victor et al. (1971), the two symptom complexes described above have often been encapsulated in recent years as a single Wernicke-Korsakow's syndrome. However, this should not be misunderstood as indicating that the two syndromes cannot occur in isolation from each other. Victor later re-emphasised this (Victor 1990). Some authors distinguish between Wernicke-Korsakow's syndrome and an alcohol-induced persistent dementia (e.g. Martin et al. 1986). The latter term describes global impairments of intellectual performance and, to a lesser extent, of memory. However, making such a distinction between persistent memory disturbance and persistent dementia has not been uncontroversial (Lishman 1990).

Atrophic changes, especially affecting the superior vermis of the cerebellum, have long been recognised (Thomas 1905). Clinically, these result in tremor, dysarthria and ataxia. This finding is more frequent than previously believed and occurs in 30%–50% of patients (Torvik and Torp 1986; Mann 1992). Improvements in functioning with abstinence have been shown (Diener et al. 1984). The reductions in volume are at least partly reversible (Ron et al. 1982; Schroth et al. 1988; Mann 1992).

Between 10% and 40% of people with alcohol dependence develop symptoms of a polyneuropathy (Victor and Adams 1953). This encompasses motor, sensory and autonomic pathways. To begin with, reduced light touch and pain sensation occur in the foot region in a sock distribution. This is followed by reduction in reflexes, beginning with the ankle jerk reflex, and by muscle atrophy. Superficial and deep sensation are impaired, and burning pains are sometimes experienced in the feet and toes. The symptoms can extend to the upper extremities, and disorders of sweat production and sexual functioning can also develop. The onset may be signalled by T-reflex prolongation (Schott et al. 1995).

3 Treatment

The treatment of hazardous or dependent drinkers is individually planned according to the stage of illness. Damaging alcohol use among young people requires different measures than treatment of severely dependent individuals with many associated problems. Treatment options will be described according to severity of dependence and its consequences. The common interventions will therefore be described first and the more uncommon and more radical measures later. The focus will also shift from residential treatments more towards out-patient forms of care.

The aim of treatment is to reduce motivation for drinking in favour of motivation for abstinence. Insight into illness, readiness for change and a subjective understanding of the cause of the dependence play a role in motivation, as do specific defence mechanisms, extent of social support and fear of sanctions such as removal of driving license, loss of partner or employment. Previously, motivation tended to be seen as an enduring aspect of personality (a trait). Such a trait model is not, however, very helpful in considering how to influence readiness for treatment or for change. More recently, a dynamic concept of motivation as an alterable state has become prominent. Thus establishing and maintaining motivation become tasks rather than conditions for treatment.

Prochaska and DiClemente (1986) distinguish stages in readiness for change. Their model is essentially applicable to all forms of dependence and can be used in all the therapeutic contexts described below. A requirement is that patients are capable of the cognitive tasks of problem recognition, decision-making and organisation of behaviour. The stages of readiness for change are divided into pre-contemplation, contemplation, preparation, action and maintenance. Any stage may persist for a long period of time (sometimes many years), and during each, a slide back to the pre-contemplation stage is possible as well as progress on to the next stage. Thus the model provides an account of an ambivalent motivational state between readiness for treatment and relapse.

3.1 Early Interventions

When harmful use or misuse of alcohol is first diagnosed, a minimal intervention is indicated. This may consist of discussion with a doctor, in which the warning symptoms already present are identified and a reduction in the level of drinking is recommended. If alcohol dependence is present, early intervention may

be successful, in the form of a discussion in which the problem is explained and confronted. The diagnostic criteria and laboratory investigations (see Chap. 18, Vol. 3, Part 2) are discussed point by point. The goal is motivation for abstinence. The following factors have been described as helpful in establishing and maintaining motivation (the *frames* presented by Bien et al. 1993):

- Give *feedback* on the personal risks associated with drinking
- Emphasise personal *responsibility* for change
- Give clear *advice* about aims
- Highlight alternative behaviours ("*menu* of behaviour change")
- Use a non-confrontational style of discussion ("*empathy*")
- Reinforce the *self-efficacy* of the patient

At this stage of the motivational process, involvement of relatives is very important, as many dependent individuals show a characteristic defensiveness and tendency to minimise problems. If all these measures fail, the next step is in-patient withdrawal or, in some exceptional cases, an out-patient detoxification programme.

3.2 Withdrawal

The first stage of withdrawal is physical detoxification, including appropriate treatment for the manifestations of withdrawal (see below). However, without work on motivation, simple physical withdrawal is associated with high rates of relapse and is a prelude to continuing treatment in only a few cases (Wieser and Kunad 1965). On the other hand, the vulnerability of patients and relatives at this stage provide particularly good conditions for building up and stabilising motivation for change. Thus the integrated withdrawal programmes which have recently been described (e.g. Mann et al. 1995) involve use of psychological interventions simultaneously with physical detoxification (e.g. "motivational interviewing" as described by Miller 1983).

Treatment goal is again dependent on severity of dependence:

- Participation in self-help groups
- Self-help groups together with primary care treatment, supplemented by anti-craving substances
- Out-patient psychological treatment (rehabilitation)
- In-patient psychological treatment (rehabilitation)

Only around 30%–50% of patients with withdrawal syndromes require pharmacological treatment. Chlormethiazole has proved to be of value for this purpose in in-patient settings. Clonidine, carbemazepine (not in a slow-release form) and benzodiazepines, sometimes combined with neuroleptics, can also be prescribed.

Clinical investigations support the effectiveness of integrated detoxification and psychological intervention (Veltrup et al. 1995; Stetter and Mann 1997). Following this, up to 50% of patients are capable 6 months later of taking up continuing treatment. The objection that a concept of motivation as externally imposed and confrontational cannot lead to success does not seem convincing, as few of the patients receiving this type of programme drop out during the first 2 weeks of entering in-patient treatment (Küfner and Feuerlein 1989; Mann et al. 1995). The decision about what therapeutic steps should follow withdrawal again depends on the severity of dependence and related problems. In the majority of cases, out-patient care, perhaps combined with adjuvant pharmacotherapy (see below), and participation in self-help groups seem to be indicated.

3.3 Primary Care Treatment and Self-Help

In-patient treatment aimed at relapse prevention has substantial success (see below), but in Germany is taken up each year by only around 1% of all people with alcohol dependence. Every year, around 2.5% of people with alcohol dependence undergo physical detoxification once or several times; 25% are treated for a variety of reasons in a general hospital setting. The great majority of people with alcohol dependence (70%) are seen by general practitioners at least once a year (Wienberg 1992).

Thus only few patients manage to find their way after withdrawal to in-patient, specialist addiction treatment. People with alcohol disorders are, however, especially vulnerable to relapse in the few weeks following withdrawal. This is the stage at which there is a particular requirement for the development of new treatments. These need to be community-based, immediately available and applicable by a doctor who is not an addiction specialist, in collaboration with addiction therapists. This requires substantially improved higher training and continuing education in addiction medicine.

The out-patient treatment of more mildly dependent and less damaged patients by general practitioners should, whenever feasible, be supplemented by self-help groups (e.g. Alcoholics Anonymous). In each form of continuing out-patient care for alcohol-dependent individuals, use of an adjuvant pharmacological treatment should be considered at least in the 6 months following withdrawal (see below).

3.4 Relapse Prevention (Rehabilitation)

3.4.1 Out-patient Psychological Treatment

Out-patient psychological treatment should begin immediately after withdrawal. The central therapeutic goal is reinforcement of the wish for abstinence. A broad spectrum of psychotherapeutic methods have proved useful to this end. In principle, the same fundamental factors influence effectiveness as in general psychological treatment, e.g. clarifying the nature of the problem, establishing motivation, activating resources and providing active help in overcoming the problem (Grawe 1995). Meyer (1990) distinguishes the following: mobilisation of hope, application of a theory as to how healing may be achieved, provision of a helping relationship, clarification or redefinition of problems and the search for constructive options for solving problems.

Despite all the common ground with psychological treatment for other disorders, a number of particular aspects of the treatment of people with addictions modify psychological treatment procedures:

- Abstinence is the central goal of the treatment and in fact a requirement for traditional psychological treatment for patients with dependence.
- The behaviour targeted in psychological treatment is an "approach behaviour" (the consumption of an addictive substance). In contrast, the psychological treatment of other disorders often mainly involves changing an avoidance behaviour (e.g. in anxiety). While in the latter case a graded hierarchy of success can be followed, in people with dependence a comprehensive alteration in behaviour needs to be achieved as quickly as possible and then needs to be stabilised.
- Self help movements implemented the first steps towards change decades ago. This success has continued to influence psychological treatment of people with dependence to the present day.

Relapse prevention programmes based on the work of Marlatt and Gordon (1985) have attracted particular interest in the last few years. According to these authors, relapse is not to be seen as an event which suddenly occurs, but as a developmental process. It involves a sequence of events affecting cognition and behaviour, which finally culminate in relapse. It follows that appropriate measures can reduce the likelihood of relapse if they can be targeted properly and applied at the right time. Despite the theoretical brilliance of the model, for relapse prevention, as for most models, a few positive evaluations have to be set against the majority of studies, in which no evidence has been found for a specific effect of relapse prevention (Veltrup 1995).

A recent meta-analysis summarises all the studies of psychological treatment for people with alcohol dependence which meet a methodological standard. It is based on around 30 experimental and 20 non-experimental studies and compares effectiveness with control conditions (Süß 1995). Investigation of differences in effectiveness suggests no statistically significant advantages for behavioural treatments compared with standard methods, which in this context also included family and couple therapy. Behavioural treatment was superior to minimal interventions and to disulphiram treatment.

Therapist variables probably play an important role in treatment outcome. Thus controlled studies have indicated that the degree of therapist empathy is correlated with the outcome of treatment (Institute of Medicine 1990). Confrontational and directive techniques led to a higher relapse rate during a 12-month observation period than supportive and accepting ones. The extent of initial motivation played no role in later treatment outcome.

Overall, the following conclusions have been made (Institute of Medicine 1990):

1. An appropriate and specific treatment of alcohol-dependent people can lead to clearly positive results. Comparisons made with waiting-list groups or other forms of treatment in controlled trials indicate that a wide range of specific forms of psychological treatment are associated with good outcomes.
2. No single form of treatment is clearly superior to other types and appropriate for all alcohol-dependent individuals. "Rather than trying to prove the superiority of a single method through examination of specific interventions in heterogeneous samples, outcome studies should investigate the characterstics of sub-groups for whom particular forms of treatment are especially successful" (Institute of Medicine 1990, p. 537ff.).
3. The importance of therapist variables in determining outcome has so far clearly been underestimated. Therapist skills and attitudes have an important influence on outcome, regardless of the nature of their training and theoretical orientation.
4. Self-help groups, especially Alcoholics Anonymous, are very widespread. There have, however, been hardly any studies evaluating their results. Nonetheless, it appears likely that, in general, self-help groups represent a positive and stabilising factor in the struggle of many alcohol-dependent people with their illness.
5. The treatment of other life problems associated with drinking can have a positive influence on the

outcome of treatment. This includes social skills training, couples and family therapy and sometimes anti-depressant pharmacological treatment, stress management and involvement in a community-based system of care.

6. Global treatment outcomes for non-selected patients appear to show no difference between in-patient and out-patient forms of treatment. The same applies to more long-term treatments compared with brief ones. However, an important limitation of this is that a longer, in-patient treatment is superior to a shorter out-patient one for greater severity of dependence, more marked psychiatric co-morbidity and more serious alcohol-related damage.

The results of the largest study of psychological treatment for people with alcohol dependence emphasise some of these points. In the Matching Alcoholism Treatments to Client Heterogenity (MATCH) Project, around 1800 patients receiving out-patient treatment were investigated. They were randomised to receive three different types of treatment: (1) a treatment based on the 12-step programme of Alcoholics Anonymous, (2) a cognitive behavioural intervention aimed at improving "coping skills" and (3) a programme aimed at increasing motivation.

The overall outcome of the study was positive. It indicated a highly significant increase in drink-free days. When drinking occurred, the volume consumed was less than before treatment. The authors were surprised to find that all three treatment modalities had similarly good outcomes. Of numerous matching hypotheses (e.g. cognitive behaviour therapy is superior to the other two in the presence of cognitive deficits), only one was confirmed: for low degrees of psychiatric severity (measured using the Addiction Severity Index, ASI), the 12-step programme was better than both other forms of treatment. Patients who initially received in-patient treatment did significantly better than those who received mainly out-patient treatment, despite more severe initial symptoms (Project MATCH Research Group 1997).

3.4.2 In-patient Psychological Treatment

The first in-patient facility in the world for the treatment of alcohol-dependent individuals was opened in 1851 in a village near Düsseldorf. It was led by the village pastor. The initial aim was to bring about improvements by providing patients with good role models and imparting ethical values, but looking back on his experience 30 years later, Pastor Hirsch was able to identify two other conditions which were decisive for success: selection of patients and a clear requirement for abstinence, reinforced by controls and sanctions.

For decades, in-patient "cures" lasting 6 months or more dominated German practice. Küfner and Feuerlein (1989) investigated the effectiveness of standard treatment in these specialist clinics and came to some noteworthy conclusions. Successful treatment facilities are characterised by the following features: selection of patients with better prognosis, use of a therapeutic community model, active involvement of spouses and people who can exert a positive influence on patients and active aftercare following therapy.

In recent years, there has been a gradual reduction in treatment times in in-patient facilities. This trend was fostered by positive outcomes from model facilities (Mann and Batra 1993) and fitted with international experiences. Thus length and modality of treatment is determined on an individual basis. Depending on the stage of illness and the personal and social resources of a patient, out-patient, partial hospitalisation, brief in-patient (2–4 months) or intermediate-length in-patient treatment (4–6 months) is recommended. Treatment lasting 6 months will continue to have a place, especially for patients with a poor prognosis and limited social support.

The outcomes of inpatient and outpatient treatment are difficult to compare, as there is only a limited number of relevant studies. An older review (Miller and Hester 1986) reached the conclusion there was no real difference in treatment outcome. Recently, this conclusion has been emphatically rejected (see above). It has been argued that the samples on which this statement was based were not comparable in severity of illness. If this is taken into account, in-patient treatment has continued to show superiority for the severely dependent (Pettinati and Belden 1996). The results of the MATCH Project (see above) also support a slight superiority for in-patient treatment and are an argument against abandoning this proven treatment modality.

3.4.3 Adjuvant Pharmacotherapy

The final development to be discussed is the introduction of anti-craving substances. This development was scarcely considered possible until a few years ago and significantly enriches the treatment options available, at least in out-patient management of alcohol dependence. These drugs can be used not only during out-patient rehabilitation treatment following a withdrawal programme (see above), but also by experienced general practitioners in the context of early interventions. Substances with effects on the cholinergic, glutamatergic, dopaminergic, serotoninergic and opioidergic systems have been clinically tested. The most

promising so far have been the glutamate modulator acamprosate and the opiate antagonist naltrexone.

Animal experiments provided the first indications that acamprosate (calcium acetyl homotaurinate) could reduce craving for alcohol and could do so without altering overall fluid intake (Boismare et al. 1984). Relapse behaviour was suppressed in rats (Spanagel et al. 1996). As the substance does not interfere with the pharmacokinetics and pharmacodynamics of alcohol (Le Magnen et al. 1987) and does not substitute for the alcohol stimulus in a discrimination task, it has been concluded that its anti-craving qualities are responsible for this effect (Spanagel and Zieglgänsberger 1997). Acamprosate reduces withdrawal symptoms (Gewiss et al. 1991). The excitatory effects on neurones of L-glutamate are suppressed by acamprosate (Madamba et al. 1996; Zeise et al. 1993). A reduction in mesolimbic dopaminergic activity during withdrawal (Diana et al. 1993) and an increase in the extracellular concentration of L-glutamate in the nucleus accumbens have also been described with acamprosate (Dahchour et al. 1996). The neuronal over-excitability observed during withdrawal is accompanied by increased expression of c-fos, an immediate-early gene. Here, too, may lie one of the effective elements in the action of acamprosate, as the substance suppresses increased c-fos expression in many neuronal structures, including the nucleus accumbens (Littleton 1995; Putzke et al. 1996). Binding probably occurs to the polyamine binding site of an NMDA-R1/R2 receptor combination (Spanagel and Zieglgänsberger 1997).

Clinical successes have been demonstrated in several placebo-controlled double-blind studies. Studies from Austria (Whitworth et al. 1996) and Italy (Poldrugo 1997) have shown increased abstinence rates for patients treated with acamprosate. In France, two different doses of acamprosate were tested, as well as placebo. The higher dose (2 g/day) was more effective than the lower one (1.3 g/day). Statistically, only the higher dose was superior to placebo (Paille et al. 1995). In a multicentre study in Germany (Saß et al. 1996), medication was administered for a period of 12 months, and patients were observed for a further 12 months after it was discontinued. The time to the first relapse and the number of days of abstinence were significantly different between the two groups. After 12 months, 42.8% of the patients in the acamprosate group were abstinent, compared with 12.7% of patients in the placebo group. The positive effects in the acamprosate group were also maintained after medication was discontinued. In a summary evaluation of 11 independent placebo-controlled studies in Europe, the effectiveness of acamprosate was demonstrated in an overall sample approaching 4000 patients. Apart from one English study, the remaining ten studies all showed positive outcomes (for reviews, see Mann 1996; Soyka 1997).

In animal experiments, multiple interactions have been demonstrated between alcohol and the endogenous opioid systems (overview by Herz, 1997). These have also received confirmation in human experiments. In probands with a high risk of alcohol dependence, the administration of alcohol led to a dose-related rise in plasma level of β-endorphin (Gianoulakis et al. 1996). Conversely, in animal experiments, small doses of opiates increase alcohol consumption (Hubbell et al. 1988). This leads to the assumption that the blockade of endogenous opioid receptors reduces alcohol intake. This was already demonstrated in the 1980s in a variety of animal experiments (e.g. Altschuler et al. 1980). Thus a direct effect might be expected from opiate antagonists such as naloxone on the rewarding effects of alcohol. Not only the μ- but also the δ- and κ-receptors appear to have a role in this (Ulm et al. 1995). Their action in humans might be connected with effects on extero- and interoceptive stimuli. These stimuli mediate conditioned alcohol-related responses, which provoke physical and psychological states such as reinforcement or conditioned withdrawal.

The effectiveness of the opiate antagonist naltrexone has been demonstrated in three studies in the United States. Volpicelli et al. (1992) investigated 70 men with alcohol disorders over a 12-week period. Patients in the experimental group showed less craving for alcohol than those in the control condition. Where relapse occurred, volumes consumed were significantly lower and episodes of excessive drinking rarer. O'Malley et al. (1992) investigated the combination of naltrexone, placebo and two psychological treatments. They compared a general supportive psychotherapy with a behaviourally oriented "coping skills" training. In a sample of 97 alcoholics, the authors were able to demonstrate that, over the course of 12 weeks, naltrexone was associated with significantly better outcomes than placebo. In addition, there was an interaction between medication and psychological treatment processes. The greatest degree of success in attainment of abstinence resulted from the combination of naltrexone and general supportive psychotherapy. Anton (1997) also found naltrexone to be superior to placebo in 136 alcohol-dependent subjects.

In Europe, the effectiveness of naltrexone has been examined in two multicentre studies in England and Germany. The results of intention-to-treat analysis were negative in both studies (Chick 1996; Gastpar and Mann 1998). On the other hand, a Swedish study (Balldin et al. 1997) did confirm the American results cited above.

On the basis of findings from animal experiments, a variety of investigations involving administration of

dopamine agonists or antagonists in humans have been carried out in the last few years. However, initial positive effects from the agonist bromocriptine could not be replicated (Borg 1983; Dongier et al. 1991). Positive effects found in open trials of the dopamine antagonist tiapridex were initially confirmed in a randomised double-blind study of 32 alcohol-dependent subjects, using a dose of 100 mg/day (Shaw et al. 1994). However, a large multicentre study showed no difference between experimental and control conditions (M. Gastpar, unpublished). Because of its high affinity for the D_1 receptor, the dopamine agonist lisuride was investigated in a placebo-controlled double-blind trial. Contrary to expectations, the relapse rate was significantly higher in the experimental than the control condition (Schmidt et al. 1997). Overall, outcomes from dopaminergic substances are not convincing. In addition, where treatment employs neuroleptic substances, the risk of extrapyramidal side-effects, especially tardive dyskinesias, must be considered.

Positive effects found for serotonin reuptake inhibitors in animal studies were not confirmed or were demonstrated only transiently in clinical studies. In a placebo-controlled double-blind trial of fluoxetine in 20 patients, significantly reduced volume of consumption and craving were observed only during the first week of treatment (Gorelick and Paredes 1992). In a 4-week placebo-controlled trial with fluoxetine, no specific reduction was demonstrated in volume of alcohol consumed, but only a general reduction in fluid intake (Naranjo et al. 1990). Kranzler et al. (1995) also found no effects. Administration of the 5-HT_2 antagonist ritanserine was not superior to placebo (Böning et al. 1997). Testing of serotonin agonists such as buspirone has proved similarly disappointing. These may, however, still be helpful in anxious alcohol-dependent patients, for whom a reduction in anxiety rather than a change in drinking patterns has been described (Kranzler et al. 1994).

A meta-analysis of studies with a treatment duration of at least 3 months has shown that some of the anti-craving substances tested so far give good results. The effectiveness is as great as that of classical antidepressants in long-term relapse prevention in unipolar depression (Engel and Schöchlin 1995). Thus combined treatment by psychological and pharmacological means has now also become possible for alcohol-dependent individuals. There may be a case for requiring that anti-craving substances should be used only in the context of psychological treatment. Otherwise, there is a danger of neglecting or even obstructing the confrontation with their illnesses and life situations which individuals in treatment need to have. Anti-craving drugs are essentially adjuvant or supportive elements in treatment.

4 References

Altschuler HL, Phillips PE, Feinhandler DA (1980) Alteration of ethanol self-administration by naltrexone. Life Sci 26: 679–688

Anton RF (1997) Naltrexone as adjunctive treatment to cognitive behavioral therapy for outpatient alcoholics. American College of Neuropsychopharmacology (ACNP), Waikoloa, Hawaii, 8–12 December 1997

APA (1994) Diagnostic and statistical manual of mental disorders, 4th edn (DSM-IV). American Psychiatric Association, Washington, DC

Athen D (1986) Syndrome der akuten Alkoholintoxikation und ihre forensische Bedeutung. Springer, Berlin Heidelberg New York

Benedetti G (1952) Die Alkoholhalluzinose. Thieme, Stuttgart

Balldin J, Berglund M, Borg S et al (1997) A randomized 6 months double-blind placebo-controlled study of naltrexone and coping skills educational programme. Alcohol Alcohol 32(3): 325

*Bien TH, Miller WR, Tonigan JS (1993) Brief interventions for alcohol problems: a review. Addiction 88: 315–336

Blass JP, Gibson GE (1977) Abnormality of a thiamine-requiring enzyme in patients with Wernicke-Korsakoff syndrome. N Engl J Med 297: 1367–1370

Boismare F, Daoust M, Moore ND, Saligaut C, Lhuintre JP, Chretien P, Durlach J (1984) A homotaurine derivative reduces the voluntary intake of ethanol by rats: are cerebral GABA receptors involved? Pharmacol Biochem Behav 21: 787–789

Böning J, Wiesbeck GA, Weijers HG (1997) Ritanserin: results of the international 6-months multicenter study. Alcohol Alcohol 32(3): 326

Borg V (1983) Bromocriptine in the prevention of alcohol abuse. Acta Psychiatr Scand 68: 100–110

Brandt J, Butters N, Ryan C, Bayog R (1983) Cognitive loss and recovery in long-term alcohol abusers. Arch Gen Psychiatry 40: 435–442

Buck KJ, Harris RA (1991) Neuroadaptive responses to chronic ethanol. Alcohol Clin Exp Res 15: 460–470

Carlen PL, Wortzman G, Holgate RC, Wilkinson DA, Rankin JG (1978) Reversible cerebral atrophy in recently abstinent chronic alcoholics measured by computed tomography scans. Science 200: 1076–1078

Chick JD (1996) Opiate receptors: role in addiction and relapse in alcohol dependence. Eur Psychiatry 11(4): 160s–161s

Charness ME (1992) Molecular mechanisms of ethanol intoxication, tolerance, and physical dependence. In: Mendelson JH, Mello NK (eds) Medical diagnosis and treatment of alcoholism. McGraw Hill, New York, pp 155–199

Colmant HJ (1965) Enzephalopathien bei chronischem Alkoholismus. Enke, Stuttgart

Courville CB (1955) Effects of alcohol on the nervous system of man. San Lucas, Los Angeles

Dahchour A, Durbin P, DeWitte P (1996) Acamprosate decreases the extracellular glutamate in the microdialysate nucleus accumbens of male withdrawn rats. Alcohol Clin Exp Res 20: 93A

Diana M, Pistis M, Carboni S, Gessa, GL, Rossetti ZL (1993) Rewarding and aversive effects of ethanol: interplay of GABA, glutamate and dopamine. Alcohol Alcohol Suppl 2: 315–319

Diener HC, Dichgans J, Bacher M, Guschelbauer B (1984) Improvement of ataxia in alcoholic cerebellar atrophy through alcohol abstinence. J Neurol 231: 258–262

Dlugos CA, Pentney RJ (1997) Morphometric evidence that the total number of synapses on purkinje neurons of old F344 rats is reduced after long-term ethanol treatment and restored to control levels after recovery. Alcohol Alcohol 32(2): 161–172

Dongier M, Vachon L, Schwartz G (1991) Bromocriptine in the treatment of alcohol dependence. Alcohol Clin Exp Res 15: 970–977

Dufour MC (1993) The epidemiology of alcohol-induced brain damage. Res Monogr 22: 3–14

Engel R, Schöchlin C (1995) Therapeutische Strategien zur Überwindung der Alkoholsucht: Metaanalyse klinischer Prüfungen. In: Mann K, Buchkremer G (eds) Suchtforschung und Suchttherapie in Deutschland. Neuland, Geesthacht, pp 102–105

Ferrer I, Fabregues I, Rairiz J, Galofre E (1986) Decreased numbers of dendritic spines on cortical pyramidal neurons in human chronic alcoholism. Neurosci Lett 69: 115–119

Flink EB, Stutzman FL, Anderson AR, Konig T, Fraser F (1954) Magnesium deficiency after prolonged parenteral fluid adminstration and after chronic alcoholism complicated by delirium tremens. J Lab Clin Med 43: 169–183

Gastpar M, Mann K (1998) A multisite placebo controlled study of naltrexone in Germany. International Society of Biomedical Research in Alcoholism (ISBRA), Copenhagen

Gewiss M, Heidebreder C, Opsomer L, Durbin P, DeWitte P (1991) Acamprosate and diazepam differentially modulate alcohol-induced behavioral and cortical alterations in rats following chronic inhalation of ethanol vapour. Alcohol Alcohol 26: 129–137

*Gianoulakis C, Krishnan B, Thavundayil J (1996) Enhanced sensitivity of pituitary β-endorphin to ethanol in subjects at high risk of alcoholism. Arch Gen Psychiatry 53: 250–257

Glass IB (1989) Alcohol hallucinosis: a psychiatric enigma – 2. follow-up studies. Br J Addict 84: 151–164

Glenn SW (1993) Sex differences in alcohol-induced brain damage. Res Monogr 22: 195–212

Goodwin DW, Crane JB, Guze SB (1969) Phenomenological aspects of the alcoholic blackout. Br J Psychiatry 115: 1033–1038

Gorelick DA, Paredes A (1992) Effect of fluoxetine on alcohol consumption in male alcoholics. Alcohol Clin Exp Res 16: 261–265

Grant KA, Woolverton WL (1989) Reinforcing and discriminative stimulus effects of Ca-acetyl homotaurine in animals. Pharmacol Biochem Behav 32: 607–611

Grawe K (1995) Grundriß einer Allgemeinen Psychotherapie. Psychotherapeut 40: 130–145

Harper C (1983) The incidence of Wernicke's encephalopathy in Australia – a neuropathological study in 131 cases. J Neurol Neurosurg Psychiatry 46: 593–598

Harper C, Corbett D (1990) Changes in the basal dendrites of cortical pyramidal cells from alcoholic patients – a quantitative Golgi study. J Neurol Neurosurg Psychiatry 53: 856–861

*Harper C, Kril J, Daly J (1987) Are we drinking our neurons away? Br Med J 294: 534–536

Harper CG, Kril JJ, Daly JM (1988) Brain shrinkage in alcoholics is not caused by changes in hydration: a pathological study. J Neurol Neurosurg Psychiatry 51: 124–127

Heimann H (1992) Geleitwort. In: Mann K (1992) Alkohol und Gehirn. Über strukturelle und funktionelle Veränderungen nach erfolgreicher Therapie. Springer, Berlin Heidelberg New York, pp V–VI

**Herz A (1997) Endogenous opioid systems and alcohol addiction. Psychopharmacology 129: 99–111

Hubbell CL, Abelson ML, Burkhardt CA, Herlands SE, Reid LD (1988) Constant infusions of morphine and intakes of sweetened ethanol solution among rats. Alcohol 5: 409–415

*Institute of Medicine (1990) Broadening the base of treatment for alcohol problems. National Academy, Washington, DC

Jensen GB, Pakkenberg B (1993) Do alcoholics drink their neurons away? Lancet 342: 1201–1204

Jernigan TL, Butters N, Di Tragila G et al (1991a) Reduced cerebral grey matter observed in alcoholics using magnetic resonance imaging. Alcohol Clin Exp Res 15: 418–427

Jernigan TL, Schafer K, Butters N, Cermak LS (1991b) Magnetic resonance imaging of alcoholic Korsakoff patients. Neuropsychopharmacology 4/3: 175–186

Julien RM (1997) Drogen und Psychopharmaka. Spektrum, Heidelberg

Kolle K (1932) Über Eifersuchtswahn bei Trinkern. Monatsschr Psychiatr Neurol 83: 224–242

Kranzler HR, Burleson JA, Del Boca FK, Babor TF, Korner P, Brown J, Bohn MJ (1994) Buspirone treatment of anxious alcoholics: a placebo-controlled trial. Arch Gen Psychiatry 51: 720–731

Kranzler HR, Burleson JA, Korner P, Del Boca FK, Bohn MJ, Brown J, Liebowitz N (1995) Placebo-controlled trial of fluoxetine as an adjunct to relapse prevention in alcoholics. Am J Psychiatry 152: 391–397

*Küfner H, Feuerlein W (1989) In-patient treatment for alcoholism. A multi-centre evaluation study. Springer, Berlin Heidelberg New York

Laux G, Reimer F (1979) Zur Pathogenese des alkoholischen Eifersuchtswahns. Nervenarzt 50: 299–301

Le Magnen J, Tran G, Durlach J, Martin C (1987) Lack of effects of Ca-acetyl homotaurinate on chronic and acute toxicities of ethanol in rats. Alcohol 4: 103–108

Lishman WA (1987) Organic psychiatry. The psychological consequences of cerebral disorder. Blackwell, Oxford

Lishman WA (1990) Alcohol and the brain. Br J Psychiatry 156: 635–644

Littleton JM (1995) Acamprosate in alcohol dependence: how does it work? Addiction 90: 1179–1188

Madamba SG, Schweitzer P, Zieglgänsberger W, Siggins GR (1996) Acamprosate (Calcium acetylhomotaurinate) enhances the *N*-methyl-D-aspartate component of excitatory neurotransmission in rat hippocampal CA1 neurons in vitro. Alcoholism Clin Exp Res 20: 651–658

*Mann K (1992) Alkohol und Gehirn – Über strukturelle und funktionelle Veränderungen nach erfolgreicher Therapie. Springer, Berlin Heidelberg New York

Mann K (1996) The pharmacological treatment of alcohol dependence. Needs and possibilities. Alcohol Alcohol 31(1): 55–58

Mann K, Batra A (1993) Die gemeindenahe Versorgung von Alkoholabhängigen – Evaluation eines kombinierten stationären und ambulanten Behandlungskonzeptes. Psychiatr Prax 20: 831–834

Mann K, Widmann U (1995) Zur Neurobiologie der Alkoholabhängigkeit: Neuropathologie und CT/NMR-Befunde. Fortschr Neurol Psychiatr 63: 238–247

Mann K, Dengler W, Nägele T, Klose U, Schroth G (1993a) Liquorvolumetrie und spektroskopische Bestimmung – eine

prospektive MR-Verlaufsstudie an Alkoholikern. In: Baumann P (ed) Biologische Psychiatrie der Gegenwart. Springer, Berlin Heidelberg New York, pp 547–550

Mann K, Mundle G, Längle D, Petersen D (1993b) The reversibility of alcohol brain damage is not due to rehydration: a CT study. Addiction 88: 649–653

Mann K, Stetter F, Günthner A, Buchkremer G (1995) Qualitätsverbesserung in der Entzugsbehandlung von Alkoholabhängigen. Dtsch Ärztebl 92(45): 2217–2221

Marlatt GA, Gordon JA (1985) Relapse prevention: maintenance strategies in the treatment of addictive behaviors. Guildford, New York

Martin PR, Adinoff B, Weingartner H, Mukherjee AB, Eckardt MJ (1986) Alcoholic organic brain disease: nosology and pathophysiologic mechanisms. Prog Neuropsychopharmacol Biol Psychiatr 10: 147–164

McMullen P, Saint-Cyr JA, Carlen PL (1984) Morphological alterations in the rat CA1 hippocampal pyramidal cell dendrites resulting from chronic ethanol consumption and withdrawal. J Comp Neurol 225: 111–118

Meyer AE (1990) Kommunale Faktoren in der Psychotherapie als Erklärung für nicht grob unterschiedliche Ergebnisse – Ein Mythos mehr in der Psychotherapieforschung? Psychother Psychosom Med Psychol 40: 152–157

Miller WR (1983) Motivational interviewing with problem drinkers. Behav Psychother 11: 147–172

Miller WR, Hester RK (1986) The effectiveness of alcoholism. What research reveals. In: Miller WR, Heather N (eds) Treating addictive behaviors. Plenum, New York, pp 121–174

Miyakawa T, Hattori E, Shikai I, Shimoji A, Nagatoshi K, Suzuki T (1977) Histopathological changes of chronic alcoholism. Folia Psychiatr Neurol Jpn 31: 253–261

Muuronen A, Bergman H, Hindmarsh T, Telakivi (1989) Influence of improved drinking habits on brain atrophy and cognitive performance in alcoholic patients: a 5-year follow-up study. Alcoholism Clin Exp Res 13/1: 137–141

Naranjo CA, Kadlec KE, Sanhuezza P, Woodley-Remus D, Sellers EM (1990) Fluoxetine differentially alters alcohol intake and other consumatory behaviors in problem drinkers. Clin Pharmacol Ther 47: 490–498

O'Malley SS, Jaffe AJ, Chang G, Schottenfeld RS, Meyer RE, Rounsaville B (1992) Naltrexone and coping scills therapy for alcohol dependence. A controlled study. Arch Gen Psych 49: 881–887

Paille FM, Guelfi JD, Perkins AC, Royer RJ, Steru L, Parot P (1995) Double-blind randomised multicentre trial of acamprosate in maintaining abstinence from alcohol. Alcohol Alcohol 30: 239–247

Parsons OA, Butters N, Nathan PE (1987) Neuropsychology of alcoholism. Implications for diagnosis and treatment. Guilford, New York

Peiffer J (1985) Zur Frage atrophisierender Vorgänge im Gehirn chronischer Alkoholiker. Nervenarzt 56: 549–657

Pettinati H, Belden P (1996) Ambulante versus stationäre Therapie bei Abhängigkeitserkrankungen: Neue Perspektiven. In: Mann K, Buchkremer G (eds) Sucht. Grundlagen, Diagnostik, Therapie. Fischer, Stuttgart, pp 265–273

Poldrugo F (1997) Acamprosate treatment in a long-term community-based alcohol rehabilitation programme. Addiction 92: 1537–1546

Prochaska JO, DiClemente CC (1986) Toward a comprehensive model of change. In: Miller WE, Heather N (eds) Treating addictive behaviors. Processes of change. Plenum, New York, pp 3–27

**Project Match Research Group (1997) Matching alcoholism treatments to client heterogeneity: project MATCH posttreatment drinking outcomes. J Stud Alcohol 58(1): 7–29

Putzke J, Spanagel R, Tölle TR, Zieglgänsberger W (1996) The anti-craving drug acamprosate reduced c-fos expression in rats undergoing ethanol withdrawal. Eur J Pharmacol 317: 39–48

Ramayya A, Jauhar P (1997) Increasing incidence of Korsakoff's psychosis in the east end of Glasgow. Alcohol Alcohol 32(3): 281–285

Ron MA, Acker W, Shaw GK, Lishman WA (1982) Computerised tomography of the brain in chronic alcoholism. A survey and follow-up study. Brain 105: 497–514

Ryback RS (1970) Alcohol amnesia: observations in seven drinking inpatient alcoholics. Q J Stud Alcohol 37: 40–45

*Saß H, Soyka M, Mann K, Zieglgänsberger W (1996) Relapse prevention by acamprosate: results from a placebo controlled study in alcohol dependence. Arch Gen Psychiatry 53: 673–680

Schmidt LG, Kuhn S, Rommelspacher H (1997) Pharmacological effects of lisuride shorten, expectations to receive the drug prolong the latency of relapse in weaned alcoholics. Pharmacopsychiatry 30: 219

Scholz H (1982) Das Ausfallssyndrom nach Unterbrechung der Alkoholabhängigkeit. Fortschr Neurol Psychiatr 50: 279–296

Schott K, Schäfer G, Bartels M, Mann K (1995) T-reflex studies in latent polyneuropathy. Alcohol Alcohol 30(4): 533

Schroth G, Remmes U, Schupmann A (1985) Computertomographische Verlaufsuntersuchungen von Hirnvolumenschwankungen vor und nach Alkoholentzugsbehandlung. Fortschr Röntgenstr 142(4): 363–369

Schroth G, Naegele T, Klose U, Mann K, Petersen D (1988) Reversible brain shrinkage in abstinent alcoholics, measured by MRI. Neuroradiology 30: 385–389

Schwoon D (1996) Nutzung professioneller Nachsorge und Selbsthilfegruppen durch Alkoholiker nach stationärer Kurzzeittherapie. In: Mann K, Buchkremer G (eds) Sucht. Grundlagen, Diagnostik, Therapie. Fischer, Stuttgart, pp 281–287

Shaw GK, Wallers S, Majumdar SK, Alberts JL, Latham CJ, Dunn G (1994) Tiapride in the prevention of relapse in recently detoxified alcoholics. Br J Psychiatry 165: 515–523

Soyka M (1990) Psychopathological characteristics in alcohol hallucinosis and paranoid schizophrenia. Acta Psychiatr Scand 81: 255–259

Soyka M (1997) Relapse prevention in alcoholism. Recent advances and future possibilities. Drug Therapy 7(4): 313–327

Spanagel R, Zieglgänsberger W (1997) Anti-craving compounds for ethanol: new pharmacological tools to study addictive processes. Trends Pharmacol Sci 18: 54–59

Spanagel R, Zieglgänsberger W, Hundt W (1996) Acamprosate and alcohol. III. Effects on alcohol discrimination in the rat. Eur J Pharmacol 305: 51–56

Stetter F, Mann K (1997) Zum Krankheitsverlauf Alkoholabhängiger nach einer stationären Entgiftungs- und Motivationsbehandlung. Nervenarzt 68: 574–581

Süß HM (1995) Zur Wirksamkeit der Therapie bei Alkoholabhängigen: Ergebnisse einer Meta-Analyse. Psychol Rundschau 46: 248–266

Thomas A (1905) Atrophie laméllaire des céllules de Purkinje. Rev Neurol 13: 917

Thomson A, Cook C (1997) Parenteral Thiamine and Wernicke's encephalopathy: the balance of risks and perception of concern. Alcohol Alcoholism 32(3): 207–209

Torvik A, Torp S (1986) The prevalence of alcoholic cerebellar atrophy. A morphometric and histological study of an autopsy material. J Neurol Sci 75: 43–51

Torvik A, Lindbö CF, Rodge S (1982) Brain lesions in alcoholics. J Neurol Sci 56: 233–248

Tsuang JW, Irwin MR, Smith TL, Schuckit MA (1994) Characteristics of men with alcoholic hallucinosis. Addiction 89: 73–78

Ulm RR, Volpicelli JR, Volpicelli LA (1995) Opiates and alcohol self-administration in animals. J Clin Psychiatry 56[Suppl 7]: 5–14

Vauhkonen K (1968) On the pathogenesis of morbid jealousy. Acta Psychiatr Scand Suppl 202–212

Veltrup C (1995) Abstinenzgefährdung und Abstinenzbeendigung bei Alkoholabhängigen nach einer umfassenden stationären Entzugsbehandlung. Waxmann, Münster

Veltrup C, Weber J, Metten D, Driessen M, John U (1995) Katamnestische Untersuchungen bei Alkoholabhängigen. In: Mann K, Buchkremer G (ed) Suchtforschung und Suchttherapie in Deutschland. Sucht Z Wiss Prax (special issue): 172–173

Victor M (1990) MR in the diagnosis of Wernicke-Korsakoff syndrome. AJNR Am J Neuroradiol 11: 895–896

Victor M, Adams RD (1953) The effect of alcohol on the nervous system. Res Publ Assoc Res Nerv Ment Dis 32: 526–573

Victor M, Adams RD, Collins GH (1971) The Wernicke-Korsakoff syndrome. Blackwell, Oxford

Volpicelli JR, Altermann AI, Hayashida M, O'Brien CP (1992) Naltrexone in the treatment of alcohol dependence. Arch Gen Psychiatry 49: 876–880

Wienberg G (1992) Die vergessene Mehrheit. Zur Realität der Versorgung alkohol- und medikamentenabhängiger Menschen. Psychiatrie, Berlin

Wieser S, Kunad E (1965) Katamnestische Studien beim chronischen Alkoholismus und zur Frage von Sozialprozessen bei Alkoholikern. Nervenarzt 36: 477–483

Whitworth AB, Fischer F, Lesch OM et al (1996) Comparison of acamprosate and placebo in long-term treatment of alcohol dependence. Lancet 347: 1438–1442

Zeise ML, Kasparov S, Capogna M, Zieglgänsberger W (1993) Calcium diacetylhomotaurinate (Ca-AOTA) decreases the action of excitatory amino acids in the rat neocortex in vitro. Eur J Pharmacol 231: 47–52

CHAPTER

20

A. Batra, G. Buchkremer

Tobacco Misuse

Translator: S. Johnson

1 History of Tobacco Consumption

Several thousand years B.C., the tobacco plant (*Nicotiana tabacum* L.) was already in use in South and Central America for both medical and religious purposes. Only with the return of the Spanish conquistadors in the sixteenth century, however, did tobacco arrive in Europe and begin its controversial triumphal progress across the continent as a new recreational substance. Smoking was initially confined to the upper echelons of society. The tobacco pipe, which was used to "drink" of tobacco smoke in the seventeenth and eighteenth centuries, required elaborate and careful manipulation and skill. Tobacco consumption therefore demanded a great deal of time and effort. Smoking was not simplified until the cigar was introduced in the nineteenth century, followed by the mass production of cigarettes in the second half of the nineteenth century. From this time onwards, tobacco products became widely available, quick and unproblematic to consume, and readily affordable even for the lower classes.

Tobacco had considerable economic importance as early as the eighteenth century. Income from tobacco tax financed the English merchant navy and the American War of Independence, for example.

A positive, mentally stimulating effect was ascribed to tobacco. Even so, in the first decades after the plant's introduction, critical voices, including both statesmen and doctors, were already warning against "dry drunkenness" (von Birken 1658) and the health risks associated with tobacco use. They recognised both the danger of addiction and the potential health risks and also complained that smoke was polluting the environment. However, they had no more success in impeding the spread of tobacco use than the societies founded around the beginning of the twentieth century in North America and Central Europe in the fight against smoking.

2 Epidemiology of Smoking

After the Second World War, there was a remarkable increase in smoking in Central Europe. In the space of a few decades, between the 1950s and the 1970s, the quantity of cigarettes purchased in Germany increased by a factor of 4. Information about the latest trends in prevalence of smoking in Germany is available from the surveys carried out by the Federal Statistical Office in Wiesbaden in 1978, 1989, 1992 and 1995 (Statistisches Bundesamt 1993, 1996) (Fig. 1). The prevalences quoted relate to the adult population aged over 15 and include all consumers of tobacco, both heavy smokers and occasional users.

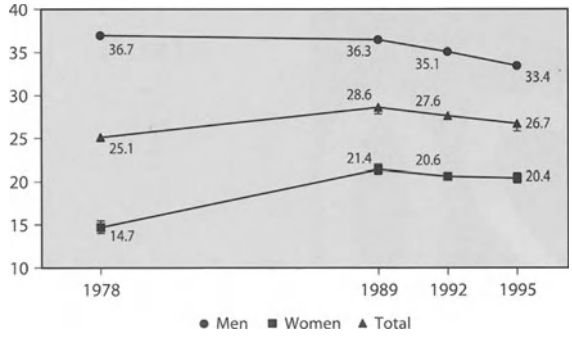


Fig. 1. Proportion of smokers in the German population, 1978–1995. (Source: Statistisches Bundesamt 1993, 1996)

The surveys document a slight decrease since 1989 in the proportion of the population smoking. The most recent data, collected in 1995, indicate that around 33.4% of all men and 20.4% of all women in Germany are smokers. Comparison with figures from the preceding surveys show a small decrease among men of 3%, while among women the reduction over the preceding 3 years was a mere 0.7%.

The validity of these figures requires critical scrutiny: the actual prevalence of smokers should be assumed to be higher than these government surveys suggest. Public campaigns have resulted in smoking increasingly becoming an undesirable social behaviour, and some smokers therefore tend to give responses about their smoking habits which gloss over reality.

Rates cited in a European comparative study (BASP 1994) are higher: the average prevalence rate for men in the European Union is 42% and for women 28%. While this survey puts the proportion of German men who are smokers slightly above the European average (former West Germany 42%, former East Germany 47%), women in Germany are somewhat less likely to smoke than in the European Union as a whole (former West Germany 27%, former East Germany 23%). The highest proportions of smokers among men are reported in Spain (46%), France (47%) and the Netherlands (48%). The peak figures for women come from Denmark (48%), the Netherlands (37%), France and Great Britain (both 30%) and Luxembourg (29%).

While the epidemiological data from the Federal Statistical Office quoted above suggest a slight decline in smoking, a contrasting picture emerges from statistics regarding purchase of tobacco products and income from taxes on tobacco (Fig. 2). In 1996, taxes were collected on 136.2 billion cigarettes, 0.9% more

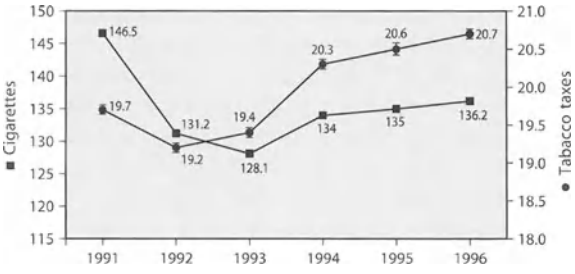


Fig. 2. Approximate levels of cigarette use (in billions of cigarettes) and income from tobacco taxes (in billions of German marks), 1991–1996. (Source: Statistisches Bundesamt 1995, 1996)

than in the preceding year. The tax income from all tobacco products in 1996 was 36.1 billion German marks, 1.4% more than in the preceding year. The constantly increasing consumption of tobacco products suggests either that smoking is continuing to become more widespread or that tobacco is being consumed more intensively by smokers. Estimates from the cigarette industry and the customs authority suggest that an additional 3.5–7 billion untaxed cigarettes are consumed per year (quoted in Junge 1995). Within Europe, cigarette smuggling is less a problem in those countries where taxes on tobacco are high than in Eastern and Southern European countries, where there is extensive street dealing in smuggled international cigarette brands (Joossens and Raw 1995).

The past few decades have seen extensive efforts to educate the public about smoking. Restrictions on tobacco advertising, repeated tax increases and campaigns, as yet not fully successful, for laws to protect non-smokers through smoking bans on public transport and in workplaces have been initiated. These initiatives have not had any enduring impact on patterns of smoking. Smoking cessation programmes offered through health insurance companies, local public health departments, doctors, psychologists and pharmacists have also failed to result in the reductions in the prevalence of smoking that were hoped for.

3
Definitions of Tobacco Dependence and Misuse

3.1
Types of Smoker

It is no longer disputed that regular tobacco consumption can lead to dependence. However, not all tobacco users are necessarily dependent smokers. A variety of motivations are associated with the use of tobacco, generally in the form of cigarettes, sometimes cigars and pipe tobacco, and rarely snuff. A frequently used typology differentiates between the opportunistic or hedonistic smoker, the conflict or stress smoker and the dependent or habitual smoker. The opportunistic smoker reaches for a cigarette in company or on particular occasions, but does not generally engage in any smoking rituals, consumes only a limited number of cigarettes and, if he or she does consume larger quantities of tobacco, often develops signs of tobacco intoxication. The hedonistic smoker has learnt to reward him- or herself in certain situations by smoking a cigarette, cigar or cigarillo, but does not consume larger quantities of tobacco. The conflict smoker, meanwhile, resorts to cigarettes in psychologically stressful or distressing situations, and he or she may therefore at times smoke large quantities of cigarettes, but is also capable of relatively long periods of abstinence. These patterns of use may be seen as instances of tobacco misuse. In contrast, characteristics of psychological and/or physical dependence are generally manifest in habitual smokers. The habitual smoker usually smokes a similar quantity of cigarettes each day, often around 20–40 or even more, inhales deeply and develops a powerful craving for cigarettes if they cease to be available. This type of smoker rarely displays manifestations of intoxication and begins to report withdrawal symptoms as soon as a significant period (usually a few hours) has passed without a cigarette.

3.2
Smokers' Motives

In surveys, more than 50% of heavy smokers say that they would like to reduce or even completely stop their consumption of cigarettes, yet only a few succeed in maintaining abstinence.

A wide variety of motivations lead people to smoke. The most frequently cited reasons are habit and pleasure. In moments of inner tension, smoking is experienced as calming, distracts from boredom and is felt to be a well-earned reward after coping with a situation which has been experienced or anticipated as aversive. Many social reinforcers (e.g. fitting in, sociability, image) encourage smoking in company.

A significant motivation for the dependent smoker to smoke is suppression of withdrawal symptoms. Withdrawal symptoms are triggered by a falling blood nicotine level and include nervousness, agitation, irritability, aggression, sleep disturbances, feelings of tiredness and exhaustion, disturbances of concentration and depressive and anxiety symptoms. Associated somatic symptoms include headaches, dizziness, faintness and gastrointestinal symptoms and are related to vagal cholinergic overactivity. The weight gain

observed in most cases is attributable to an increase in appetite, the attempt to find alternative oral gratification and altered metabolism. While the somatic and psychological symptoms are generally present only for a few days (usually between 1 day and 2 weeks), craving for cigarettes and weight gain generally persist over a period of several months.

In the early stages of abstinence, there may be a deterioration or relapse in pre-existing illnesses (e.g. depressive illnesses or ulcerative colitis), which is a further threat to abstinence.

3.3 Definition of Dependence

More than half of all heavy smokers can be described as dissonant smokers, in that they would like to give up smoking, but continue to smoke or repeatedly relapse.

According to the classification criteria for disorders related to psychotropic substances in the psychiatric section of the International Classification of Diseases (ICD-10, Chap. 5), the following features indicate the presence of tobacco dependence: a persistent strong wish or compulsion to smoke, diminished control over the initiation of tobacco consumption and over how much is consumed, the emergence of withdrawal symptoms or tolerance to the physiological effects of the substance. Other features include narrowed behavioural repertoire in patterns of use of the substance, persisting use despite awareness of obvious damaging consequences in physical, psychological or social domains and a persisting *fruitless* wish or multiple unsuccessful attempts to limit tobacco consumption.

By way of simplification, the following features can be used in practice as indicators of the presence of tobacco dependence: smoking early in the morning, consumption of more than ten cigarettes per day and evidence of multiple fruitless attempts at abstinence in the past.

Recently, the distinction between tobacco misuse and dependence has come to be seen as dimensional rather than dichotomous. Severity of dependence is indicated not only by the number of cigarettes consumed each day, but also by the pattern of consumption. Where dependence is severe, the smoker inhales deeply, lights the first cigarette soon after getting up and often smokes more heavily in the morning than in the afternoon or evening. Observing local or temporary bans on smoking is a struggle, and tobacco consumption continues during periods of illness.

Various psychometric scales allow measurement of severity of dependence. The most widely used at present is the Fagerström Test for Nicotine Dependence (FTND; Heatherton et al. 1991). Several studies have shown severity of dependence to be a significant predictor of the probability of long-term abstinence following completion of a smoking cessation treatment. Investigations in a variety of European countries also suggest a reciprocal connection between severity of tobacco dependence and prevalence of smokers. Thus the fall in the proportion of smokers in the population seems in some countries to be attributable primarily to a reduction in the number of light smokers. As it is principally heavy smokers who develop the adverse health consequences of smoking, a falling proportion of smokers in the population can unfortunately not be interpreted as a direct indicator of an improvement in the population's health.

3.4 Causes and Conditions in the Development of Dependence

Psychological and physiological factors are considered to be equally important in the development and persistence of tobacco dependence.

The psychotropic effects of nicotine are believed to be responsible for the development of physical dependence. Results from animal experiments, which show activation of the mesolimbic dopaminergic system and, in particular, release of dopamine in the nucleus accumbens, the main area involved in the self-reward system of the brain, suggest that nicotine has an addictive potential approaching that of amphetamines, alcohol or cocaine (DiChiara and Imperato 1988).

An increase in central $(\alpha_2\beta_4)$-acetylcholine receptors due to nicotine is regarded as being responsible for the emergence of withdrawal symptoms. In animal experiments, regular administration of nicotine stimulates an increase in neuronal nicotinergic acetylcholine receptors (Balfour and Fagerström 1996; Hsu et al. 1996; Wonnacott 1990).

The directly and contingently arising psychotropic effects of nicotine on the self-reward system of the brain promote intense craving for stimulation through nicotine. The formation of an association between pleasant psychotropic effects and smoking behaviour underlies classical and operant learning mechanisms. Meanwhile, maintenance of dependence occurs when the organism strives to avoid experiencing aversive withdrawal symptoms as blood nicotine level falls by renewed consumption of nicotine. Even the first manifestations of withdrawal, of which the smoker may hardly be aware, are discriminative stimuli for the maintenance of smoking (Tölle and Buchkremer 1989).

Smoking is a learned behaviour as well as a biological process, and the processes of imitative and operant learning have a role. Both positive and negative reinforcement mechanisms ensure that the addictive behaviour persists.

As smoking develops in youth and young adulthood, it is increasingly associated with certain situations: smok-

ing becomes important to relieve stress at work, leads to a reduction in tension during conversation, encourages concentration and promotes relaxation. Cues such as ashtrays, cigarette boxes lying around, the break at work, morning coffee or an evening rendezvous in a bar lead to unconscious craving for a cigarette. The smoking habit encompasses both responses to key cues and individual patterns of behaviour.

4 Development of Smoking Behaviour

Smoking generally begins in youth and early adult life. Role models are responsible – parents, teachers and peer group members, but also youth idols and cleverly used images in advertising. Many attributes make smoking attractive to young people: it represents entry into adult life, serves as a display of courage or as a "cool", socially desirable behaviour and increases the social status of the smoker. This explains the fact that the prevalence of smoking among the young has remained high and stable. Societal influences also underlie the continuing slight increases in prevalence of smoking among women. Rates of smoking among young women are clearly rising with increasing emancipation in Western European countries.

5 Adverse Effects of Tobacco on Health

While nicotine seems to have only limited aetiological significance in severe physical illness (in an adapted smoker, nicotine itself does not cause any severe health disorders: recent studies suggest that nicotine has no influence in changing the lipid distribution in the blood, and administration of pure nicotine does not increase the tendency towards developing thrombosis), the adverse effects of smoking on health can be ascribed to others of the many substances contained in tobacco. It is estimated that tobacco contains more than 3800 substances, several of which are carcinogenic and/or teratogenic.

Great methodological problems arise in calculating the frequency of tobacco-associated illnesses and deaths. However, comprehensive epidemiological surveys do allow relatively reliable estimation of tobacco-associated mortality from ischaemic heart disease (25% of deaths from this cause among men and 20% among women are thought to be caused by smoking), lung cancer (90% of deaths among men, 50% among women) and bronchitis and emphysema (75% of deaths among men, 60% among women). According to this conservative estimate, which does not take into account any other causes of death, around 80,000 smokers died of these illnesses in 1994 in Germany alone. Estimates for Europe suggest around 500,000 tobacco-associated deaths per year.

Smoking leads to an increased tendency towards developing thrombosis, altered lipoprotein fractions (high- and low-density lipoproteins, HDL/LDL) and damage to vessel endothelium, probably due to free radicals. This clearly elevated risk of cardiovascular disease should certainly not be attributed simply to the effects of nicotine: cardiac infarction is promoted by other components in tobacco smoke, such as carbon monoxide, which reduces oxygen saturation.

Smoking is not only a major aetiological factor in the development of lung cancer, but also one of the risk factors for other cancers, including laryngeal, oral cavity, pancreatic, and bladder cancer and even leukaemia (Table 1). More than 40 carcinogenic substances in tobacco (e.g. *N*-nitrosamine, vinyl chloride formaldehyde, benzol, polycyclic aromatic hydrocarbons, free radicals, polonium-210, lead, cadmium) promote the development of malignancies.

Not only the smoker, but also his or her environment is contaminated by the substances contained in tobacco smoke. Passive smoking is a term which has become increasingly prominent in recent years in discussions about the effects of smoking on health and in campaigns for legal regulations to protect non-smokers. It describes the exposure of non-smokers to tobacco smoke in the workplace, at home and during leisure activities. The connection between passive smoking and increased risks of lung cancer (EPA 1992) and heart disease (Glantz and Parmley 1995) is accepted as established. A particularly important association is the tendency of children from households where there is heavy smoking to develop respiratory illnesses. The newborn children of mothers who smoke tend to have reduced body size and low birth weight (Behavioral Risk Factor Surveillance System 1989; Conter et al. 1995). They are also at statistically greater risk of developing respiratory illnesses. Passive smoking has been shown to be a significant factor in sudden infant death syndrome (Cnattingius et al. 1988; DiFranza and Lew 1995).

6 Parkinson's Disease, Alzheimer's Disease and Tobacco Misuse

Smokers are actually at reduced risk of developing Parkinson's or Alzheimer's disease (Graves and

Table 1. Tobacco smoking and cancer mortality (Newcomb and Carbone 1992)

Carcinoma	Sex	Relative risk [a]	Proportion of mortality attributable to smoking (%)
Lung	Men	22.4	90
	Women	11.9	79
Larynx	Men	10.5	81
	Women	17.8	87
Oral cavity	Men	27.5	92
	Women	5.6	61
Oesophagus	Men	7.6	78
	Women	10.3	75
Pancreas	Men	2.1	29
	Women	2.3	34
Bladder	Men	2.9	47
	Women	2.6	37
Kidney	Men	3.0	48
	Women	1.4	12
Stomach	Men	1.5	17
	Women	1.5	25
Leukaemia	Men	2.0	20
	Women	2.0	20
Cervix	Women	2.1	31

[a]Risk for non-smokers, 1.0.

Mortimer 1994). The effects of nicotine include release of dopamine in the extrapyramidal motor system (Balfour and Fagerström 1996). The reduced relative risk of Alzheimer's disease has been explained in terms of the increase in cortical and subcortical nicotinergic acetylcholinergic receptors (Benwell et al. 1988). In addition, a neuroprotective effect mediated by stimulation of production of nerve growth factor (NGF) has been described (Freedman et al. 1993).

7 Tobacco Misuse Among Psychiatric Patients

Psychiatric patients are a high-risk population for heavy smoking. Among alcohol- and drug-dependent patients, in particular, but also those with schizophrenia or personality disorder, the prevalence of smoking has been shown in a variety of studies in both inpatient and outpatient settings to be far above general population rates (Hughes et al. 1986).

Depending on the study, the prevalence of comorbid tobacco dependence among people with schizophrenic illnesses is between 50% and 90%. Various hypotheses have been advanced to account for the high prevalence of smoking among patients who take neuroleptics. Smoking may be fostered by situational factors such as boredom or lack of diversion during an inpatient stay, by the influence on generally young schizophrenic patients of fellow patients who smoke or by anhedonia and apathy. Schizophrenic patients with severe negative symptoms seem to be particularly at risk, as they use the stimulation associated with tobacco consumption to increase motivation.

An advantage of tobacco consumption to patients taking neuroleptics is that it reduces the biological half-life of highly potent neuroleptics, and the experience of extrapyramidal side-effects therefore becomes less unpleasant (Erdmann 1995). The explanation lies in the greater concentration and increased release of dopamine in the mesolimbic system. Recent genetic studies on the genesis of schizophrenia have shown a connection between the expression of a cerebral α_7-acetylcholine receptor coded for on chromosome 15 (Freedman et al. 1997) and tobacco misuse: the measurable neuropsychological defect associated with this genetic characteristic can be influenced by administering nicotine.

An increased prevalence of tobacco dependence has also been recognised among people who suffer from depression. Conversely, smokers are more likely than non-smokers to have a history of depressive illness. In addition, smokers who have experienced at least one depressive episode are less likely to attain abstinence.

Some authors go so far as to see causal connections between a tendency to depression and smoking. They explain this in terms of tobacco-associated suppression and inhibition of monoamine oxidase A and B among smokers (Berlin et al. 1995).

8
Strategies for Reducing the Prevalence of Smoking

The damaging effects of smoking and resulting costs to the healthcare system make measures to reduce smoking essential. Two goals are relevant: people who do not smoke must be prevented from becoming smokers, and current smokers must be helped to quit.

Primary prevention includes measures such as restrictions or bans on tobacco advertising, legal regulation of trade in and consumption of tobacco products, limits on smoking in public buildings or on the public transport system, protection for non-smokers in the workplace and increases in tobacco taxes. Education about the dangers of smoking needs to be directed at young people who have not yet become smokers and at people who are already heavy smokers.

Smoking continues to be linked with too many positive attributes: in advertising, it is associated with freedom, adventure, attractive appearance, sociability and good taste. Only comprehensive information about the mechanisms which make advertising effective and the conditions in which dependence develops can help to combat this image.

9
Smoking Cessation

Finally, the numbers of tobacco-associated deaths expected to occur over the next few years will only be reduced if regular smokers stop smoking.

Smoking cessation treatment has a long history. However, only since the 1960s have interventions that draw on scientifically based concepts been developed and investigated empirically.

The smoking cessation interventions which have so far proved most successful are based on the methods of behaviour therapy and may be offered in individual or group form. Pharmacological interventions are also available with some degree of effectiveness in supporting withdrawal, particularly when combined with psychological interventions (APA 1996).

9.1
Nicotine Replacement

Since 1984, nicotine has been available as a medication, to be used above all as an adjunct to behaviour therapy aimed at smoking cessation. Nicotine replacement takes the form of a patch, chewing gum or a nasal spray and is intended as a temporary, bridging strategy to reduce the withdrawal symptoms which endanger abstinence. Other potential forms of replacement (nicotine inhalers or tablets) are not available in Germany. Nicotine replacement preparations are primarily intended to alleviate withdrawal symptoms at the beginning of smoking cessation treatment. The greatest success has been achieved with nasal sprays and skin patches (Silagy et al. 1994). The principle on which these preparations work is that chewing nicotine gum or continuous administration of nicotine via a patch reduces the manifestations of withdrawal and the intense craving to have a cigarette. This facilitates the initial process of withdrawal and increases the likelihood that psychological treatment will be successful.

Nicotine chewing gum and nasal spray, which rapidly satisfy craving, are unduly similar to the previous addictive behaviour. Thus from the perspective of behavioural psychology, it is preferable to use a nicotine patch in order to break the association between replacement of the addictive substance and the previous pattern of behaviour.

9.2
Behaviour Therapy

In behaviour therapy for smoking cessation, the often still fragile motivation of the smoker for abstinence first needs to be reinforced. This is followed by a self-monitoring phase which involves identification of the old behaviour patterns associated with smoking. This allows these patterns to be erased and replaced with new ones.

After a short period of preparation, the smoker should, if possible, endeavour to give up smoking completely, rather than to gradually reduce the number of cigarettes smoked. This prevents the smoker from achieving a temporary reduction to five to ten cigarettes and deciding that this is an acceptable and attractive target, thus endangering the original goal of total abstinence. As nicotine replacement preparations are not intended to be used concurrently with cigarettes, giving up completely makes the use of these preparations less problematic than with gradual reduction.

Relapse prevention strategies include using role playing to recognise and cope with situations where there is a high risk of relapse, and regular booster sessions aimed at reinforcing motivation for abstinence.

Current combination treatments combine these various behavioural interventions with nicotine replacement, generally in the form of a nicotine patch (Buchkremer and Rath 1989).

Variants of the behaviour therapy for smoking cessation used in groups allow smokers to try to

achieve smoking cessation on their own with the help of written material.

A major target in current research on smoking cessation is the development of risk group-specific smoking cessation programmes. These are tailored to the specific needs and characteristics of particular subgroups of smokers, such as highly dependent smokers or those who are in serious danger due to associated medical illnesses.

9.3 Other Smoking Cessation Interventions

A plethora of alternative smoking cessation interventions (laying on of hands, natural healing products) are offered and utilised by many patients, but cannot be considered valid smoking cessation treatments, as scientific evidence is insufficient. Dubious treatments frequently offer smokers the prospect of remarkable treatment results without it being at all clear what therapeutic principle is being applied.

Interventions such as acupuncture, treatments based on suggestion such as hypnosis and aversion therapies are also worth mentioning. Acupuncture and hypnosis have the disadvantage that they equip the smoker with few strategies for self-help when their abstinence is under threat. The "rapid smoking" procedures used in aversion therapy may involve significant dangers to health.

Other pharmacological treatments used for smoking cessation include antagonism of the effects of nicotine (mecamylamine) and use of silver nitrate to spoil the taste of cigarettes. Very promising initial results have been reported for use of the antidepressant bupropion (300 mg/day, not available in Germany), which appears to be superior to nicotine patches. Results from investigations of the effectiveness of other antidepressants, beta blockers or neuroleptics in assisting cessation have been contradictory or not very impressive.

9.4 Outcome of Smoking Cessation Treatment

The great difficulty of achieving smoking cessation is highlighted by the fact that figures for prevalence of smoking in Germany have remained consistently high, despite extensive initiatives in the areas of education and smoking cessation. This reflects not only the small proportion of smokers to whom smoking cessation treatment is made available, but also the success rates of the treatments, which have thus far been low (Batra 1996). Whereas merely making a resolution to give up smoking leads to long-term success rates of less than 3%, prospects of success are increased to 5%–15% if a nicotine patch is used without the provision of any other treatment. The most successful smoking cessation method currently appears to be behaviour therapy combined with nicotine replacement via a patch. This achieves 1-year abstinence rates of around 25%–40%. Risk group-specific smoking cessation treatments which are geared to the needs of particular patient groups (e.g. coronary patients, pregnant women) achieve 1-year abstinence rates of up to 50% in some cases.

Greater success in reducing smoking can only be expected if extensive measures can be implemented in primary care and if smoking cessation treatments are made available to a wider section of the population.

10 References

**APA (1996) Practice guidelines for the treatment of patients with nicotine dependence. Am J Psychiatry 153[Suppl 10]: 1–31

Balfour DJ, Fagerström KO (1996) Pharmacology of nicotine and its therapeutic use in smoking cessation and neurodegenerative disorders. Pharmacol Ther 72: 51–81

BASP (ed) (1994) Tobacco and health in the European Union: an overview. European Bureau for Action on Smoking Prevention, Brussels

Batra A (1996) Raucherentwöhnung. Aktueller Stand und künftige Entwicklungen. Prax Klin Verhalt Med Rehab 34: 73–77

Behavioral Risk Factor Surveillance System (1989) Cigarette smoking among reproductive-aged women. MMWR Morb Mortal Wkly Rep 40: 719–723

Benwell MEM, Balfour DJK, Andreson JM (1988) Evidence that smoking increases the density of nicotine bindings sites in human brain. J Neurochem 50: 1243–1247

Berlin I, Said S, Spreux-Varaquaux O, Olivares R, Launay J-M, Puech AJ (1995) Monoamine oxidase A and B activities in heavy smokers. Biol Psychiatry 38: 756–761

Buchkremer G, Rath N (1989) Raucherentwöhnung. Psychologische und pharmakologische Methoden. Thieme, Stuttgart

Cnattingius S, Haglund B, Meirik O (1988) Cigarette smoking as a risk factor for the late fetal and early neonatal death. Br Med J 297: 258–261

Conter V, Cortinovis I, Rogari P, Riva L (1995) Weight growth in infants born to mothers who smoked during pregnancy. Br Med J 310: 768–771

DiChiara G, Imperato A (1988) Drugs abused by humans preferentially increase synaptic dopamine concentrations in the mesolimbic system of freely moving rats. Proc Natl Acad Sci USA 85: 5274–5278

DiFranza JR, Lew RA (1995) Effect of maternal cigarette smoking on pregnancy complications and sudden infant death syndrome. J Fam Pract 40: 385–394

EPA (1992) Respiratory health effects of passive smoking: lung cancer and other disorders. U.S. Environmental Protection Agency, Washington, DC (EPA/600/6–90/006F)

Erdmann R (1995) Neuroleptika und Nikotin. Psychiatr Prax 22: 223–227

Freedman R, Wetmore C, Stromberg I, Leonard S, Olson L (1993) Alpha-bungarotoxin binding to hippocampal interneurons: immunocytochemical characterization and effects on growth factor expression. J Neurosci 13: 1965–1975

Freedman R, Coon H, Myles-Worsley M et al (1997) Linkage of a neurophysiological deficit in schizophrenia to a chromosome 15 locus. Proc Natl Acad Sci USA 94: 587–592

Glantz SA, Parmley WW (1995) Passive smoking and heart disease: mechanisms and risk. JAMA 273: 1047–1053

Graves AB, Mortimer JA (1994) Does smoking reduce the risk of Parkinson's and Alzheimer's disease? J Smok Disord 5[Suppl 1]: 79–90

Heatherton TF, Kozlowski LT, Frecker RC, Fagerström KO (1991) The Fagerström Test for Nicotine Dependence: a revision of the Fagerström Tolerance Questionnaire. Br J Addict 86: 1119–1127

Hsu YN, Amin J, Weiss DS, Wecker L (1996) Sustained nicotine exposure differentially affects alpha 3 beta 2 and alpha 4 beta 2 neuronal nicotinic receptors expressed in xenopus oocytes. J Neurochem 66: 667–675

Hughes JR, Hatsukami DK, Mitchell JE, Dahlgren LA (1986) Prevalence of smoking among psychiatric outpatients. Am J Psych 143: 993–997

Joossens L, Raw M (1995) Smuggling and cross border shopping of tobacco in Europe. Br Med J 310: 1393–1397

Junge B (1995) Tabak. In: Deutsche Hauptstelle gegen Suchtgefahren e.V. (DHS) (ed) Jahrbuch Sucht 1996. Neuland, Geesthacht, pp 69–83

Newcomb PA, Carbone PP (1992) The health consequences of smoking. Cancer. Med Clin North Am 76: 305–331

**Silagy C, Mant D, Fowler G, Lodge M (1994) Meta-analysis on efficacy of nicotine replacement therapies in smoking cessation. Lancet 343: 139–142

Statistisches Bundesamt (ed) (1993) Fragen zur Gesundheit 1992. Metzler-Pöschel, Stuttgart (Fachserie 12, Reihe S3)

Statistisches Bundesamt (ed) (1995) Absatz von Tabakwaren 1994. Metzler-Pöschel, Stuttgart (Fachserie 14: Finanzen und Steuern, Reihe 9.1.1)

Statistisches Bundesamt (ed) (1996) Fragen zur Gesundheit 1995. Metzler-Pöschel, Stuttgart (Fachserie 12, Reihe S3)

**Tölle R, Buchkremer G (1989) Zigarettenrauchen. Epidemiologie, Psychologie, Pharmakologie und Therapie, 2nd edn. Springer, Berlin Heidelberg New York

von Birken S (1658) Die truckene Trunckenheit. Eine aus Jacobi Balde Lateinischem gedeutschte Satyra oder Straff-Rede wider den Mißbrauch des Tabaks. Samt einem Discurs von dem Nahmen, Ankunft, Natur, Krafft und Würkung deses Krauts. Endter, Nuremberg

Wonnacott S (1990) The paradox of nicotinic acetylcholine receptor upregulation by nicotine. Trends Pharmacol Sci 11: 216–219

CHAPTER

21

F.R. Levin, H.D. Kleber

Drug Abuse: Overview and New Research Directions

1 Introduction

With the burgeoning growth of knowledge, no single chapter can adequately address all aspects of drug abuse. Therefore, this chapter provides a general overview with a special focus on current "state-of-the-art" methods of evaluation and treatment. Emphasis is placed on areas which are currently being researched as well as those that need further scientific exploration.

2 Epidemiology

2.1 Surveys to Assess Substance Use Patterns in the General Population

A variety of surveys are routinely carried out in the United States and are useful for discerning trends in the pattern of drug and alcohol use. Although each has flaws, taken together they can be important for detecting patterns of change in the populations. They are generally less useful for providing estimates of heavy or addicted use.

The High School Survey is a yearly questionnaire given to approximately 50,000 eighth-, tenth-, and twelfth-graders in school with a representative sample chosen for follow-up. Based on statistics collected in 1995, there have been several trends of concern. Following a decline of over 50% between 1979 and 1991, marijuana use has increased since the early 1990s, with a 2.5-fold increase since 1991; the use of LSD has also increased. The rise in adolescent use has been accompanied by a decrease in perceived risk and peer disapproval of marijuana and LSD use. Until these turn around, it is likely that use will continue to rise. The 1995 rate, however, still remains substantially below the 1979 peak.

The National Household Survey on Drug Abuse, in which over 20,000 individuals are interviewed yearly, provides important data on both adult and adolescent use. The 1995 report showed that the number of adult illicit drug users has remained steady since 1992.

In an average month, approximately 5% of Americans used illicit drugs, 4% used marijuana, 1.4 million used cocaine, 5% had five or more drinks per occasion on 5 or more days in the month, and 24% – including 4 million adolescents aged 12–17 – smoked cigarettes.

Since 1985, "casual" (non-addicted) cocaine use has sharply declined among those at high school and in the population at large, but heavy use has not declined, as shown by a gradual rise in emergency room surveys (the Drug Abuse Warning Network data, DAWN). This may reflect both the large numbers already in the pipeline and the shortage of adequate treatment services. Heroin abuse is the most difficult to track. A variety of indicators suggest a gradual rise in such use accompanying the increased availability of cheaper and purer drug on the streets. Between 1990 and 1995, the number of emergency room episodes related to heroin doubled. Traditionally, the purity of heroin in a "bag" has averaged less than 10%; in 1996, the national average was 35%, with purity over 50% not uncommon in New York and other areas (Office of National Drug Control Policy 1997).

2.2 Surveys to Assess Psychiatric and Substance Use Disorders in the General Population

The Epidemiologic Catchment Area Program (ECA) was carried out in the early 1980s to determine the prevalence rates of various DSM-III disorders, including substance abuse and dependence, within five communities in the United States. It used structured interviews to ascertain whether or not individuals had lifetime or current psychiatric conditions. Using the ECA data, Robins et al. (1984) found that 11%–16% of individuals gave a lifetime history of alcohol abuse or dependence. Approximately 5% met criteria for lifetime drug abuse or dependence. Not surprisingly, alcoholics were more likely to abuse drugs and have additional psychiatric comorbidity. Compared to men, alcoholic women were more likely to suffer from another psychiatric disorder, suggesting that problem drinking may be more deviant for women (Helzer and Pryzbeck 1988). Young adults who experienced a major depression or anxiety disorder had twice the likelihood of developing a substance use disorder (Christie et al. 1988), suggesting that certain psychiatric conditions impart an increased vulnerability for substance abuse.

The National Comorbidity Study (NCS) is a more recently completed large-scale survey to assess the prevalence of psychiatric disorders, especially the extent of psychiatric comorbidity, in the general population. Lifetime alcohol and drug dependence rates were fairly similar for both the ECA Study and the NCS. Kessler and colleagues (1996) found that between 41% and 66% of respondents with a lifetime substance use disorder also had a lifetime history of another psychiatric disorder, while 51% of individuals with a lifetime psychiatric disorder also had at least one lifetime substance use disorder. The one situation where substance use disorders were more

likely to precede the psychiatric disorder involved co-occurring affective disorders and alcohol use disorders in men.

Elevated rates of psychiatric disorders – both axis I (e.g. depression, attention-deficit hyperactivity disorder, ADHD) and axis II (e.g. antisocial personality) – have also been documented among substance abusers seeking treatment, and up to 50% of those entering psychiatric treatment have a current or lifetime history of substance use problems. Regardless of the specific psychiatric disorder, individuals seeking help for a psychiatric disorder are more likely than the general population to have a comorbid substance use disorder. Those who abuse substances are more likely to be rehospitalized, be noncompliant with treatment, have a greater utilization of health resources, and have less satisfying familial relationships (Drake and Wallach 1989).

3 Theories of Causality

There have been several models to "explain" chemical addiction, including the moral, learning, disease, self-medication, and social model (Brower et al. 1989). Each employs a different etiology to explain addiction behavior and has different treatment strategies and goals.

3.1 Biologic Model

During the past 20 years, there have been many adoption studies to support the suggestion that alcohol dependence may be biologically transmitted. Goodwin et al. (1973) found that adopted men of biologic alcoholic fathers who were raised by nonalcoholic fathers were four times more likely to develop alcoholism than adopted sons of nonalcoholic biologic fathers. Schuckit and Smith (1996) found that individuals who experienced a low level of reaction to alcohol (i.e. less body sway, less hormonal effects) were more likely to develop an alcohol-related problem than those with a high level of reaction, suggesting that individuals who experience less bodily cues related to alcohol ingestion may be at greater risk for alcoholism.

Another intriguing area of research is the potential link between novelty-seeking and drug use behavior. Both novelty-seeking and taking abusable substances activate the mesolimbic dopaminergic system, the brain's reward pathway. This suggests that there may be a similar biologic predisposition for novelty-seeking and drug abuse (Bardo et al. 1993).

Since the majority of individuals with positive family histories do not develop alcoholism, other factors are likely to impede the development of alcohol dependence.

There are less available data regarding familial transmission of drug use disorders because of the changing availability and purity of various illicit substances. One study by Rounsaville et al. (1991) compared the prevalence of psychiatric disorders, including substance use disorders, in first-degree relatives of a large sample of opiate addicts compared to a sample of normal controls. Compared to relatives of normal subjects, opiate addicts' relatives had substantially higher rates of alcoholism, drug abuse, depression, and antisocial personality disorder. Because alcohol remains readily available, and therefore exposure to it remains relatively stable, contributory biologic factors are easier to assess.

3.2 Psychologic Model

There is ample evidence to suggest that certain individuals may be predisposed to drug abuse due to underlying psychopathology or psychological vulnerabilities. Individuals may self-medicate with specific addictive substances in order to deal with painful or disturbing affects. Drug abusers often have great difficulties in dealing with both positive and negative feelings about themselves and in relationship to others.

Several large epidemiologic surveys have found that individuals with both psychiatric and substance use disorders are more likely to develop a psychiatric illness first, suggesting that psychopathology may lead to substance abuse (Kessler et al. 1996; Christie et al. 1988). However, such an association does not necessarily prove causality; another coexisting factor may need to be present. For example, epidemiologic studies have found high rates of childhood and adult ADHD in substance-abusing populations. However, being impulsive, a common characteristic found among individuals with ADHD, may not be enough; instead, having an antisocial peer group or drug availability may be important intervening variables.

Risk factors for future alcoholism include childhood antisocial behavior, achievement-related difficulties, greater activity level, and interpersonal deficits. There are less data regarding whether early identity problems predict the occurrence of substance abuse in patients (usually female) who are eventually diagnosed with borderline personality disorder and may use chemicals to deal with painful feelings associated with real or perceived rejection.

3.3
Social Model

Steinberg (1991) outlines the individual, interpersonal, cultural, and societal factors which may lead to increased drug use in adolescents. Individual risk factors include school failure and academic difficulties, precocious sexual activity, truancy, or non-drug-related criminal or delinquent behavior. Interpersonal risk factors include distant or hostile relations with parents or guardians, familial disruption, and membership in a peer group that encourages or tolerates substance use. Institutional risks include school transitions into less personal, less protected environments, lack of access to meaningful roles in the community, and growing up in poverty. Adolescents who work more than 20 h per week may be at higher risk for drug use because they are paying less attention to schoolwork and have access to money to pay for drugs. Clearly, some of these factors might be associated with or due to underlying psychopathology. Therefore, both psychological and social factors might have an additive or synergistic effect on drug use.

Whereas most drug use begins in adolescence, for some individuals problematic use may not occur until mid-adulthood or later. Job dissatisfaction or loss, disruption in relationships, medical problems, or other significant losses may increase the risk for substance abuse, particularly among those who do not have adequate supportive relationships or who have never learned how to effectively cope with stress.

Cultural influences may also make drug use more or less likely. For example, Irish Americans may be more likely to develop drinking problems than Italian, Jewish, or Chinese Americans (Valliant and Milofsky 1982). It may be that for groups which allow limited drinking, but do not tolerate "loss of control" drinking, alcohol dependence may be less likely (Peele 1988). Similarly, women may have lower rates of alcohol and drug dependence not simply because of lower genetic/biologic vulnerability, but because women face disapproval for excessive drinking/drug use, particularly during pregnancy.

Societal factors can also have a substantial impact on the extent of use. With increased availability, drug use tends to increase. Thus it is not surprising that nicotine and alcohol are the most highly abused substances in the United States. Even if the government attempts to prohibit the use of certain substances by making them illegal, e.g. marijuana, or not allowing the sale of certain drugs to minors, e.g. alcohol and tobacco, if the society has permissive attitudes regarding the use of these substances, the magnitude of use will increase.

3.4
Biopsychosocial Model

A more inclusive model favored today is the biopsychosocial one, which encourages psychiatrists to examine the biological, psychological, and social factors that initiate, exacerbate, or ameliorate the symptoms associated with substance abuse. For any particular individual, certain factors might be more influential than others in promoting chemical dependency.

4
Course of Illness and Recovery

Perhaps the most well-known work regarding the course of illness has been Dr. George Valliant's work on alcoholism (1983–1996). In the late 1930s, researchers at Harvard began to study adult development. Although not all of the initial subjects were evaluated regarding their alcohol use, the extraordinary length of follow-up allowed important information regarding the natural progression of alcoholism to be evaluated. For most individuals, the progression from asymptomatic social drinking to alcohol abuse/dependence occurred over a period ranging from 3 to 15 years. Progressive alcoholics were more likely to be heavy cigarette smokers compared to nonprogressive alcoholics, with the death rate among alcoholics between the ages of 40 and 70 being approximately three times that of age-matched controls. Social stability was a good predictor of short-term outcome. However, long-term outcome was not necessarily correlated with factors cited as positive predictors in short-term studies (e.g. length of education, exposure to Alcoholics Anonymous (AA), physiologic dependence on alcohol; Valliant 1996.) Because Valliant's work focused on white males and did not assess individuals for other drug use, the results can only be generalized to a limited extent.

Cloninger et al. (1981) have proposed that there are two types of alcoholism, which may have different clinical courses. Using a Swedish adoption sample, Cloninger delineated type I and type II alcoholics. Type I occurs in both sexes, begins in the third decade, is likely to have significant environmental influences, and has limited destructiveness. Type II occurs more often in men, begins in adolescence, is less influenced by environmental factors, and is more malignant. More often, clinicians divide alcoholics into those with and without a family history of alcoholism, and those with a family history may have an earlier age of problematic use. Less information is available regarding the genetic

risks associated with alcoholism in women. Similar to men, women who have an alcoholic father may be at an increased risk for alcoholism (Lex et al. 1994). However, once a woman develops an alcohol problem, data suggests that women become "sicker quicker" (Hill 1984), perhaps related to biologic factors such as decreased amounts of alcohol dehydrogenase, a gastric enzyme that metabolizes alcohol before it enters the bloodstream, lower weight, and higher concentration of fat. Consequently, the same amount of alcohol will produce higher blood levels in women and ultimately greater organ damage.

With the epidemic rise in cocaine use in the 1980s, clinicians became interested in the natural history of its use. Similar to the patterns described for marijuana use, Siegel (1984) described cocaine users as having one of five patterns: experimental, social/recreational, circumstantial/situational, intensified, and compulsive. Using these categories, Siegel found that all users had episodes of decreased or no use. A limited supply and legal risks, along with other factors, may allow some individuals to use cocaine without progressing to a more debilitating pattern of use. However, when individuals switch to more potent routes of administration, i.e. from using cocaine intranasally to smoking crack cocaine, they usually become addicted and get into trouble faster. Not surprisingly, as cocaine use becomes more severe, antisocial behaviors increase and work and relationships with others decrease.

Recovery can occur spontaneously or through treatment. Vaillant (1983) estimated that approximately 1%–2% of alcoholics will become abstinent with each successive year. More recent work by Vaillant's suggests that those with alcohol abuse, compared to dependence, may be less likely to become abstinent. One explanation is that individuals who are alcohol dependent may be more likely to die earlier from their disease than alcohol abusers and thus have less opportunity to become abstinent. Another possibility is that abusers, compared to alcohol-dependent individuals, may suffer less negative consequences and therefore have less impetus to stop using alcohol. Recovery rates for clinical samples may be different from individuals who cease using drugs without treatment.

Various clinicians have proposed the use of developmental stages to describe the recovery process. DiClemente and Proschaska (1982) describe several stages of recovery, including commitment to change, initial change of addictive behavior, and maintenance of change. Gorski and Miller (1986) emphasizes that each stage has a major task or goal, has an approximate duration, and has predictable relapse causes. Minkoff (1989) emphasizes that recovery from psychiatric illness and from addiction have a parallel process and include progressive stages of acute stabilization, engagement, prolonged stabilization, and rehabilitation. Working with severely mentally ill patients with co-occurring substance abuse, it becomes clear that recovery is a long-term process, measured in years rather than weeks.

In the past, it was not uncommon to view any return to alcohol or drug use as a treatment failure. Marlatt and Gordon (1985) stress the importance of distinguishing between a "lapse" and a relapse: "A lapse provides useful information about both the cause of the event and how to correct for its occurrence in the future." Washton and Stone-Washton (1990) state that "the major difference between slips and relapses does not depend on the quantity of drugs used or how often the drug use occurs, but on the outcome. If the patient fails to immediately reestablish abstinence, continues active drug use, and rejects offers of help (i.e. drops out of treatment), it is a relapse." Several researchers have explored whether there are specific variables that can predict short-term relapse. Generally, length of prior abstinence periods and involvement by supportive significant others in the patient's treatment are correlated with improved treatment retention. Not surprisingly, patients who achieve abstinence during treatment are more likely to remain abstinent at follow-up than patients who use drugs during treatment (Kosten et al. 1992).

While income and marital status may not predict outcome conclusively, they are often reflections of a level of social support that is correlated with improved likelihood of a good outcome. It does not seem that the amount or frequency of alcohol or drug use predicts treatment outcome. Depressed or anxious mood are common reasons given for relapse, while situational factors are given less often (Pickens et al. 1985). Severity of psychiatric problems in alcohol- and drug-addicted patients is also associated with poor treatment outcome, supporting the need for adequate assessment and treatment of psychiatric illness in addiction settings in order to improve treatment outcome (McLellan et al. 1983).

5
Diagnostic Assessment

5.1
Defining Substance Use Disorders

There are various ways that clinicians and lay people describe individuals who use substances. Unfortunately, this often leads to miscommunications and the mistaken belief that "addiction" is a poorly defined concept, somehow different from other well-defined medical disorders. In fact, there are widely accepted

definitions which provide reliable criteria for making an accurate diagnosis. Appendix A shows the commonly accepted definitions of abuse and dependence (American Psychiatric Association 1994).

Making the diagnosis of abuse requires having *repeated* negative consequences as a result of regular use, while the diagnosis of substance dependence focuses on the personal impact of the use, the compulsivity of use, and the associated physiological changes. The diagnosis of abuse or dependence is not drug specific; although specific withdrawal symptoms or drug effects may vary, the same diagnostic criteria can be applied regardless of the drug. Second, a distinction is made between physiological and non-physiological dependence, emphasizing that there may (or may not) be physical aspects to being dependent on a psychoactive substance. Third, the amount or frequency of use is not critical when making the diagnosis. Given the highly variable patterns of substance use, i.e. daily versus weekend binge use, using frequency or amount of use as part of the criteria would be difficult. However, this contrasts with how clinicians often "decide" whether someone has a substance use problem.

The ICD-10 classification uses a similar approach to diagnosing substance use disorders as DSM-IV Dependence, but differs with regard to DSM-IV Abuse (American Psychiatric Association 1994; World Health Organization 1992). The term used in ICD-10 is "harmful use", which, as in DSM-IV Abuse, requires that the substance use has caused damage to the physical or mental health of the patient. However, an individual cannot be diagnosed as having "harmful use" if the pattern of substance use has only resulted in negative social consequences, i.e. marital arguments or arrest. This contrasts with the DSM-IV diagnosis of abuse, which allows repeated negative social consequences to serve as an adequate criterion of abuse. The diagnosis of dependence is more similar between DSM-IV and ICD-10. Both diagnostic systems focus on the compulsion to use, difficulties in controlling use, physical withdrawal and tolerance associated with chronic use, neglect of other interests, and harmful consequences of use. Neither diagnostic system requires the presence of physiologic withdrawal or tolerance to make the diagnosis of dependence. Unlike "harmful use" versus "abuse", DSM-IV and ICD-10 appear to be "speaking the same language" when diagnosing an individual with substance dependence.

5.2 Initial Interview

In order to obtain reliable and accurate information, rapport needs to be established. What often gets in the clinician's way is his or her attitude toward substance abuse. Clinicians may believe that substance abuse "is not a medical problem". Additionally, they may be moralistic, or even punitive, toward substance abusers. Part of this may be related to the fact that clinicians are frequently exposed to individuals with end-stage addictive disease during their professional training. Because these individuals may require repeated medical intervention as a result of their chronic drug use, clinicians may hold the mistaken belief that all individuals with substance use problems refuse treatment or invariably relapse despite any treatment provided.

These attitudes and misconceptions need to be addressed in order to effectively diagnose and treat this population. Substance abusers manifest various levels of denial and may be unaware of the negative impact that drug use has on their lives. Physicians who recognize this will be more likely to view substance-abusing patients not as "withholders" of information, but as individuals who need to receive information and understand how drug use has affected their lives and that of others.

Unfortunately, in general psychiatric evaluations, substance use histories are often obtained in a cursory manner. Simply asking patients whether or not they use alcohol or drugs may lead patients to answer no or grossly underreport their actual use. Instead, asking patients (in a nonjudgmental fashion) about their pattern of drug use over the course of their life, first starting in the past and eventually moving towards the present, may provide useful data both for assessment and for making an appropriate treatment referral. Specific questions might include the following:

- When did you initiate use of tobacco/alcohol/drugs?
- When did you begin using regularly? What factors led you to use?
- When did you use most heavily? How did/does your use impact on your life?
- How often do you drink/use drugs in a typical week? How much do you use on heavier use days?

Generally, it is better to weave questions regarding a patient's substance use pattern into the clinical interview with specific focus on the three areas of dependence mentioned previously, namely the salience and compulsivity of the drug use and whether the individual is physiologically dependent on the abused substance. In addition, it is particularly important that a comprehensive medical and social history is taken in order to assess the impact of the substance abuse on the patient's physical, occupational and social functioning. Although repeated injuries, pancreatitis, or gastritis might signal the clinician to consider substance abuse, even vague symptoms such as insomnia or weakness may be due to an alcohol or drug

problem. A significant loss, a change in peer group, or deterioration in functioning at school or work may have triggered or been triggered by substance use. Physical, emotional, or sexual traumas often precede substance abuse and need to be explored in a sensitive fashion. Recent estimates suggest that approximately half of all women substance abusers and 40% of male adolescents in substance abuse treatment have been abused. Legal difficulties such as disorderly conduct, driving while intoxicated, and selling/carrying drugs invariably signal problematic use.

5.3 Screening Tests

Several reliable screening tests are available to enhance a clinician's ability to detect possible substance abuse. Perhaps the most widely used screening tool is the CAGE (Ewing 1984). The CAGE is an acronym used to remember four questions:

C Have you ever felt you should *Cut* down on your drinking/drug use?
A Have other people *Annoyed* you by criticizing your drinking/drug use?
G Have you ever felt bad or *Guilty* about your drinking/drug use?
E Have you ever had a drink/used drugs first thing in the morning to steady your nerves (or other drug-specific withdrawal symptoms) or to get rid of a hangover (*Eye-opener*)?

For individuals endorsing two or more items, problematic alcohol or drug use is likely. The CAGE has been found to have both good sensitivity and specificity (91% and 77%, respectively). Other screening instruments available include the Michigan Alcoholism Screening Test (MAST), which consists of 25 questions about drinking behavior, consequences of use, and prior treatment (Selzer 1971); the brief MAST, which consists of ten items (Pokorny et al. 1972); and the Trauma Scale, which consists of five items regarding prior injuries. None of the screens are as simple to use and have as good a sensitivity and specificity as the CAGE.

An instrument commonly used in Europe is the Munich Alcoholism Test (Feuerlein et al. 1980). This consists of seven items to be assessed by a physician and 24 items which are assessed by the patient as being "true" or "not true". The physician's rating is based on information gleaned from physical examination, laboratory findings, the patient's history, and information collected from family and friends. The physician's assessment is weighted more heavily (i.e. each positive response by the physician receives 4 points) than the self-report items (i.e. each item is given 1 point). A total score greater than 10 is indicative of alcoholism.

5.4 Laboratory Tests

There are basically two types of laboratory tests, those to indicate that alcohol or drugs have been recently used or those which suggest that alcohol/drug use has produced some physiologic damage. Urine toxicological screens are the most common approach used for detection. There are a wide array of tests available to detect substance use, ranging from tests with lower specificity, e.g. thin-layer chromatography (TLC) or enzyme immunoassay (RIA), which are inexpensive and provide rapid results, to those which are highly specific, yet more expensive and more labor-intensive. These tests are usually employed to provide confirmation of positive results from other less expensive tests.

Detection of a particular drug depends on the frequency and amount of use, the patient's metabolism and ability to excrete the substance, the half-life of the drug, and the manner of urine collection (i.e. ensuring that the specimen has been taken from the patient and has not been tampered with). Generally, urine toxicology screens will detect cocaine via the metabolite benzoylecognine up to 12–24 h after use. Stimulants might be detected slightly longer, up to 48 h after use, and opiates up to 96 h after the last use. Cannabis often lasts 1–3 days for occasional users, but may persist up to 1 month for chronic heavy users after the last use. Depressants such as barbiturates may be detected 1 week or longer after the last use, but longer-acting benzodiazepines may produce metabolites which can be detected weeks to months in chronic heavy users. Whereas LSD is usually detected only up to 2 days, PCP may be detected for several days to weeks after the last use because it is fat soluble. One important factor influencing detection is the cutoff value used. Laboratories may vary such that samples with benzoylecognine concentrations less that 300 mg/ml may be reported negative by one laboratory but positive by another laboratory using lower cutoff values.

Urine testing has generally been the major method used to detect drug use. However, several disadvantages exist. Because urine collection is frequently done without direct observation, it can be easily adulterated or substituted, i.e. the patient may give the clinician or researcher someone else's urine. In addition, as described above, there is a short period of detection with most abusable substances. Therefore, other options which are currently being investigated include the use of hair, saliva, and sweat. Hair has the longest

length of possible detection, ranging from 20–90 days. However, laboratory studies have shown the potential for positive results based on passive exposure. Further, there are differences in the amount of cocaine and other drugs absorbed depending on hair type, with Black African permed hair having the greatest absorption rate.

An important caveat to the use of toxicological screens is that these tests are designed to assess patients for use of psychoactive substances and do not, in and of themselves, provide diagnoses of substance dependence or abuse. Unlike methods used for detection of substance use, there are other less direct laboratory tests which may indicate problematic use. These include tests which may show disturbances in hematologic or liver functioning (see Table 1). For example, heavy alcohol use with concomitant nutritional deficiencies may result in elevated mean corpuscular volume (MCV) values or cytopenias. Both inhalants and alcohol may produce elevations in SGOT and/or SGPT. Gamma glutamyltransferase (GGT) is the most commonly used laboratory marker of heavy drinking. Serum GGT is elevated in 75%–80% of heavy drinkers. The GGT level will generally return to normal after a 4- to 5-week period of abstinence (Miller and Gold 1991). Cocaine may produce elevations in creatinine phosphokinase as a result of muscle damage. Using several sources of information (the clinical interview, information provided by friends and family, the physical examination, screening instruments, and laboratory results), a fairly accurate picture of the patient's substance use problems is possible.

5.5 Differential Diagnosis

It is often extremely difficult to differentiate an underlying psychiatric condition from a substance-induced disorder. The presentation of psychiatric symptoms will depend on the substance used and whether the individual is intoxicated, overdosed,

Table 1. Physical and laboratory findings related to substance abuse

Substance	Possible physical findings	Relevant laboratory studies
Cocaine	Rhinitis, rash around the nasal area, perforation of the nasal septum, hypertension, tachycardia. *Crack:* hoarseness; parched lips, tongue and throat; singed eyebrows or eyelashes: IV stigmata	Immunoassay; chromatography; mass spectroscopy, urine toxicology screen (up to 12–48 h after use
Stimulants	Worn-down teeth (from tooth grinding), scratches, skin ulcers, dyskinesia	Urine toxicology screen (up to 24–48 h after use)
Inhalants (hydrocarbons)	Halitosis, rash around nose or mouth; mental status changes	CBC, liver function tests, kidney function tests
Cannabis	Conjunctival suffusion; distinct odor of burnt leaves on breath and clothes; dilated, poorly reactive pupils	Urine toxicology screen (occasional user, 1–3 days; chronic heavy users may have a positive screen for 1 month or more after cessation
Opioids	Track marks, skin lesions, constricted pupils, swollen nasal mucosa, thrombosis, lymphadenopathy	Urine toxicology screen (up to 2–4 days after use)
Depressants	Slurred speech without the odor of alcohol; "track" marks if IV user (especially barbiturates); pupillary constriction with glutethimide (Doriden)	Urine toxicology screen (detection time after last use: up to 1 week; barbiturates, chronic users. up to several weeks; long-acting benzodiazepines, chronic users: weeks to months
Hallucinogens	Myopathy; renal failures with PCP	Urine toxicology screen (detection time after last use: PCP, several days to several weeks; LSD, up to 12 h; LSD metabolites, 2 days); myoglobinuria elevated CPK creatinine/BUN with PCP

If a screening assay is positive for opioids, a confirmatory test specific for morpine or codeine is necessary. Illcit versus clinical use of these substances cannot be determined by test results alone. Further, it is impossible to distinguish whether heroin, codeine or morphine has been taken when low concentrations of morphine or codeine are found in the urine. Ingestion of a large quantity of poppy seeds will produce a positive immunoassay result up to 60 h after ingestion. To distinguish between poppy seed ingestion or heroin use, use the GC/MS to test for 6-*O*-acetylmorphine (a heroin metabolite). Methadone must be analyzed separately; fentanyl and its analogues are not detected by routine methods. Xanax (alprazolam) may not be detected in routine assays. If suspected, the laboratory test should be instructed to use GC/MS for specific testing for alprazolam.
From Project ADEPT (1989).
CBC, complete blood count; PCP, phencyclidine; CPK, creatinine phosphokinase; BUN, blood urea nitrogen; GC/MS, gas chromatography/mass spectroscopy.

withdrawing, or chronically using. Clinicians often have the greatest difficulties disentangling a "true" psychiatric disorder from a substance-induced disorder when a patient is withdrawing or chronically using a psychoactive substance. Complicating this picture is the increasing phenomenon of multiple substance use, in which combinations of drugs which might exacerbate psychiatric symptoms, i.e. chronic cocaine and alcohol use, may produce a severe depression. Alternatively, combined use may produce diagnostic dilemmas. For example, a young adult who combines marijuana and phencyclidine develops a prolonged psychosis lasting longer than 1 month such that the clinician is uncertain whether the psychosis is drug induced or a "first-break" psychotic disorder.

There are some guidelines suggested by DSM-IV to help guide the clinician in reaching a diagnosis (Appendix B). As these recommendations imply, the clinician needs to have a good understanding both of when the psychiatric symptoms began and when the substance use developed. Generally, the disorder which appears first is considered the primary one and the disorder which follows after is considered the secondary one. Clinicians also rely on periods of abstinence to distinguish whether psychiatric symptoms are persistent, and perhaps independent of drug use, or whether they diminish when the drug use ceases. A psychiatric condition which occurs during abstinent periods may also be referred to as the "primary" disorder. In order to facilitate whether periods of regular use, heavy use, and abstinence coincided with an increase/decrease or presence/absence of psychiatric symptoms, a time line can be helpful.

Unfortunately, clinical situations are rarely as clear as the DSM-IV guidelines would suggest. First, some psychiatric disorders may be episodic, whereas others are more likely to be continuous. With episodic illnesses, such as depression, obtaining reliable diagnoses can be particularly difficult. Evaluating an individual after a 1-month period of abstinence can usually help distinguish a *current* substance-induced depression from an underlying depressive disorder. However, it is often less clear when assessing individuals for prior depressive episodes.

Because depression may be episodic, the absence of depression during a prolonged period of abstinence, e.g. 1 month, does not necessarily mean that an individual's depression is secondary to the alcohol use. Thus "stable use periods" along with periods of abstinence may be useful when assessing the relationship between the patient's psychiatric symptoms and substance use. For example, if an individual reports a period of depression during a period of "stable psychoactive use", e.g. three beers a day, the clinician may "decide" that the patient had a major depressive episode, particularly if the patient had been drinking in this manner for a period of time prior to having a substantial change in his or her mood state. In assessing cocaine abusers for past and present psychiatric symptomatology, Rounsaville et al. (1991) counted psychiatric symptoms as not being substance induced if they took place during stable substance use periods. However, the research group made several noteworthy exceptions such as not making a diagnosis based on paranoiac, anxiety, or depression symptoms that occur during regular heavy use of stimulants (amphetamine and cocaine).

Other disorders which are less likely to totally remit, such as schizophrenia or social phobia, may be exacerbated (or even alleviated) by certain psychoactive substances and are less likely to remit after 1 month of abstinence. These and other long-standing disorders are therefore less likely to pose diagnostic dilemmas. Exceptions include situations in which the first presentation of symptoms, e.g. paranoia, is accompanied by the use of a drug which can produce psychosis, e.g. amphetamines. In this case, it is generally the persistence or absence of symptoms after the drug use has ceased which helps to differentiate an underlying psychiatric illness from a substance-induced psychosis.

5.6 General Considerations

Whenever a clinician is asked to evaluate the psychiatric status of a patient, drug use must be evaluated. Each class of drug can induce psychiatric symptomatology: depression, psychosis, anxiety, or delirium are all common. Whenever possible, several approaches should be used in tandem in order to reach the proper diagnosis. These include:

1. Always ask about alcohol and drug use. If possible, do not solely rely on the patient for the substance use history. Elicit information from family, significant others, and friends.
2. Obtain a urine/blood drug screen and carry out a breathalyzer test to probe for recent alcohol use. The urine test is more useful to determine whether any drugs have been used during the past 48 h, whereas blood is more useful on current status.
3. Examine the patient for physical signs of drug use (see Table 1); monitor vital signs and neurologic status.
4. Obtain appropriate laboratory studies (see Table 1); abnormal liver, renal, or hemopoietic tests may all indicate chronic use.
5. Observe the patient over time. Drug effects may diminish within a few hours, a few days, or a few weeks.

6 Treatment

6.1 General Remarks

Because drug dependence is due to a variety of factors, including biological, psychological, and social determinants, it makes sense that a combination of treatment approaches works best. Kleber (1989) emphasizes several important points:

1. Psychological problems or comorbid psychiatric illnesses may worsen treatment outcome. Some investigators have found that treatment outcome is worse among individuals who have high severity of psychological problems. Individuals with previous episodes of trauma or character pathology may require more intensive treatment with experienced clinicians.
2. The patient's motivation for treatment may affect treatment retention as well as other measures of treatment outcome.
3. Polydrug use is common. Treatment programs rarely find individuals who are abusing only one psychoactive substance.
4. Medications to treat substance dependence need to be given in the context of appropriate psychosocial intervention. Without such support, the patient is likely either to discontinue the medication or to take it while still using their drug of choice.

All of these points address the importance of "treatment-matching". McLellan and colleagues (1993) have repeatedly shown that, by providing adequate and appropriate services, patients can substantially improve in treatment. One particular treatment program or clinician may not have adequate resources to provide help for all of a patient's problem areas. However, by exploring a wide range of problem areas, either through structured instruments such as the Addiction Severity Index (ASI) (McLellan et al. 1980) or a psychiatric interview, the areas of greatest concern can be determined. These areas may be initially addressed within the treatment program, or concomitant treatment or referral to another program may be necessary. What is clear, however, is that a "one size fits all" kind of treatment program will fail with a substantial percentage of its patients.

6.2 Nonpharmacologic Treatment Strategies

6.2.1 Brief Interventions

Although treatment may be necessary for individuals with moderate or severe dependence, for some individuals who are heavily using, misusing, or are mildly dependent, a brief intervention, particularly from a family physician, may be all that is necessary. Repeatedly, large-scale studies have shown that a 5- to 10-min intervention in which information regarding the deleterious effects of alcohol and drugs are discussed and the patient's destructive drug-using behavior is confronted can dramatically reduce problematic use.

For those who continue to have problematic use, formal treatment may be required. Unfortunately, patients are often unmotivated to enter and comply with treatment. In order to understand why a certain patient is difficult to engage in treatment, the clinician needs to understand how a patient conceptualizes his or her addictive behavior and what motivated the patient to enter treatment. Although leverage from family, an employer, or the legal system may cause a patient to enter treatment, unless this leverage is consistently maintained, the patient is likely to discontinue treatment. Therefore, early on in treatment it is important to address any misconceptions that a patient might have about substance abuse and to work with the patient toward becoming more "internally" motivated for change.

One of the new techniques being used, referred to as "motivational enhancement therapy" (MET) is a treatment strategy designed to produce rapid internally motivated change. The MET approach is based on the stages of recovery model described by DiClemente and Prochaska (1982). Understanding these stages can help motivational enhancement therapists to empathize with their patients and move them toward the next stage in recovery. Miller and Rollnick (1991) describe five important aspects of MET: (1) express empathy, (2) develop discrepancy, (3) avoid argumentation, (4) roll with resistance, and (5) support self-efficacy. MET stresses active listening rather than telling the patient what to do. Persuasion is based on the assumption that change is up to the patient. Motivation to change may require the therapist to point out discrepancies of how the patient wants to function in his or her life and exploring how alcohol and/or drug use impacts on his or her current ability to function. A well-trained motivational enhancement therapist avoids confrontation with the patient; instead, the therapist helps the individual accurately see the consequences of his or her drug use. Resistance is not perceived as pathological, but rather is understood as a normal process based on the patient's ambivalence toward change. Further, optimism is imbued in the patient by emphasizing that change is possible and that the patient has the ability to accomplish this change.

More recently, clinicians working with substance-abusing patients with severe mental illness have developed a substance abuse treatment scale (McHugo and Drake 1995). This scale uses concrete, behaviorally

based descriptions for the stages of recovery rather than simply inferring psychological states. This allows clinicians to objectively determine an individual's current engagement in treatment. The level of engagement may reflect the individual's motivation for change.

6.2.2 Self-Help

Perhaps the most widely available form of help for alcohol and drug problems in the United States are the various self-help groups, e.g. Alcoholics Anonymous (AA), Narcotics Anonymous (NA), and Cocaine Anonymous (CA). The program consists of the 12 steps and 12 traditions. The steps provide a way for the alcoholic to learn to live effectively off alcohol. The 12 traditions describe the purpose of AA and the purview of the organization. AA meetings are autonomous and provide support for individuals striving to remain drug-free. Individuals are encouraged to form bonds with other members and to seek out a sponsor (another AA member with a sustained period of recovery) who can guide the individual through the program. Unlike the relationship established with a clinician, members develop a closeness based on their shared addiction. Outcome studies suggest that approximately 26%–40% of AA members report being abstinent for at least 1 year, and 20%–30% report being sober for at least 5 years (Nace 1992). It appears that those individuals who have greater participation in AA have a more favorable outcome. Increased involvement in AA is associated with better social adjustment, greater occupational satisfaction, and greater concern for others.

Although it is clear that self-help groups can be quite effective for some individuals, it is less obvious who these individuals are. Most clinicians will encourage patients to attend self-help groups and attend a variety of meetings in order to find a group in they which feel most comfortable. However, if a patient refuses these programs, it may be more pragmatic on the part of the clinician to explore other treatment options with the patient.

A recently developed alternative to traditional self-help programs, for example, is rational emotive therapy. Initially started by a clinical social worker, groups consist of eight to ten members led by lay coordinators and have a cognitive orientation. Because of legal issues related to the use of the term "Rational Recovery" (RR), a new organization, Self-Management and Recovery Training (SMART), was instituted. This treatment approach differs from AA in that (a) crosstalk is encouraged, and members provide feedback to one another with the past 2 weeks and next 2 weeks being stressed; (b) labeling of individuals as "addicts" or alcoholics is discouraged; (c) recovery is viewed as a process which has an end point; individuals may be "recovered" not exclusively in perpetual recovery; (d) it is recognized that individuals are ambivalent about being abstinent; therefore, supportive, nonconfrontational, motivating approaches are used; (e) a lapse or relapse does not mean that the patient has to "start over." Instead, it is an experience to learn from and lead to new, more effective coping strategies (Bishop et al. 1995).

RR may be a reasonable alternative to AA, particularly for those who have difficulties with the spiritual orientation of AA. The utility of this treatment approach deserves further study. This form of self-help is much less available than AA or other traditional self-help groups both in terms of cities where it is offered and number of meetings per week.

6.2.3 Cognitive and Behavioral Treatment

In the last few years, there has been much greater emphasis on cognitive and behavioral treatment approaches for addiction. Marlatt and Gordon (1985) stress the importance of the patient recognizing his or her high-risk situations or feelings that lead back to drug use as well as early warning signals and then utilizing various strategies to avoid relapse. Many ongoing clinical research studies have used manualized versions of this form of treatment to ensure the treatment is provided in a consistent manner and to allow it to be compared to other methods. In order for this method of treatment to be effective, the patient must be fairly cognitively intact and be willing to be an active participant in his or her treatment. For some patients who cannot view themselves as "powerless" over their addiction, both rational emotive and cognitive/behavioral approaches are often preferred.

The other behavioral approach which is being actively researched and applied in treatment settings is contingency management. This treatment seeks to increase prosocial, positive behaviors through the use of reinforcement. Higgins et al. (1994) have successfully used a system of vouchers for cocaine-negative urine which could be exchanged for a variety of goods and services among patients in a small metropolitan location. The reinforcement for "clean" urine was combined with therapy designed around the community reinforcement approach. This research was extended into an urban methadone program with good results (Silverman et al. 1996). Three key factors which may increase the likelihood of success are the following: (1) providing positive reinforcers rather than negative ones, (2) increasing the value of the reinforcer with each successive negative urine, and (3) administering reinforcement temporally close to the desired behavior. Although this technique was shown to be effective in reducing drug use, a large number of

patients did not respond to this method, particularly those with the heaviest drug use, and most relapsed after the positive contingencies were removed. It may be necessary, therefore, to maintain abstinence contingencies until other lifestyle changes are implemented, internal motivation for change is adequate, and other techniques, such as coping with relapse triggers, have been learned.

6.2.4 Network Therapy

Up until recently, it has been assumed that clinicians working in office-based practices have little success in treating addicted individuals. Galanter (1993) describes "network therapy", an approach consisting of cognitive/behavioral and psychodynamic techniques while incorporating individuals significant in the patient's life into the treatment process. Family members and peers meet with the patient and therapist at periodic intervals. Orchestrated by the therapist, the therapeutic network provides support and cohesiveness, confronts denial, and promotes the patient's compliance. Within the context of this approach, ambulatory detoxification, monitoring of compliance with medications such as disulfiram or naltrexone, or contingency contracting can be accomplished. Although research data demonstrating network therapy's efficacy is not yet available, clinicians have begun to employ this approach within their private practices.

6.3 Pharmacological Treatment

With the meteoric rise in cocaine use in the late 1970s through to the mid-1980s, there has been great interest in finding medications which might reduce drug craving and use. Not unexpectedly, clinicians turned to medications which were readily available and administered these medications to various groups of substance abusers, sometimes on a slim theoretical basis. Unfortunately, with the exception of alcoholism, this strategy has not yet proven effective. However, there is much reason for optimism. Novel medications for the treatment of opiate and cocaine addiction are currently being explored as well as other biologic approaches which may impact on the metabolism of certain abusable substances. Similar to the recent plethora of new medications that have become available to treat both depression and psychosis, it would not be surprising if new medications for the treatment of various drug dependencies soon become available.

There is an extensive literature on the pharmacologic treatment of opiate addiction. Generally, the first approach is to attempt to detoxify the opiate-dependent patient using various pharmacologic strategies. Some clinicians substitute longer-acting opiates, e.g. methadone, for shorter-acting ones, e.g. heroin, and slowly lower the dose over a period of 5–7 days. A second method is to use the alpha adrenergic agonist clonidine, which ameliorates opiate withdrawal symptoms by decreasing the activity of the locus coeruleus and thus the noradrenergic outflow that is responsible for much of the withdrawal signs and symptoms. Clonidine-assisted detoxification has the advantage of not using a narcotic, but does not shorten the duration of the heroin withdrawal.

Another approach which is gaining greater acceptance is to carry out rapid detoxification (2–3 days) using a combination of clonidine, naltrexone, and benzodiazepines. A withdrawal syndrome is precipitated by using low doses of naltrexone and simultaneously treating with clonidine to block some of the physiological symptoms, ondansetron to treat the nausea/vomiting, and benzodiazepines such as oxazepam to alleviate the anxiety and insomnia. The advantage of this approach is that detoxification can be done quickly, and by the end of the 48- to 72-h period, the patient is already maintained on naltrexone, an opiate antagonist, which can be continued in order to help the patient maintain his or her abstinence. The procedure can be done by experienced personnel on a day basis on day 1 and then on an outpatient basis. All three methods have been recently reviewed in detail (Kleber 1996).

Naltrexone mentioned above is a long-acting orally effective antagonist used to help patients refrain from using illicit opiates. It can be taken daily or as little as three times a week. Although naltrexone is quite effective, most patients are unwilling to take it. Often those who have the "most to lose", i.e. physicians whose licenses are contingent on taking the medication, are the individuals who comply with this treatment. Other good candidates include those individuals with strong family support and those under parole or probation supervision. A longer-acting injectable form of naltrexone is currently under development and may help improve patient compliance.

Most patients who have repeatedly failed detoxification prefer to be placed on an opiate agonist such as methadone rather than naltrexone. Many studies have shown that individuals maintained on methadone have substantial reductions in illicit drug use and other illegal activities, with improvement in both occupational and social functioning (Ball and Ross 1991). Methadone is well absorbed orally, does not produce euphoria by that route, and has a relatively long half-life, enabling it to be given once a day. This contrasts with heroin, which is not effective orally, is short-acting, and requires two to four doses a day. The practical implications of this are that patients can gain greater

stability in their lives via methadone maintenance and make use of other psychosocial interventions.

The major problems with methadone are difficulty in withdrawal and failure to decrease use of drugs such as cocaine and alcohol. Because of these disadvantages, new medications are becoming available in the United States, namely buprenorphine and levo-alpha-acetylmethadol (LAAM). Opiate-dependent patients report having less withdrawal symptoms when detoxified with buprenorphine compared to methadone. Buprenorphine is a partial mu agonist, producing less respiratory depression at high doses, and is less likely to be diverted than methadone, since opiate-dependent individuals taking additional buprenorphine may experience withdrawal rather than euphoria.

LAAM, a long-acting methadone derivative, was approved for maintenance by the Food and Drug Administration (FDA) in 1993. Because it may be given every 2–3 days, patients who are stable can come to the clinic less often. In addition, since take-home doses are unnecessary, diversion may be less.

Although methadone was originally proposed as a "life-long" treatment, there has been increased interest in tapering individuals off methadone after a 1- to 2-year period of maintenance treatment. Individuals who might be good candidates for this option are those who have attained occupational and interpersonal stability and have remained drug-free. Surprisingly, there are few data regarding which individuals will continue to do well after being tapered off methadone (or other opiate agonist agents) and which patients will be most likely to relapse. Given the risks associated with relapse, e.g. possible increased exposure to hepatitis and human immunodeficiency virus (HIV) infection and increase in illegal activities, the discontinuation of methadone must be approached cautiously.

The oral effectiveness and long duration of action along with lack of euphoria during chronic administration has made methadone the drug of choice for maintenance of heroin addicts. Prior to the advent of methadone maintenance in the mid-1960s, heroin was used for maintenance in countries such as England. While initially available from any physician, abuses by a few doctors increased heroin availability and led to heroin maintenance being restricted to specially licensed physicians and clinics. In the 1970s, the continued criminal activities of those maintained on heroin and the advantages of methadone maintenance led to a turning away from heroin maintenance. At present, there are an estimated 150,000 heroin addicts in England, of whom 17,000 are maintained on methadone and less than 400 on heroin, even though over 100 physicians are licensed to prescribe heroin.

In Switzerland, after the failure of the "Needle Park" experiment, where use and sale of heroin was permitted openly in the middle of Zurich and led to up to 20,000 individuals shooting up and hundreds of overdoses a day, the Swiss authorities opened up a heroin maintenance demonstration program. The program has been targeted to the most socially marginal addicts to try to bring them into the mainstream. Approximately 1000 patients are being maintained in this way, and data so far suggest mixed results with continued high rates of cocaine use and unemployment, but improvement in housing and physical health. Controlled experiments rather than demonstration programs would be needed to see whether heroin maintenance has any advantages over the known benefits of the longer-acting agonist methadone or the partial agonist buprenorphine, which is expected to become available in the sublingual form for maintenance in the United States in 1998.

Withdrawal from chronic cocaine or other stimulants may result in anhedonia, depression, and fatigue. Because of the associated dysphoria with prolonged cocaine use, researchers started using a variety of antidepresssants, most notably desipramine, to treat cocaine abuse. In early double-blind trials with cocaine-snorting addicts, desipramine reduced both cocaine craving and use. However, later double-blind trials with cocaine-abusing methadone patients or crack addicts have not shown clear-cut efficacy. Interestingly, desipramine-treated subjects did show greater reductions in cocaine use than the placebo group when individuals with antisocial personality disorder were excluded from the analysis. This suggests that individuals with certain additional psychopathology included in clinical trials may obfuscate any potential therapeutic benefits. Alternatively, certain medications may only prove beneficial for specific psychiatric subpopulations. For example, methylphenidate may be useful for cocaine abusers with adult ADHD (Khantzian et al. 1983, 1984; Levin et al. 1998a), but not among those without this disorder (Gawin et al. 1985).

Cocaine craving has been hypothesized to be the subjective response to dopamine depletion, leading to trials of a variety of medications to correct this supposed dopamine imbalance. These include bromocriptine, pergolide, amantadine, and neurotransmitter precursors. Other researchers have suggested alternative mechanisms for cocaine craving, i.e. the "kindling hypothesis", which has led to the use of carbamazepine. While early open-label trials suggested that carbamazepine may be effective, double-blind trials have not shown such success. Because serotonin receptors have been implicated as sites which contribute to the euphorogenic effects of cocaine, several studies have evaluated the impact of fluoxetine, a selective serotonin uptake inhibitor. Improvements have not been robust. The partial opiate agonist buprenorphine has shown mixed results in reducing cocaine use, and larger trials are necessary.

A third approach has been to develop agents to block the euphorigenic effects of cocaine in several ways: (a) block cocaine's binding to the dopamine transporter system, (b) to block dopamine binding, and (c) to increase cocaine's breakdown into inactive compounds. It is believed that the addictive properties of cocaine may be related to its ability to inhibit the dopamine transporter protein. Because cocaine binding is blocked by mazindol and bupropion, both medications have been tried, albeit with limited success. In the future, drugs may be designed which prevent cocaine binding but do not interfere with the transporter system.

Various D_2 antagonists have also been tried without any clear-cut efficacy. Animal data suggests that by blocking D_2 receptors there may actually be an increase in self-administration due to a partial masking of cocaine's effects. An open trial with flupenthixol reported a reduction in craving and use (Gawin et al. 1989), but controlled studies are needed. Other agents being considered include *N*-methyl-D-aspartate (NMDA) antagonists and gabaminergic agents.

One promising method to block cocaine's effects is to rapidly break down cocaine before it can produce euphoria. Unlike heroin, which is broken down to an active metabolite, i.e. morphine, cocaine can be deactivated by a simple cleavage reaction that produces two inactive products. Landry et al. (1993) have developed catalytic antibodies that bind to cocaine and degrade it and then repeat the sequence. The goal of ongoing research is to develop catalytic antibodies which would work fast enough to neutralize a large dose of cocaine prior to the cocaine reaching the brain. In the future, an individual may be given a passive immunization of a catalytic antibody which would remain in the body for several weeks. By blocking cocaine's effect, there might be a "window of opportunity" in which the patient would be able to become cocaine-free and perhaps be able to benefit from other behavioral or psychosocial interventions.

Unfortunately, even with the evaluation of over 30 medications in over 100 uncontrolled and controlled studies, no medication has shown clear-cut efficacy. This has led researchers to conclude that medications may need to be targeted for subpopulations of users and that medications or specific agents may need to be designed in the laboratory rather than taken from existing medications used for other clinical purposes.

In targeting subpopulations, preliminary data suggest that desipramine may be useful as an adjunctive treatment for schizophrenic cocaine abusers (Ziedonis et al. 1992) and depressed cocaine abusers (Carroll et al. 1994), methylphenidate may be useful for cocaine abusers with ADHD (Khantzian et al. 1983, 1984; Levin et al. 1998a), flupenthixol may be useful for schizophrenic cocaine abusers (Levin et al. 1998b), and mood stabilizers may be useful in reducing both manic symptoms and substance use (Brady et al. 1995). Bupropion, which was not found to be more effective than placebo in treating cocaine abuse, did show a moderate therapeutic benefit when a subsample of depressed cocaine abusers receiving it were compared with those receiving placebo (Margolin et al. 1995). Targeted treatment strategies may be important ways to improve the efficacy of certain medications.

7 Conclusion

In this chapter, the emphasis has been on the diagnosis and treatment of individuals with substance use disorders. New treatment strategies were presented, both pharmacologic and nonpharmacologic. With the development of new technologies, it is incumbent on clinicians treating substance abusers to gain greater facility with these new approaches. In contrast to the pessimism often voiced by clinicians who believe that substance abusers are "doomed to fail", the future of new, improved treatment strategies looks bright.

Appendix A

DSM-IV Criteria for Substance Abuse

A. A maladaptive pattern of substance use leading to clinically significant impairment or distress as manifested by one (or more) of the following, occurring within a 12-month period:
   1. Recurrent substance use resulting in a failure to fulfill major role obligations at work, school, or home (e.g. repeated absences or poor work performance related to substance use; substance-related absences, suspensions, or expulsions from school; neglect of children or household)
   2. Recurrent substance use in situations in which it is physically hazardous (e.g. driving an automobile or operating a machine when impaired by substance use)
   3. Recurrent substance-related legal problems (e.g. arrests for substance-related disorderly conduct)
   4. Continued substance use despite having persistent or recurrent social or interpersonal problems caused or exacerbated by the effects of the substance (e.g. arguments with spouse about consequences of intoxication, physical fights)

B. The symptoms have never met the criteria for substance dependence for this class of substances.

DSM-IV Criteria for Substance Dependence

A maladaptive pattern of substance use, leading to clinically significant impairment or distress, as manifested by three (or more) of the following, occurring at any time in the same 12-month period:

1. Tolerance, as defined by either of the following:
   a) A need for markedly increased amounts of the substance to achieve intoxication or desired effect
   b) Marked diminished effect with continued use of the same amount of substance
2. Withdrawal, as manifested by either of the following:
   a) The characteristic withdrawal syndrome for the substance
   b) The same (or closely related) substance is taken to relieve or avoid withdrawal symptoms
3. The substance is often taken in larger amounts or over a longer period than was intended.
4. There is a persistent desire or unsuccessfully efforts to cut down or control substance use.
5. A great deal of time is spend in activities necessary to obtain the substance (e.g. visit multiple doctors or driving long distances), use the substance (e.g. chain-smoking), or recover from its effects.
6. Important social, occupational, or recreational activities are given up or reduced because of the substance use.
7. The substance use is continued despite knowledge of having a persistent or recurrent physical or psychological problem that is likely to have been caused or exacerbated by the substance (e.g. current cocaine use despite recognition of cocaine-induced depression, or continued drinking despite recognition that an ulcer was made worse by alcohol consumption).

Specify if:

- With physiological dependence: evidence of tolerance or withdrawal
- Without physiological dependence: no evidence of tolerance or withdrawal

Appendix B

DSM-IV Guidelines for Diagnosis

Label the psychiatric syndrome according to the most prominent symptom pattern.

Diagnose substance-induced disorder only if:

- It developed during or within 1 month of intoxication or withdrawal.
- Symptoms are consistent with exposure to the substances patient has been using.
- Disorder not better explained by a non-substance-induced disorder.

Diagnose primary, non-substance-induced disorder if:

- It developed before the substance use.
- It has been present during abstinence lasting at least 1 month.
- Symptoms not consistent with those produced by the abused substance.
- Symptoms better explained by a non-substance-induced disorder.

8 References

American Psychiatric Association (1994) Diagnostic and statistical manual of mental disorders, 4th edn. American Psychiatric Association, Washington, DC

Ball J, Ross A (1991) The effectiveness of methadone maintenance treatment. Springer-Verlag, Berlin Heidelberg New York

Bardo M, Bowling S, Robinet P, Rowlett J, Lacy M, Mattingly B (1993) Role of dopamine D1 and D2 receptors in novelty-maintained place preference. Exp Clin Psychopharmacol 1: 101–109

Bishop F, Tate P, Horvath A, Robb H (1995) S.M.A.R.T. recovery/rational recovery update. The Addictions Newsletter, division 50, vol 2, no 2. American Psychiatric Association, Washington

Bowling S, Bardo M (1994) Locomotor and rewarding effects of amphetamine in enriched, social, and isolate reared rats. Pharmacol Biochem Behav 2(48): 459–464

Brady K, Sonne S, Anton R, Ballenger J (1995) Valproate in the treatment of acute bipolar affective episodes complicated by substance abuse: a pilot study. J Clin Psych 56: 118–121

*Brower KJ, blow FC, Beresford TP (1989) Treatment implications of chemical dependency models integrative approach. J Substance Abuse Treat 6: 147–157

**Carroll K, Rounsaville B, Gordon L, Nich C, Jatlow P, Bisighini R, Gawin F (1994) Psychotherapy and pharmacotherapy for ambulatory cocaine abusers. Arch Gen Psych 51: 177–187

Christie K, Burke J, Regier D, Rae D, Boyd J, Locke B (1988) Epidemiologic evidence for early onset of mental disorders and higher risk of drug abuse in young adults. Am J Psych 145(8): 971–975

Cloninger C, Bohman M, Sigvardsson S (1981) Inheritance of alcohol abuse: cross-fostering analysis of adopted men. Arch Gen Psych 38: 861–868

**DiClemente C, Prochaska J (1982) Self-change and therapy change of smoking behavior: a comparison of processes of change in cessation and maintenance. Addict Behav 7: 133–142

Drake R, Wallach M (1989) Substance abuse among the chronic mentally ill. Hosp Comm Psych 40: 1041–1046

Ewing J (1984) Detecting alcoholism: the CAGE questionnaire. JAMA 252(14): 1905–1907

Feuerlein W, Ringer C, Kufner H, Antons K (1980) Diagnosis of alcoholism: the Munich Alcoholism Test (MALT). In: Galanter M (ed) Currents in alcoholism: recent advances in research and treatment, vol VII. Grune and Stratton, New York, pp 137–147

Galanter M (1993) Network therapy for addiction: a model for office practice. Am J Psych 150: 28–36

Gawin F, Riordan C, Kleber H (1985) Methylphenidate treatment of cocaine abusers without attention deficit disorder: A negative report. Am J Drug Alc Abuse 11: 193–197

Gawin F, Allen D, Humblestone B (1989) Outpatient treatment of "crack" cocaine smoking with flupenthixol decanoate. Arch Gen Psych 46: 322–325

*Goodwin D, Schulsinger F, Hermansen L, Guze S, Winokur G (1973) Alcohol problems in adoptees raised apart from alcoholic biological parents. Arch Gen Psych 28: 238–243

Gorski T, Miller M (1986) Staying sober: a guide for relapse prevention. Independence Press, Independence

Helzer JE, Pryzbeck TR (1988) The co-occurrence of alcoholism with other psychiatric disorders in the general population and its impact on treatment. J Stud Alcohol 49: 219–224

*Higgins ST, Delaney DD, Budney et al (1994) A behavioral approach to achieving initial cocaine abstinence. Am J Psychiatry 148: 1218–1224

Hill SY (1984) Vulnerability to the biomedical consequences of alcoholism and alcohol-related problems among women. In: Wilsnack SC, Beckman LJ (eds) Alcohol problems in women: antecedents, consequences and intervention. Guilford, New York

**Kessler R, Nelson C, McGonagle K, Edlund M, Frank R, Leaf P (1996) The epidemiology of co-occurring addictive and mental disorders: implications for prevention and service utilization. Am J Orthopsychiatry 66: 17–31

Khantzian E (1983) An extreme case of cocaine dependence and marked improvement with methylphenidate treatment. Am J Psych 140: 484–485

Khantzian E (1985) The self-medication hypothesis of addictive disorders: focus on heroin and cocaine dependence. Am J Psychiatry 142: 1259–1264

Khantzian E, Gawin F, Riordan C, Kleber H (1984) Methylphenidate treatment for cocaine dependence: a preliminary report. J Subst Abuse Issues 1: 107–112

Khantzian EJ, Gawin FH, Riordan C, Kleber HD (1984) Methylphenidate treatment for cocaine dependence: a preliminary report. J Substance Abuse Issues 1: 107–112

Kleber HD (1989) Treatment of drug dependence: What works. Int Rev Psychiatry 1: 81–100

Kleber HD (1996) Outpatient detoxification from opiate. Primary Psychiatry 1: 42–52

Kosten T, Gawin F, Kosten T, Morgan C, Rounsaville B, Schottenfeld R, Kleber H (1992) Six-month follow-up of short-term pharmacotherapy for cocaine dependence. Am J Addict 1(1): 40–49

Landry D, Zhao K, Yang GX-Q, Glickman M, Georgiadis T (1993) Antibody-catalyzed degradation of cocaine. Science 259: 1899–1901

Levin F, Evans S, McDowell D, Kleber H (1998a) Efficacy of methylphenidate treatment for cocaine abusers with adult attention-deficit/hyperactivity disorder: a pilot study. J Clin Psychiatry

Levin FR, Evans S, Coomaraswammy S, Collins ED, Regent N, Kleber HD (1998b) Flupenthixol treatment of cocaine abusers with schizophrenia/schizoaffective illness: a pilot study. Am J Drug Alcohol Abuse

Lex B, Rhoades EM, Teoh SK, Mendelson JH, Greenwald NE (1994) Divided attention task performance and subjective effects following alcohol and placebo: differences between women with and without a family history of alcoholism. Drug Alcohol Depend 35: 95–105

Margolin A, Kosten T, Avants S, Wilkins J, Ling W, Beckson M, Arndt I, Cornish J, Ascher J, Li S, Bridge P (1995) A multicenter trial of bupropion for cocaine dependence in methadone-maintained patients. Drug Alcohol Depend 40: 125–131

**Marlatt G, Gordon J (1985) Relapse prevention: maintenance strategies in the treatment of addictive behaviors. Guilford, New York

McHugo GJ, Drake RE (1995) A scale for assessing the stage of substance abuse treatment in persons with severe mental illness. J Nerv Ment Dis 183: 762–767

*McLellan A, Luborsky L, Woody G, O'Brien C (1980) An improved diagnostic evaluation instrument for substance abuse patients: the addiction severity index. J Nerv Ment Dis 168: 26–33

McLellan A, Luborsky L, Woody G, O'Brien C, Druley K (1983) Predicting responses to alcohol and drug abuse treatments: role of psychiatric severity. Arch Gen Psych 40: 620–625

**McLellan A, Arndt I, Metzger D, Woody G, O'Brien C (1993) The effects of psychosocial services in substance abuse treatment. JAMA 269: 1953–1959

Miller NS, Gold MS (1991) Alcohol. Drugs of abuse: a comprehensive series for clinicians, vol 2. Plenum, New York

Miller WR, Rollnick S (1991) Motivation interviewing. Guilford, New York

Minkoff K (1989) An integrated treatment model for dual diagnosis of psychosis and addiction. Hosp Comm Psych 40(10): 1031–1036

Nace E (1992) Alcoholics anonymous. In: Lowinson J, Ruiz P, Millman R, Langrod J (eds) Substance abuse: a comprehensive textbook. Williams and Wilkins, Baltimore

Office of National Drug Control Policy, Executive Office of the President (1997) Heroin: facts and figures. Office of National Drug Control Policy, Washington, DC

O'Malley SS, Jaffe AJ, Chang G, Schottenfeld RS, Meyer RE, Rounsaville B (1992) Naltrexone and coping skills therapy for alcohol dependence: a controlled study. Arch Gen Psychiatry 49: 881–887

Peele S (1988) A moral vision of addiction: how people's values determine whether they become and remain addicts. In: Peele S (ed) Visions of addiction: major contemporary perspectives on addiction and alcoholism. Lexington Books, Heath, Lexington/MA, pp. 201–233

Pickens R, Hatsukami D, Spicer J, Svikis D (1985) Relapse by alcohol abusers. Alcohol Clin Exp Res 9(3): 244–247

Pokorney A, Miller B, Kaplan H (1972) The brief MAST: a shortened version of the Michigan Alcoholism Screening Test. Am J Psych 129: 342–348

Project ADEPT (1989) Assessment and diagnosis instructor's guide. The Project ADEPT Curriculum for Primary Care Physician Training. Brown University, Providence

Robins L, Helzer J, Weissman M, Orvaschel H, Gruenberg E, Burke J, Regier D (1984) Lifetime prevalence of specific psychiatric disorders in three sites. Arch Gen Psych 41: 949–958

Rounsaville BJ, Tierney T, Crits-Christoph K, Wiessman MN, Kleber HD (1982) Predictors of treatment outcome in opiate addicts: evidence of multidimentionality and addicts' problems. Compr Psychiatry 23: 462–478

Rounsaville BJ, Kosten TR, Weissman MM, Prusoff B, Paula D, Antol SF, Merikangas K (1991) Psychiatric disorders in relative of probands with opiate addiction. Arch Gen Psychiatry 48: 33–42

*Schuckit M, Smith T (1996) An 8-year follow-up of 450 sons of alcoholic and control subjects. Arch Gen Psychiatry 53: 202–210

Selzer M (1971) The Michigan Alcoholism Screening Test: the quest for a new diagnostic instrument. Am J Psychiatry 127: 1653–1658

Siegel R (1984) Changing patterns of cocaine use: longitudinal observations, consequences, and treatment. NIDA Res Monogr 50: 92–110

Silverman K, Chutuape MA, Bigelow GE, Stitzer ML (1996) Voucher-based reinforcement of attendance by unemployed methadone patients in a job skills training program. Drug Alcohol Dependence 41: 197–207

Steinberg L (1991) Adolescent transitions and alcohol and other drug use prevention. In: Goplerud E (ed) Preventing adolescent drug use: from theory to practice. OSAP Prevention Monograph. U.S. Department of Health and Human Services, Rockville, pp. 13–51

**Vaillant G (1983) The natural history of alcoholism: causes, patterns, and paths to recovery. Harvard University Press, Cambridge

Vaillant GE (1996) A long-term follow-up of male alcohol abuse. Arch Gen Psychiatry 53: 243–249

Vaillant G, Milofsky E (1982) The etiology of alcoholism: a prospective viewpoint. Am Psychol 37: 216–232

Volpicelli JR, Alterman AI, Hayashida M, O'Brien CP (1992) Naltrexone treatment in alcohol dependence. Arch Gen Psychiatry 49: 876–860

Washton A, Stone-Washton N (1990) Abstinence and relapse in outpatient cocaine addicts. J Psychoactive Drugs 22(2): 135–147

World Health Organization (1992) Internal statistical classification of diseases and related health problems, 10th revision, vol 1. World Health Organization, Geneva

Ziedonis D, Richardson T, Lee E, Petrakis I, Kosten D (1992) Adjunctive desipramine in the treatment of cocaine abusing schizophrenics. Psychopharmacol Bull 28: 309–314

Zucker R, Lisansky X, Gomberg E (1986) Etiology of alcoholism reconsidered: the case for a biopsychosocial process. Am Psychol 41(7): 783–793

CHAPTER

22

N.D. Volkow, J.S. Fowler, G.-J. Wang

Imaging Studies in Substance Abuse

This work was supported by DOE DE-ACO2-CH00016, NIDA ROIDA06891, NIDA ROIDA09490, and NIDA ROIDA06278.

1 Introduction

Imaging technologies developed over the past 15–20 years such as positron emission tomography (PET), single photon emission computed tomography (SPECT), and magnetic resonance imaging (MRI) have bridged the gap between the basic neurosciences and clinical research. Their application to the investigation of substance abuse has allowed pharmacological mechanisms that pertain to the reinforcing and toxic properties of drugs abuse and the brain metabolic and neurochemical changes associated with addiction to be investigated directly in the human brain.

Most of the imaging work in substance abuse has been done using PET and SPECT. These are nuclear medicine instruments that detect and measure the spatial distribution and movement of radioisotopes in tissues of living subjects (Mullani and Volkow 1992; Rogers and Ackerman 1992). The tracers measured with PET are compounds labeled with positron-emitting radioisotopes, and those with SPECT are compounds labeled with single photon-emitting radioisotopes. Because there are positron emitters for the natural elements of life (^{11}C, ^{15}O, ^{24}N, and ^{18}F, which can be used to substitute for hydrogen), one can label compounds without affecting their pharmacological properties. SPECT radiotracers are usually labeled with ^{123}I or ^{99m}Tc. Although SPECT cannot be used to measure the pharmacokinetics of drugs, because of the need to use these isotopes many iodine-substituted radiotracers have been developed which have high biological selectivity and affinity for specific molecular targets (Kung 1993). The positron emitters used for imaging have shorter half-lives than the single photon emitters. Both of these types of isotopes can be used to label ligands for specific receptor, transporter and/or enzymatic systems to be used with PET or SPECT to quantify these parameters in the living human brains. In addition, PET using ^{15}O-labeled water, ^{15}O-labeled butanol, or ^{13}N-labeled ammonia and SPECT using ^{99m}Tc-hexamethyl propyleneamine oxime (HMPAO) can be used to assess cerebral blood flow (CBF), and PET using ^{18}F- or ^{11}C-labeled deoxyglucose can be used to assess glucose metabolism in brain. These parameters provide an index for regional brain function since, under normal conditions, glucose metabolism and CBF are tightly coupled to brain activity.

Recent advances in MRI have expanded its application from that of an anatomical imaging device to one that enables functional brain imaging (fMRI) to be carried out with a much higher spatial and temporal resolution than that achievable with either PET or SPECT (Andrew 1994). Functional imaging with MRI takes advantage of the differences in magnetic properties of oxygenated versus deoxygenated blood. During activation, there is an increase in CBF with a consequent increase in oxygenation of the activated region (decrease in the ratio of deoxyhemoglobin to oxyhemoglobin), which then gives rise to the activation signal. Because it is such a new technique, no studies have yet been published using fMRI in substance abuse. However, this will change as fMRI becomes a more widely available research tool.

Multiple studies have been performed with PET and SPECT to investigate the neurobiology of drugs of abuse from two perspectives, one involving an investigation of the pharmacological properties of the drugs of abuse themselves and the other investigating the consequences of these drugs on brain function and neurochemistry. These strategies, as they pertain to the various types of drugs of abuse, are summarized in Table 1. In this chapter, we discuss some of the main findings resulting from these studies.

2 Pharmacological Properties of Drugs of Abuse in the Human Brain

The investigation of the pharmacological properties of drugs of abuse entail studies of both their pharmacokinetics and their pharmacodynamics. Furthermore, because the studies are done in awake human subjects, one can investigate the temporal relation between the effects of drugs and their uptake and clearance from brain.

2.1 Pharmacokinetics

Pharmacokinetic properties of drugs of abuse are ideally studied with PET, because the availability of the positron emitter ^{11}C makes it possible to directly label drugs of abuse without changing their pharmacological properties. In this way, PET can be used to measure their absolute uptake, their regional distribution, and their kinetics in the human brain. The labeled drug and whole-body PET can also be used to determine the target organs for the drug and its labeled metabolites and thus provide information on potential organ toxicity as well as tissue half-lives. Table 1 shows the various addictive drugs which have been labeled with positron emitters and whose distribution has been evaluated with PET methodology.

Table 1. Various addictive drugs that have been labeled with positron emitters and whose distribution has been evaluated by positron emission tomography (PET)

Drug class	Specific drug	Parameter measured	Labeled drug or tracer	Reference
Psychostimulants	Cocaine	Pharmacokinetics	[^{11}C]Cocaine	Fowler et al. 1989; Volkow et al. 1995c
		Metabolism	[^{18}F]FDG	Volkow et al. 1993b; London et al. 1990b
		Blood flow	$H_2{}^{15}O$, ^{99m}Tc-HMPAO, [^{123}I]amphetamine	Volkow et al. 1988a; Levin et al. 1994; Weber et al. 1993
		DA receptors	[^{11}C]Raclopride, ^{18}F-labeled NMS, [^{123}I]IBZM	Volkow et al. 1990a, 1993b; 1994a
		DA transporters	[^{11}C]Cocaine, [^{11}C]d-threo-MP, [^{123}I]β-CIT	Volkow et al. 1996a
		DA metabolism	^{18}F-labeled DOPA	Baxter et al. 1988
		NE transporter	^{18}F-labeled NE, [^{11}C]HED	Fowler et al. 1994; Melon et al. 1992
	Methylphenidate	Pharmacokinetics	[^{11}C]MP	Volkow et al. 1995b
		Blood flow	$H_2{}^{15}O$	Wang et al. 1994
		DA responsiveness	[^{11}C]Raclopride	Volkow et al. 1994a; Laruelle et al. 1995
	Amphetamine	Metabolism	[^{18}F]FDG	Wolkin et al. 1987
Sedative hypnotics	Alcohol	Metabolism	[^{18}F]FDG	Volkow et al. 1990b; DeWit et al. 1990
		Blood flow	$H_2{}^{15}O$, ^{99m}Tc-HMPAO	Volkow et al. 1988b; Nicolas et al. 1993
	Benzodiazepines	Pharmacokinetics, pharmacodynamics	[^{11}C]Flumazenil, [^{123}I]iomazenil	Pappata et al. 1988; Abi-Dargham et al. 1994
		Metabolism	[^{18}F]FDG	Volkow et al. 1993a, 1995a; Buchsbaum 1987; DeWit et al. 1991
		Blood flow	$H_2{}^{15}O$	
	Barbiturates	Metabolism	[^{18}F]FDG	Theodore et al. 1986
Opiates	Morphine, heroin, buprenorphine	Pharmacokinetics	[^{11}C]Morphine, [^{11}C]heroin, [^{11}C]codeine, [^{11}C]pethidine, [^{11}C]buprenorphine	Hartvig et al. 1984; Galynker et al. (1996)
		Metabolism	[^{18}F]FDG	London et al. 1990a
		Binding site distribution	[^{11}C]Carfentanil	Frost et al. 1985
Cannabinoids	THC	Metabolism	[^{18}F]FDG	Volkow et al. 1991b, 1996b
Nicotine	Nicotine	Pharmacokinetics	[^{11}C]Nicotine	Mazière et al. 1976; Bergstrom et al. 1995
		Metabolism	[^{18}F]FDG	Stapleton et al. 1993
Caffeine	Caffeine	Blood flow	$H_2{}^{15}O$	Cameron et al. 1990

FDG, fluorodeoxyflucose; HMPAO, hexamethyl proplyeneamine oxime; DOPA, dihydroxy-phenylalanine; NMS, *N*-methylspiroperidol; IBZM, iodobenzamide; MP, methylphenidate.

An example of the power of this strategy is illustrated by its application to the investigation of the in vivo pharmacokinetics of cocaine in brain. Using ^{11}C as a label, PET has been used to investigate the pharmacokinetics and distribution of cocaine in the human brain (Fowler et al. 1989) and in the human body (Volkow et al. 1992a). These studies showed binding of cocaine in the human brain mainly to dopamine (DA) transporters in basal ganglia, as had previously been shown by animal studies. They also documented a high brain uptake of cocaine (8%–10% injected dose) and a very fast uptake (peaking 4–6 min after injection) and clearance (half-life, 20 min) from brain. Furthermore, for the first time it was possible to evaluate the relation between the kinetics of an abused drug and the temporal relation of the behavioral effects. In the case of cocaine, a parallelism was found between the kinetics of uptake and clearance of cocaine in the brain and cocaine-induced "high."

PET has also been used to compare the pharmacokinetics and brain distribution of cocaine with that of methylphenidate, another DA transporter inhibitor drug which is prescribed in the treatment of attention deficit disorder (Wilens and Biederman 1992; Barkley 1977) and of narcolepsy (Meyer et al. 1980). Comparison of the regional distribution and the pharmacokinetics of these two drugs in the living human brain using [^{11}C]cocaine and [^{11}C]methylphenidate (Volkow et al. 1995b) showed that, while the percentage uptake (7%–10% of the injected dose) and regional distribution of these two drugs in brain was almost identical, their pharmacokinetics were not (Fig. 1). They were both taken up rapidly (peak uptake, 4–8 min) by the human brain, but they differed markedly in their rate of clearance. Methylphenidate clearance from striatum (half-life, >90 min from peak uptake) was significantly slower than that of cocaine (half-life, 20 min from peak uptake) (Fig. 2). For both drugs, the fast uptake in striatum paralleled the experience of the "high." However, whereas for cocaine the rate of clearance paralleled the decline in the "high," for methylphenidate the "high" declined as fast as that for cocaine, while there was still significant binding of the drug in brain. Because it was the rate of uptake that was associated with the "high" for both drugs and not the presence of the drug in brain, it was postulated from this observation that the rate of clearance may affect the propensity of a drug to promote frequent, repeated administration. In the case of methylphenidate, its relatively slow clearance from brain may interfere with its frequent, repeated administration as occurs with cocaine.

2.2 Pharmacodynamics

Different labeled tracers can be used to assess drug pharmacodynamics, which include the biological processes involved in the drugs effects. For example, with appropriate radiotracers, the effects of a drug on metabolism and CBF, on neurotransmitter activity, and on enzyme activity have been probed. These parameters can be assessed directly in the human body both

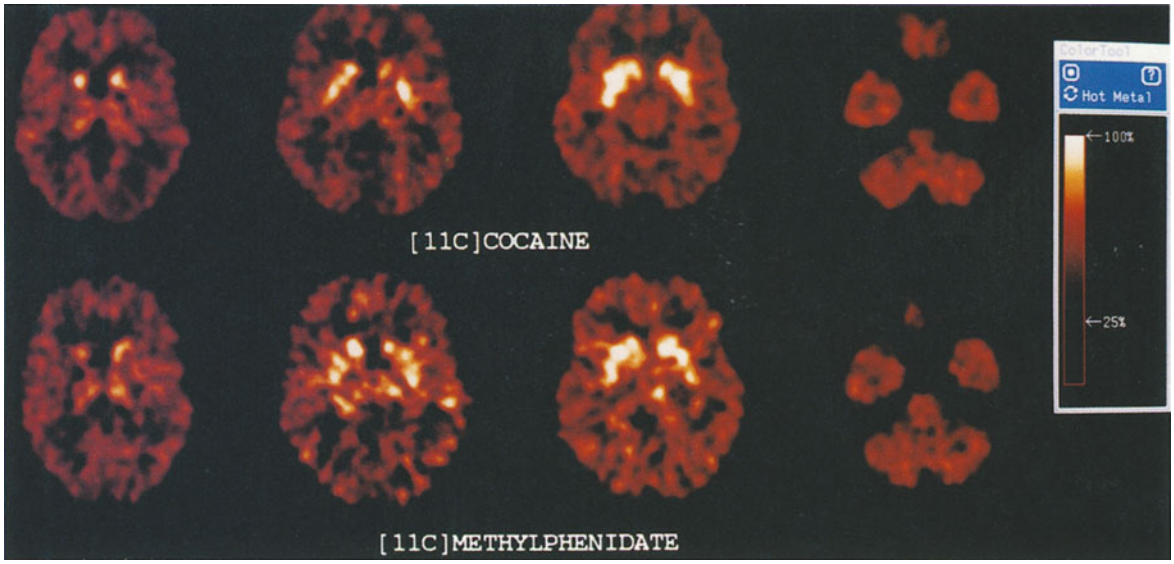


Fig. 1. Brain images obtained with [^{11}C]cocaine and with [^{11}C]methylphenidate. Notice the similar distribution of the two radiotracers in brain with the highest uptake in striatum

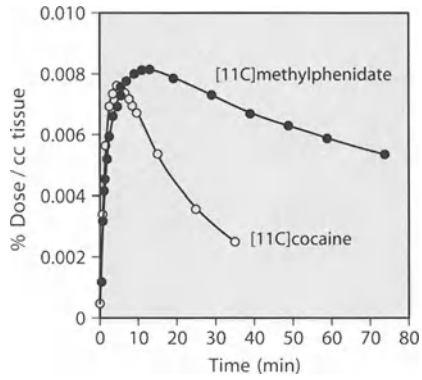


Fig. 2. Averaged time activity curves for [^{11}C]cocaine ($n = 20$) and for [^{11}C]methylphenidate ($n = 10$) in striatum. Notice that, while both drugs enter the brain almost at the same rate, cocaine clears much faster than methylphenidate does. Notice also that the peak uptake for the two drugs is almost identical

in normal controls and in addicted subjects to determine whether there are differences in the responses between addicted and nonaddicted subjects. The most widely utilized approach has been to assess the effects of acute drug administration on brain glucose metabolism and on CBF. This allows analysis of the brain regions that are most sensitive to the effects of the drug. Because the studies are done in awake human subjects, they allow an analysis to be performed of the relation between regional changes in metabolism or flow and the behavioral effects of the drug. This strategy has been used to investigate the effects on brain glucose metabolism and/or CBF for most of the drugs of abuse (Table 1). The effects of acute pharmacological drug administration on regional brain metabolism and CBF can also be used to determine the brain regions associated with drug effects. For example, heroin-induced changes in scores related to euphoria were associated with heroin-induced changes in metabolic activity in lateral occipital and primary visual cortex (London and Morgan 1993). While most drugs of abuse decrease regional brain glucose metabolism, their effects on CBF are increased by some drugs and decreased by others. This discrepancy between metabolism and CBF is probably an indication of the vasoactive properties of many of these pharmacological agents, a property that is relevant for understanding their toxicity as it relates to cerebrovascular pathology. In addition, the fact that changes in metabolism reflect an average of the changes which occur over the uptake period of fluorodeoxyglucose (FDG), which is usually 35–45 min, while the temporal resolution of blood flow measurements is higher, needs to be taken into account.

The effects of drugs of abuse on neurotransmitter activity have also been investigated with PET and SPECT. These studies have been targeted to assess the effects of psychostimulant drugs on DA activity (Volkow et al. 1994a; Laruelle et al. 1995). Changes in synaptic DA concentration in response to psychostimulant drugs have been measured using DA receptor ligands which are sensitive to the endogenous concentration of DA. For this purpose, subjects are scanned twice, at baseline and after administration of the psychostimulant drug known to increase DA concentration. The changes in binding of the ligand with the stimulant drug are an indication of the relative changes in synaptic concentration of DA (Fig. 3). Studies showed that the response between subjects was quite variable, that it decreased with age, and that the intensity of the behavioral effects was closely correlated with the magnitude of the changes in DA concentration.

Imaging technologies can also be used to investigate the relation between behavioral effects and degree of receptor or transporter occupancy. These studies use as receptor radioligand one which is the target of the drug actions, and the studies are performed at baseline and after the drug of interest is given. For example, SPECT was used to evaluate receptor occupancy by the benzodiazepine agonist lorazepam and showed that only a very small fraction of the receptors are occupied at pharmacological doses (Sybirska et al. 1993), findings which corroborate the existence of receptor reserve for benzodiazepine receptors in humans.

The effects of drugs of abuse on enzyme activity have also been assessed with PET for the concentration of monoamine oxidase B (MAO B) in brain. These studies used the PET radiotracer [^{11}C]L-deprenyl-D_2, which binds specifically to the enzyme MAO B. Using this ligand, the effects of cigarette smoke on MAO B have been investigated in human brain (Fowler et al. 1996). This study showed that cigarette smokers have a reduction in brain MAO B of about 40% relative to nonsmokers and former smokers. MAO B inhibition is associated with enhanced activity of DA, a neurotransmitter involved in reinforcing and motivating behaviors, in movement, and in decreased production of hydrogen peroxide, a source of reactive oxygen species (Reiter 1985). Inhibition of MAO by cigarette smoke could be one of the mechanisms accounting for the lower incidence of Parkinson disease in cigarette smokers.

2.3 Drug Toxicity

The toxicity from drugs can be assessed both for brain and for other organs. Because imaging technologies provide functional information, they are very sensitive tools to detect brain pathology secondary to drug use.

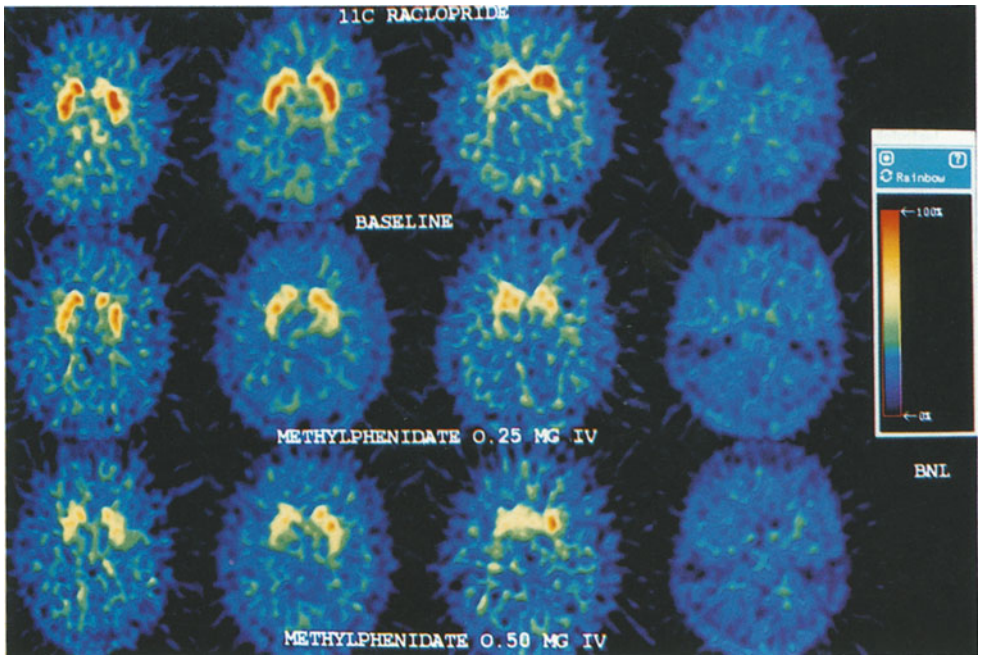


Fig. 3. Brain images obtained with [^{11}C]raclopride after placebo (baseline) and after methylphenidate (0.25 mg/kg i.v. and 0.5 mg/kg i.v.). Methylphenidate decreased the binding of [^{11}C]raclopride in the striatum, and the decrements were larger for the 0.5 mg/kg dose than for the 0.25 mg/kg dose. Binding of [^{11}C]raclopride to the D_2 receptors is sensitive to competition with endogenous dopamine (DA); thus methylphenidate-induced increases in dopamine concentration decrease its binding

Such sensitivity is illustrated by the PET studies which documented for the first time abnormalities in CBF in cocaine abusers (Fig. 4; Volkow et al. 1988a). These findings served to corroborate the clinical reports on the occurrence of cerebral strokes and hemorrhages associated with cocaine consumption. The defects in CBF appear to be secondary to the vasoactive properties of cocaine.

Imaging can also be used to evaluate mechanisms of enhanced toxicity when more than one drug is consumed at the same time, as is the case for the combined use of cocaine and alcohol (Grant and Harford 1990; Boag and Havard 1985). Although the factors mediating the increased risk associated with the combined use of cocaine and alcohol are not fully understood, it was postulated that cocaethylene, a metabolite of cocaine formed in the presence of alcohol, may contribute to the enhanced morbidity and mortality (Hearn et al. 1991a). Using PET, the pharmacokinetics of [^{11}C]cocaethylene was compared with [^{11}C]cocaine (Fowler et al. 1992a). The distribution of [^{11}C]cocaethylene in brain and heart was very similar to that of [^{11}C]cocaine, and the two drugs differed only in their rate of brain clearance, which was 20%–25% slower for [^{11}C]cocaethylene than for [^{11}C]cocaine. To evaluate the possibility that the enhanced toxicity resulted from changes in bioavailability when drugs are taken in combination, PET studies were also performed during alcohol intoxication to assess the effects of alcohol on the brain uptake and pharmacokinetics of [^{11}C]cocaine (Fowler et al. 1992b). Neither the distribution of cocaine nor its pharmacokinetics were affected when [^{11}C]cocaine was given while the subjects were intoxicated with alcohol, a finding that serves to rule out the possibility that the synergistic toxicity of alcohol and cocaine is due to changes in bioavailability. These studies showed that the enhanced toxicity is unlikely to be explained by unique pharmacokinetics properties of [^{11}C]cocaethylene or by changes in cocaine bioavailability. Rather, it is more likely that it reflects the direct toxic effects of the combined use of the two drugs.

Organ toxicity from drug use can be evaluated from several perspectives. In the heart, for example, functional imaging strategies can be used to assess whether there are changes in myocardial blood flow and metabolism in substance abusers. The labeled drug can be used to measure its distribution in the various organs, which is important when determining whether a given drug can have direct toxic effects on a given organ. In the case of cocaine, for example, PET studies done with [^{11}C]cocaine have shown significant accumulation in human heart (Volkow et al. 1992a). Also using [^{18}F]-labeled norepinephrine as a ligand to monitor the function of the peripheral norepinephrine transporter, cocaine was shown to inhibit the trans-

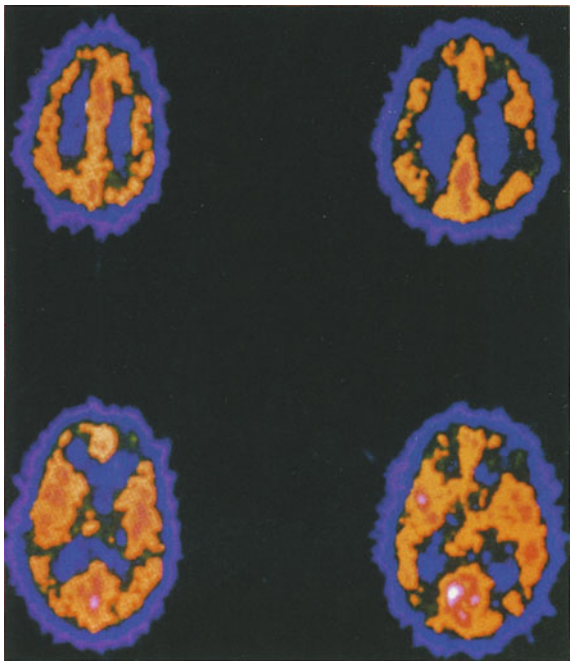


Fig. 4. Images obtained with [^{15}O]water to measure cerebral blood flow (CBF) in a cocaine abuser. CBF is scaled as red > yellow > green > blue. Notice the patchy distribution of the tracer in the cortex due to the decreased flow into both the frontal and parietal cortical areas

porter in heart (Fowler et al. 1994). This is of particular interest because of the documented cardiotoxic properties of cocaine. Cocaine is a local anesthetic, and its accumulation in heart could result in direct myocardial toxicity. At the same time, inhibition of the norepinephrine transporter by cocaine interferes with a protective mechanism of the heart to remove circulating catecholamines.

3
Imaging and Addictive Processes: Evaluation of the Addicted Subject

PET and SPECT have been used to assess neurochemical and functional changes in the brain of addicted subjects and their changes as a function of withdrawal and/or drug treatment. For example, imaging studies recently documented that the decrements in CBF in cocaine abusers improved with the administration of buprenorphine (Levin et al. 1995). Functional imaging strategies can also be used to assess the pattern of brain activation during drug-related states triggered by behavioral interventions such as those that induce drug craving. For example, it was recently reported that, in cocaine abusers in whom craving was elicited by exposure to a video showing cocaine paraphernalia, activation of limbic brain regions was induced (Childress et al. 1995). This area of research is likely to benefit tremendously from fMRI, by means of which multiple repeated studies can be performed in the same subjects without the limitation of radiation exposure in PET and SPECT. Imaging studies have been done in cocaine, marijuana, heroin, and alcohol abusers, tobacco smokers, and polysubstance abusers.

3.1
Cocaine Abusers

The effects of acute cocaine administration (30 mg i.v.) on regional brain glucose metabolism have been investigated with PET and FDG in cocaine abusers (London et al. 1990b). Cocaine decreased brain metabolism in all brain regions investigated, though to a less extent in cerebellum. The magnitude of cocaine-induced decreases in metabolism were correlated with the degree of ventricular enlargement. Subjects with enlarged ventricles were less sensitive to the effects of cocaine on brain metabolism than those that had less ventricular enlargement. This study postulates that one of the mechanisms of reinforcement of cocaine may involve the reduction of brain metabolism.

The effects of cocaine on the striatal DA system have been investigated with PET, since it has been postulated the decreased DA activity could underlie cocaine addiction (Dachis and Gold 1985). A multitracer approach to monitor DA D_2 receptors and metabolic activity in the same subjects was used to assess the relation between changes in DA parameters and regional brain metabolism during early and late detoxification. Regional brain glucose metabolism in recently detoxified cocaine abusers (<1 week) was significantly higher in orbitofrontal cortex and in striatum than in healthy nonabusing controls (Volkow et al. 1991a). The metabolic activity in these brain regions was found to be correlated with the days since last use of cocaine. The highest values were observed in subjects tested during the initial 72 h. Metabolic activity in orbitofrontal cortex and in striatum was also significantly correlated with the intensity of cocaine craving. Cocaine abusers who had the highest subjective ratings for craving were those with the highest metabolic values in orbitofrontal cortex and striatum. Studies done with ^{18}F-labeled *N*-methylspiroperiodol (NMS) on these same cocaine abusers revealed decreases in DA D_2 receptor availability (Volkow et al. 1990a). In contrast, studies done in cocaine abusers tested between 1 and 4 months of detoxification showed significant reductions in metabolic activity in prefrontal cortex, orbitofrontal cortex,

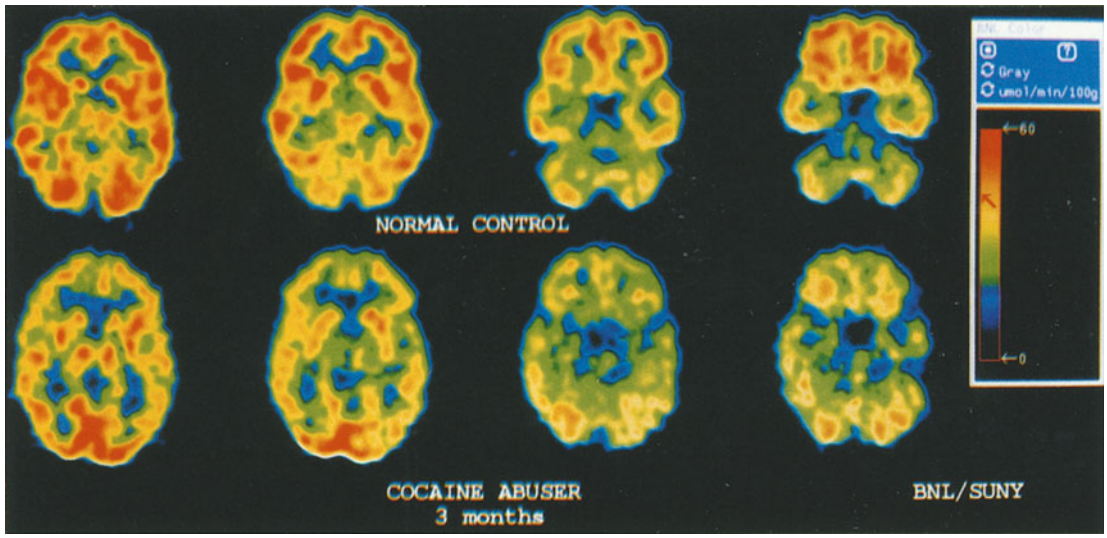


Fig. 5. Regional brain metabolic images in a normal control and in a cocaine abuser scanned 3 months after last use of cocaine. Notice the marked decrements in metabolic activity in the cocaine abuser. Decreases are most pronounced in frontal areas

temporal cortex, and cingulate gyrus (Fig. 5; Volkow et al. 1992c). Measures of DA D_2 receptor availability in these patients also showed a significant reduction, as had been observed during early detoxification. Measures for D_2 receptors correlated significantly with measures of metabolic activity in orbitofrontal cortex, cingulate gyrus, and prefrontal cortex (Volkow et al. 1993b). Decrements in regional brain metabolism and the reductions in DA D_2 receptor availability persisted in the follow-up studies performed 3 months after completing the inpatient detoxification program. The correlation between DA D_2 receptor availability and metabolic activity in orbitofrontal cortex and cingulate gyrus suggests an association between DA activity and the function of these brain regions. Lower values for D_2 receptor concentration were associated with lower metabolism in these brain regions.

This pattern of metabolic abnormalities is similar to that reported in patients with obsessive/compulsive disorders (OCD) (Baxter et al. 1987). Although these two clinical populations are distinct disorders, they share the pattern of compulsive repetitive behaviors. In OCD patients, this manifests as a behavioral ritual, and in cocaine abusers as a repetitive pattern of cocaine self-administration. Experiments in animals have shown that destruction of the orbitofrontal cortex leads to the emergence of repetitive behaviors that cannot be easily terminated (Kolb 1977), and a similar syndrome can be generated by the destruction of the mesocortical DA pathway (LeMoal and Simon 1991). Hence we postulate that DA disruption of the orbitofrontal cortex may be one of the mechanisms responsible for the compulsive administration of cocaine during a "binge" and for the loss of control experienced by the drug abusers when exposed to cocaine and/or cocaine-related cues. Thus DA involvement in addiction may be mediated by its interactions with frontal circuits involved in the control of repetitive and impulsive behaviors.

Findings from these studies have served to demonstrate that addicted individuals have neurochemical changes in their brains that may underlie their inability to control their impulses to take the drug and may explain relapse.

3.2 Alcoholics

PET studies in alcoholics have been done to measure CBF, brain glucose metabolism (baseline and with pharmacological challenges), benzodiazepine receptors, DA D_2 receptors, and DA transporters in brain. These studies have been done to determine whether the brains of alcoholics differ from those of nonalcoholic normal controls and to assess whether there are changes in the brains of subjects who have a genetic predisposition for alcoholism.

PET has been used to investigate brain metabolic changes in chronic alcoholics with and without neurological impairment (Sachs et al. 1987; Wik et al. 1988; Samson et al. 1986; Kessler et al. 1984; Volkow et al. 1992b). Alcoholics with Korsakoff's encephalopathy showed decreased metabolism in prefrontal, parietal, and temporal cortices, and alcoholics with neurological symptoms other than Korsakoff's encephalopathy showed decreased metabolism in frontal and parietal cortices. Studies in alcoholics with no evidence

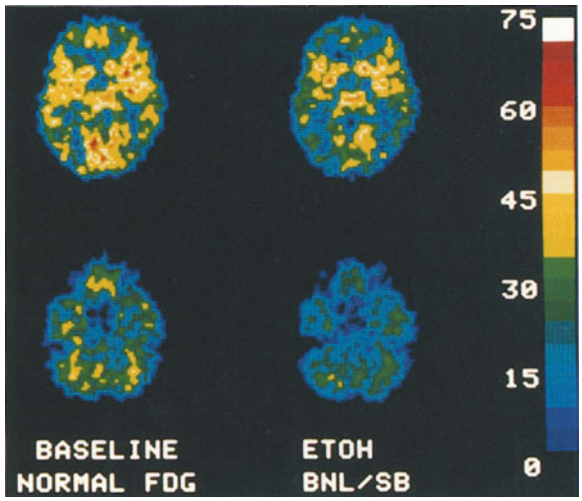


Fig. 6. Brain metabolic images of a normal control at the level of the striatum and the cerebellum at baseline (left-side images) and during alcohol intoxication (right-side images). Notice the decrements in metabolism during alcohol intoxication. FDG, fluorodeoxyglucose; ETOH, ethyl alcohol

of neurological impairment have also shown evidence of frontal abnormalities. Frontal dysfunction in alcoholics has also been consistently found in non-PET studies which measured regional CBF (for reviews, see Mathew and Wilson 1991; Lofti and Meyer 1989). The degree of brain metabolic recovery after detoxification was evaluated with PET in alcoholics who were evaluated at three different times after detoxification (Volkow et al. 1994b). This study showed that brain metabolism increased significantly during detoxification, predominantly during the first 16–30 days of detoxification. Regional increases in metabolism were larger in frontal regions. In alcoholics, metabolism in frontal, parietal, and left temporal cortices was negatively correlated with years of alcohol use and with age, whereas in the comparison group they were not. Decrements in metabolism were most accentuated in the older alcoholics with the longer histories of alcohol consumption.

The effects of ethanol on brain glucose metabolism (DeWit et al. 1990; Volkow et al. 1990b) and CBF (Volkow et al. 1988b) have been evaluated with PET. Such studies have shown that acute alcohol administration decreased brain glucose metabolism (Fig. 6). In contrast, acute alcohol administration increases CBF. When compared with normal controls, alcoholics showed a larger metabolic response to ethanol, despite showing a lower subjective response to the intoxicating properties of ethanol. In normals, the subjective response to the intoxicating effects of ethanol was significantly correlated with metabolic response (Volkow et al. 1988b). The paradoxical metabolic and behavioral responsiveness in alcoholics could reflect tolerance of the brain to ethanol-induced metabolic changes.

The acute (DeWit et al. 1991; Volkow et al. 1993a, 1995a) and chronic (Buchsbaum et al. 1987) effects of benzodiazepine drugs on brain glucose metabolism have also been evaluated with PET. Such studies have shown that, similar to ethanol, benzodiazepines decrease brain glucose metabolism and that the effects are more pronounced in the occipital cortex, the area of the brain with the highest density of benzodiazepine receptors (Braestrup et al. 1977). Studies comparing the response to benzodiazepines between normals and alcoholics have shown that, whereas normal and alcoholic subjects showed a similar response to lorazepam in occipital and cerebellar metabolism, the alcoholics showed a significantly lower response in thalamus, basal ganglia, and orbitofrontal cortex (Volkow et al. 1993a). These studies indicate that alcoholics have a blunted response to lorazepam. This is regionally specific and could reflect either chronic alcohol administration, withdrawal, and/or genetic differences.

In order to assess whether there is a genetic component in the differences in the sensitivity to benzodiazepines, the regional brain metabolic response to lorazepam was evaluated in subjects with a family history of alcoholism (FHP) and compared with that of subjects without a family history of alcoholism (FHN). At baseline, FHP subjects showed lower cerebellar metabolism than FHN subjects, and when challenged with lorazepam they also showed a blunted response in cerebellum and in cingulate gyrus. Lorazepam-induced changes in cerebellar metabolism were significantly correlated with motor impairment. These changes could account for the decreased sensitivity to the motor effects of alcohol and benzodiazepines in FHP subjects and indicate regional brain metabolic differences in children of alcoholics.

Imaging studies have also been done to measure receptors and transporters in alcoholics. Receptors studies have been done to assess changes in DA D_2 receptors as well as on benzodiazepine receptors in alcoholic subjects. Benzodiazepines were measured using [^{11}C]flumazenil and showed that, while there were no changes in receptor numbers between normals and alcoholics, the latter had a significantly larger variability for receptor concentration (B_{max}) measures than controls (Litton et al. 1993). DA D_2 receptors have also been evaluated in alcoholics with PET and [^{11}C]raclopride. This study showed that alcoholics had a 19.7% decrease in B_{max}/K_d and a 12% decrease in D_2 receptor numbers when compared with controls (Hietala et al. 1994). No significant correlations were found between days of detoxification (1–68 weeks) and D_2 receptor numbers. Because persistent reductions in

D_2 receptor availability have also been reported in cocaine abusers (Volkow et al. 1993b), D_2 decrements may represent a variable related to vulnerability for addictive disorders.

DA transporter availability has also been measured in alcoholics with SPECT and [^{123}I]2β-carbomethoxy-3β-(4-iodophenyltropane) (β-CIT). This study showed that, while violent alcoholics had increases (8%), nonviolent alcoholics had decreases in DA transporters (25%) when compared with nonalcoholics controls (Tiihonen et al. 1995). The interpretation for these two findings is not the same; the decreases in D_2 receptors in alcoholics result mainly from changes occurring in striatal GABAergic neurons, whereas assessment of DA transporters reflects DA cells.

3.3 Marijuana Abusers

Marijuana is the most widely used illicit drug of abuse in the United States (Goodman and Gilman 1990). Despite its widespread use, the mechanisms by which Δ-9-tetrahydrocannabinol (THC), the main psychoactive substance of marijuana, exerts its psychoactive effects are still not known (Martin 1986). Non-PET imaging has been used to assessed the effect of THC intoxication on CBF in chronic marijuana users (Mathew et al. 1989; Tunuing et al. 1986). Acute marijuana administration led to decreases in CBF in subjects who were not experienced marijuana smokers, whereas it increased CBF in subjects who were experienced smokers. Chronic marijuana users showed a temporary reduction in CBF that reverted to normal with abstinence. Interpretation of the effects of THC on CBF is confounded by the vasoactive properties of THC (Nahas 1986). Thus it is difficult to separate the effects of THC that are related to its action on nervous tissue from those that are related to its vasoactive effects. This problem is obviated when using deoxyglucose to measure brain glucose metabolism, since it is insensitive to fluctuations in CBF (Sokoloff et al. 1977).

The effects of THC on regional brain glucose metabolism have been evaluated in nonabusing controls (Volkow et al. 1991b) and in marijuana abusers (Volkow et al. 1996b). The whole-brain metabolic response to the effects of THC was variable among individuals; in some subjects it increased, in some it decreased, and in some it did not show change. Despite these variable responses in whole-brain metabolism, there was a very consistent pattern of metabolic activation by THC. In other words, under THC intoxication, most of the subjects showed activation of the cerebellum. The cerebellar activation by THC was significant both for the absolute and the relative measures. Apart from cerebellar activation, the only other region that showed a trend toward activation during THC intoxication was the prefrontal and the orbitofrontal cortex. In contrast to the activation of cerebellum, which was seen across most of the subjects investigated, activation of the orbitofrontal cortex occurred predominantly in the subjects who had a history of frequent marijuana use.

The regional localization of the effects of THC on brain glucose metabolism do not support a diffuse nonspecific mechanism of action for THC. These findings are more in agreement with the hypothesis that proposes interaction of THC with regionally localized receptor sites in the human brain (Howlett et al. 1990). The highly localized concentration of cannabinoid receptors in the cerebellum (Herkenham et al. 1990) supports involvement of the cannabinoid receptors in the metabolic response during THC intoxication. Activation by THC of the cerebellum, which is a system involved in motor coordination (Eccles 1973), proprioception (Brodal 1981), and learning (Derety et al. 1990), could explain some of the behavioral effects seen during THC intoxication. It could, for example, explain the disruption in motor coordination and proprioception during THC intoxication. Cannabinoid receptors are also localized in other discrete areas, namely, hippocampus, substantia nigra pars reticulata, and globus pallidus. The latter are too small to be measured with the spatial resolution of the PET instrument utilized.

Attempts to investigate THC in the living brain with PET by using the labeled drug with a positron emitter have been unsuccessful because of the highly lipophilic nature of THC. This was also a limitation for Δ-8-THC, an analogue of Δ-9-THC, which was labeled with ^{18}F (positron emitter with a half-life of 110 min). Its uptake and distribution were investigated in the baboon brain with PET, and widespread uptake in the brain with no particular pattern of localization was found (Marciniak et al. 1990). This pattern of distribution probably represented mainly nonspecific binding, since pretreatment with Δ-8-THC did not affect the uptake of [^{18}F]-Δ-8-THC in brain. A promising alternative may be the use of THC antagonists with high receptor affinities (Gatley et al. 1996).

4 Conclusion

Imaging technologies have started to document mechanisms of reinforcement of addictive substances and to delineate neurochemical changes in the brain of the addicted subject. Although these findings are still of a

preliminary nature, they provide an indication of the potential of imaging techniques in the area of substance abuse, including the following

1. The behavior of drugs of abuse in the human brain can be assessed. This is relevant both because drug pharmacokinetics and pharmacodynamics may vary across animal species. It also enables the assessment of drug behavior directly in the drug addict.
2. Because imaging studies are done in awake human subjects, the relation between behavior and regional brain effects can be determined with regard to both neurotransmitters and function as assessed by measures of glucose metabolism or CBF. Studies can also be done to assess the relation between pharmacokinetics of a given drug and the time course of its pharmacological effects.
3. Imaging technologies have the unique ability to view neurochemical and functional changes from many perspectives directly in the addicted individual.
4. The pharmacokinetics of the effects of drugs on regional brain activity can be assessed by taking advantage of the fast imaging capabilities of fMRI.
5. They can be applied in the development of new therapeutic interventions.

5 References

Abi-Dargham A, Laruelle M, Seibyl J, Rattner Z, Baldwin RM, Zoghbi SS, Zea-Ponce Y. Bremner JD, Hyde TM, Charney DS, Hoffer PB, Innis RB (1994) SPECT measurement of benzodiazepine receptors in human brain with iodine—iomazenil: kinetic and equilibrium paradigms. J Nucl Med 35: 228–239

Andrew RF (1994) Introduction to nuclear magnetic resonance. In: Gillies RJ (ed) NMR in physiology and biomedicine. Academic, San Diego, pp 1–24

Barkley RA (1977) A review of stimulant drug research with hyperactive children. J Child Psychol Psychiatry 8: 137–165

Baxter LR, Phelps ME, Mazziotta J, Guze BH, Schwartz JM, Selin CE (1987) Local cerebral glucose metabolic rates in obsessive compulsive disorder: a comparison with rates in unipolar depression and normal controls. Arch Gen Psychiatry 44: 211–218

Baxter LR Jr, Schwartz JM, Phelps ME et al (1988) Localization of neurochemical effects of cocaine and other stimulants in the human brain. J Clin Psychiatry 49: 23–26

Bergstrom M, Nordberg A, Lunell E, Antoni G, Langstrom B (1995) Regional deposition of inhaled 11C-nicotine vapor in the human airway as visualized by positron emission tomography. Clin Pharmacol Ther 57: 309–317

Boag F, Havard CWH (1985) Cardiac arrhythmia and myocardial ischaemia related to cocaine and alcohol consumption. Postgrad Med J 61: 997–999

Braestrup C, Albrechsten R, Squires RF (1977) High densities of benzodiazepine receptors in human cortical areas. Nature 269: 702–704

Brodal A (1981) Neurological anatomy. Oxford University Press, New York

Buchsbaum MS, Wu J, Haier R, Hazlett E, Ball R, Katz M, Sokolski K, Lagunas-Solar M, Langer D (1987) Positron emission tomography assessment of effects of benzodiazepines on regional glucose metabolic rate in patients with anxiety disorder. Life Sci 40: 2393–2400

Cameron OG, Modell JG, Hariharan M (1990) Caffeine and human cerebral blood flow: a positron emission tomography study. Life Sci 47: 1141–1146

Childress AR, Mozley D, Fitzgerald J, Reivich M, Jaggi J, O'Brien CP (1995) Limbic activation during cue-induced cocaine craving. Abstr Soc Neurosci 767.1

*Dachis CA, Gold MS (1985) New concepts in cocaine addiction: the dopamine depletion hypothesis. Neurosci Biobehav Rev 9: 469–477

Derety J, Sjoholm H, Ryding E, Stenberg G, Ingvar D (1990) The cerebellum participates in mental activity: tomographic measurements of regional cerebral blood flow. Brain Res 535: 313–317

DeWit H, Metz J, Wagner N, Cooper M (1990) Behavioral and subjective effects of ethanol: relationship to cerebral metabolism using PET. Alcohol Clin Exp Res 14: 482–489

DeWit H, Metz J, Wagner N, Cooper M (1991) Effects of diazepam on cerebral metabolism and mood in normal volunteers. Neurospyschopharmacology 5: 33–41

Eccles JC (1973) The cerebellum as a computer: patterns in space and time. J Physiol 229: 1–32

*Fowler JS, Volkow ND, Wolf AP, Dewey SL, Schlyer DJ, MacGregor RR, Hitzemann R, Logan J, Bendriem B, Christman D (1989) Mapping cocaine binding in human and baboon brain in vivo. Synapse 4: 371–377

Fowler JS, Volkow ND, MacGregor RR, Logan J, Dewey SL, Gatley SJ, Wolf AP (1992a) Comparative PET studies of the kinetics and distribution of cocaine and cocaethylene in baboon brain. Synapse 12: 220–227

Fowler JS, Volkow ND, Logan J, MacGregor RR, Wang G-J, Wolf AP (1992b) Alcohol intoxication does not change cocaine pharmacokinetics in human brain and heart. Synapse 12: 228–235

Fowler JS, Ding YS, Volkow N, Martin T, Mac Gregor R, Dewey SL, King P, Pappas N, Alexoff D, Shea C, Gatley J, Schlyer D, Wolf A (1994) PET studies of cocaine inhibition of the myocardial norepinephrine uptake. Synapse 16: 312–317

*Fowler JS, Volkow ND, Wang G-J, Pappas N, Logan J, MacGregor R, Alexoff D, Shea C, Wolf AP, Warner D, Zezulkova I, Cilento R (1996) Inhibition of MAO B in the brains of smokers. Nature 379: 733–736

Frost JJ, Wagner HN, Dannals RF, Ravert HT, Links JM, Wilson AA, Burns HD, Wong DF, McPherson RW, Rosenbaum AE, Kuhar MJ, Snyder SH (1985) Imaging opiate receptors in the human brain by positron tomography. J Comp Assist Tomogr 9: 231–236

Galynker I, Schlyer DJ, Dewey SL, Fowler JS, Logan J, Gatley SJ, MacGregor RR, Ferrieri RA, Holland MJ, Brodie J, Simon E, Wolf AP (1996) Opioid receptor imaging and displacement studies with [6-O-[11C]methyl]-buprenorphine in baboon brain. Nucl Med Biol 23(3): 325–331

Gatley SJ, Gifford A, Makriyannis A, Volkow ND (1996) Iodine-labeled AM 251: a radioiodinated ligand with binding in vivo

to mouse brain CB1 cannabinoid receptors. Eur J Pharmacol 307: 331–338

Goodman LS, Gilman A (1990) The pharmacological basis of therapeutics, 8th edn. Pergamon, New York, p 549

Grant BF, Harford TC (1990) Concurrent and simultaneous use of alcohol with cocaine: results of national survey. Drug Alcohol Depend 25: 97–104

Hartvig P, Bergström K, Lindberg B, Lundberg PO, Lundqvist H, Långström B, Svärd H, Rane A (1984) Kinetics of ^{11}C-labeled opiates in the brain of rhesus monkeys. J Pharmacol Exp Ther 230: 250–255

Hearn WL, Rose S, Wagner J, Ciarleglio A, Mash DC (1991a) Cocaethylene is more potent than cocaine in mediating lethality. Pharmacol Biochem Behav 39: 531–533

Hearn WL, Flynn DD, Hime GW, Rose S, Cofino JC, Mantero-Atienza E, Wetli CV, Mash DC (1991b) Cocaethylene: a unique cocaine metabolite displays high affinity for the dopamine transporter. J Neurochem 56: 698–701

Herkenham M, Lynn A B, Little MD, Ross Johnson M, Melvin LS, DeCosta BR, Rice KC (1990) Cannabinoid receptor localization in brain. Proc Natl Acad Sci USA 87: 1932–1936

Hietala J, West C, Syvälahti E, Nagren Leikoinen P, Sonninen P, Ruotsalainen U (1994) Striatal D2 dopamine receptor binding characteristics in vivo in patients with alcohol dependence. Psychopharmacology 116: 285–290

Howlett AC, Champion TM, Wilken GH, Mechoulam R (1990) Stereochemical effects of 11-OH 8 delta tetrahydrocannabinol-dimethylneptyl to inhibit adenylate cyclase and bind to the cannabinoid receptor. Neuropharmacology 29: 161–165

Kessler RM, Parker ES, Clark CM (1984) Regional cerebral glucose metabolism in patients with alcoholic Korsakoff's syndrome. Soc Neurosci (Abstr) 10: 541

Kolb B (1977) Studies on the caudate putamen and the dorsomedial thalamic nucleus: implications for mammalian frontal lobe function. Physiol Behav 18: 237–244

Kung HF (1993) SPECT and PET ligands for CNS imaging. Neurotransmissions 9: 1–8

Laruelle M, Abi-Dargham A, van Dick C et al. (1995) SPECT imaging of striatal dopamine release after amphetamine challenge. J Nucl Med 36: 1182–1190

**Le Moal M, Simon H (1991) Mesocorticolimbic dopaminergic network: functional and regulatory notes. Physiol Rev 71: 155–234

Levin JM, Holman BL, Mendelson JH, Teoh SK, Garada B, Johnson KA, Springer S (1994) Gender differences in cerebral perfusion in cocaine abuse: technetium-m-HMPAO SPECT study of drug-abusing women. J Nucl Med 35: 1902–1909

Levin JM, Mendelson JH, Holman LB, Teoh SK, Garada B, Schwartz RB, Mello NK (1995) Improved regional cerebral blood flow in chronic cocaine polydrug users treated with buprenorphine. J Nucl Med 36: 1211–1215

Litton JE, Neiman J, Pauli S, Farde L, Hindmarsh T, Halldin C, Sedvall G (1993) PET analysis of [11C]flumazenil binding to benzodiazepine receptors in chronic alcohol-dependent men and healthy controls. Psych Res 1–13

Lofti J, Meyer JS (1989) Cerebral hemodynamics and metabolic effects of chronic alcoholism. Cerebrovasc Brain Metab Rev 1: 2–25

London ED, Morgan JM (1993) Positron emission tomographic studies on the acute effects of psychoactive drugs on brain metabolism and mood. In: London ED (ed) Imaging drug action in the brain. CRC Press, Boca Raton, pp 265–280

London ED, Broussolle EPM, Links JM, Wong DF, Cascella NG, Dannals RF, Sano M, Herning R, Snyder FR, Rippetoe LR, Toung TJK, Jaffe JH, Wagner HN Jr (1990a) Morphine-induced metabolic changes in human brain: studies with positron emission tomography and [fluorine-18]fluorodeoxyglucose. Arch Gen Psychiatry 47: 73–81

London ED, Cascella NG, Wong DF, Phillips RL, Dannals RF, Links JM, Herning R, Grayson R, Jaffe JH, Wagner HN (1990b) Cocaine induced reduction of glucose utilization in human brain. A study using positron emission tomography and [fluorine 18]fluorodeoxyglucose. Arch Gen Psychiatry 47: 567–574

Marciniak G, Charalambous A, Schiue CY, Dewey SL, Schlyer DJ, Makriyannis A, Wolf AP (1990) 18F labelled tetrahydrocannabinol: synthesis, distribution in mice and PET studies in a baboon. J Nucl Med 31: 902

Martin BR (1986) Cellular effects of cannabinoids. Pharmacol Rev 38: 45–74

Mathew RJ, Wilson WH (1991) Substance abuse and cerebral blood flow. Am J Psychiatry 148: 292–305

Mathew RJ, Wilson WH, Tant SR (1989) Acute changes in cerebral blood flow associated with marijuana smoking. Acta Psychiatr Scand 79: 118

Mazière M, Comar D, Marazano C, Berger G (1976) Nicotine-^{11}C: synthesis and distribution kinetics in animals. Eur J Nucl Med 1: 255–258

Melon PG, DeGrado TR, Nguyen N, Caraher J, Tooronagian SA, Hutchins GD, Schwaiger M (1992) Non-invasive assessment of regional myocardial sympathetic neuronal function following intravenous cocaine injection in dogs. J Nucl Med 33: 994

Meyer JS, Sakai F, Karacan I, Derman S, Yamamoto M (1980) Sleep apnea, narcolepsy, and dreaming: regional cerebral hemodynamics. Ann Neurol 7: 479–485

**Mullani NA, Volkow ND (1992) Positron emission tomography instrumentation: a review and update. Am J Physiol Imaging 7: 121–135

Nahas GG (1986) Toxicolology and pharmacology. Nahas GG (ed) Marijuana in science and medicine. Raven, New York, p 109

Nicolas JM, Catafau AM, Estruch R, Lomena FJ, Salamero M, Herranz R, Monforte R, Cardenal C, Urbano-Marquez A (1993) Regional cerebral blood flow-SPECT in chronic alcoholism: relation to neuropsychological testing. J Nucl Med 34: 1452–1459

Pappata S, Samson Y, Chavoix C, Prenant C, Maziere M, Baron JC (1988) Regional specific binding of [^{11}C]RO 15 1788 to central type benzodiazepine receptors in human brain: quantitative evaluation by PET. J Cereb Blood Flow Metab 8: 304–313

Reiter RJ (1985) Oxidative processes and antioxidative defense mechanisms in the aging. FASEB J 9: 526–533

Rogers LW, Ackermann RJ (1992) SPECT instrumentation. Am J Physiol Imaging 7: 105–120

Sachs H, Russell JAG, Christman DR, Cook B (1987) Alteration of regional cerebral glucose metabolic rate in non-Korsakoff chronic alcoholism. Arch Neurol 44: 1242–1251

Samson Y, Baron JC, Feline A, Bories J, Crouzel C (1986) Local cerebral glucose utilization in chronic alcoholics: a positron tomographic study. J Neurol Neurosurg Psychiatry 49: 1165–1170

**Sokoloff L, Reivich M, Kennedy C, Des Rosiers MH, Patlak CS, Pettigrew KD, Sakurada O, Shinohara M (1977) The 14C deoxyglucose method for the measurement of local cerebral

glucose utilization: the conscious and anesthetized albino rat. J Neurochem 28: 897–916

Stapleton JM, Henningfield JE, Wong DF, Phillips RL, Grayson RF, Dannals RF, London ED (1993) Nicotine reduces cerebral glucose utilization in humans. NIDA Res Monogr 132: 106

Sybirska E, Seibyl JP, Bremner JD, Baldwin RM, Al-Tikriti MS, Bradberry C, Malison RT, Zea Ponce Y, Zoghbi S, During M, Goddard AW, Woods SW, Hoffer PB, Charney DS, Innis RB (1993) [123-I]Iomazenil SPECT imaging demonstrates significant benzodiazepine receptor reserve in human and non human primate brain. Neuropharmacology 32: 671–680

Theodore WH, DiChiro G, Margolin R, Fishbein D, Porter RJ, Brooks RA (1986) Barbiturates reduce human cerebral glucose metabolism. Neurology 36: 60–64

Tiihonen J, Kuikka J, Bergström K, Hakola P, Larhu J, Ryynänen OP, Föhr J (1995) Altered striatal dopamine re-uptake site densities in habitual violent and non-violent alcoholics. Nature Med 1: 654–657

Tunving K, Thulin S, Risberg J, Warkentin S (1986) Regional cerebral blood flow in long term heavy cannabis use. Psychiatry Res 17: 15–21

**Volkow ND, Fowler JS (1992) Neuropsychiatric disorders: investigation of schizophrenia and substance abuse. Semin Nucl Med XXII: 254–267

*Volkow ND, Mullani N, Gould KL, Adler S, Krajewski K (1988a) Cerebral blood flow in chronic cocaine users: a study with positron emission tomography. Br J Psychiatry 152: 641–648

Volkow ND, Guynn R, Marani S, Adler S, Gould L (1988b) Effects of alcohol on cerebral blood flow. Psychiatry Res 24: 201–209

Volkow ND, Fowler JS, Wolf AP, Schlyer D, Shiue C-Y, Dewey SL, Alpert R, Logan J, Christman D, Bendriem B, Hitzemann R, Henn F (1990a) Effects of chronic cocaine abuse on postsynaptic dopamine receptors. Am J Psychiatry 147: 719–724

Volkow ND, Hitzemann R, Wolf AP, Logan J, Fowler JS, Christman D, Dewey SL, Schlyer D, Burr G, Vitkun S, Hirschowitz J (1990b) Acute effects of ethanol on regional brain glucose metabolism and transport. Psychiatry Res 35: 39–48

Volkow ND, Fowler JS, Wolf AP, Hitzemann R, Dewey SL, Bendriem B, Alpert R, Hoff A (1991a) Changes in brain glucose metabolism in cocaine dependence and withdrawal. Am J Psychiatry 148: 621–626

Volkow ND, Gillespie H, Mullani N, Tancredi L, Grant L, Ivanovic M, Hollister L (1991b) Cerebellar metabolic activation by delta-9-tetrahydrocannabinol in human brain: a study with positron emission tomography and F-18-2-fluoro-2-deoxyglucose. Psychiatry Res 40: 69–78

Volkow ND, Fowler JS, Wolf AP, Wang G-J, Logan J, MacGregor R, Dewey SL, Schlyer DJ, Hitzemann R (1992a) Distribution of 11C-cocaine in human heart, lungs, liver and adrenals. A dynamic PET study. J Nucl Med 33: 521–525

Volkow, ND, Hitzemann R, Wang G-J (1992b) Decreased brain metabolism in neurologically intact healthy alcoholics. Am J Psychiatry 149: 1016–1022

Volkow ND, Hitzemann R, Wang G-J, Fowler JS, Wolf AP, Dewey SL (1992c) Long-term frontal brain metabolic changes in cocaine abusers. Synapse 11: 184–190

Volkow ND, Wang G-J, Hitzemann R, Fowler JS, Wolf AP, Pappas N, Biegon A, Dewey SL (1993a) Decreased cerebral response to inhibitory neurotransmission in alcoholics. Am J Psychiatry 150: 417–422

**Volkow ND, Fowler JS, Wang G-J, Hitzemann R, Logan J, Schlyer D, Dewey DL, Wolf AP (1993b) Decreased dopamine D2 receptor availability is associated with reduced frontal metabolism in cocaine abusers. Synapse 14: 169–177

**Volkow ND, Wang G-J, Fowler JS, Logan J, Schlyer D, Hitzemann R, Lieberman J, Angrist B, Pappas N, MacGregor R, Burr G, Cooper T, Wolf AP (1994a) Imaging endogeneous dopamine competition with [11C]raclopride in the human brain. Synapse 16: 255–262

Volkow ND, Wang G-J, Hitzemann R, Fowler JS, Overall JE, Burr G, Wolf AP (1994b) Recovery of brain glucose metabolism in detoxified alcoholics. Am J Psychiatry 151: 178–183

Volkow ND, Wang G-J, Hitzemann R, Fowler JS, Pappas N, Loremier P, Burr G, Pascani K, Wolf AP (1995a) Depression of thalamic metabolism by lorazepam is associated with sleepiness. J Neuropsychopharmacol 12: 123–132

**Volkow ND, Ding U, Fowler JS, Wang G-J, Logan J, Gatley SJ, Dewey SL, Ashby C, Lieberman J, Hitzemann R, Wolf AP (1995b) Is methylphenidate like cocaine? Studies on their pharmacokinetics and distribution in human brain. Arch Gen Psychiatry 52: 456–463

Volkow ND, Fowler JS, Logan J, Gatley SJ, Dewey SL, MacGregor RR, Schlyer DJ, Pappas N, King P, Wolf AP (1995c) Carbon-11-cocaine binding compared at sub-pharmacological and pharmacological doses: A PET study. J Nucl Med 36: 1289–1297

Volkow ND, Wang G-J, Begleiter H, Hitzemann R, Pappas N, Burr G, Pascani K, Wong C, Fowler JS, Wolf AP (1995d) Regional brain metabolic response to lorazepam in subjects at risk for alcoholism. Alcohol Clin Exp Res 19: 510–516

Volkow ND, Wang G-J, Fowler JS, Logan J, Hitzemann R, Gatley SJ, MacGregor RR, Wolf AP (1996a) Cocaine binding is decreased in the brain of detoxified cocaine abusers. J Neuropsychopharmacol 14: 159–168

Volkow ND, Gillespie H, Mullani N, Tancredi L, Grant C, Valentine A, Hollister L (1996b) Brain glucose metabolism in chronic marijuana users during baseline and during marijuana intoxication. Psychiatry Res 67: 29–38

Wang G-J, Volkow ND, Fowler JS, Wolf AP, Ferrieri R, Schlyer DJ, Alexoff D, Pappas N, Lieberman J, King P, Warner D, Wong C, Hitzemann RJ (1994) Methylphenidate decreases regional cerebral blood flow in normal human subjects. Life Sci 54: PL143–146

Weber DA, Franceschi D, Ivanovic M et al. (1993) SPECT and planar brain imaging in crack abuse: iodine-123-iodoamphetamine uptake and localization. J Nucl Med 34: 899–907

Wik G, Borg S, Sjögren I, Wiesel F, Blomquist G, Borg J, Greitz T, Nybäck H, Sedvall G, Stone Elander S, Widen L (1988) PET determination of regional cerebral glucose metabolism in alcohol-dependent men and healthy controls using ^{11}C-glucose. Acta Psychiatrica Scand 78: 234–241

Wilens TE, Biederman J (1992) The stimulants. Pediatr Psychopharmacol 15: 191–222

Wolkin A, Angrist B, Wolf A, Brodie J, Jordan B, Jaeger J, Cancro R, Rotrosen J (1987) Effects of amphetamine on local cerebral metabolism in normal and schizophrenic subjects as determined by positron emission tomography. Psychopharmacology 92: 241–246

CHAPTER

23

A. Uchtenhagen

Substitution Treatment for Opiate Dependence

1 Definition and Objectives

Substitution treatment is a therapeutic option that includes the prescription of agonist substances as an adjunct to medical and psychosocial care. The indication and the duration of treatment are determined on the basis of an individual assessment.

Substitution treatment is at present available for opiate-dependent individuals only. As a rule, the opiate of abuse is diacetylmorphine (heroin). Earlier, prescribing alcohol to alcohol-dependent individuals proved to be ineffective and was abandoned. Medical prescription of stimulants as substitution treatment is sporadically practiced, not adequately evaluated, and controversial.

The objectives of prescribing opioid agonists as a substitution for heroin are the following:

- To replace the unhygienic, uncontrolled, and risky forms of illegal heroin administration (especially intravenous injections)
- To reduce all activities linked to the purchase of illegal heroin
- To facilitate a reorientation of heroin dependents toward a socially integrated, healthy, responsible lifestyle
- To improve the chances for permanent abstinence from opiates

The objectives of substitution treatment do not diverge essentially from those of abstinence-oriented treatment. However, there is a difference regarding the timing of intermediate objectives: while abstinence-oriented treatment regards abstinence as a prerequisite for better health and social status (and therefore prefers residential settings), substitution treatment attempts to improve health and social status in order to better achieve permanent abstinence (and consequently prefers outpatient settings). Substitution treatment also allows intermediate goals to be reached where the ultimate objective of abstinence cannot yet be reached or cannot be reached at all, where it serves as a palliative treatment or a risk-reducing measure.

2 Pharmacologic Basis

The substitution of diacetylmorphine by opioid agonists is based on their affinity to μ-receptors which are responsible for euphoric effects and which enhance dependence. This is the case for morphine, methadone, levo-alpha acetylmethadol (LAAM), and also for mono-acetylmorphine, a metabolite of diamorphine. Buprenorphine and, to a lesser degree, codeine also bind to μ-receptors. When these receptors are occupied by substitution agonists, no withdrawal symptoms may develop and additional heroin is not effective.

When opioids are administered continuously, multiple mechanisms induce tolerance; continually higher dosages are needed to achieve the same effect. At the same time, higher dosages can be tolerated which might be lethal in nontolerant individuals. There is cross-tolerance between the above-mentioned opioids. Prescribed opioids induce tolerance for the effects of heroin. Discontinuation of substitution is followed by a withdrawal state comparable to the state after cessation of heroin administration, and tolerance gradually diminishes. After complete withdrawal, the same lethal dosages apply as in nontolerant individuals.

Not all opioids have exactly the same effects. For substitution purposes, the following criteria apply: longer duration of action than heroin, less euphoria and oral administration. These characteristics make the synthetic compound methadone the preferred substitution opioid.

3 Indication

Substitution treatment was developed as an alternative if abstinence-oriented treatment was not available, not feasible or had repeatedly failed. When Dole and Nyswander introduced methadone substitution in 1965, they set up strict entry criteria: minimal age of 21 years, minimal duration of heroin dependence of 4 years, no serious psychopathology, no problems with alcohol or other substances of abuse, and multiple failures in other treatment approaches. As polydrug use and polydrug dependence became more frequent, it was unrealistic to exclude these patients from substitution treatment. The human immunodeficiency virus (HIV) epidemic, the rapid increase of acquired immunodeficiency syndrome (AIDS) cases among i.v. drug users and the transmission of HIV infections to their sexual partners signaled such a threat to public health that substitution treatment became an option to reduce such risks and the access to treatment was facilitated.

At present, the following general guidelines indicate substitution therapy:

- Confirmed opiate dependence (following the criteria of ICD-10 or DSM-IV)
- Informed consent to accept the conditions of substitution treatment
- No contraindication

Further conditions also have to be considered:

- Detailed history (biography, health and social history, history of addictive behavior, including consumption patterns, consequences of substance use, former treatment attempts etc.)
- Present living situation (housing, occupation, income, network, perspectives)
- Complementary medical and/or psychiatric examination regarding indicators for a comorbidity (not in order to exclude such patients, but in the interests of comprehensive treatment planning)
- Extensive information on all available treatment options and advise on the most appropriate option for the situation and the problem at hand
- Shared identification of treatment goals (detoxification, interim substitution, mid- or long-term substitution)
- Identification of the most appropriate substitution medication
- Discussion and identification of the treatment conditions and modalities, preferably using a written informed consent statement

Detoxification using substitution medication with a reduction scheme relieves withdrawal symptoms, but eventually prolongs the withdrawal syndrome (especially when using long-acting medications such as methadone).

Interim substitution and other forms of short-term substitution treatment may be helpful in specific situations, such as pregnancy, urgent hospitalization (which would not be accepted without substitution), and physical conditions necessitating an interruption of risky injections. The duration of interim substitution is determined by the nature of the specific situation.

Long-term substitution is to be considered if opiate dependence is already chronic and if abstinence-oriented treatment is not accepted (e.g. in order for the patient not to lose his or her job or if a somatic or psychiatric condition impedes participation in a residential program).

General conditions such as minimal age, duration of dependence, and number of previous treatments that have failed are no longer considered as exclusion criteria. However, substitution treatment must be considered carefully if dependence is a recent development, especially in minors.

The duration of long-term substitution should not be limited administratively. Routine time limits that do not respect the individual situation increase the risk for relapse after the termination of treatment (Ward et al. 1998).

Substitution treatment is not an emergency treatment, with the exception of some physical conditions such as morbus embolicus or myocardial infection. However, the assessment should not take too much time, in order not to undermine positive motivation for treatment. It has been shown that a 2-week time delay results in significant patient dropout, and this does not only involve those patients with a poor prognosis (Maddux 1995).

When the initial prognosis for the feasibility of substitution treatment is unclear, the indication may be limited to a trial period, delaying the definite indication for 3 months, for example. It is recommended that the indication be reevaluated periodically, at least every 2 years.

The indication for substitution treatment should be made by a physician who has adequate experience in the management of heroin dependents and who is familiar with both abstinence-oriented and substitution treatment. In some countries, the right to determine the indication is reserved for physicians with adequate experience or those who receive continued education on the problems and modalities of addiction treatment.

Absolute contraindications are not known (except failing opiate dependence). Relative contraindications are as follows: acute hepatic conditions, severe hepatic insufficiency (risk of overdose from retarded metabolism), respiratory deficiency (risk of oxygen deficit due to respiratory depression from opioids), individual oversensitivity to a specific opioid, age under 18 years. Partial detoxification is recommended prior to starting substitution treatment if the patient is dependent on alcohol or benzodiazepines (risk of a respiratory deficiency due to combined action). Psychotic and otherwise mentally impaired patients (e.g. patients with AIDS encephalitis) will need careful assessment with regard to their coping ability for substitution treatment, possibly in collaboration with relatives or other significant individuals. Pregnancy is not a contraindication for substitution treatment.

4 Implementation

Substitution treatment is not without risks. However, such risks can be minimized by taking appropriate measures (Table 1).

4.1 Program of Care

Ancillary care is of the utmost importance for the outcome of substitution treatment. It must be tailored to the individual needs of the patient. Comparative

Table 1. Risks in substitution treatment

Risks	Guidelines on how to avoid risks
Primary dependence	Opiate dependence established before starting treatment
Weakening of motivation for abstinence	Periodic assessment of motivation for abstinence
Simultaneous use of illegal heroin	Urine tests; adequate dosage of substitution drug
Simultaneous use of other dependence-producing substances	Urine tests; restrictive prescription of dependence-producing medication (benzodiazepines)
Passing the substitution drug to others; overdose	Visually monitored intake

analysis of methadone programs has demonstrated that a comprehensive care program allows for better retention rates and better outcome (Ball and Ross 1991; McLellan et al. 1993). A program answering the needs for medical treatment and for social and rehabilitative measures is just as important as the prescription of a substitution medication. Clearly formulated agreements between the doctor and the patient help to prevent and to manage problematic situations. Such an agreement should in some cases include other parties concerned (psychotherapist, social worker, pharmacist, parole officer, partner).

Psychosocial care includes advice in financial and legal matters, help in job placement or job seeking, in housing problems, and in all kinds of crisis situations. It includes advice for spouses or relatives and contact with other agencies. Psychosocial care must adapt to changing situations in the course of treatment. The same holds for the frequency of consultations: during the initial period and in situations of crisis, more frequent and more intensive consultation is needed. Consultation mainly focuses on improving living conditions and lifestyle, provides therapeutic support, and monitors the patient's physical and psychological health.

Psychotherapy may enhance the treatment process considerably. A specific indication is needed which takes into consideration the individual condition and the suitability of a given psychotherapeutic method, be it individual or systemic treatment. When carefully chosen, psychotherapy may improve the outcome of substitution treatment (Woody et al. 1995). Group therapy is an integrated element in many substitution treatment programs; the need for individual indications is less important than the need to assess possible contraindications (e.g. in psychotic or severely depressive patients).

Good psychosocial care requires professional competence in various fields. It is therefore often provided on an interdisciplinary basis, in a multiprofessional team, or in collaboration between physician and a specialized consultation agency. The medical responsibility lies with the prescribing physician. Coordination of the various care elements by case management is recommended.

Psychosocial care must also consider the patient's motivation for treatment and readiness for change. Forced consultations may lead to treatment being broken off completely, especially when the patient is healthy and well integrated and sees no reason for regular intensive care. Indeed, substitution treatment in well-integrated patients was found to reduce illegal drug consumption and its risks even without much ancillary care (Raschke 1994). On the other hand, many patients need motivational support in order to accept treatment and care. Motivation techniques have been developed and implemented primarily in the United States (Simpson et al. 1997).

Examples include reward systems (bonuses, reduction of monitoring, take-home privileges for methadone) for compliant behavior, such as regular participation in group meetings and individual consultation sessions (Chutuape et al. 1998). Another technique focuses on enhancing problem awareness by linking individual problems to addictive behavior (node-link mapping technique; Dees et al. 1997).

Some experiments with low-threshold methadone prescription in the Netherlands and in Switzerland have demonstrated the feasibility of a safe prescription practice without ancillary psychosocial care if controlled daily methadone intake is provided and measures against overdosing are taken. When offered to heroin addicts who are reluctant to engage in a care program, results seem to be doubtful, although transition to a well-structured methadone program is encouraged. In addition, a reduction of infection risks in a low-threshold regime has not been shown (Ward et al. 1998). The regime seems to be better justified in patients without major health or social deficits. There are no studies on the mid- and long-term outcome at present.

4.2 Management and Treatment of Comorbidity

Medical conditions that preceded drug dependence or have emerged as their consequence must be properly assessed in order to receive adequate treatment by the prescribing physician or the patient's family doctor or specialist. Two problem areas have to be kept in mind:

- Interactions between substitution medication and other medications
- Interference between the various therapeutic regimes if they are not clearly defined and coordinated, leading to irritation and inviting manipulative behavior

The main somatic conditions associated with heroin dependence are a range of infectious diseases, malnutrition, caries etc. When adequately treated during a substitution regime, they may improve rather quickly, due to the discontinuation of risky injections and an improvement in living conditions. The immune system improves, and the risk of an HIV or hepatitis infection diminishes (Ward et al. 1998).

Interactions between medications must be considered. Tuberculostatic drugs such as rifampicin and rifabutin, antiepileptic drugs such as phenytoin, and certain barbiturates and benzodiazepines, for example, enhance the breakdown of methadone by enzyme induction (P450 oxidase system). This may be compensated for by increasing the frequency of oral methadone intake, not by increasing the dosage. On the other hand, chinidine, clozapine, beta blockers, and anti-arrhythmic agents may impede breakdown through competitive inhibition, thereby increasing the risk of an overdose (for an overview, see Ward et al. 1998).

Psychiatric conditions associated with heroin dependence (dual diagnoses) include (in order of frequency of prevalence) depressive disorders, anxiety disorders, post-traumatic stress syndrome, personality disorders, and schizophrenic psychosis. They may be a consequence of dependence; if they precede dependence, heroin use may be understood as a form of self-medication. Stress may be an underlying factor contributing to both heroin use and the psychiatric disorder. The psychiatric symptoms can be improved in the course of substitution treatment, especially if the addictive behavior has contributed to its manifestation. Often, however, they need to be treated according to the specific diagnosis. The psychological status should be reassessed when the methadone dosage has become stable; if psychiatric symptoms persist, specific treatment has to be initiated (Ward et al. 1998). Patients with schizophrenic symptoms need antipsychotic medication and a well-structured regime without the confrontation often used in the treatment of substance dependence. There are no general rules for the management of opiate dependents with personality disorders, but a good therapeutic climate and well-trained therapists are essential.

Drug interactions, such as the competitive inhibition of methadone metabolism by tricyclic and tetracyclic antidepressants, serotinin reuptake inhibitors, and some neuroleptics, must be taken into account (Ward et al. 1998).

4.3 Polydrug Dependence

Simultaneous dependence on more than one drug is a major impediment to successful substitution treatment. It is an obstacle to implementing positive changes in lifestyle, and it increases risks of respiratory depression, on the basis of synergistic pharmacologic action. Death during methadone substitution treatment is mainly due to an additional intake of alcohol, barbiturates, or benzodiazepines. The risk of hepatitis or HIV infection is increased in patients who are also dependent on cocaine (especially if cocaine is injected, but also if crack is used in conjunction with unsafe sex practices). Continued contact with the illegal drug market with the temptation and risks this entails is another negative consequence of cocaine dependence.

Polydrug use is frequently reduced in methadone substitution treatment if dosage and psychosocial care are adequate. This effect may be enhanced through behavioral therapy measures (bonuses if urine samples are clean, reversal of privileges if they are not). Polydrug dependence may require for additional measures. Partial detoxification treatment or drug-specific combination treatments are recommended. Sanctions, however, especially exclusion from treatment after a given number of positive urine tests, increase the risks of relapse and mortality and are not recommended (Ward et al. 1998).

In order to reduce cocaine consumption, an antidepressant medication may be prescribed, especially in the presence of depressive symptoms. A positive outcome was reported for imipramine, while controlled studies with desipramine, amantadine, and bromocriptine had no effect on cocaine use (for an overview, see Mattick et al. 1998). Anticraving medication (acamprosate) has not yet been systematically tested for its efficacy against cocaine use.

Alcohol dependence in opiate addicts has attracted less attention from therapists and researchers than cocaine dependence has. However, alcohol dependence makes a methadone regime difficult, and it increases the mortality risk in treated and untreated heroin-dependent patients. It may persist even in cases of

successfully treated heroin dependence and increases the risk of subsequent relapse. All therapeutic measures developed for the treatment of alcohol dependence and for hazardous alcohol use can be applied during methadone substitution treatment. These include early outpatient intervention techniques, aversive therapies, participation in self-help groups, and other relapse prevention methods. As an aversive medication, disulfiram shortens the metabolism of methadone and can result in a higher frequency of intake.

Benzodiazepine dependence and dependence on other sedatives is a frequent and serious complication in heroin addicts. Most prevalent is the use of flunitrazepam. An adequate methadone dose can reduce its use.

4.4 Control Measures

An essential element, especially to be implemented during the initial phase and during crisis periods, and possibly during the entire treatment, is the visually monitored intake of methadone. The daily controlled intake must be combined with reliable identification of each individual client. Controlled intake may not be feasible where the service or the pharmacy is closed on weekends; in this case, one or two daily doses have to be taken home on Fridays.

At a later stage, when dosages and the living conditions of a patient have become stable, a take-home policy for a couple of days may be applied.

Take-home policies and the frequency of consultations do not follow uniform rules; they depend on the attitude and the experience of the prescribing physician, on official regulations, and on local traditions. The advantages of a take-home policy include the fact that patients who have become stable are more likely to remain in the program (Rhoades et al. 1998) and show better compliance (Stitzer et al. 1992). It also facilitates the vocational rehabilitation of patients.

Disadvantages include the additional risks for patients who are not sufficiently stable, as they may lose or sell their methadone and replace it by illegal injections, or they may use more than one daily dose at once. In particular, the mortality risk is increased in such cases, due to overdosing. In addition, prescribed methadone may be found on the illegal market and leads to substitution treatment being discredited. All these disadvantages and risks can be avoided if only those patients who are stable enough are allowed to take doses home (Senay et al. 1994).

Take-away doses of methadone should not be replaced if claimed to be lost. Administration through third parties is not recommended, with the exception of reliable relatives or other significant individuals well known to the prescribing physician, and only if the patient is physically unable to visit the methadone clinic regularly.

Urine tests are traditionally another essential element of a methadone program. In many instances, they are legally required, while in some countries they are merely recommended. Urine tests are carried out without prior warning. Visual monitoring, checking of the temperature, or testing for substances added to the oral methadone suspension are utilized in order to minimize urine substitution. The rationale for urine tests is the recognition of nonprescribed substances and evidence of the intake of prescribed methadone.

The high costs of urine tests and the unpopularity of this type of monitoring among staff and patients alike have encouraged research on the usefulness of such measures. It has been shown that frequent and regular tests are not necessary. Infrequent and irregular tests are not linked to an increase in illegal drug use or to less favorable outcomes (Ward et al. 1998).

Urine tests still have a role to play in everyday practice. Their frequency is highly variable. Tests should not be abandoned completely, even in patients with good compliance, because an objectively established drug-free status may considerably reinforce the patient's self-esteem and self-control. Frequent tests are performed in the initial treatment phase and in crisis situations.

Heroin and other opioids and sometimes also benzodiazepines, barbiturates, and cocaine are tested for.

If illegal or nonprescribed substances are detected, this must be discussed openly with the patient. Circumstances connected to the use of such substances, ensuing risks, motivation for better compliance, and measures to be expected if noncompliant behavior continues are the main topics of discussion. The results have to be documented in the case history. Without this process, urine tests have little value.

Urine tests are also used for documenting the evolution of the patient's behavior, for administrative or legal purposes, e.g. in order to obtain a driver's license. Informed consent that relevant information be passed on to the authorities concerned must be assured.

4.5 Duration and Termination of Treatment

The duration of substitution treatments must be determined for each individual patient. Rigid and schematic limitations of their duration are counterproductive and increase the risk of relapse (Caplehorn 1994).

Improvement in health status and reduction in illegal heroin use and in heroin-related delinquency

are observed within months. It takes more time to build up new social contacts and to become reintegrated into the labor market. It usually takes 1–2 years for living conditions to become stable and behavior changes to become permanent. In many instances, further improvements are still made after a few years. However, after 3–4 years, the chances of improvements being made tend to become lower, and there may be a risk of secondary deterioration and resignation.

Many patients are impatient and ask for an early termination of substitution treatment. They want to prove, to themselves and to others, that they are perfectly well prepared for a drug-free life. It is an essential therapeutic task to motivate these patients to accept prolonged treatment, comparable to that needed for prolonged antipsychotic medication for schizophrenia or aversive treatment with disulfiram for alcohol dependence. Less experienced doctors often consider short substitution treatment to be preferable, as they are unaware of the risk of relapse after short treatment and of negative effects on the patient's motivation to resume treatment again if it has failed once. On the other hand, some patients cannot find the courage to terminate substitution treatment after years of satisfactory consolidation and reintegration. These patients may need encouragement to gradually terminate treatment. There are also patients who are unable to mature or grow out of a dependence-prone state of mind and who may need substitution treatment for many years. The basis for such a condition has not been clarified yet.

The termination of treatment must be well planned with the patient and involves clearly defined steps. This includes timing and extent of dose reductions, modality of final detoxification, continued aftercare, management of later crisis situations, and options for restarting substitution treatment if needed.

Treatment should be discontinued if the patient does not comply with treatment plans, takes methadone on an irregular basis, behaves aggressively toward staff, or is selling prescribed substances or engages in drug trafficking. Continued consumption of nonprescribed drugs involving a risk of overdose is a reason for discontinuing treatment. In any case, such a far-reaching decision should be guided by weighing up the risks of keeping the patient in treatment against those of discontinuing it.

Whenever and however substitution treatment is terminated, care should be taken to keep the patient's options open for returning to treatment in the same program or in another service. Coming back to treatment should not be labeled as a failure for the patient or for the therapist. It has been shown that returning to treatment is often a real chance and may lead to better results.

5
Special Problems

5.1
Driving

The ability and permission to drive a motor vehicle is important for practical reasons and especially for vocational rehabilitation. Many patients have lost their driver's license because of their heroin dependence. In order to regain their license, they must present no unusual risks to road safety. Such unusual risks can be excluded on the basis of an assessment of their health status, of documented drug-free urine samples over a defined period of time, and of the patient's general behavior. Assessment is made independently, not by the prescribing physician, but the responsible agency usually consults him or her for observations and prognosis.

The individual assessment of the patient's driving ability does not use uniform criteria. As a rule, a minimal duration of 6–12 months on substitution medication at a stable dosage is required. Other requirements are adequate physical and psychological health, stable housing conditions, absence of delinquent and otherwise irresponsible behavior, and abstinence from illegal and nonprescribed substances (documented by urine tests).

Driving ability is not impaired by substitution medication if the dosage is stable and taken regularly and full opiate tolerance is reached. In the case of oral methadone, attention, coordination, and reactivity are only minimally impaired under steady-state conditions (Seidenberg and Honegger 1998). The use of short-acting substitution drugs is more problematic.

An overdose of methadone may lead to impaired driving if opiate tolerance is not fully developed. Underdosing on the other hand encourages collateral use of other substances that may have a negative impact on driving.

In case of relapse into illegal or nonprescribed substance use, the treating physician must determine whether the patient is still able to drive or whether he or she has to report the patient's inability to the authority concerned.

5.2
Vacations

A normal lifestyle includes the possibility of going on holiday. Social integration and engaging in social relations with friends and partners is difficult if someone is not allowed to travel to a holiday resort or to another country. The wish to go on holiday is often encountered during substitution treatment.

The treating physician must decide whether it is acceptable to hand out substitution drugs in the amounts needed for a vacation on the basis of the individual circumstances (stabilization of dosage and living circumstances, reliability of patient, duration of vacation).

In many places and countries, it is possible to arrange for controlled intake of substitution drugs, but the options have to be explored and there may be some administrative problems to be dealt with. The national laws and regulations concerned must be respected, and the availability of the drug must be checked. As a rule, in countries where methadone substitution treatment is an accepted therapeutic modality, such an arrangement can be made. A medical certificate confirming that the patient is undergoing substitution treatment can be useful. Taking drugs into another country is not recommended, as being detected at the border may cause serious problems.

5.3 Prison

In most countries where heroin dependence is a problem, there are major difficulties in the prison milieu both for dependent inmates and for the institutions and their staff. One difficulty concerns the withdrawal syndromes of newly admitted dependents. Another frequent difficulty is continued opiate use while imprisoned; this presents specific health problems, as needle and syringe sharing may lead to blood-borne infections with HIV and hepatitis. The occasional overdose deaths are also a problem.

In some countries, ongoing substitution treatment can be continued after imprisonment, and in others such treatment can even be started while the patient is in prison. The practice of using methadone in the detoxification of inmates is more widespread (Dolan et al. 1998). Substitution treatment is practiced as a special regime for some individuals or else in special units. In any case, such treatment should be voluntary. Ideally, inmates should be able to choose between substitution and drug-free treatment. Evaluation of such practices is limited at present.

While on remand, withdrawal symptoms may impair the prisoner's fitness for examination. This is not the case if the prisoner is under substitution treatment with adequate doses; the suicide risk is also reduced.

Ideally, substitution treatment in prison should be based on consultation with an external specialized agency which has already treated the patient or is going to take over after his or her release. Organizing proper aftercare is essential in preventing relapse.

5.4 Hospital

Ongoing substitution treatment should be continued if a patient is temporarily hospitalized, be it in a somatic or a psychiatric hospital, unless there are any medical contraindications. If this is not guaranteed, many patients in need of hospital treatment will not agree to being hospitalized or leave the hospital before treatment of their illness has a chance to improve their situation.

Substitution treatment should be initiated if hospitalization would otherwise not be feasible. Hospital physicians must be competent to carry out such a program, and clear rules on the regime must exist which are known to staff and to the patient concerned. Successful substitution treatment is enhanced if good contact is established with external agencies that have already treated the patient or will be involved in his or her aftercare.

5.5 Pregnancy

A patient's health status often improves considerably in the course of substitution treatments. Women patients will begin to menstruate within months. If not adequately instructed or willing to use contraceptive measures, they may become pregnant unintentionally, especially since they did not conceive while on illegal drugs even without contraceptive measures.

If a patient becomes pregnant, the following options must be considered:

- If the woman wants to terminate substitution treatment, the risks of withdrawal syndromes must be discussed, as spontaneous abortion may occur (Finnegan et al. 1991). In any case, the dosage has to be carefully reduced, preferably between weeks 14 and 32 in order to avoid abortion or premature delivery (Ward et al. 1998). Benzodiazepines should not be used to avoid withdrawal symptoms. There is also a major risk of relapse into illegal heroin use after detoxification (Finnegan et al. 1991), and retention in a follow-up drug-free treatment is much lower than in substitution (Svikis et al. 1997).
- If the woman decides to continue substitution treatment, the advantages and potential disadvantages must also be considered. In comparison with illegal heroin (Finnegan et al. 1991), methadone has less risks for the child in terms of birth weight and chromosome damage. A disadvantage is the withdrawal syndrome in the newborn child and the risk

for mid- and long-term effects on the child's development – risks that have not yet been sufficiently explored. Neuropsychological impairments and behavior disorders cannot be ruled out, although milieu factors such as the mother's lifestyle and living conditions seem to have more impact on development than pharmacologic factors (Lifschitz et al. 1985). It is important to note that doses of methadone should not be minimal but sufficient to prevent withdrawal symptoms and thereby collateral illegal or nonprescribed use of drugs. Daily doses should be split in two to be taken at an appropriate interval in order to facilitate stable blood levels and to avoid peak concentrations. After childbirth, methadone substitution is compatible with breastfeeding; problems arise only in cases of harmful use of alcohol or other drugs. Experience with substitution drugs other than methadone is still insufficient.

In conclusion, continuing methadone substitution during pregnancy has more advantages than disadvantages, but this should be determined in each individual case. However, methadone treatment should be considered only on the condition that there is no parallel use of illegal heroin. The main reasons for starting substitution during pregnancy are to avoid the risks of illegal heroin use and to increase the chances for prenatal care.

6 Substances Used in Opiate Substitution

In some Asian countries (Laos, Pakistan) where opium used to be an important substance for both medical and leisure purposes, opium was made available for dependent individuals. These systems were abolished under international pressure by the Single Convention of 1961.

At present, the following substances are prescribed as substitutes for heroin:

- Methadone (oral, intravenous)
- LAAM (oral)
- Morphine (oral, intramuscular, intravenous)
- Codeine, dihydrocodeine (oral)
- Buprenorphine (oral, intramuscular, intravenous)
- Diacetylmorphine (heroin) (oral, intravenous)

6.1 Methadone

Oral methadone is by far the most frequently used substance in opiate substitution treatment. It is usually available as a mixture of the pharmacologically active l-form and the inactive d-form. Preparations containing only the d-form were used in Germany, but are increasingly rare. Intravenous application is traditionally a therapeutic option in England, where it is still an accepted modality (Metrebian et al. 1998). This modality was also tested in a Swiss study and found little acceptance, due to frequent and severe side effects with comparatively high doses (Uchtenhagen et al. 1999).

When given orally, methadone has a bioavailability of 90%–100%. Initial doses are determined by opiate tolerance in a given patient. They should be handled cautiously in order to avoid overdose. If tolerance is not clearly evident, doses should not exceed 30 mg/day, and in cases of evident tolerance 60 mg/day in two doses should not be exceeded. Dosage is increased following the observation of withdrawal symptoms. It is advisable to see the patient once or twice a day in order to prevent unpredictable complications. It is only rarely necessary to hospitalize a patient to determine the adequate dosage.

Maintenance doses must be individually determined. The recommended medium range is between 60 and 100 mg per day. The maximum dosages formulated for nontolerant individuals do not apply. The general rule is to prescribe doses that are sufficient to avoid withdrawal symptoms. Randomized trials have demonstrated that low doses of around 30 mg per day are more frequently followed by treatment discontinuation, by collateral use of illegal substances, and by less favorable outcomes than dosages of around 80 mg per day (for an overview, see Ward et al. 1998). On the other hand, a high dosage that is inappropriate for a particular individual may also encourage the use of nonprescribed benzodiazepine or illegal cocaine. If daily intake is interrupted, the dose has to be reduced in order to avoid overdose; for each day that methadone was not taken, a dosage reduction of 20% is recommended (Seidenberg and Honegger 1998).

No strict correlations have been found between dosage, plasma level, and treatment outcome. It should be noted that plasma levels and pharmacologic action can be influenced by intervening illness and by interactions with other medications (see Sect. 4.2) and necessitate dosage adjustment. It should also be remembered that some individuals metabolize methadone more quickly than average and therefore need higher doses.

Before ending the treatment, a dosage reduction is recommended, usually involving a weekly reduction of the daily dose by 5–10 mg. The final reduction may use even smaller steps or else is made in a residential facility in order to avoid an early relapse. Many patients would prefer a quicker reduction, but in order to avoid relapse each individual case must be examined.

The most frequent clinical side effects of methadone

are constipation, sweating, sleep disorders, and a reduced libido. They can often be managed by dose adjustments and only rarely lead to premature termination of treatment. Side effects only detected in laboratory testing are clinically irrelevant (for an overview, see Seidenberg and Honegger 1998). A comprehensive presentation of the pharmacologic and toxicologic aspects is provided in Cooper et al. (1983).

Death during methadone treatment occurs in the initial phase if there is insufficient tolerance or if initial doses are too high. Death in the course of treatment is usually due to a combination of methadone and other substances (benzodiazepines, barbiturates, alcohol), while methadone alone is rarely found in autopsies of patients who have died of an overdose (Ward et al. 1998; Scott et al. 1999).

6.2 Levo-alpha Acetylmethadol

LAAM is preferred for the long-term substitution treatment of patients with stable doses because of its longer half-life (2.6 days, active metabolites up to 4 days). Intake frequency can be reduced to three times a week, with a dose of 100 mg each time. This has practical advantages, and patients are less restricted in their mobility.

On the other hand, the longer duration of pharmacologic action also means less flexibility in controlling dose effects. The risk of overdosing may be higher. In cases of increased liver function, the rate of metabolism and the level of active metabolites are also increased (Kreek 1996). LAAM should be considered in patients who are already well stabilized on oral methadone. The recommended doses are 1.2–1.3 times higher than the methadone doses (Kreek 1996). Switching from methadone to LAAM may lead to temporary withdrawal symptoms due to its retarded action. Once a stable dosage is reached, LAAM is regarded to be as safe for patients as oral methadone (Tennant et al. 1986).

As a consequence of the retarded action, there is no subjective perception of a rapid onset (flush). This may help to explain why in randomized trials higher dropout rates were found in the LAAM groups than in the methadone groups (Savage et al. 1976; Ling et al. 1980). However, these findings were not confirmed in later studies. An overview of the clinical results indicates a satisfactory therapeutic outcome (Tennant et al. 1986).

6.3 Morphine

Morphine has no important place in substitution treatment. The oral slow-release preparation MST Continus (morphini sulfas pentahydricus) is used as an alternative to oral methadone in patients who experience unpleasant side effects. The recommended dose is 600–900 mg per day (Seidenberg and Honegger 1998).

When intravenously applied, morphine provokes histamine-like local or generalized side effects. The Amsterdam pilot project prescribing intravenous morphine to a group of highly problematic heroin addicts was discontinued, in part on the basis of such side effects (van Brussel 1995). The Swiss project on medical prescription of narcotics also included intravenous morphine, which proved to be poorly accepted due to frequent and occasionally serious side effects (Uchtenhagen et al. 1999).

6.4 Codeine

Between 5% and 20% of codeine is metabolized into morphine. Around 10% of the general population do not have the isoenzyme needed for this process (cytochrome P450 2D6). In these individuals, codeine does not have the intended pharmacologic effect.

The action of codeine is easier to control than methadone, as codeine has a shorter half-life. It also has less potential for causing dependence. It can be reduced and discontinued with fewer problems. A disadvantage is its irritating effects on the stomach. Its usage is limited to patients who need low dosages of substitution drugs.

Clinical experience with codeine was obtained in Germany in particular when substitution with methadone was discouraged and restricted. Therapeutic results have been described as being equivalent to those in methadone substitution treatment (Verthein et al. 1996).

6.5 Buprenorphine

Buprenorphine is a synthetic drug derived from thebaine. It has mixed agonist–antagonist properties. With increased doses, the opioid-like effects are reduced, probably due to the increased antagonist effects. When applied sublingually, the maximum plasma levels are reached after 2–4 h. Half-life values for elimination are 3–5 h. Duration of action increases with dosage. Respiratory depression is minimal, and the mortality risk is comparatively low (Walsh et al. 1994). Bioavailability is much lower when applied orally as compared to sublingual application. Dose reduction and termination of treatment are followed by comparatively few withdrawal symptoms (Mattick et al. 1998).

These findings have led to an increased interest in buprenorphine as an alternative to oral methadone. It is the substitution drug of choice in France, where every physician is entitled to prescribe it. The most frequently used dose is 8 mg per day. The highest dose without an increased risk is reported to be 32 mg per day (Mattick et al. 1998). Prescribing a double dose every second day therefore seems feasible without disadvantages.

Therapeutic outcome is considered to be equivalent to oral methadone treatment, measured by retention rates and by a reduction of illegal substance use. Questions remain in patients who need higher doses of substitution drugs (Mattick et al. 1998). Randomized trials resulted in divergent findings. These may partially be explained by the use of rigid dose schemes and insufficient doses (for an overview, see Mattick et al. 1998).

Misuse of buprenorphine was reported in New Zealand, Australia, and Scotland. Attempts to prevent such misuse have been made by prescribing a combination of buprenorphine and naloxone (Robinson et al. 1993).

6.6 Diacetylmorphine (Heroin)

Diacetylmorphine is a highly effective analgesic with a short duration of 4–5 h and an elimination half-life of half an hour. Monoacetylmorphine and morphine are the active metabolites. If applied intravenously, the plasma half-life is 20–30 min and the duration of effect 3–4 h. It passes the blood–brain barrier easily.

Substitution of street heroin by pharmaceutical heroin has been practiced in England for decades. Special clinics were set up in the late 1960s in urban areas, especially in London. Heroin was handed out by pharmacies on the basis of prescriptions issued by the clinics. Since the early 1970s, this practice has been continuously reduced in favor of prescribing oral and, less frequently, injectable methadone (Strang and Gossop 1996).

In a national Swiss cohort study, injectable heroin is being prescribed to chronic patients in whom other treatments have failed. The individual doses are on average 471 mg per day, administered in two to four intravenous injections under visual control in the clinics. Injectable heroin is often combined with oral methadone in order to reduce the number of injections per day. This facilitates the vocational and social integration of patients, as their day is not interrupted. Side effects include histamine-like reactions as is the case with morphine. In some cases, cerebral seizures are observed following the injection in patients with or without previous episodes of epilepsy. Patients who are unable to inject because their veins are obliterated or inflamed have been treated in a pilot study with heroin-impregnated tobacco-free cigarettes. However, because of their low bioavailability of only 10%, these cigarettes were replaced by slow-release heroin tablets (Uchtenhagen et al. 1999).

In the Netherlands, a randomized trial of prescribing oral methadone with and without injectable or inhalable heroin has been started for a predefined period of time (van Brussel 1995). Results are expected in 2001.

6.7 Criteria for the Choice of Substitution Drugs

The present evidence does not allow well-founded indication criteria to be formulated. Oral methadone is still the standard, and other substances are mainly used if methadone cannot be used for some reason or if it has proved to be unsatisfactory. The only exception thus far is buprenorphine, which is preferred by some for substitution treatment of patients who need fairly low doses, of adolescents and young adults, and of noninjecting heroin dependents.

7 Outcome

Most outcome studies concern heroin substitution by oral methadone. Overviews of research results have been presented by the National Institute of Drug Abuse (NIDA) in the United States (Cooper et al. 1983), by the WHO (Senay and Uchtenhagen 1990), and in an Australian reference book (Ward et al. 1998).

Effects of long-term substitution treatment using oral methadone were examined in a large number of studies. The early randomized trials and later clinical studies (for an overview, see Ward et al. 1998) have documented the following positive results: reduction of illegal drug use, reduction of needle/syringe sharing, reduction in morbidity (especially in blood-borne infections) and mortality, improvements in health status, reduction of contacts with drug users, reduction of drug-related delinquency, reintegration into the labor market, and an overall normalization of lifestyle. These positive effects have been confirmed by the WHO (1998).

The "classical" evaluation studies using a control group design have documented significant reductions in daily heroin consumption and in arrest rates. Naturalistic studies without control groups often give inferior results; this is explained by the more selective procedures for participants and by a better treatment

and care quality in the randomized studies. In addition, in earlier years, the rules for running methadone programs were rigid and strictly applied.

Comparative cohort studies without a control group design on methadone substitution treatment, drug-free outpatient treatment, and residential treatment have been implemented on a large scale especially in the United States; these include the Drug Abuse Reporting Program (DARP; Simpson and Sells 1990), the Treatment Outcome Prospective Study (TOPS; Hubbard et al. 1989), and the Drug Abuse Treatment Outcome Study (DATOS; Simpson and Curry 1997). Similar national cohort studies have been performed in Switzerland (Uchtenhagen and Zimmer-Höfler 1985) and recently in England (National Treatment Outcome Research Study, NTORS; Gossop 1998). The main findings of these studies are that methadone substitution treatment attracts and serves many more patients than drug-free treatments, and fewer patients drop out of the program. There are no consistent findings on differences in therapeutic effects as measured by a reduction in heroin use and in drug-related delinquency. Residential drug-free long-term treatment (mostly in therapeutic communities) yield the best results in achieving a socially integrated drug-free lifestyle. The least favorable results were found in outpatient drug-free treatment.

Recent research focused on comparisons of various types of methadone programs with the aim of identifying predictors of outcome and improving the programs. Comparative analysis has identified insufficient methadone doses as a major reason for therapeutic failure and especially for continued heroin use. Retention and reduction in heroin use could be predicted on the basis of dosage policy. Other relevant factors are the availability of ancillary care services (Ball and Ross 1991; McLellan et al. 1993) and a flexible duration of treatment. Bureaucratic limitation of treatment duration proved to be counterproductive in terms of relapse rates (Caplehorn 1994). It is worth mentioning that improvements may still occur after years on substitution treatment. Factors which impair the prognosis are polydrug dependence, personality disorders, and non-drug-related early delinquency.

New research also concerns the prescription of pharmaceutical heroin as a substitution for street heroin. There has been little research into the British practice. The one randomized study that has been carried out found lower dropout rates in the group receiving injectable heroin, but no clear difference in therapeutic results as compared to oral methadone (Hartnoll et al. 1980). A recent study documents the feasibility of prescribing injectable heroin and methadone (Metrebian et al. 1998). An overview of the programs prescribing injectable heroin or morphine was presented by Mino (1990).

A national Swiss cohort study was started in 1994 comparing groups receiving injectable morphine, methadone, or heroin in the framework of a comprehensive assessment and care program. Entry criteria included multiple failed previous treatment attempts in drug-free programs or methadone substitution. Short randomized periods were used in order to identify differences in pharmacologic effects and side effects. The overall cohort was compared with a cohort of new entries into oral methadone programs using the same research protocol. A total of 1035 patients were enrolled in the cohort receiving injectable drugs, with an average age of 30.8 and an average duration of heroin use of 10.5 years.

The main results of the study were as follows (Uchtenhagen et al. 1999):

- In the randomized substudies, the heroin group was found to have better acceptability (recruitment rates), retention rates, and reduction in illegal drug use.
- Injectable heroin allows for stable dosages as well as oral medication.
- In the total cohort, there was a rapid improvement in health status and a reduction in illegal and nonprescribed substance use, in contacts with drug users, and in drug-related delinquency.
- There was no fatal overdose from prescribed substances.
- Illegal heroin and cocaine use was reduced to a significantly greater extent in comparison to the oral methadone cohort.
- Heroin prescription programs met with good acceptance from authorities and in the general population, as shown by repeated local and national referenda.

The research protocol, its implementation, and the results were examined by an international expert committee appointed by the WHO. The positive results for patients and the feasibility and safety of the programs were confirmed; however, the specific role of heroin prescription could not be identified due to the lack of consistent randomization. The committee made recommendations for further research in the field (Ali et al. 1998).

Based on the Swiss study, on a randomized trial which started in the Netherlands (van Ree 1998), and on the intention of the German Government to start another randomized trial, the international debate over the prescribing of pharmaceutical heroin has intensified. Among those engaging in such projects, the main objective is to supplement the treatment modalities already available by an additional one for those who have failed other treatments. Oral methadone is still perceived to be the drug of choice for substitution purposes. The scientific debate focuses on the place of

heroin prescription within the range of treatment options (Bammer et al. 1999).

8 International Perspective

An overview on policy and practice in methadone substitution treatment in an international perspective was first prepared by the WHO (Arif and Westermeyer 1990). A more recent overview on the legislation, regulations, and practice of methadone substitution in 12 Member States of the European Union was commissioned by the European Commission (Farrell et al. 1996). The latter contains detailed descriptions on the basis of information provided by key informants.

Europe has seen a major increase in methadone substitution treatment and in the respective treatment programs since 1988 as a consequence of the HIV epidemic and especially as a consequence of an increased prevalence of AIDS cases among i.v. drug users and their sexual partners. The total amount of prescribed methadone increased in the various countries by 100%–500%. The level of consumption and the speed of increase differ considerably, however. At present, Spain shows the highest increase rates. The estimated total number of patients receiving methadone has increased fourfold over the past few years.

The number of treatment slots in EU countries in proportion to the total population is highest in England, Holland, Italy, and Spain. Between 10 and 200 slots per 100,000 inhabitants (aged 16–60) are available. There are differences regarding regional distribution and waiting lists.

Another worldwide survey was made on behalf of Health Canada (Ruel 1996), including information from 24 countries. In relation to population figures, methadone substitution was found to be most frequently available in Switzerland and Hong Kong, followed by Australia, Belgium, and the Netherlands.

The international debate on substitution treatment has changed over time. Strict refusal to prescribe substitution drugs on the basis of ideological arguments has become quite rare, and awareness of the requirement for safe and effective implementation has increased. The WHO produced an expert report on the options for the use of substitution treatment in detoxification and maintenance, listing advantages and disadvantages (WHO 1989). According to this report, the most important advantages were the high potential to recruit dependents into treatment, the comparatively low dropout rates, and the low selectivity, while (avoidable) disadvantages were a diversion of prescribed drugs into the illegal market, endless duration of treatment, and the costs involved.

The specific risks of prescribing a narcotic substance as a substitute for heroin were already mentioned in the first treatment protocol (Dole and Nyswander 1965). The authors especially mentioned respiratory depression from overdose and deviation into the illegal market. The regulations in each case were respected in the first methadone programs and included the following: visually monitored intake of methadone, adequate medical and psychosocial care, and urine tests to check for nonprescribed substances. These regulations have been reduced or completely abandoned for economic reasons or in order to lower the threshold for patients. Especially after the arrival of the HIV epidemic, so-called low-threshold methadone substitution was introduced in some places. At present, substitution practices differ in so many respects that the report by the European Commission stated, as its most significant finding, a lack of binding treatment protocols, regulations, monitoring systems, and evaluation guidelines (Farrell et al. 1996). An overview of regulations in force by end of 1995 is provided in Table 2.

The most frequently applied rules are restrictions for handing out methadone on a take-out basis and checks for nonprescribed substance use by urine analysis. Less frequent rules include dose restrictions, special authorization needed for prescribing, and a central registration of patients receiving methadone. Only a minority of countries have detailed national treatment standards or recommendations for good substitution practice. Such provisions have been made in Switzerland, for example (Bundesamt für Gesundheit 1985), USA (Parrino 1991) and Germany (Bühringer et al. 1995).

9 Place of Substitution Treatment in a Therapeutic Network

The old debate on whether or not substitution treatment is acceptable has now been replaced by a debate on how substitution treatment can be implemented in a helpful way among other treatment approaches for heroin dependents.

Originally conceived as a treatment of second choice (in cases where drug-free treatment failed), substitution treatment is now one of the main pillars in the management of heroin dependence. More patients prefer substitution treatment compared to residential drug-free treatment, especially those who still have jobs and a satisfactory social network. Other groups

Table 2. Regulations for methadone substitution treatment (after Ruel 1996)

	Australia	Canada	Denmark	England	Finland	France	Germany	Hong Kong	Hungary	Israel	Italy	Mexico	Netherlands	New Zealand	Spain	Sweden	Switzerland	USA
Central registration of patients	x	x	x	x			x	x					x			x	x	
Urine controls mandatory	x	x	x	x	x	x	x	x		x	x	x	x	x	x	x	x	x
Maximum dose per day		x	x	x		x	x			x		x	x	x	x		x	x
Restricted take-out policy	x	x	x	x	x		x	x		x	x	x	x	x	x	x	x	x
Special authorization required	x	x			x		x		x	x		x		x	x	x	x	x
Maximum patient load per prescribing physician	x	x					x			x	x	x		x				
Maximum patient load per clinic	x									x				x		x		
Authorization for specialized clinics only					x				x	x		x			x	x		x
Authorization only for physicians specializing in addiction treatment					x				x	x		x			x	x		

who favor substitution are those who feel discriminated or exposed in the context of residential programs (e.g. young females, dual-diagnosis patients, patients from ethnic minorities with language problems) and also patients who have tried drug-free treatment without positive results. Wherever heroin dependence is a problem for public health and order, substitution treatment is one of the most effective measures to improve the situation of patients and to reduce the public problems.

Substitution treatment does not necessarily lead to reduced applications for drug-free treatment. The more patients receive substitution treatment, the more will eventually need detoxification treatment, and the more will look for drug-free treatment if they are not satisfied with substitution treatment. On the other hand, many who were unsuccessful in drug-free treatment can be well cared for in substitution treatment. In addition to a meaningful coexistence of therapeutic approaches, there are also programs which offer a mix of approaches: therapeutic communities accepting methadone patients, and methadone programs using methods developed in therapeutic communities especially for day programs or semi-residential programs.

A well-functioning integration of approaches and services in a regional network is feasible if there is the political will to support this tendency in order to reach a maximum number of patients. This requires education to be offered to all staff concerned in order to improve knowledge of the potential of all therapeutic options and to improve coordination and collaboration.

10 Unsolved Problems

One of the most frequent problems arising in substitution treatment is the simultaneous use of nonprescribed substances. In the forefront are benzodiazepines, cocaine, and alcohol. Polydrug use increases the mortality risk considerably. It also undermines all efforts to reorient the lifestyle of patients. New ways of dealing with polydrug use need to be developed and tested.

The high relapse rates which are sometimes found after termination of substitution treatment also constitute a problem. Relapse is more frequent after a short duration of treatment or if the duration is determined on the basis of administrative rules rather than the individual needs of patients. The majority of patients need adequate psychosocial and rehabilitative measures in order to prevent relapse, but this is often in conflict with the need to keep down treatment costs.

Most treatment failures seem to be due to insufficient treatment quality. This includes inadequate diagnostic assessment when starting treatment, deficits in psychosocial care, and deficient dosage. Many programs need improved conditions and increased resources for coping with these demands, but at the same time they suffer from reduced budgets. A major improvement would be the implementation of binding treatment standards, as is the case in other medical fields.

Ambivalence in attitudes towards substitution treatment also has other roots, however. There is no motivation to substitute a drug of abuse if use and abuse are mainly considered to be irresponsible behavior that must be punished or at least not gratified. There is also little motivation if the subjective expectations of users for positive effects of their treatment are ignored. An evidence-based policy looking for pragmatic approaches and the proper evaluation of such approaches is needed in order to overcome the problems outlined here.

11 References

Ali R, Auriacombe M, Casas M, Cottler L, Farrell M, Kleiber D, Kreuzer A, Ogborne A, Rehm J, Ward P (1998) Report of the external panel on the evaluation of the swiss scientific studies of medically prescribed narcotics to drug addicts. Report to the World Health Organization, Geneva

Arif A, Westermeyer J (1990) Methadone maintenance in the management of opioid dependence. An international review. Praeger, New York

Ball JC, Ross A (1991) The effectiveness of methadone maintenance treatment. Springer, Berlin Heidelberg New York

Bammer G, Dobler-Mikola A, Fleming PH, Strang J, Uchtenhagen A (1999) The heroin prescribing debate: integrating science and politics. Science 284: 1277–1278

Bühringer G, Gastpar M, Heinz W, Kovar KA, Ladewig D, Naber D, Täschner KL, Uchtenhagen A, Wanke K (1995) Methadon-Standards. Vorschläge zur Qualitätssicherung bei der Methadonsubstitutionim Rahmen der Behandlung von Drogenabhängigen. Enke, Stuttgart

Bundesamt für Gesundheit (1985) Methadonbericht. Suchtmittelersatz in der Behandlung Heroinabhängiger in der Schweiz. Bundesamt für Gesundheit, Bern

Caplehorn JR (1994) A comparison of abstinence-oriented and indefinite methadone maintenance treatment. Int J Addict 29: 1361–1375

Chutuape MA, Silverman K, Stitzer ML (1998) Survey assessment of methadone treatment services as reinforcers. Am J Drug Alcohol Abuse 24: 1–16

**Cooper JR, Altman F, Brown BS, Czechowicz D (1983) Research on the treatment of narcotic addiction. The state of the art. National Institute of Drug Abuse, Rockville

Dees SM, Dansereau DF, Simpson DD (1997) Mapping-enhanced drug abuse counseling: urinalysis results in the first year of methadone treatment. J Subst Abuse Treat 14: 45–54

Dolan K, Hall W, Wodak A (1998) The provision of methadone within prison settings. In: Ward J, Mattick RP, Hall W (eds) Methadone maintenance treatment and other opioid replacement therapies. Harwood, Amsterdam, pp 379–396

*Dole V, Nyswander M (1965) A medical treatment for diacetylmorphine (heroin) addiction. JAMA 193: 146–150

*Farrell M, Neeleman J, Gossop M, Griffiths P, Buning E, Finch E, Strang J (1996) Drug prevention. A review of the legislation, regulation and delivery of methadone in 12 member states of the European Union. Final report. Office for Official Publications of the European Communities, Brussels

Finnegan LP, Hagan T, Kaltenbach KA (1991) Scientific foundation of clinical practice: opiate use in pregnant women. Bull NY Acad Med 67: 223–239

Gossop M (1998) NTORS at one year. The National Treatment Outcome Research Study. Department of Health, London

Hartnoll R, Mitcheson M, Battersby A, Brown G, Ellis M, Fleming P, Hedley N (1980) Evaluation of heroin maintenance in a controlled trial. Arch Gen Psychiatr 37: 877–884

Hubbard RL, Marsden ME, Rachal JV, Haarwood HJ, Cavanaugh ER, Ginzburg HM (1989) Drug abuse treatment. A national study of effectiveness. University of North Carolina Press, Chapel Hill

Kreek MJ (1996) Long-term pharmacotherapy for opiate (primarily heroin) addiction: opioid agonists. In: Schuster CH, Kuhar MJ (eds) Pharmacological aspects of drug dependence. Toward an integrated neurobehavioral approach. Springer, Berlin Heidelberg New York, pp 487–562

Lifschitz MH, Wilson GS, Smith EO, Desmond MM (1985) Factors affecting head growth and intellectual function in children of drug addicts. Pediatrics 75: 269–274

Ling W, Klett JC, Gillis RC (1980) A cooperative clinical study of methadyl acetate. Arch Gen Psychiatry 37: 908–911

Maddux JF, Desmond DP, Esquivel M (1995) Rapid admission and retention on methadone. Am J Drug Alcohol Abuse 21: 533–547

Mattick RP, Oliphant D, Ward J, Hall W (1998) The effectiveness of other opioid replacement therapies: LAAM, heroin, buprenorphine, naltrexone and injectable maintenance. In Ward J, Mattick RP, Hall W (eds) Methadone maintenance treatment and other opioid replacement therapies. Harwood, Amsterdam, pp 123–160

*McLellan AT, Arndt IO, Metzger IS, Woody GE, O'Brien CP (1993) The effects of psychosocial services in substance abuse treatment. JAMA 269: 1953–1959

Metrebian N, Shanahan W, Wells B, Stimson GV (1998) Feasibility of prescribing injectable heroin and methadone to opiate-dependent drug users: associated health gains and harm reductions. Med J Aust 168: 596–600

Mino A (1990) Analyse scientifique de la litérature sur la remise contrôlée d'héroïne ou de morphine. Institutions universitaires de psychiatrie, Geneva

Parrino MW (1991) State methadone treatment guidelines. U.S. Department of Health and Human Services, Rockville

*Raschke P (1994) Substitutionstherapie. Ergebnisse langfristiger Behandlung von Opiatabhängigen. Lambertus, Freiburg im Breisgau

Rhoades HM, Creson D, Elk R, Schmitz J, Grabowsk J (1998) Retention, HIV risk, and illicit drug use during treatment: methadone dose and visit frequency. Am J Public Health 88: 34–39

Robinson GM, Dukes PD, Robinson BJ, Cook RR, Mahoney GN (1993) The misuse of buprenorphine and buprenorphine-naloxone combination in Wellington, New Zealand. Drug Alcohol Dependence 33: 81–86

Ruel JM (1996) International survey on the use of methadone in the treatment of narcotic addiction. Bureau of Drug Surveillance, Ottawa

Savage C, Karp EG, Curran SF, Hanlon TE, McCabe OL (1976) Methadone/LAAM maintenance: a comparison study. Compr Psychiatry 17: 415–424

Scott RTA, Jay MJH, Keith R, Oliver JS, Cassidy MT (1999) A confidential enquiry into methadone-related deaths. Addiction 94: 1789–1794

*Seidenberg A, Honegger U (1998) Methadon, Heroin und andere Opioide. Medizinisches Manual für die ambulante opioidgestützte Behandlung. Huber, Bern

Senay EC, Uchtenhagen A (1990) Methadone in the treatment of opioid dependence. In: Arif A, Westermeyer J (eds) Methadone maintenance in the management of opioid dependence. Praeger, New York, pp 19–54

Senay EC, Barthwell A, Marks R, Bokos PJ (1994) Medical maintenance: an interim report. J Addict Di 13: 65–69

Simpson DD, Curry SJ (eds) (1997) Drug Abuse Treatment Outcome Study DATOS. Psychol Addict Behav 11 (special issue): 4

Simpson DD, Sells SB (1990) Opioid addiction and treatment: a 12-year follow-up. Krieger, Malabar

Simpson DD, Joe GW, Dansereau DF, Chatham LR (1997) Strategies for improving methadone treatment process and outcomes. J Drug Issues 27: 239–260

Stitzer ML, Iguchi MY, Felch LJ (1992) Contingent take-home incentive: effects on drug use of methadone. J Consulting Clin Psychol 60: 927–934

Strang J, Gossop M (1996) Heroin prescribing in the British system: a historical review. Eur Addict Res 2: 185–193

Svikis DS, Lee JH, Haug NA, Stitzer ML (1997) Attendance incentives for outpatient treatment: effects in methadone- and nonmethadone-maintained pregnant drug dependent women. Drug Alcohol Dependence 48: 33–41

Tennant FS, Rawson RA, Pumphrey E, Seecof R (1986) Clinical experiences with 959 opioid-dependent patients treated with levo-alpha-acetylmethadol (LAAM). J Subst Abuse Treat 3: 195–202

*Uchtenhagen A, Zimmer-Höfler D (1985) Heroinabhängige und ihre normalen Altersgenossen. Herkunft, Lebenssituation, Zweijahresverlauf im Quervergleich. Haupt, Bern

Uchtenhagen A, Dobler-Mikola A, Steffen T, Gutzwiller F, Blättler R, Pfeifer S (1999) Prescription of narcotics for heroin addicts. Main results of the Swiss National Cohort Study. Karger, Basel

van Brussel G (1995) Le programme de distribution de morphine a Amsterdam: experiences pratiques. In: Rihs-Middel M, Clerc J, Stamm R (eds) La préscription de stupéfiants sous contrôle médical. Editions Médecine et Hygiène, Geneva, pp 161–167

van Rhee JM (1998) Study on the effectiveness of medically coprescribed heroin. Progress report. Central Committee on the Treatment of Heroin Addicts, Utrecht

Verthein U, Degkwitz P, Haasen C, Raschke P, Krausz M (1996) Die Substitutionsbehandlung Opiatabhängiger mit Codein/Dihydrocodein und Methadon – ein Kontrollgruppenvergleich. Sucht 42: 108–115

Walsh SL, Preston KL, Stitzer ML, Cone EJ, Bigelow GE (1994) Clinical pharmacology of buprenorphine. Ceiling effects at high doses. Clin Pharmacol Ther 55: 569–580

**Ward J, Mattick R, Hall W (1998) Methadone maintenance treatment and other opioid replacement therapies. Harwood, Amsterdam

WHO (1989) Options for the use of methadone in the treatment of drug dependence. World Health Organization, Geneva (WHO/MNH/DAT/89/2)

WHO (1998) WHO Expert Committee on Drug Dependence. Thirtieth Report. World Health Organization, Geneva (Technical report series 873)

Woody GE, McLellan AT, Luborsky LO, Brien CP (1995) Psychotherapy in community methadone programs: a validation study. Am J Psychiatr 152: 1302–1308

CHAPTER

24

F. Creed

Co-morbidity of Psychiatric Disorders

1 Introduction

Co-morbidity has been broadly defined as the presence of more than one disorder in a person in a defined period of time (Wittchen 1996). The concept has gained increasing interest with the widespread use of DSM-III diagnostic criteria. "Two or three DSM-III-based diagnoses are the rule rather than the exception" (Van Praag 1996). Some researchers find the concept compelling, while others are critical of it. This chapter will describe the research which indicates the extent of co-morbidity and some of the explanations for the concept and will indicate the clinical importance of recognising forms of co-morbidity.

There are a number of general principles. Firstly, the chances of two disorders occurring concurrently varies according to the rates of each in the population under study; the higher the base rates of the two conditions, the greater the chance of them occurring together. Secondly, there are problems with independent diagnosis. The diagnoses of anxiety and depression may not be truly independent, whereas depression and alcohol dependence are separate diagnoses. Thirdly, different authors use different definitions. Some work has concerned "lifetime" diagnosis, where the chance of co-morbidity is greater than that occurring with current disorders.

Studies of co-morbidity have been performed in population-based, primary and secondary care settings. Most weight is now being given to population-based studies, because they demonstrate the true nature of co-morbidity, rather than that which might be attributed to artefacts of treatment-seeking or referral.

There are three main types of co-morbidity: co-morbidity of related disorders (e.g. depressive and anxiety disorders), co-morbid disorders that are less directly related (e.g. depressive and somatoform disorders or psychotic disorder and substance misuse) and co-morbid physical and psychiatric disorders.

There are practical and theoretical issues surrounding co-morbidity. On the one hand, the finding that two forms of disorder frequently co-exist may throw light on their aetiology; on the other hand, co-morbidity may be associated with greater disability and require special efforts to achieve successful treatment.

2 Extent of Co-morbid Psychiatric Disorders

2.1 Population-Based Studies

The National Co-morbidity Survey (NCS) of the United States (Kessler et al. 1994) documented psychiatric diagnoses in a population-based sample of 8098 individuals aged 15–55 years. This major study has documented co-morbidity in the general population more clearly than any other. The response rate to the Composite International Diagnostic Interview (CIDI) was 82.6%, and the results have been adjusted for non-responders and weighted so that the sample is representative of the U.S. population.

In terms of DSM-III-R disorders, 52% respondents had never had any such psychiatric disorder. During their lifetime, 21% had one, 13% had two and 14% had three or more disorders. The fact that 14% of the sample had a lifetime history of three or more disorders indicates that the major burden of psychiatric disorder is concentrated in a group of highly co-morbid people who constitute about one sixth of the population.

The most common lifetime diagnoses were major depressive episode (17.1%), social phobia (11.3%), simple phobia (11.3%) and alcohol dependence (14.1%). During the last 12 months (i.e. current/recent disorders), the commonest conditions were major depressive episode (10.3%), simple phobia (8.8%), social phobia (7.9%) and alcohol dependence (7.2%).

The features associated with co-morbid (three or more) disorders were: female gender, younger age-groups, low income, low educational level and living in a large city. These variables are associated with an increased prevalence of psychiatric disorder in general and may not be specific to co-morbidity.

The prevalence of psychiatric disorder and, therefore, co-morbidity was higher in this survey than in the previous Epidemiologic Catchment Area (ECA) study (Robins et al. 1991). It is not clear whether this difference may be mostly attributed to the type of interview and probe questions used to determine prevalence in the two studies, or whether there has been a real increase in the prevalence of disorder and extent of co-morbidity (Kessler et al. 1994).

2.2 Primary Care

In the World Health Organization (WHO) study of psychological problems in general healthcare, it was reported that 24% of patients attending primary care facilities around the world received an ICD-10 diagnosis and 9.5% received multiple diagnoses, i.e. co-morbidity (Sartorius et al. 1996). The most common diagnosis was current depressive episode (10.4%), followed by generalised anxiety (7.9%), dysthymia (2.1%), panic disorder (1.1%) and agoraphobia (1.5%). There was considerable overlap between the depressive and anxiety disorders: depressive disorder was

recorded alone in 7.5% of patients, anxiety disorder alone in 5.6% and co-morbid anxiety/depression in 4.6%.

This study documented the co-occurrence of depressive with other disorders: with alcohol dependence (odds ratio [OR], 2.1), anxiety disorders (OR, 9.3) and other psychological disorder (neurasthenia, somatisation disorder, hypochondriasis or dysthymia; OR, 10.6). The authors of this study noted the very frequent overlap of anxiety and depressive symptoms (as opposed to disorders), raising questions about the exact way that psychiatric diagnoses are made in relation to co-morbidity.

2.3
Affective Disorders Co-morbid with Anxiety and Substance Misuse

An international task force attempted to identify patterns of co-morbidity across several countries (Merikangas et al. 1996). Their attempt was partly successful, but also demonstrated problems of symptom-based research. Across five countries, the association between major depression and other psychiatric disorders was consistently greater for anxiety disorders (OR, 2.7–14.9) than for alcohol misuse/dependence (OR, 1.1–2.0). Depression was more often reported in association with panic disorder (OR, 2.7–30) than with phobic states (OR, 2.5–12.2). The authors admit that the very high co-occurrence of depressive and panic disorders reflects, in part, an artefact of the syndromal approach to diagnosis, in which the large overlap of symptoms between these disorders are allowed to contribute to two diagnoses.

2.4
"Dual Diagnosis"

Numerous studies from the United States have demonstrated an increased prevalence of substance misuse in patients with schizophrenia or bipolar affective disorders, so-called "dual diagnosis" (Regier et al. 1990). The clinical importance of this form of co-morbidity is that patients with psychotic disorder who also abuse drugs or alcohol have a worse outcome (Lehman et al. 1993). These patients are often heavy users of emergency services and in-patient care, and they may also have high rates of forensic histories and homelessness. There has been relatively little work on this particular form of co-morbidity in Europe, but approximately one quarter of patients admitted in Munich with schizophrenia also had a diagnosis of substance misuse or dependence (Soyka et al. 1993). In London, the 1-year prevalence rate for any substance misuse problem among patients with psychotic illnesses in contact with a mental health service was 36.3%. These patients had spent almost twice as many days in hospital as those without such a problem over the previous 2 years (Menezes et al. 1996).

It is unlikely that substance abuse is merely a response to schizophrenic symptoms, as no clear relationship has been found. Non-compliance with medication is associated with a poor outcome, more disturbed behaviour and longer hospital stays in people with schizophrenia who also abuse medication.

3
Co-morbidity of Psychiatric and Physical Disorders

3.1
Population-Based Studies

A number of population-based studies have assessed the prevalence of psychiatric disorders in respondents who have physical disorders. These have been summarised by Weyerer (1990), who documented an increased prevalence of psychiatric disorder in those with physical disorder (risk ratio of between 1.4 and 3.5). The main difficulty with such studies in the community is that they rely on the respondents' self-report of physical illness.

The most specific data were recorded in the ECA study (Wells et al. 1988b). The 6-month prevalence of any psychiatric disorder was 19.7% (most commonly phobic, depressive disorders, dysthymia and alcohol abuse), and the prevalence of a current chronic medical condition was 34% (most commonly arthritis and hypertension). Just over 7% of the population had co-morbid psychiatric and physical disorders. The prevalence of psychiatric disorders was greater in respondents with physical disorders than the remainder: a psychiatric disorder was present in 24.7% of those with and 17.5% of those without a chronic medical condition. Corresponding figures for anxiety disorders were 12% and 6%, and for depressive disorders 9.4% and 5.8%. These differences were significant even after controlling for demographic factors. The medical conditions with the greatest association with psychiatric disorders were arthritis, cancer, chronic lung disease, neurological disorder and heart disease. The exception was hypertension, which was the commonest chronic medical disease in the ECA sample, but was not

associated with an increased risk of psychiatric disorder.

3.2
Primary and Secondary Care

These associations between psychiatric and physical disorders have been repeatedly demonstrated in patients treated at general hospitals (Rodin et al. 1991; Creed and Guthrie 1996). The highest prevalence rates were in patients with cancer, arthritis, chronic respiratory disease and neurological disorders, notably stroke and Parkinson's disease. The principal difficulty in these studies is that of measuring psychiatric disorder independent of the physical illness. Many of the symptoms of depression overlap with those of physical disease (e.g. weight loss, fatigue, impaired sleep, pain; Creed 1997). It has been noted that the risk factors for psychiatric disorder in the physically ill are similar to those in the general population (female gender, living alone, single status, previous psychiatric history), together with aspects of the physical illness, notably its disabling effect and its treatment (Rodin et al. 1991).

The prevalence of psychiatric disorder in hospital samples depends on the location. Approximately 15% of medical in-patients have anxiety or depressive disorders; the same holds for out-patients with recognised physical disorders. The prevalence is higher (35%–60%) in out-patients with medically unexplained symptoms or syndromes (e.g. irritable bowel syndrome, chronic fatigue, headache). Approximately 18% of men and 4% women have alcohol problems, and dementia and delirium are common among older individuals in the general hospital setting.

3.3
Hypochondriasis in Patients Attending Secondary Care Settings

In a series of studies concerning medical out-patients, Barsky has demonstrated that 4%–6% have DSM-III-R hypochondriasis, compared to 13% with major depression, 8.5% with panic disorder and 5% with alcohol abuse/dependence. Co-morbidity of hypochondriasis and panic disorder is common, so these patients have three diagnoses – a physical disorder, hypochondriasis and panic disorder. Treatment of the panic disorder may lead to a reduction of hypochondriasis, emphasising the close relationship of the two diagnoses, but a majority of patients with hypochondriasis do not have panic disorder (Bass and Murphy 1996). One fifth also have an axis II (personality) disorder or somatisation disorder (Barsky et al. 1992).

4
Effect of Co-morbidity on Severity and Disability

4.1
Disability Associated with Co-morbid Conditions

Depressive disorders and chronic medical conditions have additive effects on patient functioning and well-being (Wells et al. 1989). In one study, respondents with a chronic medical condition who also had a psychiatric disorder were three times as likely to have spent 1 or more days in bed and twice as likely to have restricted activity because of illness than respondents with chronic illness who do not have psychiatric disorder (Wells et al. 1988a).

Patients with mixed anxiety and depressive disorders experience greater disability then those with "pure" anxiety or "pure" depression. Similarly, depressed patients with co-morbid panic disorder have poorer role and social functioning than those depressed patients without co-morbid panic disorder (Noyes 1990). The presence of anxiety co-morbid to depressive or medical disorders leads to a significant reduction of the functioning and well-being of patients, with an additional 3.8 bed days per month.

In the WHO Primary Care Study, there was a significant correlation between disability, whether rated by the doctor or the patient, and the extent of psychiatric disorder. A total of 31% of patients with a single ICD-10 psychiatric diagnosis rated themselves as moderately or severely occupationally disabled; in those with two or more psychiatric diagnoses, the figure was 50%. The study also found a less striking increase in disability in patients who had a physical disorder in addition to their psychological disorder. Further analysis indicated that both psychiatric and physical disorders were independently associated with disability. In relation to occupational disability, psychiatric disorder was somewhat more important. Both physical and psychiatric disorder contributed approximately equally to self-rated disability (Ormel and Costa e Silva 1995).

Patients with physical illness who also have a co-morbid psychiatric disorder have a worse outcome than similar patients without psychiatric disorder. For example, heart disease patients with concurrent depressive or anxiety disorders experience more chest pain and a worse quality of life than the remainder and are slower to return to work after a heart attack. Attention has been focused recently on patients with a recent heart attack, where concurrent depression is

an independent risk factor for early death, even after physical risk factors have been controlled for (Musselman et al. 1998).

4.2 Severity of Co-morbid Disorders

When depressive disorder is secondary to a prior anxiety disorder, it tends to be more persistent and more severe than depression without anxiety features. The increased severity is associated with greater self-perceived interference with life and activities, more suicide attempts and hospitalisation. In his longitudinal study, Angst (1996) noted that suicide attempts were particularly common in those who had either combined depressive disorders or a combination of major depressive disorder and anxiety disorder.

Data from the ECA study demonstrated that panic attacks and major depression occurred together 11 times more frequently than expected by chance. Patients with both disorders had worse symptoms than those who only had one disorder, but the pattern of symptoms was similar. This suggests that co-morbidity, in this setting, may represent a more severe form of depressive disorder.

4.3 Health Care Utilisation

In the NCS (Kessler et al. 1994), 22.5% of those with three or more disorders had attended a mental health treatment facility in the previous 12 months compared to 11.5% of those who had a single disorder. During the course of a lifetime, the figures were 41.0% and 26.2%, respectively. It is clear that those with co-morbid disorder consult more frequently than those with a single disorder, but these data indicate that only a minority of people with two or more disorders attend for treatment.

Similarly, in the general medical sector, the presence of depressive disorder in patients who also have organic disorder leads to increased health care utilisation (Barsky et al. 1986).

Paradoxically, although co-morbidity leads to additional demands and need for treatment, in certain settings it also reduces the chance that additional treatment is provided. Depression co-morbid with physical disorder is less likely to be recognised and treated than when depression occurs alone, as the attention of the doctor and patient is focused on the physical illness. When depressive disorder occurs with another psychological disorder, e.g. anxiety, the depressive disorder is more likely to be recognised because the depression is more severe. However, the concurrent anxiety is often missed leading to inadequate treatment. The same is true of depressive disorder in association with schizophrenia or substance abuse. There is also a considerable problem with recognition of substance abuse disorders among patients with schizophrenia (Citrome and Volavka 1999).

4.4 Response to Treatment

Depressed patients in primary care with a co-morbid anxiety disorder recover more slowly than the remainder and tend to terminate treatment prematurely more frequently than patients with major depression alone. A lifetime history of panic disorder in addition to current major depression led to a poor response to standard psychopharmacological and psychotherapeutic antidepressant treatments.

Panic disorders with co-morbid anxiety (including agoraphobia), depressive and personality disorders are less likely to improve with treatment than "pure" panic disorders (Page 1998). It has been recommended that co-morbidity in panic disorder be routinely sought, as the combination of disorders may respond particularly well to cognitive behaviour therapy. The high co-morbidity of neurotic disorders among panic disorder patients decreases after cognitive behaviour therapy, which is compatible with the concept that co-morbidity represents a more severe disorder.

5 Theories Surrounding the Aetiology of Co-morbid Anxiety and Depressive Disorders

There are different theoretical models of co-morbidity of anxiety and depressive disorders (Wittchen 1996; Kessler et al. 1998):

- One disorder (e.g. anxiety) may predispose a person to develop a second disorder (depression). For example, one hypothesis states that long-standing anxiety disorders predispose a person to develop later depressive disorders. The reverse may be true: depression could lead to anxiety.
- One or more common aetiological factors may be responsible for both anxiety and mood disorders. Such aetiological factors may be distant (e.g. genetic factors or lasting behavioural responses learned in childhood) or recent, such as life events or more situational behavioural responses.
- The "disorders" (anxiety and depression) may be different manifestations of the same disorder.

5.1 Primary or Secondary Depression

In the NCS, most cases of major depressive disorder were secondary, in the sense that they occurred in people with a prior history of another DSM-III-R disorder. Two thirds of the respondents in the NCS with a lifetime history of major depressive disorder had a previous history of another disorder (usually anxiety), i.e. this could be regarded as secondary depression. A further 12% of patients had "primary depression", i.e. the anxiety disorder only developed after the depressive disorder. The remaining subjects had "pure" depression, i.e. without any anxiety or other psychiatric disorder; this was more common among women than men (30% of depressive episodes in women and 19% in men).

The association between panic attack and depression is clear. The odds ratios between lifetime panic disorder and depression are statistically significant and substantial: for panic disorder with depression, the odds ratio is 6.2, and for panic disorder with depression, 6.8, i.e. a person with depressive disorder is six times more likely than others to experience a panic attack, and someone with panic disorder is 6.8 times more likely to develop depression. There are approximately equal numbers of those who experience depression first or panic first (Kessler et al. 1998).

When depression is secondary to anxiety or panic disorder, its onset occurs in very close association with the timing of anxiety disorder. However, when the primary disorder is agoraphobia, substance misuse or conduct disorder, the secondary depression occurs after a longer time lag (5–12 years), suggesting a much less close aetiological relationship.

5.2 Social Factors in the Aetiology of Anxiety and Depression

Brown et al. (1996) examined the role of social factors in co-morbidity of anxiety and depressive disorder. In a community sample of 400 women with children (a high-risk group for the development of depression), the 1-year prevalence was 16% for depression alone, 7% for anxiety alone and 16% for mixed anxiety and depression. The study indicated that both anxiety and depression were associated with the experience of childhood adversity (sexual or physical abuse or marked parental neglect) and with the experience of severe life events. The differences lay in the type of life event and current social situation. Whereas the onset of depression was associated with "loss" events (e.g. bereavement, marital or other family separation), the onset of anxiety was associated with a "danger" event (e.g. starting an affair or discovering a breast lump, which carries the risk of future threat). Women with co-morbid anxiety and depression had an excess of events that carried both elements of loss and danger (Brown et al. 1996).

The absence of adequate social support appears to be crucial to the development of depression, whereas in anxiety it appears to be unimportant. This suggests that anxiety disorders may be less responsive to current psychosocial factors.

In the second analysis, Brown et al. (1996) calculated the chances of anxiety and depressive disorders occurring together; this is likely: (a) as both are common in the population under study and (b) because the two disorders share a common antecedent factor – childhood adversity. However, even after taking these two factors into account, 47% of the co-morbidity of anxiety and depression remained unexplained; genetic or constitutional factors are suggested as the factors most likely to explain this remaining variance.

The last analysis in this study concerned the development of a new episode of depressive disorder during the year of the study. A depressive episode was three times more likely to develop in the context of an ongoing anxiety disorder than in a period free of anxiety. On the other hand, the reverse was not true: onsets of anxiety were not more common in the presence of ongoing depressive disorder.

Others have reported that the onset of anxiety disorders tends to precede or occur contemporaneously with depression, whereas the onset of alcohol misuse is equally likely to precede or follow the onset of depression (Merikangas et al. 1996).

Kessler and colleagues (1997) also found a link between childhood adversity and an increased chance of psychiatric disorder in adulthood. The clearest direct association lay between childhood adversity and mood disorders. Respondents who had co-morbid conditions reported a higher rate of childhood adversity than the remainder. The relationships between childhood adversity and addictive disorders were more likely to be attributable to an indirect link, i.e. through prior lifetime co-morbid disorders; in other words, the addictive disorders were more likely than other disorders to be secondary to another psychiatric disorder.

5.3 Genetic Factors in Major Depression and Generalised Anxiety Disorder

Using bivariate twin analyses, Kendler (1996) has tackled the question of whether genetic factors account for the common co-occurrence of major depressive

disorder and anxiety disorders. In a series of studies, he has demonstrated that the relationship between major depression and generalised anxiety disorder is greatly affected by the model of diagnostic scheme employed. If a non-hierarchical diagnostic classification was used, there was clear evidence of co-morbidity – if major depression was present, the odds ratio for generalised anxiety disorder was 8.9. If a diagnostic hierarchical model was used (so that generalised anxiety disorder that occurred only in the presence of major depression was not diagnosed), then the odds ratio fell to 3.9. Further analyses led to the conclusion that all the genes that influenced the liability to major depression also influenced the liability to generalised anxiety disorder. These results suggest that, although some environmental factors are both angiogenic and depressogenic, other environmental factors are relatively specific in their impact on the liability either to major depression or generalised anxiety disorder.

Further twin studies have suggested that common genetic factors are clustered into three groups. In one group, the genetic factors are shared by phobias, panic disorder and bulimia; a second set of genetic factors are shared by unipolar depression and generalised anxiety disorder; and a third genetic factor was specific to alcoholism (Kendler 1996). These results are not compatible with the idea that these six disorders are fundamentally distinct disease entities (Sullivan and Kendler 1998). They are also not consistent with the existence of a general neurotic syndrome, in which most or all of these disorders would result from the same genetic factors (Tyrer 1989).

6
Concepts of Co-morbidity: Dimensional Models for Mental Disorders

Some researchers reject the concept of co-morbidity between anxiety and depression, claiming that the same symptom contributes to both diagnoses (Goldberg 1996). Using latent trait analysis to visualise symptoms in dimensions rather than diagnostic categories, Goldberg and colleagues have found that two correlated dimensions accounted for all the common variance shared between symptoms. The more depressive symptoms were recorded, the more likely it was that a diagnosis of depressive disorder was made. The same was true of anxiety symptoms, but the majority of symptoms occupied shared space, i.e. they contributed to both diagnoses. The size of "co-morbidity" between anxiety and depression is not a measure of the frequency with which two independent morbid conditions co-exist; it merely indicates the extent to which symptoms contribute to both "disorders".

This model emphasises that anxiety and depression share a common cause, including the common genetic basis, the shared risk factor of parental neglect and abuse during childhood and the fact that pre-existing anxiety is a risk factor for depression. Certain social variables, such as dangerous events and inadequate social support, are seen to modify the symptom presentation rather than support the concept of independent disorders. Genetic factors are only seen as having a role in determining general vulnerability towards common mental disorders, but not at all in determining why one person develops anxiety and another depression (Goldberg and Huxley 1992).

An even broader view is taken in the concept of a general "neurotic syndrome", which has been suggested to describe a subset of patients whose symptoms are remarkably protean and who present at different times with varying symptom complexes of anxiety and depression. Consistent with this conceptualisation, psychiatric conditions approximating to these "neuroses" have a remarkable tendency to co-occur in epidemiological samples (Tyrer 1989). This model is similar to the view of a genetically determined general vulnerability, but is incompatible with the genetic findings of Kendler described above.

7
Co-occurrence of Physical and Psychiatric Disorder

The ways that physical and psychiatric disorders are associated show some similarities but some differences to the co-occurrence of psychiatric disorders. One difference lies in the category of psychiatric disorders which arise directly from physical disease (Creed and Guthrie 1996; Royal College of Physicians/Royal College of Psychiatrists 1995). There may be an obvious and direct link following cerebral involvement in disease (e.g. stroke, encephalitis, temporal lobe epilepsy), or there may be a less direct link. For example, there is an increased prevalence of depression in cancer, which may occur prior to the cancer becoming manifest.

The temporal relationship between the onset of the physical and psychiatric disorders has been intensely studied. This has led to the development of two distinct models. The first emphasises the importance of pre-existing life stress and depression, which is later followed by the onset of a physical disorder. This relationship is particularly close if the disorder lacks evidence of obvious organic disease, e.g. irritable bowel syndrome, chronic fatigue syndrome (Creed 1993).

A psychological disorder may develop following the onset of physical disease, and it has often been assumed that the pain, disability and potentially life-threatening quality of the illness leads to the later development of depression. This model is too simplistic, however, as other factors, including social support and prior history of psychiatric disorder, are involved (Creed 1990).

Part of the association between physical and psychiatric disorders can be understood in terms of an underlying, common aetiology. One example is the abuse of alcohol, which may lead to gastric or hepatic diseases associated with depression. Eating disorders may also lead to secondary mood disorders and physical illness.

Chronic fatigue syndrome and related disorders frequently co-occur with anxiety or depressive disorders. Whether this is true co-morbidity or whether they are different manifestations of the same disorder presents the same issues described above in relation to anxiety and depressive disorders. The close association between chronic fatigue and anxiety/depressive disorders may be similar to the anxiety and depression dimensions described by Goldberg (1996). Conventional explanations of the aetiology of these disorders characterised by pain or fatigue have assumed genetic, developmental, cognitive and social factors (Hickie et al. 1998). Recent research provides evidence for the role of health anxiety, treatment-seeking and neurobiological and physiological mechanisms in some of these patients.

The leading symptom in these disorders, apart from fatigue, is pain. Pain and depression are intimately related, and there are several ways in which pain and depression may be related (von Korff and Simon 1996). Neurotransmitters implicated in depressive illness also play a critical role in pain modulation. Chronic pain has been viewed as a form of somatisation in which negative emotions are expressed as bodily complaints. Somatosensory amplification occurs in depression, leading to an increased propensity to experience and report dysphoric symptoms, including pain. Depression may occur secondary to impaired social role performance and reduced activity levels as a form of learned helplessness.

8 Conclusion

Co-morbid disorders are frequent in population studies and clinical practice at primary and secondary care levels. Their importance lies in the fact that co-morbid disorders generally represent more severe forms of the constituent disorders compared to "pure" disorders. The co-morbidity is often overlooked because the doctor focuses on one diagnosis and may ignore the second disorder. Undetected co-morbid disorders tend to impair the prognosis of the first disorder.

The theoretical discussions will continue until such time as a combination of genetic, epidemiological and neurobiological studies indicates more clearly the relationship between disorders. It seems likely that there is very considerable overlap between anxiety, panic and depressive disorders and those concerning somatic expression of distress. The relationship between psychotic disorders and substance misuse and between physical and psychological disorders is more complex, as these may or may not be related disorders. There are a variety of reasons underlying the association. The practitioner should be guided by the statistical fact that if one disorder is present, then there is an increased, not reduced likelihood of a second disorder also being present. This must be detected and treated in addition to the primary disorder.

9 References

Angst J (1996) Comorbidity of mood disorders: a longitudinal prospective study. Br J Psychiatry 168(30): 31–37

Barsky AJ, Wyshak G, Klerman GL (1986) Medical and psychiatric determinants of outpatient medical utilization. Med Care 24(6): 548–560

Barsky AJ, Wyshak G, Klerman GL (1992) Psychiatric comorbidity in DSM-III-R hypochondriasis. Arch Gen Psychiatry 49: 101–108

Barsky AJ, Fama JM, Bailey D, Ahern DK (1998) A prospective 4- to 5-year study of DSM-III-R hypochondriasis. Arch Gen Psychiatry 55: 737–744

Bass C, Murphy M (1996) Somatisation, somatoform disorders and factitious illness. In: Guthrie E, Creed FH (eds) Seminars in liaison psychiatry. Gaskell, London, pp 103–156

Brown GW, Harris TO, Eales MJ (1996) Social factors and comorbidity of depressive and anxiety disorders. Br J Psychiatry 168(30): 50–57

Citrome L, Volavka J (1999) Schizophrenia: violence and comorbidity. Curr Opin Psychiatry 12: 47–51

Creed FH (1990) Psychological disorders in rheumatoid arthritis: a growing consensus? Ann Rheum Dis 49: 808–812

Creed FH (1993) Stress and psychosomatic disorders. In: Goldberger L, Berznitz S (eds) Handbook of stress. Free Press, New York

Creed FH (1997) Assessing depression in the context of physical illness. In: Robertson MM, Katona CLE (eds) Perspectives in psychiatry, vol 6. Depression and physical illness. Wiley, London, pp 3–19

Creed FH, Guthrie E (1996) The classification of psychiatric disorders and their relationship to physical disorders. In: Guthrie E, Creed FH (eds) Seminars in liaison psychiatry. Gaskell, London, pp 53–74

Goldberg D (1996) A dimensional model for common mental disorders. Br J Psychiatry 168(30): 44–49

Goldberg D, Huxley P (1992) Common mental disorders: a biosocial model. Routledge, London

Hickie IB, Scott EM, Davenport TA (1998) Somatic distress: developing more integrated concepts. Curr Opin Psychiatry 11: 153–158

Kendler KS (1996) Major depression and generalised anxiety disorder. Same genes, (partly) different environments – revisited. Br J Psychiatry 168(30): 68–75

**Kessler RC, McGonagle KA, Zhao S, Nelson CB, Hughes M, Eshleman S, Wittchen H-U, Kendler KS (1994) Lifetime and 12-month prevalence of DSM-III-R psychiatric disorders in the United States. Arch Gen Psychiatry 51: 8–19

Kessler RC, Davis CG, Kendler KS (1997) Childhood adversity and adult psychiatric disorder in the US national comorbidity survey. Psychol Med 27: 1101–1120

Kessler RC, Stang PE, Wittchen H-U, Ustun TB, Roy-Burne PP, Walters EE (1998) Lifetime panic-depression comorbidity in the National Comorbidity Survey. Arch Gen Psychiatry 55: 801–808

Lehman AF, Myers CP, Thompson JW et al (1993) Implications of mental and substance abuse disorders: a comparison of single and dual diagnosis patients. J Nerv Ment Dis 181: 365–370

Menezes PR, Johnson S, Thornicroft G, Marshall J, Prosser D, Bebbington P, Kuipers E (1996) Drug and alcohol problems among individuals with severe mental illnesses in south London. Br J Psychiatry 168: 612–619

Merikangas KR, Angst J, Eaton W, Canino G, Rubio-Stipec M, Wacker H, Wittchen H-U, Andrade L, Essau C, Kraemer H, Robins LN, Kupfer DJ (1968) Comorbidity and boundaries of affective disorders with anxiety disorders and substance misuse: results of an international taskforce. Br J Psychiatry 168(30): 58–67

Merikangas KR, Angst J, Eaton W, Canino G (1996) Comorbidity and boundaries of affective disorders with anxiety disorders and substance misuse: results of an international task force. Br J Psychiatry Suppl 30: 58–67

Musselman DL, Evans DL, Nemeroff CB (1998) The relationship of depression to cardiovascular disease. Arch Gen Psychiatry 55: 580–592

Noyes R (1990) The comorbidity and mortality of panic disorder. Psychiatr Med 8: 41–66

Ormel J, Costa e Silva JA (1995)The impact of psychopathology on disability and health perceptions. In: Üstün TB, Sartorius N (eds) Mental illness in general health care. An international study. Wiley, New York, pp 335–346

Page AC (1998) Assessment of panic disorder. Curr Opin Psychiatry 11: 137–141

Regier DA, Farmer ME Rae DS et al (1990) Comorbidity of mental disorders with alcohol and other substances: results from the Epidemiological Catchment Area. JAMA 264: 2511–2518

Robins LN, Locke BZ, Regier DA (1991) An overview of psychiatric disorders in America. In: Robins LN, Regier DA (eds) Psychiatric disorders in America: The Epidemiologic Catchment Area Study. Free Press, New York, pp 328–366

Rodin R, Craven J, Littlefield C (1991) Depression in the medically ill. An integrated approach. Brunner/Mazel, New York

Royal College of Physicians and Royal College of Psychiatrists (1995) The psychological care of medical patients. Royal College of Physicians/Royal College of Psychiatrists, London

Sartorius N, Üstün TB, Lecrubier Y, Wittchen H-U (1996) Depressions comorbid with anxiety: results from the WHO study on psychological disorders in primary health care. Br J Psychiatry 168(30): 38–43

Soyka M, Albus M, Kathmann N et al (1993) Prevalence of alcohol and drug abuse in schizophrenic inpatients. Eur Arch Psychiatry Clin Neurosci 242: 362–372

Sullivan PF, Kendler KS (1998) Genetic epidemiology of 'neurotic' disorders. Curr Opin Psychiatry 11: 143–147

Tyrer P (1989) Classification of neurosis. Wiley, New York

Van Praag HM (1996) Comorbidity (psycho) analysed. Br J Psychiatry 168(30): 129–134

von Korff M, Simon G (1996) The relationship between pain and depression. Br J Psychiatry 168(30): 101–108

Wells KB, Golding JM, Burnam MA (1988a) Psychiatric disorder and limitations in physical functioning in a sample of the Los Angeles general population. Am J Psychiatry 145(6): 712–717

Wells KB, Golding JM, Burnam MA (1988b) Psychiatric disorder in a sample of the general population with and without chronic medical conditions. Am J Psychiatry 145(8): 976–981

Wells KB, Stewart A, Hays RD, Burnam A, Rogers W, Daniels M, Berry S, Greenfield S, Ware J (1989) The functioning and well-being of depressed patients. Results from the Medical Outcomes Study. JAMA 262: 914–919

Weyerer S (1990) Relationships between physical and psychological disorders. In: Sartorius N, Goldberg D, de Girolamo G, Costa e Silva JA, Lecrubier Y, Wittchen H-U (eds) Psychological disorders in general medical settings. Hogrefe and Huber, Toronto, pp 34–46

*Wittchen H-U (1996) Critical issues in the evaluation of comorbidity of psychiatric disorders. Br J Psychiatry 168(30): 9–16

Subject Index

Subject Index

Subject Index